JMIR mHealth and uHealth

Contents

Viewpoints

Reviews

Original Papers

XSL•FO

RenderX

Corrigenda and Addenda

Viewpoint

Prescribing Behavior Change: Opportunities and Challenges for Clinicians to Embrace Digital and Mobile Health

Anish Agarwal[1,2,3,4*], MD, MPH, MSHP; Mitesh Patel[2,3,4,5,6*], MD, MBA, MSHP

[1]Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA, United States

[2]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[3]Leonard Davis Institute for Health Economics, University of Pennsylvania, Philadelphia, PA, United States

[4]Center for Health Care Innovation, Penn Medicine, Philadelphia, PA, United States

[5]Nudge Unit, Penn Medicine Center for Health Care Innovation, University of Pennsylvania, Philadelphia, PA, United States

[6]Department of Medicine, Crescenz Veterans Affairs Medical Center, Philadelphia, PA, United States

[*]all authors contributed equally

Corresponding Author:
Anish Agarwal, MD, MPH, MSHP
Department of Emergency Medicine
University of Pennsylvania
Blockley Hall, Room 428
423 Guardian Drive
Philadelphia, PA, 19146
United States
Phone: 1 610 304 2318
Email: anish.agarwal@pennmedicine.upenn.edu

Abstract

Individual behaviors impact physical and mental health. Everyday behaviors such as physical activity, diet, sleep, and tobacco use have been associated with a range of acute and chronic medical conditions. Educating, motivating, and promoting sustained healthy behaviors can be challenging for clinical providers attempting to manage their patients' health. The ubiquity and integration of mobile and digital health devices (eg, wearable step counters, smartphone-based apps) allow for individuals to generate and record enormous amounts of patient-generated health data. Research studies have begun to reveal how mobile and digital devices offer promise in motivating individual behavior change but they have not had consistent results. In this viewpoint, we discuss the potential synergy of digital health modalities and behavioral strategies as an approach for clinicians to prescribe, motivate, monitor, and sustain healthy behaviors. We discuss the strengths, challenges, and opportunities for the future of promoting health behaviors.

Everyday health behaviors such as diet, physical activity, tobacco use, and medication adherence contribute significantly toward long-term health [1]. While clinicians certainly recognize the importance of these behaviors, they often feel paralyzed when it comes to addressing them [2,3]. Most commonly, clinicians will attempt to change behavior through education (eg, highlighting the importance of a healthy diet and regular exercise) [4]. While it may be helpful for a patient to hear this from their clinician, it is often not enough to lead to sustained behavior change [5].

What if clinicians could more easily prescribe, motivate, and sustain behavior change, similar to how they prescribe medications and monitor changes in laboratory test results for diabetes or high cholesterol? Regular exercise is just as important for both of these conditions, yet clinicians lack an easy mechanism to offer therapy or motivation as well as monitor progress. Previously, the only way to obtain information on daily health behaviors was through self-reporting, which is often unreliable and inconsistent. Nearly 80% of US adults now carry a smartphone [6] and the ubiquity of these digital technologies as well as wearable devices allows for seamless, passive, and remote tracking of behaviors such as physical activity or sleep patterns [7]. Emerging literature has begun to describe the applications of digital health modalities across

varying specialties and patient populations [8,9]. The strength of digital health may lie in its ability to offer clinicians a means to individualize, engage, and support an individual's behavior change continuously as compared with intermittent office visits or activity logs. Moreover, these strategies may be used worldwide, including in low- and middle-income countries where mobile phone ownership is high and access to care may be limited. Effective implementation of these approaches, combined with strategies rooted in behavioral science, could improve health across populations. However, they remain relatively unexplored and underutilized by clinicians.

Several recent studies reveal promising opportunities. In a randomized trial of patients with ischemic heart disease, those provided with a wearable device for 6 months had no meaningful change in physical activity [10]. However, patients with personalized goals coupled with loss-framed financial incentives accumulated more than 100 miles of increased steps with sustained activity levels during the following months, when incentives were stopped. In another study, families from the Framingham Cohort were randomly assigned to use an activity tracker, select a step goal, and receive daily feedback or do each of those in addition to being enrolled in a game that incorporated behavioral insights (eg, precommitment, loss aversion, and peer support) into the design [11]. Families that played the game achieved activity goals at significantly higher rates during the 12-week intervention with sustained differences relative to control during the 12 weeks after the game ended. These trials reveal an important insight. Simply providing an activity tracker did not lead to meaningful changes in activity. However, combining the technology with a behavioral strategy led to significant changes that were sustained in the postintervention period. Although longer-term studies are needed, these trials demonstrate the potential use of digital health to effectively motivate populations toward long-term healthy behaviors such as physical activity, dietary changes, or smoking cessation. These studies offer early insights on an approach that, when integrated into practical clinical workflows, offers patients an additional method of capturing and managing their own data. In addition, this approach would enable clinicians to deploy evidence-based behavior change strategies.

How can clinicians incorporate these types of approaches within their practice? First, clinicians must be able to interpret and manipulate data collected on patient behaviors within the electronic health record (EHR). Granular data reflecting daily step counts are likely to become overwhelming. Instead, longitudinal trends and inflection points may reveal opportunities for intervention and longer-term monitoring. Some institutions enable patients to send their information to their EHR through Apple's Healthkit [12], but these data need to be presented in more actionable ways and integration models are needed to reflect the growing number of devices produced by different companies (eg, FitBit, Samsung, Google). Clinicians need the ability to automate and individualize tailored behavior change techniques to patient needs in real time. Behavior change will likely require an approach which evolves with the patient over time to maintain long term results. Digital interfaces have a unique capacity for personalization due to patient-level sociodemographic factors. Additionally, accessing and understanding these data and the methods of integrating behavior change into practice must begin during medical training. Medical trainees, now adept with digital technologies, are primed to learn how to best weave in behavior change strategies when managing individual or population health.

Second, clinicians and patients must be able to trust the data collected to inform health decisions [13]. Many consumer-grade health technologies have not been extensively evaluated for large-scale use, and their accuracy, validity, and reliability remain unknown. Some types of data may require a higher level of validity. For example, algorithms that not only monitor heart rate but also attempt to identify abnormal heart rhythms [14] may require larger clinical trials than studies testing the accuracy of tracking steps. Data security is also important, particularly for more sensitive patient information.

Third, once the infrastructure is in place to collect, display, and interpret data, then clinicians need a mechanism and reimbursement structure for prescribing and monitoring behavior change. This will depend on sufficient evidence to identify effective approaches and EHRs to allow for ordering of behavioral interventions to be just as simple as ordering a medication. Reimbursement mechanisms will need to be in place to support these initiatives. There are examples of insurance coverage for effective behavioral interventions such as cardiac rehabilitation and the diabetes prevention program [15,16]. Yet, even these interventions are often not offered to all patients who could benefit from them. The additional time and effort clinicians use for digital health monitoring and the accompanying decision making will need to be accounted for and will vary. Similar to "interactive" automobile or life insurance policies, which use smartphone data to offer incentives for safe driving practices or physical activity, health insurers could stand to benefit from accessing enrollees' data from behavior change programs.

Finally, to achieve success in prescribing behavior change, clinicians will need to use effective, evidence-based methodologies to motivate and sustain behaviors. Prescribing behavior change through digital health methods will need real-time, seamless, and continuous quality improvement to evaluate which mechanisms are working (or failing) to produce meaningful outcomes. As the capabilities of technology evolve and the approaches to monitoring health data grow, research will need to prioritize efforts in investigating the sustained effectiveness of varying behavioral approaches [17].

Motivating lifestyle-related changes in patients remains a critical yet challenging task for clinicians. Not all behavior change strategies are created equal and studies have demonstrated heterogenous results in the ability to maintain long-term results. Coupling these strategies with digital health may provide a means to build a multimodal approach to sustained outcomes. However, there are some obstacles to utilizing digital health to monitor, motivate, and support these behavior changes. Despite the seeming ubiquity of technology, not all individuals have access to smartphones or wearable devices. Many digital health companies have created partnerships with health insurers to expand access and provide a means to monitor chronic disease and drive healthy habits. Furthermore, the technologic landscape

XSL•FO

RenderX

is rapidly changing and will require clinicians to be nimble in their understanding of the use and limitations of digital health technologies. Data validity and integration will need to be continuously studied, upheld, and monitored to ensure safe and effective use. Qualitative research from key stakeholders highlights the need for good data management and quality assessment [18]. The digitization of medicine will offer clinicians an additional mode of delivering care and monitoring health. The role of other stakeholders (eg, insurers or employers) will also be important to study. As patient-generated digital data continues to evolve, health systems must focus their efforts on embedding these digital modalities into clinical practice alongside behavioral approaches. Medical training should incorporate these approaches and empower clinicians with an additional toolkit to motivate and monitor behavior change in their patients.

The predominant approach to promoting lifestyle modification has relied on educating patients. The exponential growth, capacity, and functionality of digital health may help provide a pathway toward prescribing behavior change just like a clinician would prescribe a statin. While several challenges exist, technology will continue to evolve rapidly. Developing behavioral strategies which are evidence-based and practically scalable will help clinicians focus on implementation in their local environments. The critical next step is to better integrate and test behavioral strategies to support clinicians' efforts to improve their patients' long-term health.

Conflicts of Interest

MP is the founder of Catalyst Health LLC, a consulting firm, and is an advisory board member for Life.io, Healthmine Services, and Holistic Industries. AA has no conflicts to declare.

References

1. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington (DC), USA: National Academies Press; 2001.

2. Stange KC. The paradox of the parts and the whole in understanding and improving general practice. Int J Qual Health Care 2002 Aug;14(4):267-268. [doi: 10.1093/intqhc/14.4.267] [Medline: 12201184]

3. Greco PJ, Eisenberg JM. Changing physicians' practices. N Engl J Med 1993 Oct 21;329(17):1271-1273. [doi: 10.1056/NEJM199310213291714] [Medline: 8413397]

4. Paterick TE, Patel N, Tajik AJ, Chandrasekaran K. Improving health outcomes through patient education and partnerships with patients. Proc (Bayl Univ Med Cent) 2017 Jan;30(1):112-113 [FREE Full text] [doi: 10.1080/08998280.2017.11929552] [Medline: 28152110]

5. Cohen DJ, Tallia AF, Crabtree BF, Young DM. Implementing health behavior change in primary care: lessons from prescription for health. Ann Fam Med 2005 Aug;3 Suppl 2:S12-S19 [FREE Full text] [doi: 10.1370/afm.334] [Medline: 16049075]

6. Pew Research Center. Mobile Fact Sheet. Demographics of Mobile Device Ownership and Adoption in the United States Internet. URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-08-27]

7. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]

8. Asbjørnsen RA, Smedsrød ML, Solberg Nes L, Wentzel J, Varsi C, Hjelmesæth J, et al. Persuasive System Design Principles and Behavior Change Techniques to Stimulate Motivation and Adherence in Electronic Health Interventions to Support Weight Loss Maintenance: Scoping Review. J Med Internet Res 2019 Jun 21;21(6):e14265 [FREE Full text] [doi: 10.2196/14265] [Medline: 31228174]

9. Flores MG, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]

10. Chokshi NP, Adusumalli S, Small DS, Morris A, Feingold J, Ha YP, et al. Loss-Framed Financial Incentives and Personalized Goal-Setting to Increase Physical Activity Among Ischemic Heart Disease Patients Using Wearable Devices: The ACTIVE REWARD Randomized Trial. J Am Heart Assoc 2018 Jun 13;7(12) [FREE Full text] [doi: 10.1161/JAHA.118.009173] [Medline: 29899015]

11. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, et al. Effect of a Game-Based Intervention Designed to Enhance Social Incentives to Increase Physical Activity Among Families: The BE FIT Randomized Clinical Trial. JAMA Intern Med 2017 Nov 01;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]

12. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. Am Heart J 2019 Jan;207:66-75 [FREE Full text] [doi: 10.1016/j.ahj.2018.09.002] [Medline: 30392584]

13. Reading MJ, Merrill JA. Converging and diverging needs between patients and providers who are collecting and using patient-generated health data: an integrative review. J Am Med Inform Assoc 2018 Jun 01;25(6):759-771 [FREE Full text] [doi: 10.1093/jamia/ocy006] [Medline: 29471330]

XSL•FO

RenderX

14. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. Am Heart J 2019 Jan;207:66-75 [FREE Full text] [doi: 10.1016/j.ahj.2018.09.002] [Medline: 30392584]

15. Servey JT, Stephens M. Cardiac Rehabilitation: Improving Function and Reducing Risk. Am Fam Physician 2016 Jul 01;94(1):37-43 [FREE Full text] [Medline: 27386722]

16. Weber MB, Narayan KMV. Health Insurance for Diabetes Prevention Confers Health Benefits and Breaks Even on Cost Within 2 Years. Diabetes Care 2019 Sep;42(9):1612-1614. [doi: 10.2337/dci19-0022] [Medline: 31431495]

17. Granja C, Janssen W, Johansen MA. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]

18. Abdolkhani R, Gray K, Borda A, DeSouza R. Patient-generated health data management and quality challenges in remote patient monitoring. JAMIA Open 2019 Dec;2(4):471-478 [FREE Full text] [doi: 10.1093/jamiaopen/ooz036] [Medline: 32025644]

Abbreviations

EHR: electronic health record

XSL•FO
RenderX

<u>Review</u>

Effectiveness of Mobile App-Assisted Self-Care Interventions for Improving Patient Outcomes in Type 2 Diabetes and/or Hypertension: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Kaifeng Liu[1], PhD; Zhenzhen Xie[1], MSc; Calvin Kalun Or[1], PhD

Department of Industrial and Manufacturing Systems Engineering, University of Hong Kong, Hong Kong, Hong Kong

Corresponding Author:
Calvin Kalun Or, PhD
Department of Industrial and Manufacturing Systems Engineering
University of Hong Kong
Room 8-7, 8/f., Haking Wong Building,
University of Hong Kong
Hong Kong
Hong Kong
Phone: 852 39172587
Email: <u>klor@hku.hk</u>

Related Article:

This is a corrected version. See correction statement: <u>http://mhealth.jmir.org/2020/8/e23600/</u>

Abstract

Background: Mobile app-assisted self-care interventions are emerging promising tools to support self-care of patients with chronic diseases such as type 2 diabetes and hypertension. The effectiveness of such interventions requires further exploration for more supporting evidence.

Objective: A systematic review and meta-analysis of randomized controlled trials (RCTs) were conducted to examine the effectiveness of mobile app-assisted self-care interventions developed for type 2 diabetes and/or hypertension in improving patient outcomes.

Methods: We followed the Cochrane Collaboration guidelines and searched MEDLINE, Cochrane Library, EMBASE, and CINAHL Plus for relevant studies published between January 2007 and January 2019. Primary outcomes included changes in hemoglobin A_{1c} (HbA_{1c}) levels, systolic blood pressure (SBP), and diastolic blood pressure (DBP). Changes in other clinical-, behavioral-, knowledge-, and psychosocial-related outcomes were included as secondary outcomes. Primary outcomes and objective secondary outcomes that were reported in at least two trials were meta-analyzed; otherwise, a narrative synthesis was used for data analysis.

Results: A total of 27 trials were identified and analyzed. For primary outcomes, the use of mobile app-assisted self-care interventions was associated with significant reductions in HbA_{1c} levels (standardized mean difference [SMD] –0.44, 95% CI –0.59 to –0.29; $P<.001$), SBP (SMD –0.17, 95% CI –0.31 to –0.03, $P=.02$), and DBP (SMD –0.17, 95% CI –0.30 to –0.03, $P=.02$). Subgroup analyses for primary outcomes showed that several intervention features were supportive of self-management, including blood glucose, blood pressure, and medication monitoring, communication with health care providers, automated feedback, personalized goal setting, reminders, education materials, and data visualization. In addition, 8 objective secondary outcomes were meta-analyzed, which showed that the interventions had significant lowering effects on fasting blood glucose levels and waist circumference. A total of 42 secondary outcomes were narratively synthesized, and mixed results were found.

Conclusions: Mobile app-assisted self-care interventions can be effective tools for managing blood glucose and blood pressure, likely because their use facilitates remote management of health issues and data, provision of personalized self-care recommendations, patient–care provider communication, and decision making. More studies are required to further determine which combinations of intervention features are most effective in improving the control of the diseases. Moreover, evidence

regarding the effects of these interventions on the behavioral, knowledge, and psychosocial outcomes of patients is still scarce, which warrants further examination.

Introduction

Type 2 diabetes mellitus and hypertension are two common, serious medical conditions that can lead to the development of other disabling and life-threatening health problems such as stroke and heart attack. The two diseases are closely interlinked and frequently coexist. Globally, approximately 80% of type 2 diabetic patients have hypertension [1]. US statistics indicate that type 2 diabetes is 2.5 times more prevalent in hypertensive individuals than in normotensive individuals [2]. In Hong Kong, 58% of diabetic patients exhibit increased blood pressure (BP), whereas 56% of hypertensive patients have hyperglycemia [3]. These figures emphasize that the treatment and management of both of these conditions are essential.

Diabetes and hypertension management requires lifelong self-care by patients, which can be demanding and overwhelming because patients are often unskilled or unaware of self-care and also lack the necessary knowledge, tools, and support [4]. Technology is increasingly being used to help address these challenges. In particular, mobile app-assisted interventions that capitalize on smart and networking features are suggested to facilitate patient–care provider communication, information exchange, health literacy, decision making, and peer support, without the constraints of time and geography [5-11], all of which are important for self-care.

However, the effectiveness of mobile app-assisted self-care interventions developed for type 2 diabetes and/or hypertension requires more supporting evidence and thus warrants a systematic review. First, previous reviews on the use of mobile app-assisted self-care interventions for diabetes [12-15] and hypertension [16] have mainly focused on the effects of these interventions on hemoglobin A_{1c} (HbA$_{1c}$) levels or BP and have paid relatively little attention to other variables important for effectiveness evaluation such as behavioral, knowledge, and psychosocial outcomes. Second, several randomized controlled trials (RCTs) [17-19] have recently been conducted to test the effects of such interventions on patient outcomes, and these studies must be reviewed. Third, little is known about the features of such technologies that are effective at improving blood glucose (BG) and BP management. In light of such knowledge gaps, this study systematically reviewed the existing evidence on the effectiveness of mobile app-assisted self-care interventions developed for type 2 diabetes and/or hypertension in improving patient outcomes. In this review, mobile health apps refer to mobile device-based software programs that provide health-related resources and support for the self-care of patients with type 2 diabetes and/or hypertension.

Methods

We followed the Cochrane Collaboration guidelines for conducting this review [20]. Screening of studies for eligibility, data extraction, risk of bias assessment, and assessment of quality of evidence were performed by KL (author) and ZX (author)/MJ (researcher) independently, and any disagreement was resolved through discussion and consensus.

Search Strategy

MEDLINE, Cochrane Library, EMBASE, and CINAHL Plus were searched for relevant studies published between January 2007 and January 2019. According to some reviews [12,14], the majority of mobile apps were released after the launch of the first generation of iPhone in 2007 and the main app stores (IOS and Android Market) in 2008; therefore, we used 2007 as the starting year in our search. The following search terms were used: (phone* or tablet*) and (monitor* or manag* or care or control) and (examin* or evaluat* or assess* or compar*) and (diabetes or diabetic* or hyperten*).

Selection Criteria

Studies were included in the review if they (1) were RCTs, (2) examined the effects of mobile app-assisted self-care interventions relative to those of usual care on patient outcomes, (3) studied type 2 diabetic and/or hypertensive patients, and (4) were published in English-language, peer-reviewed journals. We excluded review articles, case reports, and studies that only provided an abstract.

Study Selection

Two researchers independently read the titles and abstracts of the citations identified in the literature search, excluded clearly irrelevant studies, and reviewed the full text of the remaining articles for inclusion. The reference lists of the included studies and relevant review papers were also examined to identify missed articles.

Data Extraction

Two researchers independently extracted the following study characteristics from each included trial: authors, publication year, study location, disease studied, sample size, HbA$_{1c}$/BP eligibility, mean age of participants, sex ratio, trial length, features of the interventions, and changes in patient outcomes from baseline to the end of the trial in both intervention and control groups. For an RCT with multiple intervention groups relevant to this review, we split the "shared" control group into two or more groups (with smaller sample size) to apply two or more pair-wise comparisons in the meta-analysis [20]. For 2-arm cross-over RCTs, data from only the first period were extracted and analyzed.

Outcome Measures

Primary outcomes included changes in HbA$_{1c}$ levels, systolic BP (SBP), and diastolic BP (DBP) at the end of the trial. Changes in other outcomes, including clinical (eg, fasting BG [FBG]), behavioral (eg, medication adherence), knowledge (eg, diabetes knowledge), and psychosocial (eg, distress) outcomes, were included as secondary outcomes.

Risk of Bias Assessment

Following the Cochrane Collaboration's tool for risk of bias assessment [20], two researchers independently assessed the risk of bias of included trials for seven aspects: sequence generation, allocation concealment, blinding of participants and health care providers (HCPs), blinding of outcome assessors, incomplete outcome data, selective outcome, and other sources of bias. Other sources of bias included significantly different baseline characteristics between groups, presence of co-interventions, unacceptable compliance with the intervention, and different outcome assessment timings.

Data Analysis

Primary outcomes and objective secondary outcomes were meta-analyzed when they were reported in at least two trials. We pooled data across trials using random effects models and calculated the standardized mean difference (SMD) for each outcome. The I^2 statistic was calculated to measure the percentage of variation across trials due to heterogeneity, with values of 25%, 50%, and 75% indicating low, moderate, and high levels of heterogeneity, respectively [21]. The possibility of publication bias was assessed using the Egger test [22]. The meta-analysis was performed using Comprehensive Meta-Analysis version 2 (Biostat Inc) statistical software. We narratively synthesized outcomes that were reported in only one trial or were self-reported (because of the differences in the scales used across trials). In the synthesis, for each outcome we counted the numbers of trials reporting significant positive effects, no significant effects, and significant negative effects. Subgroup analyses were performed for primary outcomes when they were reported in at least two trials in each subgroup. These analyses were stratified by (1) disease type to examine the effects of the interventions in different disease populations and (2) intervention feature to identify which features are effective in glycemic and BP control.

Assessment of Quality of Evidence

The quality of evidence for the primary and objective secondary outcomes was assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system [23]. For each of the outcomes, the quality of evidence was downgraded from high by one level for each serious issue identified in the domains of risk of bias, imprecision, indirectness, inconsistency, and publication bias.

Results

Study Selection

The study selection process (see Figure 1) identified 24 eligible publications [17-19,24-44]. Of them, the study by Holmen et al [30] had two intervention groups and the study by Quinn et al [29] had three intervention groups, all of which were relevant to this review; therefore, the control groups in these studies were split into more groups accordingly. This rendered a total of 27 independent trials for inclusion in data analysis.

Figure 1. Flow diagram of the study selection process.
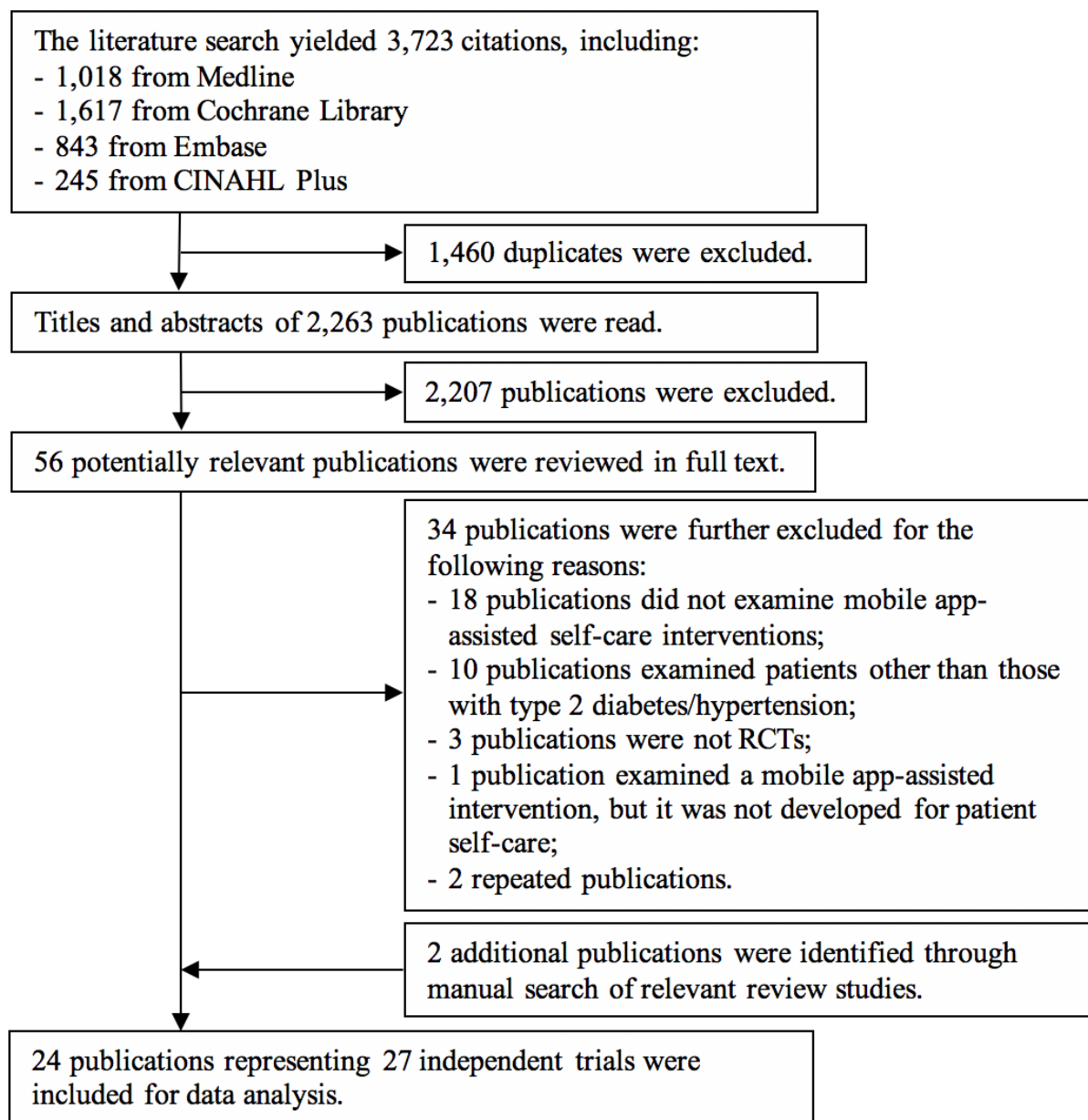
Trial Characteristics

Table 1 summarizes the characteristics of the 27 trials, and Table 2 presents the details of the trials. Fourteen features were identified in the interventions examined in the trials (see Multimedia Appendix 1 for a summary and Multimedia Appendix 2 for details). According to the taxonomies of previous studies [12,45], features were grouped into five categories: logging (ie, monitoring of BG, BP, medication, body weight, diet, physical activity, and mood), personalized feedback (ie, automated feedback, medication adjustment aid, personalized goal setting, and reminders), communication with HCPs, education materials, and data visualization.

Table 1. Summary of the characteristics of the 27 trials.

Characteristics	Value
Year of publication, n (%)	
2007-2009	2 (7)
2010-2012	4 (15)
2013-2015	10 (37)
2016-2019	11 (41)
Study location, n (%)	
North America	13 (48)
Europe	7 (26)
Asia	5 (19)
Africa	2 (7)
Disease studied, n (%)	
Type 2 diabetes	19 (70)
Hypertension	6 (22)
Type 2 diabetes and/or hypertension	1 (4)
Coexisting type 2 diabetes and hypertension	1 (4)
Sample size, median (range)	75 (14-250)
Mean age of participants in years, mean (range)	57.3 (48.4-69.5)
Proportion of male participants in %, median (range)	54 (28-76)
Trial length in months, median (range)	6 (2-12)

Table 2. Details of the 27 trials.

Trial, publication year, study location	Trial length	Sample	HbA$_{1c}$[a]/BP[b] eligibility	Intervention	Comparison treatment
Type 2 diabetes					
Anzaldo-Campos et al, 2016, Mexico	10 months	IG[c]: n=102; CG[d]: n=100; mean age 52.0 years; male 38%; diabetes duration 8.3 years	HbA$_{1c}$≥ 8%	A mobile app to facilitate self-monitoring of health-related data (eg, BG[e] and diet) and support from clinicians, nurses, and peer educators for care management	Usual care and the provision of educational classes and health evaluation in monthly medical group visits
Bender et al, 2017, US	6 months	IG: n=22; CG: n=23; mean age 57.6 years; male 38%; diabetes duration not reported	No limit for HbA$_{1c}$	A mobile app for behavior tracking, a Fitbit for steps monitoring, and social media for social support and education	Usual care and a Fitbit only for daily wear
Greenwood et al, 2015, US	6 months	IG: n=45; CG: n=45; mean age 55.7 years; male 53%; diabetes duration 8.2 years	HbA$_{1c}$: 7.5%-10.9%	A tablet-based app and a portal to support patients' BG monitoring and diabetes education and enable certified diabetes educators' access to patient data for telemonitoring	Usual care, booklets and referrals for diabetes education, and evaluation of patient self-reported glucose data by certified diabetes educators
Hansen et al, 2017, Denmark	8 months	IG: n=83; CG: n=82; mean age 58 years; male 64%; diabetes duration 12.3 years	HbA$_{1c}$> 7.5%	A tablet-based app to enable reporting of health-related data (eg, BG and BP) and monthly communication with HCPs[f] via video-conferencing	Usual care
Holmen et al (1), 2014, Norway	12 months	IG: n=51; CG: n=25; mean age 57.7 years; male 64%; diabetes duration 10.6 years	HbA$_{1c}$≥ 7.1%	A mobile phone-based system to enable vital sign monitoring, goal management, and motivational feedback	Usual care
Holmen et al (2), 2014, Norway	12 months	IG: n=50; CG: n=25; mean age 56.9 years; male 53%; diabetes duration 9.5 years	HbA$_{1c}$≥ 7.1%	A mobile phone-based system (to enable vital sign monitoring, goal management, and motivational feedback) and health counseling delivered by diabetes specialist nurses	Usual care
Hsu et al, 2016, US	12 weeks	IG: n=20; CG: n=20; mean age 53.6 years; male sex not reported; diabetes duration 9.3 years	HbA$_{1c}$: 9%-14%	A cloud-based diabetes management app supporting BG self-monitoring, insulin initiation/titration, shared decision making, and communication	Usual care with interim face-to-face visits and telephone/fax communication with educators, physicians, and/or nurses
Karhula et al, 2015, Finland	12 months	IG: n=180; CG: n=70; mean age 66.3 years; male 56%; diabetes duration not reported	HbA$_{1c}$> 6.5%	A mobile app for self-monitoring of health parameters (eg, BG and BP) and remote health coaching	Usual care
Kleinman et al, 2017, India	6 months	IG: n=44; CG: n=46; mean age 48.4 years; male 70%; diabetes duration 9.2 years	HbA$_{1c}$: 7.5%-12.5%	A smartphone app for patients and a web portal plus an app for HCPs for receiving reminders, data visualization, and providing care support to enhance self-care and collaborative care decisions	Usual care

XSL•FO
RenderX

Trial, publication year, study location	Trial length	Sample	HbA$_{1c}$[a]/BP[b] eligibility	Intervention	Comparison treatment
Nagrebetsky et al, 2013, UK	6 months	IG: n=7; CG: n=7; mean age 58 years; male 71%; diabetes duration 2.6 years	HbA$_{1c}$: 8%-11%	A mobile phone-based tele-health platform (for self-monitoring of BG and self-titration of oral glucose-lowering medication) and monthly telephone calls (for lifestyle monitoring and change)	Usual care and monthly telephone calls for lifestyle monitoring and change
Orsama et al, 2013, Finland	10 months	IG: n=24; CG: n=24; mean age 61.9 years; male 54%; diabetes duration not reported	HbA$_{1c}$: 6.5%-11%; SBP[g] >140 mm Hg or DBP[h] >90 mm Hg	A diabetes lifestyle and self-care promotion program based on a mobile app to allow patients to report their conditions and receive system-generated feedback on health behaviors	Usual care, diabetes education, and HCP counseling
Quinn et al, 2008, US	3 months	IG: n=13; CG: n=13; mean age 51.0 years; male 35%; diabetes duration 9.3 years	HbA$_{1c}$ ≥ 7.5%	A mobile phone-based software to provide real-time feedback on patient BG levels, display medication instructions, incorporate hypo- and hyperglycemia treatment algorithms, and request data for diabetes management	Usual care and instructions to patients about reporting BG levels to HCPs via phone calls or fax once every 2 weeks
Quinn et al, 2011, US (1)	12 months	IG: n=23; CG: n=19; mean age 53 years; male 52%; diabetes duration 8.3 years	HbA$_{1c}$ ≥ 7.5%	A mobile app and patient care provider web portal to support patient self-monitoring and enable HCPs to receive health data shared by patients	Usual care
Quinn et al, 2011, US (2)	12 months	IG: n=22; CG: n=19; mean age 53.5 years; male 46%; diabetes duration 7.8 years	HbA$_{1c}$ ≥ 7.5%	A mobile app and patient care provider web portal to support patient self-monitoring and allow HCPs to access unanalyzed patient data	Usual care
Quinn et al, 2011, US (3)	12 months	IG: n=62; CG: n=18; mean age 52.3 years; male 50%; diabetes duration 8.5 years	HbA$_{1c}$ ≥ 7.5%	A mobile app and patient care provider web portal to support patient self-monitoring and allow HCPs to access analyzed patient data and evidence-based care guidelines	Usual care
Sun et al, 2019, China	6 months	IG: n=44; CG: n=47; mean age 68.0 years; male 41%; diabetes duration 11.4 years	HbA$_{1c}$: 7.0%-10.0%	A mobile app for self-monitoring of BG, diet, and physical activity; sharing of measurement records; and receiving HCP-provided care recommendations	Usual care
Takenga et al, 2014, Congo	2 months	IG: n=20; CG: n=20; mean age 53.3 years; male 73%; diabetes duration not reported	No limit for HbA$_{1c}$	A mobile system to support patients' tracking of health conditions (eg, BG, BP, and body weight) and communication with HCPs	Usual care
Waki et al, 2014, Japan	3 months	IG: n=27; CG: n=27; mean age 57.3 years; male 76%; diabetes duration 9.1 years	No limit for HbA$_{1c}$	A smartphone-based system for self-monitoring of health conditions (eg, BG, BP, and diet), communication with HCPs, and receiving system's auto-generated feedback	Usual care

XSL•FO
RenderX

Trial, publication year, study location	Trial length	Sample	HbA$_{1c}$[a]/BP[b] eligibility	Intervention	Comparison treatment
Wayne et al, 2015, Canada	6 months	IG: n=67; CG: n=64; mean age 53.2 years; male 28%; diabetes duration not reported	HbA$_{1c}$ ≥ 7.3%	A mobile phone-supported health coach program allowing patients to track their conditions (eg, BG, diet, physical activity, and mood) and communicate with HCPs	Usual care, exercise education, and health coach support in goal setting and progress monitoring through in-person meetings/telephones
Hypertension					
Kim et al, 2016, US	6 months	IG: n=52; CG: n=43; mean age 57.6 years; male 32%; hypertension duration not reported	No limit for BP	A mobile app (equipped with a BP monitoring device, electronic reminders, and a web-based disease management program for patient self-monitoring) and a reach-out program (delivered by nursing staff for education about medication, disease prevent, and chronic disease management)	Usual care and a reach-out program of the same type used in the IG
Lakshminarayan et al, 2018, US	90 days	IG: n=34; CG: n=22; mean age 65.0 years; male 68%; hypertension duration not reported	No limit for BP	A smartphone app supporting BP self-monitoring, nurse-delivered education, and HCP-provided feedback	Usual care and education on hypertension management
Logan et al, 2012, Canada	12 months	IG: n=55; CG: n=55; mean age 62.9 years; male 56%; hypertension duration not reported	SBP ≥130 mm Hg	A smartphone app (for BP telemonitoring and self-care) and a booklet (with information about self-management, treatments, and therapy goals)	Usual care and a booklet of the same type used in the IG
Márquez Contreras et al, 2019, Spain	12 months	IG: n=73; CG: n=75; mean age 57.5 years; male sex 48%; hypertension duration not reported	No limit for BP	A smartphone app to promote education about hypertension and provide patients with reminders of appointments and medication	Usual care
Moore et al, 2014, US	12 weeks	IG: n=20; CG: n=22; mean age 50.0 years; male 60%; hypertension duration not reported	BP: 140/90-180/120 mm Hg	A tablet-based app, virtual visits, instant messaging, and a nurse health coach to facilitate self-monitoring of BP and medication intake; visualization of information on actions, outcomes, and medication adjustment; and discussion about care management and goal settings	Usual care together with office visits, phone calls, and emails with HCPs for discussing care management, goal settings, and medication adjustment
Sarfo et al, 2019, Ghana	9 months	IG: n=30; CG: n=30; mean age 55 years; male 65%; hypertension duration not reported	SBP ≥140 mm Hg	A smartphone app for monitoring and reporting of BP and medication intake, provision of motivational text messages generated based on patients' medication adherence, and sharing of patients' health reports with clinicians	Usual care and text messages about healthy lifestyle management and clinicians' monthly review of patients' BP
Type 2 diabetes and/or hypertension					

Trial, publication year, study location	Trial length	Sample	HbA$_{1c}$[a]/BP[b] eligibility	Intervention	Comparison treatment
Or and Tao, 2016, Hong Kong SAR, China	3 months	IG: n=33; CG: n=30; mean age 69.5 years; male 32%; diabetes duration 12.5 years; hypertension duration 10.2 years	No limit for HbA$_{1c}$ and BP	A tablet-based self-monitoring app allowing automated recording and monitoring of BG and BP values and providing educational materials and decision aids	Usual care
Type 2 diabetes and hypertension					
Yoo et al, 2009, Korea	3 months	IG: n=57; CG: n=54; mean age 58.2 years; male 59%; diabetes duration 6.6 years; hypertension duration 3.7 years	HbA$_{1c}$: 6.5%-10%; BP >130/80 mm Hg	An internet-enabled, cellphone-based system coupled with a BG measuring device, an automatic BP monitor, a body weight scale, and a database providing reminders, health recommendations, and data sharing for self-care	Usual care

[a]HbA$_{1c}$: hemoglobin A$_{1c}$.

[b]BP: blood pressure.

[c]IG: intervention group.

[d]CG: control group.

[e]BG: blood glucose.

[f]HCP: health care provider.

[g]SBP: systolic blood pressure.

[h]DBP: diastolic blood pressure.

Risk of Bias Assessment

Figures 2 and 3 present the results of the risk of bias assessment.

Figure 2. Risk of bias of the 27 trials.
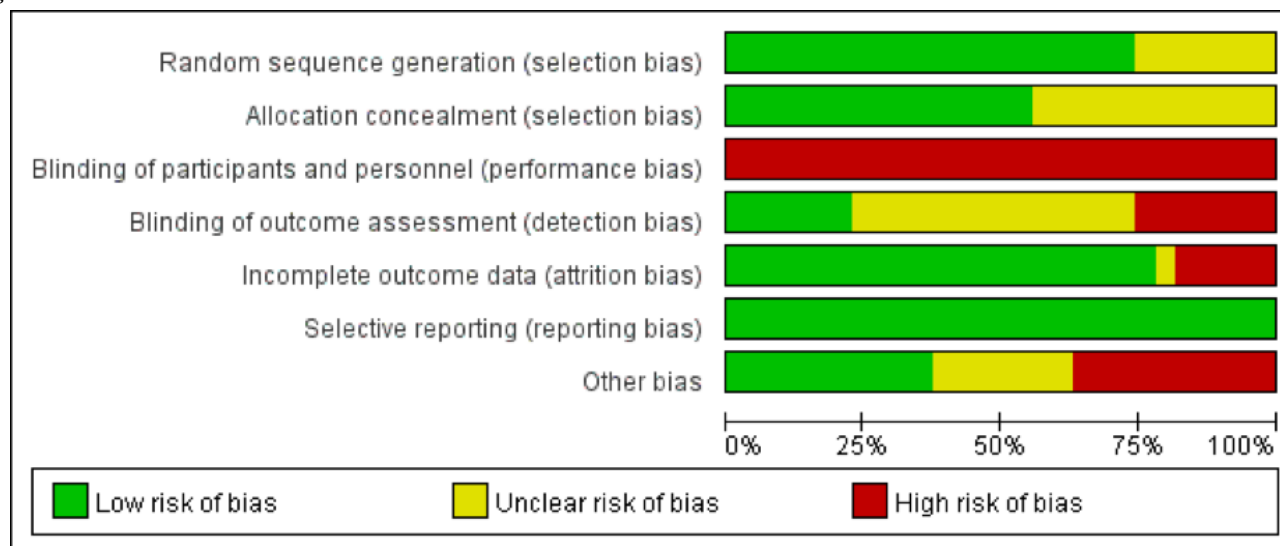
Figure 3. Risk of bias for each trial.
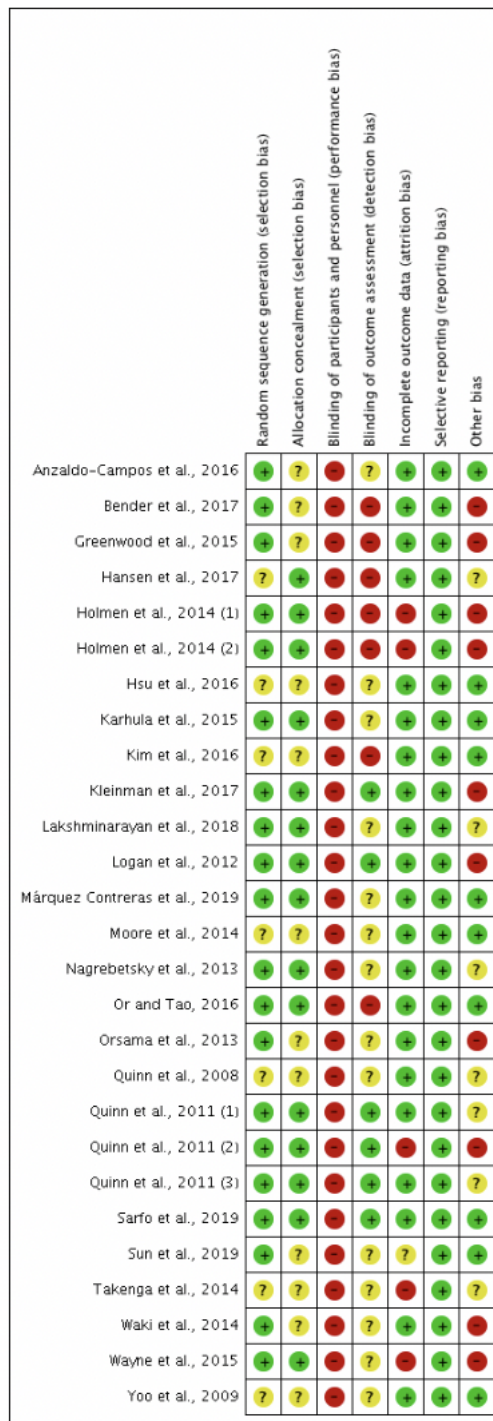
Meta-Analysis of the Effects on Primary Outcomes

The meta-analysis results showed that mobile app-assisted self-care interventions were associated with significant reductions in HbA$_{1c}$ levels (SMD −0.44, 95% CI −0.59 to −0.29, P<.001, corresponding to an absolute mean difference [MD] −0.49%, 95% CI −0.68 to −0.30), SBP (SMD −0.17, 95% CI −0.31 to −0.03, P=.02, corresponding to an absolute MD of −2.32 mm Hg, 95% CI −4.35 to −0.30), and DBP (SMD −0.17, 95% CI −0.30 to −0.03, P=.02, corresponding to an absolute MD of −1.53 mm Hg, 95% CI −2.78 to −0.28; Table 3). The GRADE revealed that the quality of evidence for HbA$_{1c}$ levels, SBP, and DBP was low, moderate, and moderate, respectively (Table 3). Figure 4 presents the forest plots for the primary outcomes.

Table 3. Results of meta-analysis and Grading of Recommendations Assessment, Development and Evaluation assessments for hemoglobin A1c levels, systolic blood pressure, and diastolic blood pressure.

Outcomes	Trials included	Sample size	SMD[a] (95% CI)	P value	I^2	Egger test		Quality of evidence (GRADE)[b]
						t value	P value	
HbA$_{1c}$[c] levels	21	1671	−0.44 (−0.59 to −0.29)	<.001	50	1.15	.26	▣ Low[d,e]
SBP[f]	16	1433	−0.17 (−0.31 to −0.03)	.02	41	0.52	.61	▣ Moderate[d]
DBP[g]	14	1292	−0.17 (−0.30 to −0.03)	.02	25	0.09	.93	▣ Moderate[d]

[a]SMD: standardized mean difference.

[b]GRADE: Grading of Recommendations Assessment, Development and Evaluation.

[c]HbA$_{1c}$: hemoglobin A$_{1c}$.

[d]Downgraded by one level for indirectness (surrogate outcome).

[e]Downgraded by one level for inconsistency (moderate heterogeneity level, $I^2 = 50\%$).

[f]SBP: systolic blood pressure.

[g]DBP: diastolic blood pressure.

XSL•FO
RenderX

Figure 4. Forest plots for hemoglobin A1c (top), systolic blood pressure (middle), and diastolic blood pressure (bottom).

Trials	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Anzaldo-campos et al., 2016	-0.561	0.143	0.021	-0.842	-0.280	-3.910	0.000
Bender et al., 2017	-0.458	0.302	0.091	-1.050	0.134	-1.515	0.130
Greenwood et al., 2015	-0.607	0.216	0.046	-1.030	-0.184	-2.815	0.005
Hansen et al., 2017	-0.301	0.157	0.025	-0.608	0.006	-1.922	0.055
Holmen et al., 2014 (1)	-0.150	0.271	0.073	-0.681	0.381	-0.553	0.580
Holmen et al., 2014 (2)	0.008	0.274	0.075	-0.528	0.545	0.031	0.976
Hsu et al., 2016	-0.679	0.325	0.106	-1.316	-0.041	-2.087	0.037
Karhula et al., 2015	-0.176	0.151	0.023	-0.473	0.120	-1.167	0.243
Kleinman et al., 2017	-0.508	0.214	0.046	-0.928	-0.088	-2.370	0.018
Nagrebetsky et al., 2013	-0.138	0.535	0.286	-1.187	0.911	-0.258	0.797
Or and Tao, 2016	0.000	0.403	0.162	-0.790	0.790	0.000	1.000
Orsama et al., 2013	-0.437	0.295	0.087	-1.016	0.141	-1.481	0.139
Quinn et al., 2008	-0.852	0.410	0.168	-1.655	-0.049	-2.079	0.038
Quinn et al., 2011 (1)	-0.723	0.320	0.102	-1.350	-0.096	-2.262	0.024
Quinn et al., 2011 (2)	-0.411	0.316	0.100	-1.031	0.209	-1.298	0.194
Quinn et al., 2011 (3)	-0.830	0.276	0.076	-1.370	-0.289	-3.010	0.003
Sun et al., 2019	-0.444	0.212	0.045	-0.860	-0.028	-2.091	0.036
Takenga et al., 2014	-1.214	0.344	0.118	-1.888	-0.539	-3.527	0.000
Waki et al., 2014	-0.685	0.280	0.078	-1.234	-0.136	-2.445	0.014
Wayne et al., 2015	0.270	0.176	0.031	-0.074	0.614	1.538	0.124
Yoo et al., 2009	-0.778	0.197	0.039	-1.164	-0.392	-3.949	0.000
	-0.439	0.076	0.006	-0.589	-0.290	-5.762	0.000

Trials	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Anzaldo-campos et al., 2016	-0.207	0.142	0.020	-0.485	0.070	-1.463	0.143
Karhula et al., 2015	-0.116	0.153	0.023	-0.416	0.184	-0.758	0.449
Kim et al., 2016	0.181	0.207	0.043	-0.223	0.586	0.878	0.380
Lakshminarayan 2018	-0.242	0.286	0.082	-0.802	0.318	-0.846	0.397
Logan et al., 2012	-0.518	0.199	0.039	-0.908	-0.129	-2.612	0.009
Márquez Contreras et al., 2019	-0.210	0.165	0.027	-0.533	0.113	-1.274	0.203
Moore et al., 2014	-0.858	0.323	0.104	-1.491	-0.225	-2.658	0.008
Or and Tao, 2016	-0.445	0.260	0.068	-0.955	0.064	-1.713	0.087
Orsama et al., 2013	-0.539	0.304	0.093	-1.136	0.057	-1.774	0.076
Quinn et al., 2011 (1)	0.137	0.310	0.096	-0.472	0.745	0.441	0.659
Quinn et al., 2011 (2)	0.000	0.313	0.098	-0.614	0.614	0.000	1.000
Quinn et al., 2011 (3)	-0.245	0.268	0.072	-0.771	0.282	-0.911	0.362
Sarfo et al., 2019	-0.063	0.258	0.067	-0.570	0.443	-0.245	0.806
Sun et al., 2019	0.512	0.213	0.045	0.094	0.930	2.401	0.016
Waki et al., 2014	-0.231	0.273	0.075	-0.766	0.304	-0.846	0.398
Yoo et al., 2009	-0.179	0.190	0.036	-0.552	0.194	-0.940	0.347
	-0.169	0.073	0.005	-0.313	-0.025	-2.307	0.021

Trials	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Anzaldo-campos et al., 2016	-0.182	0.142	0.020	-0.460	0.096	-1.285	0.199
Karhula et al., 2015	-0.048	0.153	0.023	-0.348	0.252	-0.314	0.754
Kim et al., 2016	0.341	0.208	0.043	-0.066	0.748	1.643	0.100
Logan et al., 2012	-0.379	0.197	0.039	-0.765	0.007	-1.923	0.054
Márquez Contreras et al., 2019	-0.336	0.166	0.027	-0.661	-0.012	-2.031	0.042
Moore et al., 2014	-0.612	0.316	0.100	-1.231	0.008	-1.936	0.053
Or and Tao, 2016	-0.375	0.259	0.067	-0.883	0.133	-1.447	0.148
Orsama et al., 2013	-0.440	0.302	0.091	-1.033	0.152	-1.456	0.145
Quinn et al., 2011 (1)	0.114	0.310	0.096	-0.494	0.722	0.367	0.713
Quinn et al., 2011 (2)	-0.350	0.316	0.100	-0.969	0.268	-1.110	0.267
Quinn et al., 2011 (3)	-0.185	0.268	0.072	-0.710	0.341	-0.690	0.490
Sarfo et al., 2019	0.100	0.258	0.067	-0.406	0.607	0.388	0.698
Waki et al., 2014	0.257	0.273	0.075	-0.279	0.792	0.939	0.348
Yoo et al., 2009	-0.310	0.191	0.036	-0.684	0.064	-1.622	0.105
	-0.165	0.069	0.005	-0.300	-0.030	-2.399	0.016

Subgroup Analysis for Primary Outcomes by Disease Type

The analysis of the HbA_{1c} outcome by disease type was not applicable because the outcome was only examined in diabetic patients and not hypertensive patients in the 27 included trials. The results of the subgroup analysis for SBP indicated that mobile app-assisted interventions led to significant reductions in SBP in hypertensive patients (SMD −0.28, 95% CI −0.51 to −0.04, P=.02, corresponding to an absolute MD of −4.20 mm Hg, 95% CI −7.47 to −0.93), but not in diabetic patients (SMD −0.08, 95% CI −0.29 to 0.13, P=.46, corresponding to an absolute MD of −0.82 mm Hg, 95% CI −3.51 to 1.87). No significant change in DBP was observed in either hypertensive patients (SMD −0.20, 95% CI −0.47 to 0.08, P=.17, corresponding to an absolute MD of −1.94 mm Hg, 95% CI −4.34 to 0.47) or diabetic patients (SMD −0.12, 95% CI −0.28

to 0.04, P=.16, corresponding to an absolute MD of −0.62 mm Hg, 95% CI −1.99 to 0.75).

Subgroup Analysis for Primary Outcomes by Intervention Feature

Table 4 presents the results of subgroup analysis by intervention feature in relation to reductions in HbA_{1c} levels, SBP, and DBP (details see Multimedia Appendix 3, 4, and 5). The self-care interventions with the medication monitoring feature led to a significantly greater reduction in HbA_{1c} levels than those without this feature. Interventions that allowed patient-HCP communication were associated with significant reductions in HbA_{1c} while the reduction was not significant for interventions that did not have this feature, although the difference in the reduction between the subgroups (presence of the feature vs absence of the feature) was not statistically significant. As for the personalized goal-setting feature, significant reductions in HbA_{1c} levels were observed in both subgroups, but the improvement was greater when the interventions did not have this feature. For each of the other features, reductions in HbA_{1c}

levels were found in both subgroups, but the difference was not significant between the subgroups.

The presence of BP monitoring, automated feedback, personalized goal setting, reminders, education materials, and data visualization features yielded significant reductions in SBP while the reductions were not significant for interventions that did not have these features, although the differences between the subgroups were not statistically significant. The presence of diet- and physical activity–monitoring features was not associated with reductions in SBP. For other features, changes in SBP were found to be similar between the subgroups.

Further, the presence of BG monitoring, automated feedback, and personalized goal-setting features was associated with reductions in DBP while the reductions were not significant for interventions that did not have these features, although the differences between the subgroups were not statistically significant. Diet monitoring, body weight monitoring, and data visualization were not associated with reductions in DBP. For other features, changes in DBP were found to be similar between the subgroups.

Table 4. Results of subgroup analysis by intervention feature in relation to reductions in hemoglobin A1c levels, systolic blood pressure, and diastolic blood pressure.

Features	HbA_{1c}[a] reduction	SBP[b] reduction	DBP[c] reduction
Logging			
BG[d]	—[e]	Δ[f]	•[g]
BP[h]	Δ	•	Δ
Body weight	Δ	Δ	×[i]
Medication	•	Δ	Δ
Diet	Δ	×	×
Physical activity	Δ	×	Δ
Mood	—	—	—
Personalized feedback			
Automated feedback	Δ	•	•
Medication adjustment aid	Δ	—	—
Personalized goal setting	×	•	•
Reminders	Δ	•	Δ
Communication with HCP[j]	•	—	—
Education materials	Δ	•	Δ
Data visualization	Δ	•	×

[a]HbA_{1c}: hemoglobin A_{1c}.

[b]SBP: systolic blood pressure.

[c]DBP: diastolic blood pressure.

[d]BG: blood glucose.

[e]—: Subgroup analysis was not performed for the feature because there were fewer than two trials in one of the subgroups.

[f]Δ: Similar changes were found between the two subgroups (presence of the feature vs absence of the feature).

[g]•: Presence of the feature was related to a more favorable effect on the outcome.

[h]BP: blood pressure.

[i]×: Absence of the feature was related to a more favorable effect on the outcome.

[j]HCP: health care provider.

XSL•FO
RenderX

Meta-Analysis of the Effects on Objective Secondary Outcomes

A total of 8 objective secondary outcomes were meta-analyzed (Table 5). Mobile app-assisted self-care interventions had significant lowering effects on FBG (SMD –0.29, 95% CI –0.49 to –0.10, P=.004, corresponding to an absolute MD of –0.66 mmol/L, 95% CI –1.06 to –0.26) and waist circumference (SMD –0.23, 95% CI –0.43 to –0.04, P=.02, corresponding to an absolute MD of –1.62 cm, 95% CI –2.84 to –0.40), but not on body weight, BMI, total cholesterol, low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, and triglycerides.

Table 5. Results of meta-analysis and Grading of Recommendations Assessment, Development and Evaluation assessments for objective secondary outcomes.

Outcomes	Trials included	Sample size	SMD[a] (95% CI)	P value	I^2	Egger test		Quality of evidence (GRADE[b])
						t value	P value	
FBG[c]	6	416	–0.29 (–0.49 to –0.10)	.004	2	2.27	.09	☒ Moderate[d]
Waist circumference	4	433	–0.23 (–0.43 to –0.04)	.02	0	0.60	.61	☒ Moderate[d]
Body weight	9	682	–0.09 (–0.24 to 0.07)	.97	0	0.02	.98	☒ Moderate[d]
BMI	6	575	–0.06 (–0.23 to 0.12)	.53	12	3.36	.03	☒ Low[d,e]
Total cholesterol	7	777	–0.18 (–0.37 to 0.02)	.07	35	0.23	.83	☒ Moderate[d]
LDL[f] cholesterol	7	734	–0.08 (–0.23 to 0.07)	.29	0	0.06	.95	☒ Moderate[d]
HDL[g] cholesterol	7	743	–0.10 (–0.28 to 0.07)	.24	18	1.26	.26	☒ Moderate[d]
Triglycerides	7	720	–0.13 (–0.29 to 0.02)	.09	0	0.21	.84	☒ Moderate[d]

[a]SMD: standardized mean difference.

[b]GRADE: Grading of Recommendations Assessment, Development and Evaluation.

[c]FBG: fasting blood glucose.

[d]Downgraded by one level for indirectness (surrogate outcome).

[e]Downgraded by one level for publication bias.

[f]LDL: low-density lipoprotein.

[g]HDL: high-density lipoprotein.

Narrative Synthesis of Intervention Effects

A total of 42 secondary outcomes were narratively synthesized (Table 6). The results were mixed (ie, some of the outcomes favored intervention and some other outcomes did not or for an outcome, different trials showed different results).

Table 6. Narrative synthesis results of the effects of mobile app-assisted self-care interventions.

Outcomes	Number of trials		
	Favoring intervention[a]	Showing no significant difference between intervention and control[b]	Favoring control[c]
Clinical outcomes			
Objectively measured			
Postprandial BG[d]	1 [19]		
Right brachial-ankle pulse wave velocity		1 [40]	
Left brachial-ankle pulse wave velocity		1 [40]	
Adiponectin	1 [40]		
High-sensitivity C-reactive protein		1 [40]	
Interleukin-6		1 [40]	
Homeostatic model assessment of insulin resistance		1 [40]	
Waist/hip ratio		1 [44]	
Creatinine		1 [44]	
Medication dose			1 [35]
Insulin dose		1 [34]	
Self-reported			
Quality of life		6 [24,28,30,32,44]	
Diabetes symptoms		3 [29]	
Behavioral outcomes (self-reported)			
General health-related			
Lifestyle-/health-related activity		4 [24,30,33]	
Specific disease-related			
Adherence to medication	1 [33]	5 [17,25,27,39,42]	
Adherence to physical activities	1 [43]	6 [25,27,30,39,41]	
Adherence to healthy diet	1 [39]	4 [25,30,41]	
Frequency of carbohydrate spacing	1 [25]		
Frequency of smoking		1 [27]	
Frequency of drinking		1 [27]	
Frequency of communicating with physicians		1 [33]	
Adherence to BG monitoring	2 [25,33]		
Adherence to foot care	1 [25]		
Knowledge outcomes (self-reported)			
Diabetes knowledge	1 [24]	3 [25,33,37]	
Hypertension knowledge		3 [17,35,37]	
Psychosocial outcomes (self-reported)			
Satisfaction			
Satisfaction with diabetes treatment	1 [31]	1 [33]	
Satisfaction with life		1 [28]	
Self-efficacy			
Ability to interact with health organizations and HCPs[e]		2 [30]	
Ability to monitor the conditions and having insights into living with the conditions		2 [30]	

Outcomes	Number of trials		
	Favoring intervention[a]	Showing no significant difference between intervention and control[b]	Favoring control[c]
Self-efficacy for medication taking/coping with diseases		4 [17,24,25,33]	
Emotion			
Emotional well-being		2 [30]	
Positive emotion		1 [28]	
Negative emotion		1 [28]	
Distress		4 [29,33]	
Depression		7 [24,28-30]	1 [34]
Anxiety		2 [28,34]	
Perceived behavioral control			
Comfort with self-monitoring		1 [34]	
Self-autonomous regulation		1 [17]	
Determination about not allowing illnesses to control life		2 [30]	
Positive and active engagement in life		2 [30]	
Feeling of having the skills to manage disease	1 [30]	1 [30]	
Feeling of having social support		2 [30]	

[a]Significant improvement in the outcome at the end of the trial in the intervention group compared with the control group.

[b]No significant difference in the outcome at the end of the trial between the intervention and control groups.

[c]Significant deterioration in the outcome at the end of the trial in the intervention group compared with the control group.

[d]BG: blood glucose.

[e]HCP: health care provider.

Discussion

Principal Findings

This systematic review identified 27 trials that examined the effectiveness of mobile app-assisted self-care interventions developed for type 2 diabetes and/or hypertension.

Overall, our review showed that the use of mobile app-assisted self-care interventions led to significant reductions in HbA_{1c} levels, SBP, and DBP—the fundamental clinical parameters in diabetic and hypertensive patients. For HbA_{1c} levels, we observed an SMD of −0.44 and an absolute MD of −0.49%. The effect size was clinically meaningful and similar to that reported in previous reviews that examined other similar types of health technology (ie, SMD −0.30 to −0.40 [15,46], absolute MD −0.40 to −0.49% [12-15,47]). As for BP, overall, the use of mobile app-assisted self-care interventions led to significant reductions in SBP (SMD −0.17, absolute MD −2.32 mm Hg) and DBP (SMD −0.17, absolute MD −1.53 mm Hg).

The subgroup analysis of BP by disease type showed that among hypertensive patients, the effect size for SBP (SMD −0.28, absolute MD −4.20 mm Hg) could be regarded as clinically important and was similar to that found in previous reviews that studied hypertensive patients (absolute MD −3.74 to −4.71 mm Hg) [48-50]. However, diabetic patients did not show significant reductions in SBP, consistent with previous reviews that examined changes in SBP among diabetic patients [15,47]. This observation could be explained by the reason that the diabetic patients examined might not have severe hypertension; therefore, room for BP reduction in those patients was relatively low. Significant reductions in DBP were not observed in either hypertensive patients or diabetic patients.

All of the reviewed interventions had more than one feature, and our subgroup analysis revealed that the effects of the features on patient outcomes varied, as follows. The presence of medication-, BG-, and BP-monitoring features were favorable in reducing HbA_{1c} levels, SBP, and DBP. Such result could be because patients already had a belief that the behaviors of monitoring of medication, BG, and BP were more immediately relevant to the control of the diseases, and, thereby, with the support of the features, patients' engagement in the behaviors was further developed. Also, because the features could enable the tracking and organization of the health parameters in a more structured and systematic manner [16,51], patients could be more likely to be confident in their self-care and achieve improved outcomes [52]. For diet-, physical activity-, and body weight-monitoring features, their presence yielded limited efficacy. This may be due to the reason that patients might perceive the behaviors of dietary, exercise, and body weight control less directly relevant for diabetes and hypertension control, so patients' use of the features or their engagement in the behaviors could be weak. Or even though the behaviors were considered important, it might not be easy for patients to engage in them, especially long term [53]. Education or motivational strategies may be necessary for increasing patients' awareness about importance of diet, physical activity, and body

XSL•FO
RenderX

weight control in chronic disease management. Features that enabled automated feedback and communication with HCPs were effective in improving patient outcomes. This finding could probably be explained by the fact that providing personalized feedback and suggestions based on patient health data and conditions could help patients interpret changes in their vital signs and inform them about how to deal with different situations related to the variability in their vital signs. This was especially true for patients who had low health literacy and were unable to make good use of health information. The presence of the personalized goal setting feature was favorable in reducing BP because setting specific, realistic, and timely goals could make patients more motivated to engage in planned and targeted disease management. However, this observation was not consistently reported for HbA_{1c} levels. Further evaluation is required to clarify the situations under which goal setting has a positive effect and the manner in which this feature could be used more effectively. The presence of reminder, education material, and data visualization features was associated with desirable reductions in SBP. In particular, these features could lead to higher adherence to self-care behaviors, enhanced diabetes and hypertension knowledge, and improved decision making. However, the trend was not consistently observed for HbA_{1c} levels and DBP; therefore, the efficacy of these features warrants further examination.

With respect to secondary outcomes, our meta-analyses indicated that mobile app-assisted self-care interventions had significant lowering effects on FBG and waist circumference. No significant differences were observed in body weight, BMI, total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides between the intervention and control groups, probably because the design of the interventions was less targeted for these health indexes. Our narrative synthesis indicated that in a small number of trials, the interventions were helpful in improving several clinical, behavioral, knowledge, and psychosocial outcomes. According to these trials, it appeared that such interventions have a potential to engage patients in disease management, including maintaining a healthy lifestyle, improving self-care knowledge, and addressing psychosocial needs. On the other hand, there were trials that examined these outcomes that did not show positive effects. In fact, two trials demonstrated negative effects of the mobile app-assisted self-care interventions on depression and medication dose. Given the mixed results yielded from only a small set of studies, to further understand the impacts of the interventions on these outcomes in the self-care of the diseases, more research is needed.

Implications for Research

Our review suggests several implications for research. First, limited RCTs have emphasized behavioral, knowledge, and psychosocial aspects as primary outcomes in their examination. Further RCTs should focus more on these outcomes to obtain better understanding of whether or not, how, and to what extent mobile app-assisted health interventions change the health, self-care behaviors, and health technology adoption behaviors of patients; expand their knowledge base about health decision making and care; and influence their feelings about and attitudes

toward technology-based self-care. Second, although the associations between each intervention feature and improvements in patient outcomes have been examined, information about the appropriate/optimal combinations of the features is important and limited. Future studies should further examine which combinations of intervention features are more effective for patients and disease self-care. Perhaps, the design can exploit artificial intelligence techniques to identify patients' needs and then combine and present appropriate features tailored to those needs. In addition, our results also indicated that some features, including personalized goal setting; data visualization; and monitoring of diet, physical activity, and body weight, were not always associated with (more) improvements in patient outcomes. Further studies are required to determine the possible reasons for these observations, such as variations in patients' acceptance and adoption of the features, engagement in the self-care activities that the features intended to support, or perceptions of the design and efficacy of the features. Design of the technology as well as patients' attitudes toward and acceptance of the technology determine whether the technology could demonstrate its benefits and impacts [11,54-60]. Third, the implementation duration for most of the reviewed studies was 6 months or less, which is a relatively short time period for studying health behaviors related to chronic disease management. Whether the technology would motivate patients to engage in self-care activities and achieve long-term sustained benefits remains unknown.

Implications for Practice

Our review also provides recommendations for the design and development of mobile app-assisted self-care interventions. First, our results suggest that mobile app-assisted self-care interventions should incorporate features including logging, personalized feedback, communication with HCPs, education, and data visualization in the design and implementation phases of the interventions. This suggestion is consistent with that of Greenwood et al [45] who recommended that it is important to provide a complete feedback loop between patients and their HCPs that incorporates communication, logging data, education, and personalized feedback to make the self-care process more effective. Second, some studies suggested that technical difficulties or usability problems were associated with patient withdrawals [30,32,44], whereas some other studies reported that lacking a user-friendly design is one of the most common reasons for nonadoption or low use of the technology [55,61,62]. These issues emphasize that the design and development of mobile app-assisted self-care interventions should follow human factors design methodologies and principles [55,63-66] to provide more reliable and convenient technologies for self-care. Usability tests are important in the design and development phases to identify design deficiencies. Third, some trials reported a decline in the use of such interventions over the implementation duration [34,37,41]. More effective mobile app-assisted self-care interventions should be developed to motivate patients to engage in self-care behaviors and further enhance health-related outcomes.

XSL•FO
RenderX

Strengths and Limitations

Our review has several strengths. It provides evidence regarding the effects of mobile app-assisted self-care interventions developed for type 2 diabetes and/or hypertension on patient outcomes. In addition to HbA_{1c} levels and BP, several relevant outcomes that were scarcely examined in previous reviews are also analyzed in our review. Our review also provides an evidence-based review of the features of such interventions and their associations with improvements in glycemic and BP control. Our study has limitations. First, although type 2 diabetes and hypertension overlap in population and are closely interlinked, combining the two diseases into one systematic review may cause high heterogeneity. In this study, subgroup analysis by disease type was only conducted for primary outcomes to understand the effects of the intervention in different disease population. The effects of the intervention on the secondary outcomes should be interpreted with caution due to the variability in disease type. Second, the reported effects should be interpreted with caution because control patients in some of the reviewed trials received enhanced usual care, including additional education or phone call communications with their HCPs. Third, 42 patient outcomes were examined using narrative synthesis by simply counting their statistical significance. The effect sizes and significant levels of these outcomes were not obtained. Fourth, publication bias was detected when BMI was the examined outcome. Fifth, only English language articles were included in our review, which may have introduced language and publication bias. Finally, the review lacked an a priori and published protocol.

Conclusions

For type 2 diabetic and/or hypertensive patients, performing self-care and maintaining a healthy lifestyle are necessary but also challenging. The use of mobile app-assisted self-care interventions appears to be effective in improving glycemic and BP management and control; however, this effectiveness was not consistent in some other outcomes. Hence, further investigations on the effects of the interventions on other outcomes are warranted. Moreover, it will be valuable to determine which combinations of features of such interventions are most effective in achieving improvements in the desired outcomes, as it can guide the optimal design of such interventions.

Authors' Contributions

KL and CO designed the systematic review and developed the study protocol. KL and ZX screened the studies identified in the databases and performed data extraction. KL performed the data analysis. KL, CO, ZX contributed to the writing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Features identified in the mobile app-assisted self-care interventions of the 27 trials.
[DOCX File , 20 KB - mhealth_v8i8e15779_app1.docx]

Multimedia Appendix 2
Key features of the mobile app-assisted self-care interventions.
[DOCX File , 21 KB - mhealth_v8i8e15779_app2.docx]

Multimedia Appendix 3
Effects of each intervention feature on hemoglobin A1c reduction.
[DOCX File , 16 KB - mhealth_v8i8e15779_app3.docx]

Multimedia Appendix 4
Effects of each intervention feature on systolic blood pressure reduction.
[DOCX File , 16 KB - mhealth_v8i8e15779_app4.docx]

Multimedia Appendix 5
Effects of each intervention feature on diastolic blood pressure reduction.
[DOCX File , 16 KB - mhealth_v8i8e15779_app5.docx]

References

1. Diabetes and high blood pressure. Blood Pressure UK. 2008. URL: http://www.bloodpressureuk.org/BloodPressureandyou/Yourbody/Diabetes [accessed 2020-02-26]

2. Gress TW, Nieto FJ, Shahar E, Wofford MR, Brancati FL. Hypertension and antihypertensive therapy as risk factors for type 2 diabetes mellitus. N Engl J Med 2000 Mar 30;342(13):905-912. [doi: 10.1056/NEJM200003303421301] [Medline: 10738048]

3. Cheung BMY, Wat NMS, Tso AWK, Tam S, Thomas GN, Leung GM, et al. Association between raised blood pressure and dysglycemia in Hong Kong Chinese. Diabetes Care 2008 Sep;31(9):1889-1891 [FREE Full text] [doi: 10.2337/dc08-0405] [Medline: 18556342]

4. Bayliss EA, Steiner JF, Fernald DH, Crane LA, Main DS. Descriptions of barriers to self-care by persons with comorbid chronic diseases. Ann Fam Med 2003;1(1):15-21 [FREE Full text] [Medline: 15043175]

5. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. J Diabetes Sci Technol 2013;7(1):247-262 [FREE Full text] [Medline: 23439183]

6. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]

7. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]

8. Boulos MNK, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]

9. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]

10. Cheung DST, Or CK, So MKP, Ho K, Tiwari A. The use of eHealth applications in Hong Kong: results of a random-digit dialing survey. J Med Syst 2019 Jul 23;43(9):293. [doi: 10.1007/s10916-019-1422-2] [Medline: 31338682]

11. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR Mhealth Uhealth 2018 Dec 26;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]

12. 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]

13. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]

14. Bonoto BC, de Araújo VE, Godói IP, de Lemos LLP, Godman B, Bennie M, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2017 Mar 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]

15. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016;11(11):e0166718 [FREE Full text] [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]

16. Alessa T, Abdi S, Hawley MS, de Witte L. Mobile apps to support the self-management of hypertension: systematic review of effectiveness, usability, and user satisfaction. JMIR Mhealth Uhealth 2018 Jul 23;6(7):e10723 [FREE Full text] [doi: 10.2196/10723] [Medline: 30037787]

17. Sarfo FS, Treiber F, Gebregziabher M, Adamu S, Nichols M, Singh A, PINGS Team. Phone-based intervention for blood pressure control among Ghanaian stroke survivors: a pilot randomized controlled trial. Int J Stroke 2019 Aug 22;14(6):630-638. [doi: 10.1177/1747493018816423] [Medline: 30465630]

18. Márquez Contreras E, Márquez Rivero S, Rodríguez García E, López-García-Ramos L, Carlos Pastoriza Vilas J, Baldonedo Suárez A, Compliance Group of Spanish Society of Hypertension (SEH-LELHA). Specific hypertension smartphone application to improve medication adherence in hypertension: a cluster-randomized trial. Curr Med Res Opin 2019 Jan;35(1):167-173. [doi: 10.1080/03007995.2018.1549026] [Medline: 30431384]

19. Sun C, Sun L, Xi S, Zhang H, Wang H, Feng Y, et al. Mobile phone-based telemedicine practice in older Chinese patients with type 2 diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e10664 [FREE Full text] [doi: 10.2196/10664] [Medline: 30609983]

20. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0. Hoboken: John Wiley & Sons; 2011.

21. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]

22. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [Medline: 9310563]

XSL•FO
RenderX

23. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]

24. Anzaldo-Campos MC, Contreras S, Vargas-Ojeda A, Menchaca-Díaz R, Fortmann A, Philis-Tsimikas A. Dulce wireless Tijuana: a randomized control trial evaluating the impact of Project Dulce and short-term mobile technology on glycemic control in a family medicine clinic in northern Mexico. Diabetes Technol Ther 2016 Apr;18(4):240-251 [FREE Full text] [doi: 10.1089/dia.2015.0283] [Medline: 26914371]

25. Greenwood DA, Blozis SA, Young HM, Nesbitt TS, Quinn CC. Overcoming clinical inertia: a randomized clinical trial of a telehealth remote monitoring intervention using paired glucose testing in adults with type 2 diabetes. J Med Internet Res 2015;17(7):e178 [FREE Full text] [doi: 10.2196/jmir.4112] [Medline: 26199142]

26. Takenga C, Berndt R, Musongya O, Kitero J, Katoke R, Molo K, et al. An ICT-based diabetes management system tested for health care delivery in the African context. Int J Telemed Appl 2014;2014:437307 [FREE Full text] [doi: 10.1155/2014/437307] [Medline: 25136358]

27. Kim JY, Wineinger NE, Steinhubl SR. The influence of wireless self-monitoring program on the relationship between patient activation and health behaviors, medication adherence, and blood pressure levels in hypertensive patients: a substudy of a randomized controlled trial. J Med Internet Res 2016 Jun 22;18(6):e116 [FREE Full text] [doi: 10.2196/jmir.5429] [Medline: 27334418]

28. 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;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]

29. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]

30. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014 Dec 11;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]

31. Hsu WC, Lau KHK, Huang R, Ghiloni S, Le H, Gilroy S, et al. Utilization of a cloud-based diabetes management program for insulin initiation and titration enables collaborative decision making between healthcare providers and patients. Diabetes Technol Ther 2016 Feb;18(2):59-67 [FREE Full text] [doi: 10.1089/dia.2015.0160] [Medline: 26645932]

32. Karhula T, Vuorinen A, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, et al. Telemonitoring and mobile phone-based health coaching among Finnish diabetic and heart disease patients: randomized controlled trial. J Med Internet Res 2015;17(6):e153 [FREE Full text] [doi: 10.2196/jmir.4059] [Medline: 26084979]

33. Kleinman NJ, Shah A, Shah S, Phatak S, Viswanathan V. Improved medication adherence and frequency of blood glucose self-testing using an m-health platform versus usual care in a multisite randomized clinical trial among people with type 2 diabetes in India. Telemed J E Health 2017 Sep;23(9):733-740. [doi: 10.1089/tmj.2016.0265] [Medline: 28328396]

34. Logan AG, Irvine MJ, McIsaac WJ, Tisler A, Rossos PG, Easty A, et al. Effect of home blood pressure telemonitoring with self-care support on uncontrolled systolic hypertension in diabetics. Hypertension 2012 Jul;60(1):51-57 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.111.188409] [Medline: 22615116]

35. Moore J, Marshall M, Judge D, Moss F, Gilroy S, Crocker J, et al. Technology-supported apprenticeship in the management of hypertension: a randomized controlled trial. J Clin Outcomes Manag 2014 Mar;21(3):110-122.

36. Nagrebetsky A, Larsen M, Craven A, Turner J, McRobert N, Murray E, et al. Stepwise self-titration of oral glucose-lowering medication using a mobile telephone-based telehealth platform in type 2 diabetes: a feasibility trial in primary care. J Diabetes Sci Technol 2013;7(1):123-134 [FREE Full text] [Medline: 23439168]

37. Or C, Tao D. A 3-month randomized controlled pilot trial of a patient-centered, computer-based self-monitoring system for the care of type 2 diabetes mellitus and hypertension. J Med Syst 2016 Apr;40(4):81. [doi: 10.1007/s10916-016-0437-1] [Medline: 26802011]

38. Orsama A, Lähteenmäki J, Harno K, Kulju M, Wintergerst E, Schachner H, et al. Active assistance technology reduces glycosylated hemoglobin and weight in individuals with type 2 diabetes: results of a theory-based randomized trial. Diabetes Technol Ther 2013 Aug;15(8):662-669. [doi: 10.1089/dia.2013.0056] [Medline: 23844570]

39. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 Jun;10(3):160-168. [doi: 10.1089/dia.2008.0283] [Medline: 18473689]

40. Yoo HJ, Park MS, Kim TN, Yang SJ, Cho GJ, Hwang TG, et al. A ubiquitous chronic disease care system using cellular phones and the internet. Diabet Med 2009 Jun;26(6):628-635. [doi: 10.1111/j.1464-5491.2009.02732.x] [Medline: 19538239]

41. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar 13;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]

42. Lakshminarayan K, Westberg S, Northuis C, Fuller CC, Ikramuddin F, Ezzeddine M, et al. A mHealth-based care model for improving hypertension control in stroke survivors: pilot RCT. Contemp Clin Trials 2018 Jul;70:24-34 [FREE Full text] [doi: 10.1016/j.cct.2018.05.005] [Medline: 29763657]

43. Bender M, Cooper B, Park L, Padash S, Arai S. A feasible and efficacious mobile-phone based lifestyle intervention for Filipino Americans with type 2 diabetes: randomized controlled trial. JMIR Diabetes 2017 Dec 12;2(2):e30 [FREE Full text] [doi: 10.2196/diabetes.8156] [Medline: 30291068]

44. Hansen CR, Perrild H, Koefoed BG, Zander M. Video consultations as add-on to standard care among patients with type 2 diabetes not responding to standard regimens: a randomized controlled trial. Eur J Endocrinol 2017 Jun;176(6):727-736. [doi: 10.1530/EJE-16-0811] [Medline: 28325823]

45. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A systematic review of reviews evaluating technology-enabled diabetes self-management education and support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027. [doi: 10.1177/1932296817713506] [Medline: 28560898]

46. Or CKL, Tao D. Does the use of consumer health information technology improve outcomes in the patient self-management of diabetes? A meta-analysis and narrative review of randomized controlled trials. Int J Med Inform 2014 May;83(5):320-329. [doi: 10.1016/j.ijmedinf.2014.01.009] [Medline: 24534118]

47. Marcolino MS, Maia JX, Alkmim MBM, Boersma E, Ribeiro AL. Telemedicine application in the care of diabetes patients: systematic review and meta-analysis. PLoS One 2013;8(11):e79246 [FREE Full text] [doi: 10.1371/journal.pone.0079246] [Medline: 24250826]

48. Omboni S, Gazzola T, Carabelli G, Parati G. Clinical usefulness and cost effectiveness of home blood pressure telemonitoring: meta-analysis of randomized controlled studies. J Hypertens 2013 Mar;31(3):455-468. [doi: 10.1097/HJH.0b013e32835ca8dd] [Medline: 23299557]

49. Liu S, Dunford SD, Leung YW, Brooks D, Thomas SG, Eysenbach G, et al. Reducing blood pressure with Internet-based interventions: a meta-analysis. Can J Cardiol 2013 May;29(5):613-621. [doi: 10.1016/j.cjca.2013.02.007] [Medline: 23618507]

50. McLean G, Band R, Saunderson K, Hanlon P, Murray E, Little P, et al. Digital interventions to promote self-management in adults with hypertension systematic review and meta-analysis. J Hypertens 2016 Apr;34(4):600-612 [FREE Full text] [doi: 10.1097/HJH.0000000000000859] [Medline: 26845284]

51. Dyer JS. Effects of consumer-facing technologies on patient engagement, behavior change, and type 2 diabetes-related health outcomes. Diabetes Spectr 2013 May 16;26(2):98-101. [doi: 10.2337/diaspect.26.2.98]

52. Fitzner K, Moss G. Telehealth: an effective delivery method for diabetes self-management education? Popul Health Manag 2013 Jun;16(3):169-177. [doi: 10.1089/pop.2012.0054] [Medline: 23216062]

53. Jerant AF, von Friederichs-Fitzwater MM, Moore M. Patients' perceived barriers to active self-management of chronic conditions. Patient Educ Couns 2005 Jun;57(3):300-307. [doi: 10.1016/j.pec.2004.08.004] [Medline: 15893212]

54. Casper G, Karsh B, Or C, Carayon P, Grenier A, Sebern M, et al. Designing a technology enhanced practice for home nursing care of patients with congestive heart failure. In: AMIA Annu Symp Proc. 2005 Presented at: AMIA 2005 Annual Symposium; October 22, 2005 - October 26, 2005; Washington p. 116-120 URL: http://europepmc.org/abstract/MED/16779013

55. Or C, Tao D. Usability study of a computer-based self-management system for older adults with chronic diseases. JMIR Res Protoc 2012;1(2):e13 [FREE Full text] [doi: 10.2196/resprot.2184] [Medline: 23612015]

56. Or CKL, Karsh B. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009 Aug;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]

57. Tao D, Or CKL, Li K. The influence of computer experience and screen size on use performance and perceptions of a tablet-based chronic illness self-management system for older adults. 2014 Jul 22 Presented at: 2014 International Symposium on Human Factors and Ergonomics in Health Care; March 16, 2014 - March 19, 2014; Chicago p. 58-62. [doi: 10.1177/2327857914031008]

58. Yan M, Or C. Factors in the 4-week acceptance of a computer-based, chronic disease self-monitoring system in patients with type 2 diabetes mellitus and/or hypertension. Telemed J E Health 2018 Feb;24(2):121-129. [doi: 10.1089/tmj.2017.0064] [Medline: 28737995]

59. Yan M, Or C. A 12-week pilot study of acceptance of a computer-based chronic disease self-monitoring system among patients with type 2 diabetes mellitus and/or hypertension. Health Informatics J 2019 Sep;25(3):828-843. [doi: 10.1177/1460458217724580] [Medline: 28820007]

60. Or CKL, Valdez RS, Casper GR, Carayon P, Burke LJ, Brennan PF, et al. Human factors and ergonomics in home care: current concerns and future considerations for health information technology. Work 2009;33(2):201-209 [FREE Full text] [doi: 10.3233/WOR-2009-0867] [Medline: 19713630]

61. 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]

62. 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]

63. Cheung DST, Or CKL, So MKP, Tiwari A. Usability testing of a smartphone application for delivering Qigong training. J Med Syst 2018 Sep 05;42(10):191. [doi: 10.1007/s10916-018-1048-9] [Medline: 30187139]

64. Liu K, Chan F, Or CK, Sun DT, Lai W, So H. Heuristic evaluation and simulated use testing of infusion pumps to inform pump selection. Int J Med Inform 2019 Nov;131:103932. [doi: 10.1016/j.ijmedinf.2019.07.011] [Medline: 31557700]

65. Nielsen J. Usability Engineering. San Francisco: Morgan Kaufmann; 1994.

66. Shneiderman B, Plaisant C. Designing the User Interfaces: Strategies for Effective Human-Computer Interaction. New York: Pearson Education; 2010.

Abbreviations

BG: blood glucose
BP: blood pressure
DBP: diastolic blood pressure
FBG: fasting blood glucose
GRADE: Grading of Recommendations Assessment, Development and Evaluation
HbA$_{1c}$: hemoglobin A$_{1c}$
HCP: health care provider
HDL: high-density lipoprotein
LDL: low-density lipoprotein
MD: mean difference
RCT: randomized controlled trial
SBP: systolic blood pressure
SMD: standardized mean difference

XSL•FO
RenderX

Review

Communication Technology Use by Caregivers of Adolescents With Mental Health Issues: Systematic Review

Ronelle Jansen[1], BSocSc, MSocSc, MCur; Marianne Reid[1], BCur, MSocSc, PhD

School of Nursing, Faculty of Health Science, University of the Free State, Bloemfontein, South Africa

Corresponding Author:
Ronelle Jansen, BSocSc, MSocSc, MCur
School of Nursing
Faculty of Health Science
University of the Free State
205 Nelson Mandela Drive
Bloemfontein, 9301
South Africa
Phone: 27 825544731
Fax: 27 514013435
Email: jansenro@ufs.ac.za

Abstract

Background: Caregivers of adolescents with mental health issues experience challenges that may result in the caregivers having a variety of unmet needs. There is a growing need to support these caregivers. Effective support to strengthen positive caregiving behavior in caregivers may address their challenges. Communication technologies offer novel opportunities to assist these caregivers and may contribute to strengthening caregiver behavior. However, little is known about the use of communication technologies among caregivers of adolescents with mental health issues.

Objective: The study aimed to answer the question: "What is the best evidence available to strengthen positive behavior of caregivers of adolescents with mental health issues using communication technology."

Methods: A systematic review of articles published between January 2007 and August 2018 was conducted. Searches included articles of multiple study designs from EBSCO Host and Scopus platforms with prespecified eligibility criteria. Methodological quality was evaluated using the applicable Critical Appraisal Skills Programme and Joanna Briggs Institute assessment tools.

Results: The search yielded 1746 articles. Altogether, 5 articles met the eligibility criteria and were included in the review for data synthesis. Data analysis and synthesis identified three thematic conclusions reflecting the types of communication technologies used, caregivers as the target population, and strengthening of positive behavior through determinants of the Integrated Model of Behavior Prediction.

Conclusions: The review reported the usefulness of communication technology by caregivers. Caregivers also demonstrated improvement in self-efficacy, knowledge, parent-child communication, and parental skills reflecting positive behavior. Although the use of communication technology is expanding as a supportive intervention to address caregivers' needs, the evidence for usefulness among caregivers of adolescents with mental health issues is still scarce. More research and information related to preferred methods of communication delivery among caregivers of adolescents is still needed.

Introduction

Background

> *"Nothing about me, without me"* [1]

The caregiver landscape can be successfully supported through tailored solutions by acknowledging their unique challenges. Caregivers are largely invisible and are mostly underappreciated resources in community health services [2,3]. Caregivers are vital partners who contribute to health care provisioning in communities. They contribute to alleviating resource

deficiencies and assist with task sharing in the health sector, despite not formally being part of the health sector [4-6]. Regardless of this important contribution to health care delivery, caregivers receive minimal support from governments [7,8]. Globally, caregivers share a myriad of ongoing challenges and obstacles that are burdensome [2,9], often leading to unmet physical, psychological, and financial needs [10,11].

Caregivers' own needs are mostly neglected because there is usually more attention focused on the care recipient or the family's needs [12]. It is thus vital to identify the unmet needs of the caregiver independently from those of the care recipient [9,13]. Unmet needs of caregivers merit more recognition and understanding in order to employ effective support to their prioritized needs [14]. Interventions for caregivers may be beneficial to address needs such as skills training, support, education, and access to resources [15,16].

Fortunately, there is greater awareness of mediating interventions to support caregivers who are providing care for relatives with a chronic illness or disability [17]. Various interventions have been geared toward caregivers of persons presenting with dementia [14,18,19], cancer [20,21], HIV [22,23], chronic diseases [24-26], maternal and child health [27,28], and long-term conditions in children [29,30]. Various authors [31,32] claim that there is limited research available that focuses on families caring for adolescents with mental health needs, while Cardamone-Breen et al [33] stated there is a scarcity of parental interventions to prevent them from internalizing the mental health issues of adolescents. Parent- and family-focused interventions for child and adolescent mental health care are poorly represented in research [34-36]. Authors do not use similar terminology when referring to mental health issues, problems, illnesses, or conditions. This review will refer to *mental health issues* as a broad term for mental, emotional, and behavioral problems or disorders focusing on depression, anxiety, and substance use. Depression, anxiety, and substance use disorders are more common among adolescents [37-39].

Armoiry et al [40] reported equivocal findings regarding the use of communication technologies among families of young people with long-term conditions. Rodríquez-Meirinhos et al [31] also highlighted caregivers' need for education and information in preparation for caring for an adolescent with a mental illness. These authors [31] also established the necessity for dependable interventions to support families and community services to empower this group of caregivers.

Evidence-based supportive interventions that are flexible and developed according to the needs of caregivers may reduce caregiver burden and improve mental health caregiving [9]. Kuhn and Laird [41] stated that a combination of supportive interventions would be more useful for a diverse population, allowing tailored communication. Understanding caregivers' cultural context is important to develop a tailored intervention congruent to their specific needs [6,9,42]. Moreover, caregiver interventions should be sustainable [43], easily accessible [44], available [9], culturally, ethnically, and linguistically tailored [45], and caregiver-focused [46].

Literature suggests that communication technologies are arising as an acceptable intervention to assist caregivers of children and young people in the management of conditions [40]. Communication technologies terminology is labeled interchangeably in the literature but refers to computers, the internet, electronic health (eHealth), and mobile health (mHealth), including networks and media services to transmit information [47-50]. This taxonomy of communication technologies is often used simultaneously in the research of interventions reporting on various outcomes [51]. Combined communication technology interventions often result in effective health outcomes and are feasible substitutes for traditional health promotion approaches such as printed material [52,53]. Vergunst [54] also recommended that creative interventions need to be established for caregivers in resource-poor settings such as rural areas to bridge gaps in mental health service delivery for the individual, the family, and the community. Although communication technology interventions are potentially valuable to support caregivers of children with mental and behavioral problems [55], limited data are available even though it is a major public health concern [56]. The emergence of communication technologies to deliver training for caregivers is increasing and show promise in real-world setting [57].

Evidently, the usefulness of communication technology interventions for caregivers of adolescents with mental health issues differs. DeHoff et al [58] found that parents of children with special needs got support through an online social platform to accept and manage the child's condition. Another study disclosed that parents of youth with mental health problems seem to be positive about the use of computer-based therapies [59]. A web-based health promotion program utilized by adolescent girls and their mothers for drug use, among other conditions, resulted in positive health behavior changes [60]. The study by Russell et al [61] indicated a high satisfaction among parents of children with behavior problems when engaging in a computer-based education program.

The novelty of communication technologies in the management of psychiatric and mental health illnesses may be valuable to investigate [62]. According to Casale et al [63] and Robila [64], there is a need to support parents through interventions that provide resources to improve parenting, particularly in reducing adolescent behavioral and emotional problems. It would appear that communication technologies are beneficial in supporting caregivers in improving their skills, abilities, self-efficacy, and knowledge [65,66]. A review of caregivers caring for adolescents with developmental disabilities similarly highlighted the effect of tailored interventions to augment caregivers' self-efficacy, self-esteem, positive coping skills, and supportive social networks [67].

Theoretical Framework

The Integrative Model of Behavior Prediction (IMBP) endorses determinants such as attitudes, norms, and self-efficacy, by predicting behavior. Consequently, these determinants are based on underlying beliefs that, in turn, influence the intention to perform a specific behavior. Behavior may be challenged due to a lack of skills or environmental constraints [68,69].

The IMBP guided this review in clarifying caregivers' behavioral intention for utilizing communication technologies by determining their belief, attitude, norms, self-efficacy, skills,

environment, and intention. Behavioral intention can be predicted if the populations' beliefs, attitudes, norms, self-efficacy, skills, and environment are known [68,70,71]. To our knowledge, however, there have been no systematic efforts to synthesize evidence on the usefulness of communication technologies to strengthen positive behavior in caregivers of adolescents with mental health issues. Any positive response to a determinant was regarded as strengthening [71,72] of caregivers' behavior toward the adolescent.

The authors conducted a systematic review to collect and synthesize all evidence fitting the prespecified eligibility criteria. The authors sought to answer the following question based on the PICO format [73] (see Table 1 for the application of the PICO format): "What is the best evidence available to strengthen positive behavior of caregivers of adolescents with mental health issues using communication technology?" This paper provides a critical review and synthesis of the best available evidence about the use of communication technologies by caregivers of adolescents with mental health issues.

Table 1. Application of the PICO format.

PICO	Application
Population	Caregivers of adolescents with mental health issues (Caregivers namely families, parents, sibling, carer, etc)
Intervention	Communication technology
Comparison	Routine communication technology
Outcome	IMBP[a] determinants (beliefs, attitude, norms, self-efficacy, skills, environment, and intention)

[a]IMBP: Integrated Model of Behavior Prediction.

Methods

Design

The authors undertook a systematic review that included multiple study designs [73-75]. This systematic review has been registered in PROSPERO (CRD42018094680) and was conducted according to PRISMA guidelines.

Search Strategy

A senior research librarian conducted a systematic search on electronic databases that covered articles published from January 1, 2007 to August 2, 2018. Databases from the EBSCO Host platform (PsycINFO, Academic Search Ultimate, MEDLINE with Full Text, Health Source: Nursing/Academic Edition, SocINDEX with Full Text, CINAHL with Full Text, ERIC, CAB Abstracts, MasterFILE Premier, Africa-Wide Information,

PsycARTICLES, OpenDissertations, Communication & Mass Media Complete, Business Source Ultimate, SPORTDiscus with Full Text, Health Source - Consumer Edition, Humanities Source, EconLit with Full Text, GreenFILE) and the Scopus database were included. No language or study design restrictions were applied. Search strings determined by the authors in collaboration with the research librarian were used to retrieve articles from the abovementioned databases. It was decided to make use of free text, as the search would be conducted on multiple databases, of which only MEDLINE and CINAHL utilize controlled vocabulary (MeSH and CINAHL subject headings respectively). Before conducting the search, a concept analysis was done to identify synonyms to be used in the free-text search (see Table 2 for a combination of the search string using Boolean operators). The first author (RJ) also searched the reference lists of retrieved articles. Dissertations and book chapters formed part of the retrieved articles.

Table 2. Search strings used.

Search strings
caregiver* OR "care giver*" OR family* OR families* OR parent* OR mother* OR father* OR sibling* OR carer OR carers OR "lay worker*" OR "next of kin" AND
"mental* health*" OR "mental* ill*" OR "mental* disorder*" OR "mental* diseas*" OR depress* OR anxiet* OR substance AND
teen* OR adoles* OR juvenile* OR youth* AND
belief* OR conviction* OR faith OR trust* OR norm OR norms OR custom* OR attitude* OR outlook* OR approach* OR Self-efficac* OR ability* OR Skill OR skills OR expertis* OR able OR abilit* OR talent* OR proficien* OR knowhow OR capabilit* OR knack OR competen* OR Intent* OR determination* OR planning OR resolve OR decide* OR decision* OR choose OR select* OR choice* AND
mobile* OR cell* OR smart OR sms OR "short message service*" OR text* OR device* OR mhealth* OR m-health* OR ehealth* OR e-health* OR "instant messag*" OR app OR apps OR phone* OR smartphone* OR "electronic device*" OR "portable device*" OR "phone intervention*" OR "telephon* intervention*" OR online*) not ("stem cell*" OR "sickle cell*" OR "assist" device*") AND
mobile* OR cell* OR smart OR sms OR "short message service*" OR text* OR device* OR mhealth* OR m-health* OR ehealth* OR e-health* OR "instant messag*" OR app OR apps OR phone* OR smartphone* OR "electronic device*" OR "portable device*" OR "phone intervention*" OR "telephon* intervention*" OR online* OR caregiver* OR "care giver*" OR family* OR families* OR parent* OR mother* OR father* OR sibling* OR carer OR carers OR "lay worker*"

Eligibility Criteria

Articles were eligible for inclusion if they (1) focused on caregivers such as families, parents, siblings, or relatives of adolescents with mental health issues (concentrating on depression, anxiety, and substance use disorders) as the target population, (2) reported on communication technology usage, and (3) described caregiver determinants according to the IMBP. Articles were excluded if they did not meet the inclusion criteria or (1) if they focused only on adolescents as the target population, (2) if they reported on non–communication technology interventions, and (3) if the focus was on mental health–related therapies or traditional communication interventions.

Selection Procedure

The first author (RJ) filtered all titles and abstracts obtained from the search against the review question and eligibility criteria. The second author (MR) verified articles for compliance with eligibility criteria. A large number of articles were excluded, and duplicates were removed. Full-texts for the remaining articles were obtained from the librarian. These full-text articles were independently screened for eligibility by both authors, and any discordances were resolved through discussion. Additionally, references of selected articles were screened further for relevant studies.

Quality Appraisal

The first author (RJ), together with two senior researchers, rigorously evaluated the quality of the selected articles according to the Critical Appraisal Skills Programme [76] and Joanna Briggs Institute assessment tools applicable to each study design [77]. All tools evaluated the appropriateness of methods used, clarity of focus, the recruitment process, the accuracy of measures used, data collection, presentation and analysis, clarity in the statement of the findings, and appropriateness of context. If more than two aspects were not addressed in the articles, it was excluded. This was depicted as either Level 1* (full marks obtained) or Level 1 (–1 mark) in Table 3. Table 3 includes a hierarchy classification system in terms of research design, according to the American Dietetic Association [78,79]. This critical appraisal of the full-text articles determined the selection and inclusion of articles in this systematic review.

Table 3. Summary of included articles' study designs.

Extracted data	Reference				
	Deitz et al [80]	Cardamone-Breen et al [33]	Choi et al [81]	Molleda et al [82]	Estrada et al [83]
Design	Randomized control trial	Randomized control trial	Quasi-experimental	Qualitative, descriptive	Qualitative interviews
Grading					
Level	1	1	1*	1	1
Hierarchy	A	A	C	D	D
Communication technology					
Type	Web-based	Web-based	Web-based	Internet-based, eHealth	Internet-based, eHealth, tablets
Reminders	Multiple—not specified	—	Phone calls, text messages, and emails	—	Phone or text messages
Financial support	National Institute on Mental Health	National Health and Medical Research Council	Korean government	—	Centers for Disease Control and Prevention (grant)
Recruitment	—	Online networks, social media, computer, online parent portals, and email communication	Via online communities	Emails	Online
Follow-up		Phone calls, text messages, and emails			
Targeted					
Caregivers	Parents, n=99	Parent/caregiver, n=349	Parents, n=114	Parents, n=6	Parents, n=29
Youth/ adolescents	Mental health problems or issues	With depression and anxiety disorders	Mental health problems or issues (common mental health problems (ie, bullying, depression, internet addiction, and suicide)	With drug use and risky sex behavior	With drug use and sexually risky behavior
Setting	Workplace of parents: ISA Group in Alexandria in the US	Secondary schools in Australia	Two elementary schools and four middle schools from four cities in Korea	Pediatric mobile clinic in Miami-Dade County and another clinic located on the university medical campus (University of Miami Miller School of Medicine)	Multiple middle schools relatively close to where the families lived and at the University of Miami offices.

Data Extraction, Analysis, and Synthesis

The authors extracted specific data from the selected articles in a tabular format to record the study characteristics, types, and delivery method of communication technologies and the strengthening of positive behavior according to IMBP determinants. A meta-analysis was not feasible due to the heterogeneity of the included articles. A thematic data analysis [84-86] took place and both authors discussed the themes to refine the data according to the review question. Subsequently, the data were organized into relevant thematic conclusions addressing the research question based on the PICO elements. After that, the results were synthesized across all articles.

Results

Search Results and Selection of Articles

The search identified 1746 electronic records of possible interest. Two additional articles were added, one through contacting an author and another through reference list checking. After electronic and manual removal of duplicates, 1089 records remained. Further screening for eligibility of identified records resulted in the retrieval of 59 full-text articles that were potentially relevant for analysis. The authors then conducted a comprehensive review of these full-text articles and included articles that specifically focused on communication technologies used by caregivers of adolescents with mental health issues. A total of 5 articles met all the eligibility criteria. Figure 1 presents the PRISMA flow diagram of the study, and Table 3 presents a summary of the included articles' study design evaluations.

Figure 1. PRISMA flow diagram.
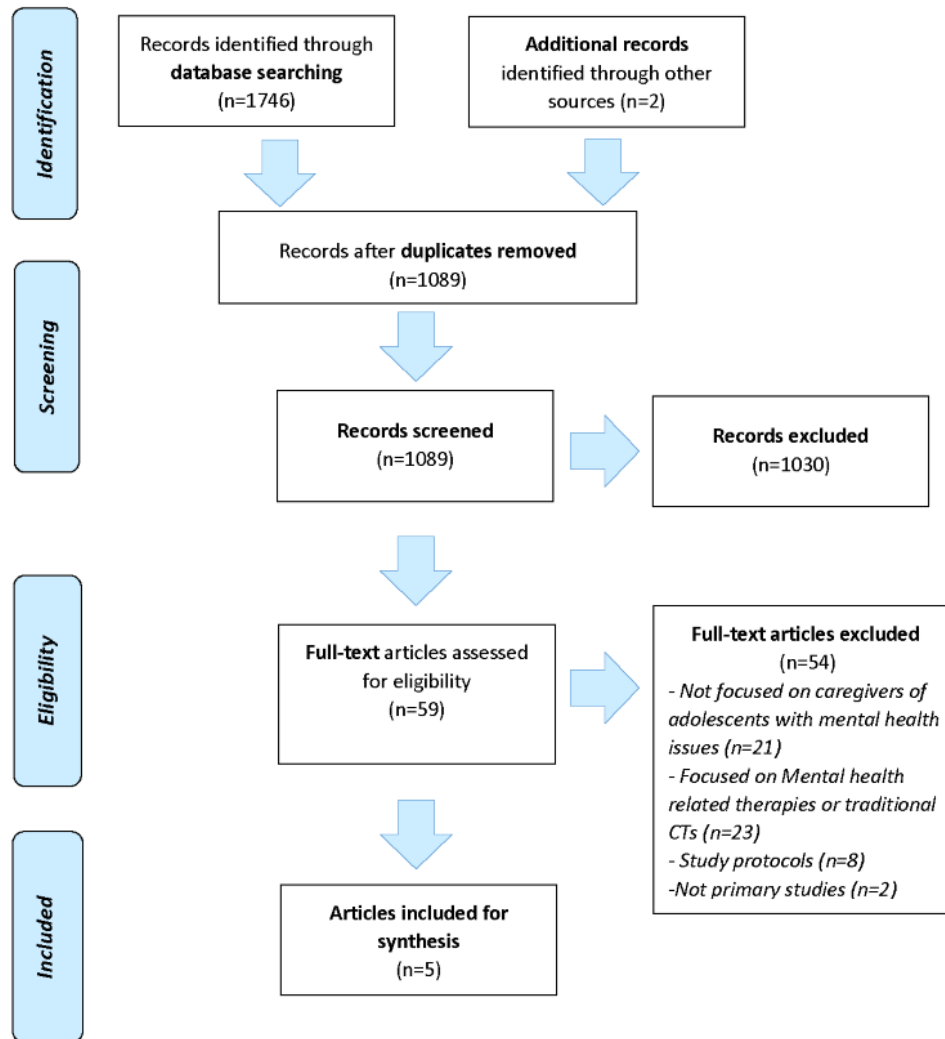
Characteristics of Included Articles

The final articles included in this systematic review consisted of multiple designs. There were 2 randomized controlled trials [33,80], 1 quasi-experimental study [81], and 2 qualitative studies [82,83]. In all 5 articles, the caregivers had been exposed to communication technologies in caring for adolescents with mental health issues. Of the 5 studies, 3 studies were performed in the United States [80,82,83], 1 was performed in Australia [33], and 1 was performed in Korea [81]. Articles varied in methodology, intervention, outcomes, measurement scales, and findings. Not all of the articles complied with all of the identified IMBP determinants leading to positive behavior. Table 4 presents a summary of the results reported by the 5 articles.

Results from the 5 articles were synthesized and thematically analyzed to formulate conclusions. The thematic conclusions are discussed according to the types of communication technologies used, caregivers as the target population, and strengthening of positive behavior through IMBP determinants.

Table 4. Summary of results reported by the 5 articles.

Reference	Strengthening positive behavior according to determinants within IMBP[a]
Deitz et al [80]	Total knowledge score: significantly greater for the experimental group than the control group ($t=-7.96$, $P<.001$)
	Knowledge of treatment: statistically significant ($t=-2.92$, $P=.006$)
	A significant difference was found between experimental and control groups on self-efficacy in handling mental health issues in their children ($F=12.73$, $P=.001$).
	Significant paired t test analysis of increases in self-efficacy from pretest to posttest among the individuals receiving the intervention was found ($t=-3.20$, $P=.003$).
	No significant differences between groups on the measures of family communication, adaptability, cohesion, and attitudes toward mental health issues were found.
Cardamone-Breen et al [33]	Chi-square analyses of the postintervention (1-month) assessment revealed a significant difference in attempts to change parenting (n=307, $\chi^2_3=19.65$, $P<.001$), with significantly more intervention group parents reporting making some changes to their parenting.
	The majority of parents (93.6%) reported they were somewhat or very satisfied with the feedback received, and 95.1% reported the feedback as either somewhat, very, or extremely useful.
	Most parents (90.2%) reported they were somewhat or very likely to change their parenting based on the feedback provided.
	Conclusion: accessible, low-cost preventive approach
Choi et al [81]	Participants in the intervention, compared to those in the control group, demonstrated preliminary evidence of improved parental knowledge.
	The effect size of parental knowledge was large (d=0.60; 95% CI 0.21-0.99).
	Parents in the intervention group showed increased parental self-efficacy, parent-child communication, and satisfaction with parent-child relationships, and decreased parent-child conflict.
	Parents reported accessibility and convenience to complete the intervention.
Molleda et al [82]	Effective parent-adolescent communication skills.
	Parents' positive experience with flexibility, accessibility, and convenience in delivery of eHealth.
	Parents also reported the ability to apply the lessons learned from eHealth *Familias Unidas* to their daily lives.
Estrada et al [83]	It was feasible to recruit, engage, and retain Hispanic families into an eHealth intervention and deliver it electronically.
	Positive feedback was provided by the parents regarding eHealth.
	Parents stated there were multiple lessons learned from engaging in eHealth *Familias Unidas*: effective parent-adolescent communication and active parental attention and involvement in an adolescent's life.
	Culturally appropriate online content allowed parents to access sessions at their convenience and minimized costs for researchers and participants alike

[a]IMBP: Integrated Model of Behavior Prediction.

Types of Communication Technologies Used

All the studies implemented web-based interventions [33,80-83]. Content employed via the communication technologies and delivery thereof varied considerably. The web-based program by Deitz et al [80] consisted of 4 multimedia modules that provided information regarding symptoms and treatment options for depression and anxiety, building parental skills, and additional resources with information about mental health issues. Parental skills included communication, relationships, and healthy lifestyles. Estrada et al [83] and Molleda et al [82] adapted an evidence-based intervention, *Familias Unidas*, into a web-based intervention consisting of 8 e-parent group sessions of video recordings and 4 online family sessions (parent-adolescent) with a facilitator. Choi et al [81] reported on a 4-week web-based intervention (Stepping-stone) containing educational sessions, media files, verbal feedback, weekly assignments, and practice sessions regarding bullying, depression, suicide, and communication skills. Cardamone-Breen et al [33] adapted a single-session web-based component of the Partners in Parenting intervention. This intervention provided personalized feedback and psychoeducation concerning guidelines on adolescent depression and anxiety after the parents completed an online survey to assess parenting practices.

Target Population

The study populations in the included articles predominantly comprised caregivers as parents or family members. Deitz et al [80] used parents or caregivers of youth between the ages 5 and 21, while Molleda et al [82] included clinic personnel, facilitators (ie, physicians, nurse practitioners, administrators, and mental health workers), and parents or primary caregivers of Hispanic adolescents (age range: 12 to 16 years) [82]. Estrada et al [83] also included Hispanic families with parents and adolescents between the ages of 12 and 16 [83]. The study by Choi et al [81] encompassed parents who had at least one child aged 11 to 16 years. Cardamone-Breen et al [33] used a community sample of parents, together with adolescents aged 12 to 15 years old. The most common problems among adolescents were mental health or behavioral issues [80,81], depression and anxiety disorders [33,81], bullying [81], suicide [81], and substance use [82,83]. The majority of participants in the studies were female, recruited from various urban settings, and educated.

IMBP and Other Determinants

Outcomes focusing on the IMBP determinants in the current review were about self-efficacy and skills such as parent-child communication: 2 studies [80,81] measured the effect of the intervention on caregiver self-efficacy that improved after the intervention; 3 studies [81-83] also demonstrated the effectiveness of communication skills between caregiver and adolescent, and 1 randomized controlled trial [80] measured no significant differences between study groups in family communication skills. Of the 5 studies, Cardamone-Breen et al's study [33] showed promising changes in parenting skills and behavior, and another study [80], a randomized control trial, reported outcomes regarding parental attitudes demonstrating no significant differences between the two study groups.

Another observation was that effective parental skills resulted in satisfactory parent-child relationships as well as a decline in parent-child conflict [81]. Caregivers similarly appreciated the feedback they received regarding their parental skills after using the intervention [33] and lessons learned to implement in their daily lives [82,83].

We identified 2 studies [80,81] that measured the effect of communication technologies use on caregivers' knowledge showing an improvement. Additionally, Estrada et al [83] found that parents were more involved and attentive in the life of adolescents after engaging with the eHealth intervention. In turn, participants in Estrada et al's study [83] gave positive feedback regarding the use of the eHealth intervention while Molleda et al [82] found that parents experienced the eHealth delivery as positive.

Discussion

Principal Findings

The review results show the potential value of communication technologies to strengthen caregivers' behavior when caring for an adolescent with mental health issues. Overall, the results yielded evidence of the usefulness of communication technologies by caregivers. Moreover, the results contributed to the limited literature on communication technology interventions for caregivers caring for the adolescent with mental health issues.

As shown by this study, the results were mostly positive, which may raise questions about selection bias. Aligned to the review question, we were interested in determinants showing improved outcomes among caregivers' when using communication technologies. It becomes particularly imperative to recognize users' preferences and acceptance of communication technologies in a health care setting. Identifying appropriate communication technologies to inform and support caregivers may be a challenge, but it is essential to understand their adoption thereof [87,88]. Information dissemination delivered through communication technologies is gradually replacing traditional approaches [89], but the implementation in health care [90] and among caregivers [14,91] is still lacking.

Communication technologies can distribute an unprecedented amount of information through technology such as electronic devices, system software, and information networks such as the internet, which provides access to resources [92,93]. Moreover, Schneider's [94] systematic review reported that utilizing communication technologies to gather and stream information received the highest proportion of use. Multimedia approaches including internet and web-based programs are a means for providing informational support to caregivers; they also found that effective parent education through communication technologies can augment the mental well-being of children [95]. Our results are consistent with earlier reports in this field that suggest that web-based interventions are effective for caregivers managing mental health issues in adolescents [91,96,97].

A prevailing expectation of communication technologies is that it might improve knowledge acquisition regarding illnesses. However, applicable information should be according to the caregivers' specific needs [98]. Statistically significant differences were found related to parental outcomes of knowledge, attitudes, and skills in efficient web-based interventions [99]. Exploiting the internet for knowledge about chronic diseases, as noted by Mahmud et al [53], can assist individuals and communities in health promotion. In contrast, low-income parents were not confident in using the internet and could not distinguish between good- or bad-quality information [100]. Sweeney et al [59] concluded that parents of youth with mental health problems demonstrated poor knowledge regarding computer-based therapies but were positive about using it.

Communication technologies are now mature enough to enable learning and knowledge exchange among caregivers in health care [101]. This functionality is attributable to the fact that communication technologies are associated with boosting the health care landscape through information exchange and transformation among large populations [102]. Our review indicated some evidence that computer-based interventions among large groups of parents are less expensive. The financial implication is particularly relevant in the current global socioeconomic climate, especially in low-resource settings that should keep cost-effectiveness of communication technologies in mind [103] when serving vulnerable populations such as caregivers in deploying health care [104]. According to Sprague et al [105], the world still experiences barriers to internet adoption despite high technology penetration because of disproportionately rural populations, low-income, illiteracy, elderly users, and female users. Therefore, it is vital to explore the best evidence available related to communication technologies barriers such as delivery methods, cost-effectiveness, caregiver characteristics, and outcomes, which will predict adoption thereof.

None of the included studies measured text messaging, phone calls, mobile apps, or social networking as a favored network delivery; however, included studies mentioned the use of communication technologies to recruit participants and send reminders. Domek et al [106] and Anderson-Lewis et al [107] suggest that text message interventions may be useful in rural families and have the potential to disseminate public health information. Mobile apps show some promise in serving families of youth with mental health issues in resource-constraint settings [108]. Breitenstein et al [109] also determined that digital delivery, such as mobile apps, might theoretically be

cost-effective, sustainable, and reach large numbers for parent training. Furthermore, some studies [110,111] have shown that social media effectively supports and informs caregivers through shared participation. Catalano et al [112] identified that parental social support and interacting improved caregiver well-being.

Collective participation is valuable, but researchers should be mindful of caregivers' individuality when exploring the usage of communication technologies. It is important to conceptualize the characteristics of caregivers in the context of communication technologies acceptance [113-115]. Caregivers of adolescents are typically a parent or family member who assumes a central role in caregiving. Transition into the caregiver role can be different for each caregiver. Recognizing the individual needs of each caregiver is crucial when investigating the usability of communication technology interventions [9,116], especially in caregivers of children or adolescents with mental health issues [67,117,118].

Caregivers' adoption of tailored web-based interventions focusing on their needs may improve their resourcefulness and mobilize effective caregiving [119,120]. Tailored communication technologies was highlighted in this review, indicating contextual relevance related to the usage of communication technologies. Besides, this review also highlighted common elements depicting their approval of utilizing communication technologies, such as the accessibility, cost-effectiveness, convenience, and flexibility thereof. Generally, study participants gave positive feedback related to communication technologies use. Modifying communication technology interventions to match caregivers' preferences might lead to favorable changes in parenting practices and satisfaction in intervention utility [121]. Overall, the findings demonstrated the acceptability of communication technologies by caregivers, and the use of communication technologies was associated with improved caregiver and mental health outcomes among adolescents.

Some studies in this review identified diverse interventions that facilitated the strengthening of caregivers' behavior, such as improved self-efficacy, enhanced knowledge, and better parent-child communication skills and practices. These findings complement those of other studies in that online tools have proved to be successful in improving caregivers' knowledge, skills, coping [122,123], and self-efficacy [120,124]. Nieuwboer et al [99] summarized that the internet supplies information, support and advice to parents with different needs that encourage changes in their parental abilities. Furthermore, technology provides access to others for building support and knowledge through positive engagement [125]. Reportedly, the potential impact of communication technologies on caregivers' well-being is ubiquitous and may guide their behavior toward the care recipient [126].

Caregivers who feel that they can perform behavior effectively will continue to repeat that specific behavior. This performance is based on knowledge, skills, and self-efficacy. A repetition of the behavior will occur if it is associated with a positive feeling and a sense of confidence; the accomplishment of the behavior, reflects self-efficacy and competency in performing the task [127,128]. People hold specific beliefs about behavior and intentionally perform according to that belief. From the IMBP, people who have the necessary skills will perform more satisfactorily, which will lead to favorable outcomes [129]. Hall and Bierman [130] reported improvement in parental knowledge and attitude through technology-based interventions. Of course, targeted interventions strengthening parental knowledge, self-efficacy, and skills may lead to improved child mental health outcomes [100] and positive parental outcomes [9].

Thus, it is legitimate to say that a communication technologies–based intervention has great potential to be used by caregivers to reinforce positive behavior when caring for an adolescent with mental health issues. Communication technologies also represent possibilities to provide support to caregivers by addressing their individual needs which could be highly convoluted. This review shows a need for further research in the area of supporting caregivers of adolescents with mental health issues.

Strengths

The authors conducted a comprehensive review following a stepwise methodological process for a systematic review. We included multiple study designs to understand how communication technologies may strengthen caregivers' behavior and improve caregiver outcomes. We used standardized checklists to appraise selected articles critically. The results provided knowledge related to the use of communication technologies among caregivers of adolescents with mental health issues.

Limitations

We might have missed potentially relevant articles, although explicit inclusion and exclusion criteria were set up. The first author initially screened titles and abstracts independently for applicability while some were verified by the second author for quality control purposes. This review was limited to articles describing caregiver outcomes according to the IMBP determinants, and this restriction may have eliminated meaningful information from other communication technology advantages.

Conclusions

The review results indicate that using communication technology is useful to strengthen caregivers' behavior by providing information and resources. Additionally, a better understanding of caregivers' attitudes and environmental constraints toward communication technologies may inform optimal usefulness thereof. Evidence for caregivers of adolescents with mental health issues using communication technologies is not readily available in the literature. There was limited empirical research outlining the methods of communication delivery, and it is worth exploring this in future research. Besides, future research should focus on the development of more innovative communication technology interventions for caregivers in different contexts and scaling up of efforts to implement them for improved mental health care among adolescents.

Acknowledgments

Authors' Contributions

RJ and MR conceptualized and designed the study. RJ and MR developed search strategies. RJ and MR conducted data extraction, analysis, and synthesis. MR contributed to the revision and approval of the manuscript.

Conflicts of Interest

None declared.

References

1. Delbanco T, Berwick D, Boufford J, Edgman-Levitan P, Ollenschläger G, Plamping D. Healthcare in a land called peoplepower: Nothing about me without me. Heal Expect 2001;4:144-150. [doi: 10.1046/j.1369-6513.2001.00145.x]

2. Margolis KL, Fosco GM, Stormshak EA. Circle of Care: Extending Beyond Primary Caregivers to Examine Collaborative Caretaking in Adolescent Development. J Fam Issues 2016 Jun;37(9):1179-1202 [FREE Full text] [doi: 10.1177/0192513X14536565] [Medline: 27453615]

3. Adelman RD, Tmanova LL, Delgado D, Dion S, Lachs MS. Caregiver Burden. JAMA 2014 Mar 12;311(10):1052. [doi: 10.1001/jama.2014.304]

4. Dookie S, Singh S. Primary health services at district level in South Africa: a critique of the primary health care approach. BMC Fam Pract 2012 Jul 2;13(1). [doi: 10.1186/1471-2296-13-67]

5. McInnis MG, Merajver SD. Global mental health: Global strengths and strategies. Asian Journal of Psychiatry 2011 Sep;4(3):165-171. [doi: 10.1016/j.ajp.2011.06.002]

6. Healy EA, Kaiser BN, Puffer ES. Family-based youth mental health interventions delivered by nonspecialist providers in low- and middle-income countries: A systematic review. Families, Systems, & Health 2018 Jun;36(2):182-197. [doi: 10.1037/fsh0000334]

7. Angothu H, Chaturvedi S. Civic and legal advances in the rights of caregivers for persons with severe mental illness related disability. Indian J Soc Psychiatry 2016;32(1):28. [doi: 10.4103/0971-9962.176764]

8. Musyimi CW, Mutiso VN, Ndetei DM, Unanue I, Desai D, Patel SG, et al. Mental health treatment in Kenya: task-sharing challenges and opportunities among informal health providers. Int J Ment Health Syst 2017 Aug 1;11(1). [doi: 10.1186/s13033-017-0152-4]

9. Talley R, Fricchione G, Druss B. Introduction: Caregiving to promote mental health and prevent mental illness. In: Talley R, editor. The challenges of mental health caregiving. Research - Practice - Policy. New York: Springer; 2014:1-14.

10. Collins LG, Swartz K. Caregiver care. Am Fam Physician 2011 Jun 01;83(11):1309-1317 [FREE Full text] [Medline: 21661713]

11. Penning MJ, Wu Z. Caregiver Stress and Mental Health: Impact of Caregiving Relationship and Gender. GERONT 2015 Apr 17;56(6):1102-1113. [doi: 10.1093/geront/gnv038]

12. Joslin D. Introduction. In: Joslin D. editor. Invisible care givers: Older Adults Raising Children in the Wake of HIV/AIDS. New York: Columbia University Press; 2002:1-26.

13. Mayberry LS, Heflinger CA. How Caregivers Make Meaning of Child Mental Health Problems: Toward Understanding Caregiver Strain and Help Seeking. Fam Soc 2013 May;94(2):105-113 [FREE Full text] [doi: 10.1606/1044-3894.4286] [Medline: 27293367]

14. Bergström AL, Hanson E. An integrative review of information and communication technology based support interventions for carers of home dwelling older people. TAD 2018 Feb 05;29(1-2):1-14. [doi: 10.3233/tad-160158]

15. Hughes S, Shuman S, Wiener J, Gould E. Background paper. In: Research Summit on Dementia Care. Washington: RTI International; Mar 2017:12.

16. Chang S, Zhang Y, Jeyagurunathan A, Lau YW, Sagayadevan V, Chong SA, et al. Providing care to relatives with mental illness: reactions and distress among primary informal caregivers. BMC Psychiatry 2016 Mar 25;16:80 [FREE Full text] [doi: 10.1186/s12888-016-0786-9] [Medline: 27016185]

17. Roth D, Fredman L, Haley W. Informal caregiving and its impact on health: A reappraisal from population-based studies. Gerontologist 2015;55(2):309-319. [doi: 10.1093/geront/gnu177]

18. Jackson D, Roberts G, Wu ML, Ford R, Doyle C. A systematic review of the effect of telephone, internet or combined support for carers of people living with Alzheimer's, vascular or mixed dementia in the community. Archives of Gerontology and Geriatrics 2016 Sep;66:218-236. [doi: 10.1016/j.archger.2016.06.013]

XSL•FO
RenderX

19. Waller A, Dilworth S, Mansfield E, Sanson-Fisher R. Computer and telephone delivered interventions to support caregivers of people with dementia: a systematic review of research output and quality. BMC Geriatr 2017 Nov 16;17(1). [doi: 10.1186/s12877-017-0654-6]

20. Lambert SD, Girgis A. Unmet Supportive Care Needs Among Informal Caregivers of Patients with Cancer: Opportunities and Challenges in Informing the Development of Interventions. Asia Pac J Oncol Nurs 2017;4(2):136-139 [FREE Full text] [doi: 10.4103/2347-5625.204485] [Medline: 28503646]

21. Luker K, Cooke M, Dunn L, Lloyd-Williams M, Pilling M, Todd C. Development and evaluation of an intervention to support family caregivers of people with cancer to provide home-based care at the end of life: a feasibility study. Eur J Oncol Nurs 2015 Apr;19(2):154-161. [doi: 10.1016/j.ejon.2014.09.006] [Medline: 25667125]

22. Mbuagbaw L, Mursleen S, Lytvyn L, Smieja M, Dolovich L, Thabane L. Mobile phone text messaging interventions for HIV and other chronic diseases: an overview of systematic reviews and framework for evidence transfer. BMC Health Serv Res 2015 Jan 22;15(1). [doi: 10.1186/s12913-014-0654-6]

23. Winskell K, Miller KS, Allen KA, Obong'o CO. Guiding and supporting adolescents living with HIV in sub-Saharan Africa: The development of a curriculum for family and community members. Children and Youth Services Review 2016 Feb;61:253-260. [doi: 10.1016/j.childyouth.2015.12.017]

24. 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. [doi: 10.2196/mhealth.7798]

25. Kim J, Lee S. Communication and cybercoping: coping with chronic illness through communicative action in online support networks. J Health Commun 2014 Feb;19(7):775-794. [doi: 10.1080/10810730.2013.864724] [Medline: 24559492]

26. Anstey Watkins J, Goudge J, Gómez-Olivé FX, Huxley C, Dodd K, Griffiths F. mHealth text and voice communication for monitoring people with chronic diseases in low-resource settings: a realist review. BMJ Glob Health 2018 Mar;3(2):e000543 [FREE Full text] [doi: 10.1136/bmjgh-2017-000543] [Medline: 29527356]

27. 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]

28. Chen H, Chai Y, Dong L, Niu W, Zhang P. Effectiveness and Appropriateness of mHealth Interventions for Maternal and Child Health: Systematic Review. JMIR Mhealth Uhealth 2018 Jan 09;6(1):e7. [doi: 10.2196/mhealth.8998]

29. Phillips JH, Wigger C, Beissbarth J, McCallum GB, Leach A, Morris PS. Can mobile phone multimedia messages and text messages improve clinic attendance for Aboriginal children with chronic otitis media? A randomised controlled trial. J Paediatr Child Health 2014 Feb 25;50(5):362-367. [doi: 10.1111/jpc.12496]

30. Smith J, Cheater F, Bekker H. Parents' experiences of living with a child with a long-term condition: a rapid structured review of the literature. Health Expect 2013 Jan 14;18(4):452-474. [doi: 10.1111/hex.12040]

31. Rodríguez-Meirinhos A, Antolín-Suárez L, Oliva A. Support Needs of Families of Adolescents With Mental Illness: A Systematic Mixed Studies Review. Arch Psychiatr Nurs 2018 Dec;32(1):152-163. [doi: 10.1016/j.apnu.2017.09.004] [Medline: 29413065]

32. O'Loughlin K, Althoff RR, Hudziak JJ. Health Promotion and Prevention in Child and Adolescent Mental Health. In: Rey JM, editor. IACAPAP e-Textbook of Child and Adolescent Mental Health. Geneva: International Association for Child and Adolescent Psychiatry and Allied Professions; 2017:1-25.

33. Cardamone-Breen MC, Jorm AF, Lawrence KA, Rapee RM, Mackinnon AJ, Yap MBH. A Single-Session, Web-Based Parenting Intervention to Prevent Adolescent Depression and Anxiety Disorders: Randomized Controlled Trial. J Med Internet Res 2018 Apr 26;20(4):e148. [doi: 10.2196/jmir.9499]

34. Lund C, Brooke-Sumner C, Baingana F, Baron EC, Breuer E, Chandra P, et al. Social determinants of mental disorders and the Sustainable Development Goals: a systematic review of reviews. The Lancet Psychiatry 2018 Apr;5(4):357-369. [doi: 10.1016/s2215-0366(18)30060-9]

35. Patel V, Chisholm D, Parikh R, Charlson FJ, Degenhardt L, Dua T, et al. Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities, 3rd edition. The Lancet 2016 Apr;387(10028):1672-1685. [doi: 10.1016/s0140-6736(15)00390-6]

36. Pedersen GA, Smallegange E, Coetzee A, Hartog K, Turner J, Jordans MJD, et al. A Systematic Review of the Evidence for Family and Parenting Interventions in Low- and Middle-Income Countries: Child and Youth Mental Health Outcomes. J Child Fam Stud 2019 Apr 29;28(8):2036-2055. [doi: 10.1007/s10826-019-01399-4]

37. Paruk S, Karim E. Update on adolescent mental health. S Afr Med J 2016 May 14;106(6):548. [doi: 10.7196/samj.2016.v106i6.10943]

38. Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, Bolton P, et al. The Lancet Commission on global mental health and sustainable development. The Lancet 2018 Oct;392(10157):1553-1598. [doi: 10.1016/s0140-6736(18)31612-x]

39. Erskine HE, Baxter AJ, Patton G, Moffitt TE, Patel V, Whiteford HA, et al. The global coverage of prevalence data for mental disorders in children and adolescents. Epidemiol Psychiatr Sci 2016 Jan 20;26(4):395-402. [doi: 10.1017/s2045796015001158]

XSL•FO
RenderX

40. Armoiry X, Sturt J, Phelps EE, Walker C, Court R, Taggart F, et al. Digital Clinical Communication for Families and Caregivers of Children or Young People With Short- or Long-Term Conditions: Rapid Review. J Med Internet Res 2018 Jan 05;20(1):e5. [doi: 10.2196/jmir.7999]

41. Kuhn ES, Laird RD. Family support programs and adolescent mental health: review of evidence. Adolesc Health Med Ther 2014 Jul;5:127-142 [FREE Full text] [doi: 10.2147/AHMT.S48057] [Medline: 25177156]

42. Maar MA, Yeates K, Perkins N, Boesch L, Hua-Stewart D, Liu P, et al. A Framework for the Study of Complex mHealth Interventions in Diverse Cultural Settings. JMIR Mhealth Uhealth 2017 Apr 20;5(4):e47. [doi: 10.2196/mhealth.7044]

43. Sheets DJ, Black K, Kaye LW. Who cares for caregivers? Evidence-based approaches to family support. J Gerontol Soc Work 2014 Aug;57(6-7):525-530. [doi: 10.1080/01634372.2014.920606] [Medline: 25105462]

44. Metzler CW, Sanders MR, Rusby JC, Crowley RN. Using consumer preference information to increase the reach and impact of media-based parenting interventions in a public health approach to parenting support. Behav Ther 2012 Jun;43(2):257-270 [FREE Full text] [doi: 10.1016/j.beth.2011.05.004] [Medline: 22440064]

45. Issel L. Health Program Planning and Evaluation. A practical, systematic approach for community health. Third edit. Burlington: Jones & Bartlett Learning books; 2014:1-613.

46. Yesufu-Udechuku A, Harrison B, Mayo-Wilson E, Young N, Woodhams P, Shiers D, et al. Interventions to improve the experience of caring for people with severe mental illness: systematic review and meta-analysis. Br J Psychiatry 2015 Apr;206(4):268-274. [doi: 10.1192/bjp.bp.114.147561] [Medline: 25833867]

47. Davis TL, DiClemente R, Prietula M. Taking mHealth Forward: Examining the Core Characteristics. JMIR Mhealth Uhealth 2016 Aug 10;4(3):e97 [FREE Full text] [doi: 10.2196/mhealth.5659] [Medline: 27511612]

48. Delponte L, Grigolini M, Moroni A, Vignetti S, Claps M. ICT in the Developing World. Scientific Foresight Unit. Brussels: European Parliamentary Research Service; 2015 Dec. URL: https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2015)563482 [accessed 2020-08-17]

49. Orlowski SK, Lawn S, Venning A, Winsall M, Jones GM, Wyld K, et al. Participatory Research as One Piece of the Puzzle: A Systematic Review of Consumer Involvement in Design of Technology-Based Youth Mental Health and Well-Being Interventions. JMIR Human Factors 2015 Jul 09;2(2):e12. [doi: 10.2196/humanfactors.4361]

50. Hemmat M, Ayatollahi H, Maleki M, Saghafi F. Future Research in Health Information Technology: A Review. Perspect Heal Inf Manag 2017 Jan 01;14(Winter):A. [Medline: 28566991]

51. Baker S. Broadening the reach of evidence-based parenting interventions: Evaluation of a brief online version of the Triple P - Positive Parenting Program. Doctoral thesis. Queensland: University of Queensland; 2016. URL: https://espace.library.uq.edu.au/view/UQ:466891 [accessed 2020-08-18]

52. Hou S, Charlery SR, Roberson K. Systematic literature review of Internet interventions across health behaviors. Health Psychol Behav Med 2014 Jan 1;2(1):455-481 [FREE Full text] [doi: 10.1080/21642850.2014.895368] [Medline: 25750795]

53. Mahmud AJ, Olander E, Eriksén S, Haglund BJ. Health communication in primary health care -a case study of ICT development for health promotion. BMC Med Inform Decis Mak 2013 Jan 30;13:17 [FREE Full text] [doi: 10.1186/1472-6947-13-17] [Medline: 23363566]

54. Vergunst R. From global-to-local: rural mental health in South Africa. Glob Health Action 2018 Jan;11(1):1413916. [doi: 10.1080/16549716.2017.1413916] [Medline: 29320947]

55. MacDonell KW, Prinz RJ. A Review of Technology-Based Youth and Family-Focused Interventions. Clin Child Fam Psychol Rev 2016 Oct 27;20(2):185-200. [doi: 10.1007/s10567-016-0218-x]

56. Nagata JM, Ferguson BJ, Ross DA. Research Priorities for Eight Areas of Adolescent Health in Low- and Middle-Income Countries. J Adolesc Health 2016 Dec;59(1):50-60 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.03.016] [Medline: 27235375]

57. Breitenstein SM, Shane J, Julion W, Gross D. Developing the CPP: Adapting an Evidence-Based Parent Training Program for Digital Delivery in Primary Care Settings. Worldviews on Evidence-Based Nursing 2015 Jan 14;12(1):31-40. [doi: 10.1111/wvn.12074]

58. DeHoff BA, Staten LK, Rodgers RC, Denne SC. The Role of Online Social Support in Supporting and Educating Parents of Young Children With Special Health Care Needs in the United States: A Scoping Review. J Med Internet Res 2016 Dec 22;18(12):e333. [doi: 10.2196/jmir.6722]

59. Sweeney GM, Donovan CL, March S, Laurenson SD. Logging into therapy: Parent attitudes and intentions to use computer-based therapies for youth mental health. Internet Interventions 2015 Nov;2(4):437-445. [doi: 10.1016/j.invent.2015.11.001]

60. Schwinn TM, Schinke S, Fang L, Kandasamy S. A web-based, health promotion program for adolescent girls and their mothers who reside in public housing. Addict Behav 2014 Apr;39(4):757-760 [FREE Full text] [doi: 10.1016/j.addbeh.2013.11.029] [Medline: 24447886]

61. Russell BS, Maksut JL, Lincoln CR, Leland AJ. Computer-mediated parenting education: Digital family service provision. Children and Youth Services Review 2016 Mar;62:1-8. [doi: 10.1016/j.childyouth.2016.01.012]

62. 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. [doi: 10.2196/mhealth.2994]

63. Casale M, Cluver L, Crankshaw T, Kuo C, Lachman J, Wild L. Direct and Indirect Effects of Caregiver Social Support on Adolescent Psychological Outcomes in Two South African AIDS-Affected Communities. Am J Community Psychol 2015 Jan 27;55:336-346. [doi: 10.1007/s10464-015-9705-3]

64. Robila M. Families , Mental Health and Well Being: Pursuing Sustainable Development Goal 3. In: Family policies and the 2030 Sustainable Development Agenda. New York: United Nations; 2016 May Presented at: Family policies and the 2030 Sustainable Development Agenda; 12-13 May 2016; New York p. 1-15 URL: https://www.un.org/development/desa/family/meetings-events/family-policies-and-the-2030-sustainable-development-agenda.html

65. Carretero S, Stewart J, Centeno C. Information and communication technologies for informal carers and paid assistants: benefits from micro-, meso-, and macro-levels. Eur J Ageing 2015 Jan 24;12(2):163-173. [doi: 10.1007/s10433-015-0333-4]

66. Jang J, Hessel H, Dworkin J. Parent ICT use, social capital, and parenting efficacy. Computers in Human Behavior 2017 Jun;71:395-401. [doi: 10.1016/j.chb.2017.02.025]

67. Isa SNI, Ishak I, Ab Rahman A, Mohd Saat NZ, Che Din N, Lubis SH, et al. Health and quality of life among the caregivers of children with disabilities: A review of literature. Asian J Psychiatr 2016 Oct;23:71-77. [doi: 10.1016/j.ajp.2016.07.007] [Medline: 27969083]

68. Fishbein M, Ajzen I. Predicting and changing behavior. The Reasoned Action Approach. New York: Psychology Press. Taylor & Francis Group; 2010:1-539.

69. Dai M, Wombacher K, Matig JJ, Harrington NG. Using the Integrative Model of Behavioral Prediction to Understand College Students' Hookup Sex Beliefs, Intentions, and Behaviors. Health Commun 2018 Sep;33(9):1078-1087. [doi: 10.1080/10410236.2017.1331306] [Medline: 28622006]

70. Montano DE, Kasprzyk D. Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In: Glanz K, Rimer BK, Viswanath K, editors. Health behavior and Health education. Theory, Research and Practice. 4th editio. San Francisco: John Wiley & Sons, Inc; 2008:67-96.

71. Rey N. Adhering to healthy behavior. In: Brannon L, Feist J, Updegraff JA, editors. Health Psychology: An Introduction to Behavior and Health. Belmont: Wadsworth: Cengage Learning; 2013:58-86.

72. Fishbein M, Yzer MC. Using Theory to Design Effective Health Behavior Interventions. Commun Theory 2003 May;13(2):164-183. [doi: 10.1111/j.1468-2885.2003.tb00287.x]

73. Gough D, Oliver S, Thomas J. Introducing systematic reviews. In: Gough D, Oliver S, Thomas J, editors. An introduction to Systematic Reviews. London: SAGE; 2012:1-16.

74. Hong QN, Pluye P. A Conceptual Framework for Critical Appraisal in Systematic Mixed Studies Reviews. Journal of Mixed Methods Research 2018 Apr 21;13(4):446-460. [doi: 10.1177/1558689818770058]

75. Petticrew M, Roberts H. Why do we need systematic reviews? In: Systematic reviews in the social sciences. Malden: Blackwell Publishing Ltd; 2006:1-26.

76. CASP. Critical Appraisal Skills Programme (CASP). https://casp-uk.net/casp-tools-checklists/. Oxford: CASP; 2018. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2018-10-01]

77. Godfrey C, Harrison M. Systematic review resource package. The Joanna Briggs Institute method for systematic review research quick reference guide. 2015. URL: http://healthindisasters.com/images/Books/Systematic-Review-Resource-Package.pdf [accessed 2016-08-28]

78. Academy of Nutrition and Dietetics. Evidence Analysis Manual: Steps in the Academy Evidence Analysis Process. Academy of Nutrition and Dietetics. Chicago: Evidence Analysis Library; 2016 Apr. URL: https://www.andeal.org/evidence-analysis-manual [accessed 2020-08-18]

79. Salmond SW. Steps in the systematic review process. In: Holly C, Salmond SW, Saimbert MK, editors. Comprehensive systematic review for advanced nursing practice. New York: Springer; 2012:13-32.

80. Deitz DK, Cook RF, Billings DW, Hendrickson A. A web-based mental health program: reaching parents at work. J Pediatr Psychol 2009 Jun;34(5):488-494 [FREE Full text] [doi: 10.1093/jpepsy/jsn108] [Medline: 18845585]

81. Choi H, Kim S, Ko H, Kim Y, Park CG. Development and preliminary evaluation of culturally specific web-based intervention for parents of adolescents. J. Psychiatr. Ment. Health Nurs 2016 Aug 08;23(8):489-501. [doi: 10.1111/jpm.12327]

82. Molleda L, Bahamon M, St George SM, Perrino T, Estrada Y, Correa Herrera D, et al. Clinic Personnel, Facilitator, and Parent Perspectives of eHealth Familias Unidas in Primary Care. J Pediatr Health Care 2017 May;31(3):350-361 [FREE Full text] [doi: 10.1016/j.pedhc.2016.11.001] [Medline: 28012799]

83. Estrada Y, Molleda L, Murray A, Drumhiller K, Tapia M, Sardinas K, et al. eHealth Familias Unidas: Pilot Study of an Internet Adaptation of an Evidence-Based Family Intervention to Reduce Drug Use and Sexual Risk Behaviors Among Hispanic Adolescents. IJERPH 2017 Mar 04;14(3):264. [doi: 10.3390/ijerph14030264]

84. Harden A, Thomas J, Cargo M, Harris J, Pantoja T, Flemming K, et al. Cochrane Qualitative and Implementation Methods Group guidance series-paper 5: methods for integrating qualitative and implementation evidence within intervention effectiveness reviews. J Clin Epidemiol 2018 May;97:70-78. [doi: 10.1016/j.jclinepi.2017.11.029] [Medline: 29242095]

85. Sandelowski M, Leeman J, Knafl K, Crandell JL. Text-in-context: a method for extracting findings in mixed-methods mixed research synthesis studies. J Adv Nurs 2012 Aug 27;69(6):1428-1437. [doi: 10.1111/jan.12000]

86. Thomas J, Harden A, Newman M. Synthesis: combining results systematically and appropriately. In: Gough D, Oliver S, Thomas J, editors. An introduction to Systematic Reviews. London: SAGE; 2012:66-82.

87.   Taherdoost H. A review of technology acceptance and adoption models and theories. Procedia Manufacturing 2018;22:960-967. [doi: 10.1016/j.promfg.2018.03.137]

88.   Brinkel J, Dako-Gyeke P, Krämer A, May J, Fobil J. An investigation of users' attitudes, requirements and willingness to use mobile phone-based interactive voice response systems for seeking healthcare in Ghana: a qualitative study. Public Health 2017 Mar;144:125-133. [doi: 10.1016/j.puhe.2016.11.017]

89.   Oldenburg B, Taylor CB, O'Neil A, Cocker F, Cameron LD. Using New Technologies to Improve the Prevention and Management of Chronic Conditions in Populations. Annu. Rev. Public Health 2015 Mar 18;36(1):483-505. [doi: 10.1146/annurev-publhealth-031914-122848]

90.   Christodoulakis C, Asgarian A, Easterbrook S. Barriers to Adoption of Information Technology in Healthcare. In: Proceedings of the 27th Annual International Conference on Computer Science and Software Engineering. 2017 Nov Presented at: ACM CASCON Conference; 2017; Toronto p. 66-75. [doi: 10.4018/978-1-60566-988-5.ch009]

91.   Park E, Kim H, Steinhoff A. Health-related internet use by informal caregivers of children and adolescents: An integrative literature review. J Med Internet Res 2016;18(3):1-10. [doi: 10.2196/jmir.4124]

92.   Celebic G, Rendulic D. Basic Concepts of Information and Communication Technology. Handbook. Zagreb: Open Society for Idea Exchange (ODRAZI); 2011:1-39.

93.   David L. Introduction to information and communication technologies. Module 1. In: Large JA, editor. ICT for Library and Information Professionals: A Training package for Developing Countries. Bangkok: UNESCO; 2001:1-192.

94.   Schneider S. Information and Communication Technologies in Public Child Welfare: A Systematic Literature Review. Master thesis. San Bernardino: California State University; 2017 Jun. URL: http://scholarworks.lib.csusb.edu/etd/473 [accessed 2018-03-22]

95.   Glascoe FP, Trimm F. Brief approaches to developmental-behavioral promotion in primary care: updates on methods and technology. Pediatrics 2014 May;133(5):884-897 [FREE Full text] [doi: 10.1542/peds.2013-1859] [Medline: 24777220]

96.   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 Aug 17;3(3):e38. [doi: 10.2196/mental.6061]

97.   Mitchell SJ, Godoy L, Shabazz K, Horn IB. Internet and Mobile Technology Use Among Urban African American Parents: Survey Study of a Clinical Population. J Med Internet Res 2014 Jan 13;16(1):e9. [doi: 10.2196/jmir.2673]

98.   Heynsbergh N, Heckel L, Botti M, Livingston PM. Feasibility, useability and acceptability of technology-based interventions for informal cancer carers: a systematic review. BMC Cancer 2018 Mar 2;18(1). [doi: 10.1186/s12885-018-4160-9]

99.   Nieuwboer CC, Fukkink RG, Hermanns JM. Online programs as tools to improve parenting: A meta-analytic review. Children and Youth Services Review 2013 Nov;35(11):1823-1829. [doi: 10.1016/j.childyouth.2013.08.008]

100. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet Use and eHealth Literacy of Low-Income Parents Whose Children Have Special Health Care Needs. J Med Internet Res 2011 Sep 29;13(3):e75. [doi: 10.2196/jmir.1697]

101. Pivec F. The global information technology report 2003–2004. OZ 2003;8(4):203-206. [doi: 10.3359/oz0304203]

102. Riley W. Theoretical models to inform technology-based health behavior interventions. In: Marsch LA, Lord SE, Dallery J, editors. Behavioral healthcare and technology. Using Science-based innovations to transform practice. New York: Oxfords University Press; 2015.

103. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. The Lancet Psychiatry 2017 Jun;4(6):486-500. [doi: 10.1016/s2215-0366(17)30096-2]

104. Galan-Caridad JM, Harel S, Arenzana TL, Hou ZE, Doetsch FK, Mirny LA, et al. Zfx controls the self-renewal of embryonic and hematopoietic stem cells. Cell 2007 Apr 20;129(2):345-357 [FREE Full text] [doi: 10.1016/j.cell.2007.03.014] [Medline: 17448993]

105. Sprague K, Manyika J, Chappuis B, Bughin J, Grijpink F, Moodley L. Offline and falling behind: Barriers to Internet adoption. Technology, Media, and Telecom Practice.: McKinsey & Company; 2014 Oct. URL: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/offline-and-falling-behind-barriers-to-internet-adoption# [accessed 2018-10-31]

106. Domek GJ, Contreras-Roldan IL, Asturias EJ, Bronsert M, Bolaños Ventura GA, O'Leary ST, et al. Characteristics of mobile phone access and usage in rural and urban Guatemala: assessing feasibility of text message reminders to increase childhood immunizations. Mhealth 2018 Apr;4:9 [FREE Full text] [doi: 10.21037/mhealth.2018.03.05] [Medline: 29780812]

107. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth Technology Use and Implications in Historically Underserved and Minority Populations in the United States: Systematic Literature Review. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e128. [doi: 10.2196/mhealth.8383]

108. Chacko A, Isham A, Cleek AF, McKay MM. Using mobile health technology to improve behavioral skill implementation through homework in evidence-based parenting intervention for disruptive behavior disorders in youth: study protocol for intervention development and evaluation. Pilot Feasibility Stud 2016 Sep;2:57 [FREE Full text] [doi: 10.1186/s40814-016-0097-4] [Medline: 27965873]

109. Breitenstein SM, Gross D, Christophersen R. Digital Delivery Methods of Parenting Training Interventions: A Systematic Review. Worldviews on Evidence-Based Nursing 2014 May 19;11(3):168-176. [doi: 10.1111/wvn.12040]

XSL•FO

RenderX

110. DeHoff BA, Staten LK, Rodgers RC, Denne SC. The Role of Online Social Support in Supporting and Educating Parents of Young Children With Special Health Care Needs in the United States: A Scoping Review. J Med Internet Res 2016 Dec 22;18(12):e333. [doi: 10.2196/jmir.6722]

111. Francomano JA, Harpin SB. Utilizing Social Networking Sites to Promote Adolescents' Health. CIN: Computers, Informatics, Nursing 2015;33(1):10-20. [doi: 10.1097/cin.0000000000000113]

112. Catalano D, Holloway L, Mpofu E. Mental Health Interventions for Parent Carers of Children with Autistic Spectrum Disorder: Practice Guidelines from a Critical Interpretive Synthesis (CIS) Systematic Review. IJERPH 2018 Feb 14;15(2):341. [doi: 10.3390/ijerph15020341]

113. Gavazzi S. Web-based resources on family programs. In: Levesque R, editor. Families with adolescents. Bridging the gaps between theory, research, and practice. New York: Springer USA; 2011:134.

114. Rimer B, Kreuter M. Advancing tailored health communication: A persuasion and message effects perspective. Journal of Communication 2006;56:S184-S201. [doi: 10.1111/j.1460-2466.2006.00289.x]

115. Rexhaj S, Leclerc C, Bonsack C, Golay P, Favrod J. Feasibility and Accessibility of a Tailored Intervention for Informal Caregivers of People with Severe Psychiatric Disorders: a Pilot Study. Front. Psychiatry 2017 Sep 21;8. [doi: 10.3389/fpsyt.2017.00178]

116. Fernandes CS, Angelo M. Family caregivers: what do they need? An integrative review. Rev. esc. enferm. USP 2016 Aug;50(4):675-682. [doi: 10.1590/s0080-623420160000500019]

117. Frauenholtz S, Conrad-Hiebner A, Mendenhall AN. Children's Mental Health Providers' Perceptions of Mental Health Literacy Among Parents and Caregivers. Journal of Family Social Work 2015 Jan 22;18(1):40-56. [doi: 10.1080/10522158.2014.974116]

118. Mendenhall E, De Silva MJ, Hanlon C, Petersen I, Shidhaye R, Jordans M, et al. Acceptability and feasibility of using non-specialist health workers to deliver mental health care: stakeholder perceptions from the PRIME district sites in Ethiopia, India, Nepal, South Africa, and Uganda. Soc Sci Med 2014 Oct;118:33-42 [FREE Full text] [doi: 10.1016/j.socscimed.2014.07.057] [Medline: 25089962]

119. Wozney L, Radomski AD, Newton AS. The Gobbledygook in Online Parent-Focused Information about Child and Adolescent Mental Health. Health Communication 2017 Apr 12;33(6):710-715. [doi: 10.1080/10410236.2017.1306475]

120. Ploeg J, Markle-Reid M, Valaitis R, McAiney C, Duggleby W, Bartholomew A, et al. Web-Based Interventions to Improve Mental Health, General Caregiving Outcomes, and General Health for Informal Caregivers of Adults With Chronic Conditions Living in the Community: Rapid Evidence Review. J Med Internet Res 2017 Jul 28;19(7):e263. [doi: 10.2196/jmir.7564]

121. Garcia-Huidobro D, Diaspro-Higuera MO, Palma D, Palma R, Ortega L, Shlafer R, et al. Adaptive Recruitment and Parenting Interventions for Immigrant Latino Families with Adolescents. Prev Sci 2018 Apr 11. [doi: 10.1007/s11121-018-0898-1] [Medline: 29644545]

122. McNabb DE, Gibson LK, Finnie BW. The Case of the Vanishing Workforce. Public Performance & Management Review 2006 Mar 1;29(3):358-368. [doi: 10.2753/PMR1530-9576290306]

123. Sajedi F, Habibi E, Shahshahanipour S, Hatamizadeh N. An Approach towards Promoting Iranian Caregivers' Knowledge on Early Childhood Development. Int J Pediatr 2018 Mar;6(3):7371-7382. [doi: 10.22038/ijp.2017.27419.2364]

124. Dumka LE, Gonzales NA, Wheeler LA, Millsap RE. Parenting self-efficacy and parenting practices over time in Mexican American families. Journal of Family Psychology 2010;24(5):522-531. [doi: 10.1037/a0020833]

125. Knight K, Hunter C. Using technology in service delivery to families , children and young people. Child Family Community Australia. Melbourne: Australian Institute of Family Studies; 2013 Oct. URL: https://aifs.gov.au/cfca/publications/using-technology-service-delivery-families-children-and-young-people [accessed 2018-10-20]

126. Onwumere J, Kuipers E. Caregiving roles: when will they be routinely recognized and supported? Journal of Mental Health 2017 Mar 29;26(2):95-97. [doi: 10.1080/09638237.2017.1301649]

127. Jaccard J, Levitz N. Parent-based interventions to reduce adolescent problem behaviors: New directions for self-regulation approaches. In: Oettingen G, Gollwitzer P, editors. Self-Regulation Adolescence. New York: Cambridge University Press; 2013:357-388.

128. Dunst C. Parent Involvement in Early Childhood Intervention: Characteristics of family capacity-building practices. In: 7th Annual Conference on Meeting the Therapy Needs of Infants, Toddlers and Children with Disabilities in Rural Communities. 2016 Mar 12 Presented at: 7th Annual Conference on meeting the Therapy needs of infants, toddlers and children with disabiities in rural communities; 2016; Lexington p. 1-38. [doi: 10.4324/9781315688442]

129. Yzer M. The Integrative Model of Behavioral Prediction as a Tool for Designing Health Messages: Theory and Practice. In: Cho H, editor. Designing Messages for Health Communication Campaigns. Thousand Oaks: SAGE; 2012:21-40.

130. Hall CM, Bierman KL. Technology-assisted interventions for parents of young children: Emerging practices, current research, and future directions. Early Childhood Research Quarterly 2015;33:21-32. [doi: 10.1016/j.ecresq.2015.05.003]

Abbreviations

eHealth: electronic health
IMBP: Integrated Model of Behavior Prediction

mHealth: mobile health

PICO: population, intervention, comparison intervention, outcome

XSL•FO
RenderX

Original Paper

Complementing Human Behavior Assessment by Leveraging Personal Ubiquitous Devices and Social Links: An Evaluation of the Peer-Ceived Momentary Assessment Method

Allan Berrocal[1], BSc, MSc; Waldo Concepcion[2], MD, FACS; Stefano De Dominicis[3], PhD; Katarzyna Wac[1,4], BSc, MSc, PhD

[1]Quality of Life Technologies Lab, Department of Computer Science, University of Geneva, Carouge, Switzerland

[2]Division Of MultiOrgan Transplantation, Stanford University Medical Center, Stanford University, Palo Alto, CA, United States

[3]Department of Nutrition, Exercise and Sports, University of Copenhagen, Copenhagen, Denmark

[4]Quality of Life Technologies Lab, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Katarzyna Wac, BSc, MSc, PhD
Quality of Life Technologies Lab
Department of Computer Science
University of Geneva
Route de Drize 7
Carouge, 1227
Switzerland
Phone: 41 222790242
Email: wac@stanford.edu

Abstract

Background: Ecological momentary assessment (EMA) enables individuals to self-report their subjective momentary physical and emotional states. However, certain conditions, including routine observable behaviors (eg, moods, medication adherence) as well as behaviors that may suggest declines in physical or mental health (eg, memory losses, compulsive disorders) cannot be easily and reliably measured via self-reports.

Objective: This study aims to examine a method complementary to EMA, denoted as *peer-ceived momentary assessment* (PeerMA), which enables the involvement of peers (eg, family members, friends) to report their perception of the individual's subjective physical and emotional states. In this paper, we aim to report the feasibility results and identified human factors influencing the acceptance and reliability of the PeerMA

Methods: We conducted two studies of 4 weeks each, collecting self-reports from 20 participants about their stress, fatigue, anxiety, and well-being, in addition to collecting peer-reported perceptions from 27 of their peers.

Results: Preliminary results showed that some of the peers reported daily assessments for stress, fatigue, anxiety, and well-being statistically equal to those reported by the participant. We also showed how pairing assessments of participants and peers in time enables a qualitative and quantitative exploration of unique research questions not possible with EMA-only based assessments. We reported on the usability and implementation aspects based on the participants' experience to guide the use of the PeerMA to complement the information obtained via self-reports for observable behaviors and physical and emotional states among healthy individuals.

Conclusions: It is possible to leverage the PeerMA method as a complement to EMA to assess constructs that fall in the realm of observable behaviors and states in healthy individuals.

KEYWORDS

XSL•FO
RenderX

Introduction

Background

The ecological momentary assessment (EMA) [1] method—a form of Experience Sampling Method (ESM) [2,3]—is used to collect a person's momentary self-assessment of a particular outcome of interest (eg, mood, pain). This method has advantages such as easy interpretability, ecological validity and information richness (as the assessment comes directly from the person), self-motivation to report, and practicality [4]. This is especially useful when the objective ground truth is not obtainable; researchers thus interpret self-assessments as the ground truth of a state (eg, anxiety) [5]. This practice has disadvantages in scenarios where the self-assessment imposes an unwanted burden (eg, annoyance), the self-assessment poses a risk of reactivity [6] (eg, questions about anxiety), the person cannot answer objectively (eg, because of mental disorders), or the person chooses to answer untruthfully (eg, answers with high social desirability, always agreeing or disagreeing regardless of the question, or always picking an extreme or random response) [4].

In clinical settings, *proxies* and *observers* are often involved to inform about a patient's condition when the patient cannot express himself or herself objectively (eg, children or patients with dementia) [7] or has limited ability to participate [8]. Outside of clinical settings, researchers have shown the value of *observers* or *peers* in identifying sources of chronic stress experienced by an individual [9] and for personality assessment [5], especially in studies where the assessment is taken only once. However, there is a lack of information regarding whether an individual's peers (defined as close, trusted friends or family members [10]) can serve as valuable sources of information about the states and observable behaviors of the individual, potentially facilitating the early detection of certain states, including mental disorders occurring in the daily life of healthy individuals.

Consider an illustrative case of Bob in the early stages of developing an obsessive-compulsive disorder (OCD) [11] that makes him tighten his shoes irrationally. It is inherently hard for Bob to realize that his conduct is socially awkward and unsafe in locations such as stairways or busy sidewalks. Fortunately, because the behavior is observable, his wife Alice will be able to detect the early symptoms and motivate him to go for a medical check. Additionally, on the therapeutic side, Alice's daily reports could be valuable in providing evidence that ultimately aids in his treatment. Aside from the diagnosis, peers can also be allies for an individual recovering from an addiction (eg, eating disorders, smoking, gambling). In this case, peer assessments could help prevent relapses and the associated negative consequences.

Inspired by the principles of EMA, we evaluated the peer-ceived momentary assessment (PeerMA) method, previously defined by Berrocal and Wac [10]. PeerMA is a form of EMA completed by a designated peer of an individual during the same time observation window when the individual is prompted for an EMA. Peers are asked to indicate their perception of the states of the individual as well as how confident they are about the

provided assessment. The PeerMA method leverages the ubiquitous availability of smartphones as a way to collect momentary, ecologically valid data in the context of a person's life. This enables researchers to further examine how peers could become a source of information to complement the individual's self-assessment under certain conditions [7,9,12].

We explored 2 research aims in this study: (1) to evaluate the feasibility of the PeerMA method for studying real-life phenomena in healthy populations and (2) to identify the critical operational aspects and human factors that influence the quality of the data collected and their potential scaling. Toward this end, we conducted 2 in-the-wild studies (ie, outside the laboratory) using the PeerMA method. *Study A* was conducted around the University of Geneva (UNIGE) in Switzerland during the autumn of 2018. In this study, 13 participants self-assessed their perceived levels of stress [13], fatigue [14], and anxiety [15] multiple times a day using EMA, whereas their peers leveraged the PeerMA method to express the participants' perceptions of the same states. *Study B* was conducted with 10 participants around Stanford University in Palo Alto, United States, in the summer of 2019. However, in this study, in addition to assessing the levels of stress, fatigue, and anxiety, participants assessed their levels of well-being [16]. Similarly, peers assessed the level of stress, fatigue, anxiety, and well-being perceived by the participant via PeerMA. Both studies lasted 28 days.

Related Work

Both the ESM and the EMA methods were introduced in psychology [1,2] and are often used to study human aspects such as emotional awareness [17], depression [18], happiness [19], or human virtues [20] via self-reports. These methods are increasingly being used in clinical psychology [21] to study various types of disorders, such as mood dysregulation, anxiety, substance use, or psychosis [22]. EMA has also been used in patients with chronic fatigue, acquired immunodeficiency syndrome, migraine, breast cancer, or kidney disease to assess mood and stress changes [23]. Additionally, the method was adopted in organizational research [24] as well as in computer science, particularly in the subfields of ubiquitous computing and human-computer interaction [25].

The Proxy, Observer, Informant, or Peer Assessments

In psychology, self-assessment and other-assessment methods (also referred to as proxy, observer, informant, or peer assessments) have been used in various research contexts. For example, Vazire [5] summarized the findings of three studies providing empirical evidence of how informant-provided assessments improve the validity of personality assessments in behavioral sciences, and how they helped researchers examine questions for which self-reports alone would be insufficient. Her findings encourage researchers to incorporate informant assessments as an additional source of information to study certain human behaviors.

Gosling et al [26] studied the differences and implications of an individual's self-reported behavioral acts (eg, expressing agreement) versus the reports made by observers of that individual after coding and judging the recorded behaviors. The

focus of this research was to understand the accuracy and bias (such as self-enhancement) related to self-reports of one's acts. Collecting observers' assessments to examine cases of agreement and disagreement was vital for researchers to study the psychological processes of social perception.

Balsis et al [12] studied the reliability of self-reports versus informant reports in the context of personality assessment. They used the self-report and informant-report version of the Revised NEO Personality Inventory [27] inventory (including overall health measures) and showed, in a large sample (n=1449), that informant reports had greater internal consistency than self-reports for personality assessment. Although the internal consistency of informant reports does not imply that those reports are more valid than the self-reports, in this particular study, informant reports (having low variability) were better predictors of overall health measures than the self-reports of personality, which was relevant in the context of the study.

In clinical settings, in a group of older adults (>60 years old), Neumann et al [8] examined the validity of proxy responses used in 24 peer-reviewed publications from 1990 to 1999. Proxies (ie, family members or caregivers) showed fairly good agreement with the subject at assessing functioning, physical health, and cognitive status. However, proxies showed less agreement with the subject, reporting slightly higher impairment in emotional well-being and functioning. Such disagreement was in fact valuable in certain cases; for example, proxies provided more negative ratings than subjects during the 6 months before a hip fracture.

Focusing more on the characteristics of observers, Watson et al [28] studied the *acquaintanceship effect* in the context of self-agreements versus other agreements in low visibility aspects such as affect traits (eg, attentiveness or serenity). They involved one-time assessments of the Positive and Negative Affect Schedule Expanded Form [29], Big Five [30], and other study-specific instruments with a sample comprising 74 married couples, 136 dating couples, and 279 friendship couples. Their analysis showed that self-other agreement was significantly higher among married couples compared with dating couples or friendship dyads. Moreover, the self-other agreement was moderate to high in several components of the scales for dating couples and friendship dyads. Their work directly highlights the value of incorporating other assessments to study certain human traits, although admitting that the reliability of the other assessments should be studied carefully in each case.

Finally, although not using any form of EMA or PeerMA as presented in this paper, other studies showed empirical evidence to support 2 assumptions underlying PeerMA. Namely, (1) people often rely on their peers and trust essential information to them [31-33] and (2) peers can play a crucial role in reporting and assisting in specific scenarios such as rehabilitation and general health [8,34-36] as well as identifying the sources of chronic stress [9].

Value of Technology to Complement Self-Assessments

Following the trend of *personal sensing* described by Mohr et al [37], researchers have been using passively-collected raw data from smartphones and wearables to complement self-assessments of various aspects such as mental health and educational outcomes [38], stress [39-43], depressive moods [44], and schizophrenia [45]. Harari et al [46] reviewed studies using smartphone-sensing methods to identify physical movement, social interactions, and other daily activities, which can be used as objective and automated measures of behavior. More recently, Gresham et al [47] leveraged objective data from activity monitors to predict the risk of adverse events, hospitalizations, and hazard for death in advanced cancer patients. In general, these approaches profit from the abundance of passively sensed data that are converted into informative features to create computational models (commonly using machine learning or deep learning algorithms) to ultimately make inferences about human states such as those already mentioned.

However, despite its notable value, passively sensed data from smartphone sensors do not always enable accurate modeling of the perceptions of highly subjective individuals [48], and the use of such data may pose privacy risks [49] if proper data protection measures are not in place. In addition, smartphone sensor data are likely to vary in time because of hardware or sensing platform differences [37]. There is also evidence that individuals would abandon smart devices after a brief period of use [50] for reasons such as poor fit to expectations, not perceiving any direct value from the collected data, or the perceived high maintenance (eg, battery charging), especially if these devices are not their own smartphones.

We examined informant, proxy, observer reports from the literature and observed 3 main characteristics:

1. They captured individual and observer assessments using long surveys or instruments.
2. These assessments are usually carried out infrequently. Sometimes, these are one time–only assessments or are carried out every few months or years.
3. Proxies are usually involved for patients in clinical settings (due to physical or cognitive impairments).

We researched the use of PeerMA instead by (1) specifically using short surveys, for example, single-item or few-item questionnaires capturing one variable; (2) conducting frequent assessments, from >1 per day to just a few per week or month (in the case of longitudinal studies); and (3) exploring its value by focusing on healthy populations (ie, not having been diagnosed with a disease).

Our research makes a unique contribution by exploring the use of PeerMA [10] (peer assessment) to complement the information obtained via EMA (self-assessment) and personal sensing methods and collecting otherwise hard to collect data on human behaviors. The potential implications of our research are manifold, including implications for personal health, ubiquitous technologies, and mobile human-computer interaction.

Methods

Approach

This section describes the experimental design of the studies. To explore research aim 1 on the feasibility of the PeerMA method to study real-life phenomena in healthy populations, we collected variables such as user retention during the study, overall agreement between the EMA and PeerMA assessments, and the experimental value of the method by enabling the study of self-assessments and observer assessments paired in time. Moreover, to explore research aim 2 on operational aspects and human factors that influence the quality of the collected data, we gathered qualitative elements such as user reflections after using the method, difficulty in using the technology, and reliability of the technologies that can influence the quality of the collected data.

As explained in this section, *study B* had minor methodological differences based on lessons learned from *study A*. The studies were not designed to be replicas of each other. Instead, both studies were part of the exploratory phase of our research, in which we did not (yet) provide interventions or treatments to either the participants or their peers.

Tools

We implemented the PeerMA method by leveraging the *mQoL Lab* platform [51,52] of the Quality of Life Technologies Research Group (UNIGE) [53]. For the 2 studies described here, we developed and published a mobile app called *mQoL Peers* available via the Google Play Store (for Android) and the Apple Store (for iOS). The mobile apps with the *mQoL Lab* platform implemented the EMA and PeerMA methods (further explained

by Berrocal et al [51]), and we configured the content and frequency of the questions that the app administers via EMA and PeerMA.

Participant and Peer Types

To be included in this study, participants and peers had to be >18 years old and own a data-enabled smartphone with Android version 8.1+ or iOS version 7+. Table 1 shows the number of participants and peers per study and the cumulative number of types of peers. We used the following 4 types of peers based on their social proximity [28]: (1) spouse, (2) dating couples, (3) relatives, and (4) friends.

For *study A*, we recruited participants around the UNIGE campus by distributing flyers, by placing advertisements at department boards, via mQoL Living Lab email distribution lists, and by word of mouth. Overall, 6 participants had 1 peer and 7 participants had 2 peers. As compensation, participants who completed the study entered a raffle for 2 Amazon gift cards worth US $50 each.

For *study B*, we recruited participants around the Stanford University campus by distributing flyers, by placing advertisements on Craigslist and department boards, via email distribution lists, and by word of mouth. A total of 7 participants had 1 peer and 3 participants had no peers (included in Table 1 and excluded from the analysis). As compensation, participants and peers who completed the study received an Amazon gift card worth US $50 each.

Study Design: Study Surveys, EMA, and PeerMA

This section explains the entry, ambulatory, and exit surveys from each study. These are summarized in Table 2.

Table 1. Study participants: type and gender distribution.

Studies	Participants, n (%)	Peers, n (%)
Study A[a]		
Male	7 (54)	12 (60)
Female	6 (46)	8 (40)
Study B[b]		
Male	2 (20)	3 (43)
Female	8 (80)	4 (57)

[a]Total number of participants is 13, and total number of peers is 20.

[b]Total number of participants is 10, and total number of peers is 7.

Table 2. Study design: surveys and ecological momentary assessment/peer-ceived momentary assessment content in each study.

Types of survey	Study A	Study B
Entry surveys	Study entry, GERT[a], PSS[b], and SDS[c]	Study entry, PSS[b], and SDS[c]
Daily surveys: EMA[d] and PeerMA[e]	Stress, fatigue, and anxiety; frequency: 8 times a day (9 AM-9 PM); silent push notification; expires after 40 min	Stress, fatigue, anxiety, and well-being; frequency: 3 times a day (8 AM-8 PM); silent push notification; expires after 30 min
Exit survey	Study exit	Study exit

[a]GERT: Geneva Emotion Recognition Test (0-42; higher scores reflect higher ability).

[b]PSS: Perceived Stress Scale (0-40; higher scores reflect higher perceived stress).

[c]SDS: Social Desirability Scale (0-10; higher scores reflect higher social approval concern).

[d]EMA: ecological momentary assessment.

[e]PeerMA: peer-ceived momentary assessment.

Entry Surveys

Participants and peers initially completed the entry surveys before beginning the daily EMA/PeerMA. The study entry survey had 2 parts: (1) socioeconomic status, including gender, age range, education, marital status, and employment status and (2) open-ended questions asking participants whether they considered themselves stressed, what causes their stress, and whether they think others notice when they are stressed. For peers, open-ended questions asked whether they noticed when people around them project stress, what signs they observe in those projecting stress, how they react when someone around projects stress, and whether they get stressed or change their behavior when exposed to someone who is stressed. Peers also indicated their relationship with the participant (eg, friend, spouse) and whether they cohabit with the participant.

We employed the 42-item Geneva Emotion Recognition Test (GERT) that measures a person's ability to recognize someone's emotions from facial, voice, and body inputs (higher scores reflect higher ability) [54]. We used the 10-item Perceived Stress Scale (PSS) to measure self-perceived stress (higher scores reflect higher perceived stress) [55]. We used the 13-item Social Desirability Scale (SDS) to measure the degree to which a person is concerned with social approval (higher scores reflect higher social approval concern) [56].

For *study B*, the entry surveys had 2 differences. First, we asked participants (not only peers) to indicate whether they noticed when individuals around them project stress. Second, we excluded the GERT test because several participants and peers in *study A* reported technical issues with its web interface (provided as a service by the hosting agency) when completing the test.

Ambulatory Monitoring: EMA/PeerMA and Well-Being

In both studies, we used single-item questions proposed by Rosenzveig et al [57] on a visual analog scale, as shown in Figure 1. We chose stress, fatigue, and (state) anxiety because they can be studied using introquestive methods [58]. They often occur among healthy adults [59]. They compromise a person's well-being (eg, high stress, bad sleep quality) [60] and are not trivially observable by peers, and presumably, early detection of these conditions could inform diagnosis or therapeutic decisions. Additionally, these states can vary during

a day and hence are good candidates to report with EMA/PeerMA triggered through a day.

For *study A*, we used signal-contingent triggers between 9 AM and 9 PM for a total of 8 per day. They were uniformly randomized and separated by ≥45 min. Participants and peers received the EMA and PeerMA via silent push notifications (no sound and no vibration), which expired after 30 min if unanswered by the participant (to prevent questions from piling up). With EMA, we asked the following questions: "How much {stress, fatigue, anxiety} are you experiencing?" Correspondingly, with PeerMA, we asked: "How much {stress, fatigue, anxiety} is your peer projecting?" in addition to the peer's confidence assessment (Figure 1), which allowed peers to indicate how confident they were with each assessment. In principle, being next to the participant does not guarantee that peers can report with high confidence. Similarly, not being beside the participant or not having had a recent face-to-face contact does not preclude peers from reporting with high confidence. The exact dynamics used by the peers to make their observations are complex in nature, as already noted by Uher et al [58]; therefore, given the exploratory nature of this study, we decided to leave them for future work. The confidence assessment provided by peers can be used to inform the data analysis, for instance, to discard zero-labeled confidence assessments.

For *study B*, we used signal-contingent triggers between 8 AM and 8 PM for a total of 3 per day. They were uniformly randomized and separated by ≥2 hours. Participants and peers received the EMA and PeerMA via silent push notifications (no sound and no vibration), which also expired after 30 min if not addressed by the participant. With EMA, we asked the following questions: "How much {stress, fatigue, anxiety} are you experiencing?" on a scale of 0 to 10 (the higher the worse), "How well are you?"—related to the overall level of well-being on a scale of 0 to 10 (the higher the better)—and an open question "If you wish, briefly describe your emotional state at this moment." Correspondingly, with PeerMA, we asked "How much {stress, fatigue, anxiety, well-being} is your peer projecting?" (using the same scale of 0-10), 1 open question "If you wish, briefly describe the state your peer projects at this moment," and the peer's confidence assessment. In this study, the *mQoL Peers* app also had a button to let participants and peers self-trigger an assessment (EMA or PeerMA correspondingly) when they wished to do so.

Because the explicit well-being question had been adopted only in *study B*, in *study A* we computed it afterward. In *study A*, we adopted the notion of computed well-being (additional to the *reported well-being* in *study B*), representing a reduced, yet semantically aligned form of well-being as defined by Huppert [16]. We used a pragmatic definition of well-being as 1 minus the arithmetic mean of stress, fatigue, and anxiety. Empirically, a high prevalence of stress, fatigue, and anxiety reduces well-being. On the contrary, low levels of stress, fatigue, and anxiety are likely to contribute to a healthy state of well-being.

Figure 1. Examples of ambulatory assessment: (a) Self-assessment (ecological momentary assessment) of stress, (b) peer- assessment (peer-ceived momentary assessment) of stress, (c) confidence assessment required from peers. Assessments of fatigue, anxiety, and well-being followed the same approach.
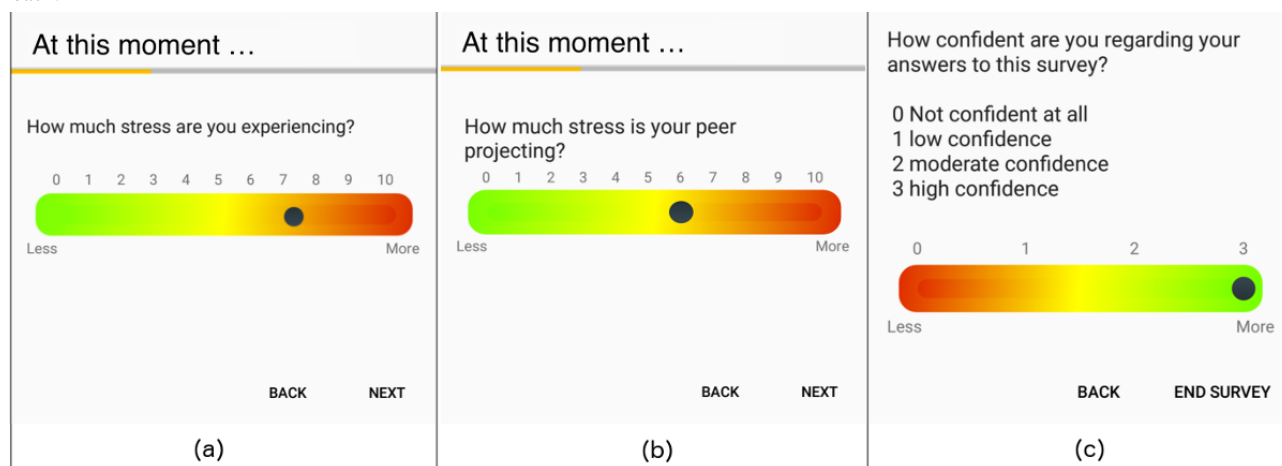
Exit Survey

At the end of the study, both participants and peers completed an exit survey commenting on usability aspects of the mobile device app (eg, usability, positive and negative aspects perceived). The survey also asked how participants felt about reflecting on their states during the day, whereas peers answered how they felt about reflecting on their peers' states during the day.

Study Protocol and Ethical Approval

At the beginning of the study, participants had a 15-min web-based or face-to-face meeting with the researcher with the following objectives: (1) explain the nature of the study, (2) hand out the informed consent, (3) train the participant to use the *mQoL Peers* app (concretely, how to invite peers, how to complete entry surveys, and how to address the push notifications), and (4) answer any questions from the participant.

During this meeting, the researcher explained to the participants that, given the nature of the study, peers had to be people with whom they had regular contact (at least daily), either face-to-face or virtually, using communication tools. We explained that peers could be spouses (significant others), close relatives (family), or friends from school or work. After the meeting, participants would complete the entry surveys, enroll their peers, and explain to them how to use the app. In these 2 studies, the researcher had no interaction with peers. After enrolling their peers and completing the entry surveys, participants pushed a button in the app to start the study and receive daily EMAs and PeerMAs. The researchers were in touch with the participants remotely to follow-up with them about these steps, if needed.

Study A was approved by the institutional review board of UNIGE under the protocol "Exploring the Value of Social Links and Human-Machine Collaboration in the Context of Stress Assessment," N. CUREG.201807. *Study B* was approved by the Panel on Human Subjects in Medical Research of Stanford University under the protocol "Studying the Subjective and Objective Momentary Perception of Quality of Life in Different Contexts of Daily Life," N.47833, Reg# 351.

Results

Structure

The first part, the section on *Feasibility of Using PeerMA,* presents the results directly associated with research aim 1 (feasibility of using PeerMA in real-life phenomena). Recalling from these methods, these results allowed us to observe variables such as user retention and completion rates in the study, the overall agreement between EMA and PeerMA assessments, and the experimental value of the method to study certain research questions. The first part is organized as follows: the section on *Collected Data Summary* introduces the data sets, including visualizations, to illustrate the contributions of the PeerMA method in the section on *Visual Explorations of Daily Dynamics.* Then it presents results from 3 analyses: directional accuracy in the section on *Mean Directional Accuracy of EMA/PeerMA,* correlations in the section on *EMA and PeerMA Correlations,* and statistical agreement between participants and peers assessments in the section on *EMA/PeerMA Statistical Agreement.*

The second part, the section on *Operational and Human Factors,* presents results associated with research aim 2 (operational and human factors). The second part of the section is organized as follows: the section on *Users' Reflections about the Studies* summarizes the qualitative results related to the users' reflections from both studies. Then, the section on *Implications of Technology Choices* briefly describes implications of the technology choices that we experienced in the studies. Finally, the section on *Suggestions from Users about Technology and*

Methods shows some recommendations or observations made by participants in the study, which are worth sharing with the research community.

Feasibility of Using PeerMA

We present the type of quantitative and qualitative data that are being obtained with PeerMA as a method and tool in the 2 observational studies, and not necessarily the strength of the results regarding stress, fatigue, anxiety, and well-being that have been explored as use cases. Nevertheless, we also present a detailed examination of the results to support the findings and observations that come along the analyses.

Collected Data Summary

The first part of the dataset, extracted from the entry surveys, describes the samples in each study. Table 3 shows the socioeconomic characteristics (gender, age, marital status, education, and employment) as well as the distribution of participants and peers who admit whether other people notice (or not) when they are stressed. Table 4 shows the test scores (GERT, PSS, and SDS) as well as the distribution of how stressed participants and peers considered themselves and how well they report to notice stress in others. The second part of

the dataset contains information about the quantity and values obtained from the EMA and PeerMA in each study.

In *study A*, which included 13 participants and 20 peers, we collected a total of 878 person-days, 380 participant-days (mean 29, SD 1.7), and 498 peer-days (mean 25, SD 5.9). Participants received 3086 EMAs (mean 237, SD 32.3), and responded to a total of 2001 EMAs (mean 154, SD 62.2) with an average response rate of 65% (SD 24.2%). Peers received 3178 PeerMAs (mean 159, SD 49.0), and responded to a total of 1328 PeerMAs (mean 66, SD 37.9) with an average response rate of 44% (SD 24.5%).

In *study B*, which included 10 participants and 7 peers, we collected a total of 373 person-days, 187 participant-days (mean 27, SD 4.5), and 186 peer-days (mean 27, SD 4.8). Participants received in total 561 EMAs (mean 80, SD 13.4), and responded to a total of 445 EMAs (mean 64, SD 16.2) with an average response rate of 80% (SD 19.6%). Peers received 558 PeerMAs (mean 80, SD 14.5), and responded to a total of 432 PeerMAs (mean 62, SD 45.4) with an average response rate of 77% (SD 48%; including peer S1P1, who provided almost twice the expected number of PeerMAs).

Table 3. Participants' socioeconomic characteristics by study.

Variables	Study A		Study B	
	Participants, n (%)	Peers, n (%)	Participants, n (%)	Peers, n (%)
Gender				
Male	6 (46)	8 (40)	2 (29)	3 (43)
Female	7 (54)	12 (60)	5 (71)	4 (57)
Age (years)				
18-20	1 (8)	2 (10)	0 (0)	0 (0)
21-29	9 (69)	8 (40)	3 (43)	3 (43)
30-39	2 (15)	4 (20)	4 (57)	4 (57)
40-49	1 (8)	6 (30)	0 (0)	0 (0)
Marital status				
Single	10 (77)	12 (60)	3 (43)	3 (43)
Married	2 (15)	5 (25)	4 (57)	4 (57)
Other	1 (8)	3 (15)	0 (0)	0 (0)
Highest education				
Undergraduate	8 (62)	11 (55)	1 (14)	3 (43)
Graduate	5 (38)	9 (45)	6 (86)	4 (57)
Currently employed				
Yes	3 (23)	9 (45)	3 (43)	6 (86)
No	10 (77)	11 (55)	4 (57)	1 (14)
Others notice my stress?				
Yes	9 (69)	15 (75)	7 (100)	6 (86)
No	4 (31)	5 (25)	0 (0)	1 (14)

XSL•FO
RenderX

Table 4. Survey scores of participants and peers by study.

Instruments and roles	Study A			Study B		
	Minimum score	Maximum score	Mean (SD)	Minimum score	Maximum score	Mean (SD)
GERT[a] (0-42)						
Participant	18	33	26 (5.2)	N/A[b]	N/A	N/A
Peer	20	31	25 (3.8)	N/A	N/A	N/A
PSS[c] (0-40)						
Participant	12	31	23 (5.9)	19	31	24 (4.6)
Peer	0	36	22 (8.2)	12	28	21 (5.9)
SDS[d] (0-13)						
Participant	3	12	7 (2.5)	0	10	5 (3.2)
Peer	0	11	6 (3.0)	1	11	6 (3.4)
Considered stressed (0-10)						
Participant	2.8	9.8	6.1 (2.4)	3.1	9.0	6.2 (2.4)
Peer	0.3	8.7	4.5 (2.6)	1.9	7.9	5.0 (1.8)
I notice other's stress (0-10)						
Participant	N/A	N/A	N/A	4.0	8.9	5.7 (1.7)
Peer	3.9	8.2	6.9 (1.07)	5.2	10	7.5 (1.5)

[a]GERT: Geneva Emotion Recognition Test.

[b]N/A: not applicable.

[c]PSS: Perceived Stress Scale.

[d]SDS: Social Desirability Scale.

Table 5 shows the number of days each person participated in the study and the response rate, computed as the percentage of EMAs/PeerMAs answered, from all triggered. For *study A*, *triggered* represents the number of EMAs/PeerMAs automatically triggered by the app. For *study B*, the response rate includes responses from *automatically triggered* plus *self-triggered* EMA/PeerMAs because, in *study B*, participants and peers were able to initiate reports voluntarily in addition to the automatic reports. The column "EMAs or PeerMAs Triggered" shows the minimum number of expected responses for each person (in this case, 3 per day). Additionally, Table 5 shows the type of relationship between the peer and the subject, as indicated by the peer. In this paper, however, given the small number of participants, we did not report results by relationship type.

Table 5. Summary of the engagement of participants and peers.

Studies, participant ID	Peer ID	Days	Ecological momentary assessments or peer-ceived momentary assessments triggered	Response rate, %	Peer-participant relationship
Study A					
S1	N/A[a]	27	220	61.8	N/A
N/A	S1P1	27	127	87.4	3: parent
S2	N/A	31	147	68.0	N/A
N/A	S2P1	31	100	61.0	3: parent
S3	N/A	29	243	97.9	N/A
N/A	S3P1	29	220	35.5	4: friend
S4	N/A	28	211	36.5	N/A
N/A	S4P1	27	165	15.2	3: parent
S5	N/A	29	265	47.9	N/A
N/A	S5P1	27	252	23.0	3: sibling
S6	N/A	29	243	97.5	N/A
N/A	S6P1	22	183	61.7	2: boyfriend
S7	N/A	29	245	70.6	N/A
N/A	S7P1	20	122	30.3	4: friend
N/A	S7P2	29	199	50.8	3: parent
S8	N/A	28	250	94.0	N/A
N/A	S8P1	26	214	32.2	4: friend
N/A	S8P2	28	213	61.5	2: boyfriend
S9	N/A	34	225	44.9	N/A
N/A	S9P1	33	130	36.9	4: friend
N/A	S9P2	33	133	45.9	3: sibling
S10	N/A	29	267	31.1	N/A
N/A	S10P1	30	102	41.2	2: girlfriend
N/A	S10P2	16	152	4.6	4: friend
S11	N/A	29	250	82.0	N/A
N/A	S11P1	22	153	86.3	4: friend
N/A	S11P2	17	100	50.0	3: sibling
S12	N/A	28	252	77.4	N/A
N/A	S12P1	26	198	13.1	3: sibling
N/A	S12P2	26	206	50.0	4: friend
S13	N/A	30	268	35.1	N/A
N/A	S13P1	15	130	9.2	3: parent
N/A	S13P2	14	79	79.7	4: friend
Study B					
S1	N/A	31	93	69.9	N/A
N/A	S1P1	31	93	172.0	4: friend
S2	N/A	28	84	78.6	N/A
N/A	S2P1	28	84	77.4	4: friend
S3	N/A	28	84	108.3	N/A

Studies, participant ID	Peer ID	Days	Ecological momentary assessments or peer-ceived momentary assessments triggered	Response rate, %	Peer-participant relationship
N/A	S3P1	28	84	60.7	1: spouse
S4	N/A	28	84	69.0	N/A
N/A	S4P1	28	84	59.5	1: spouse
S5	N/A	28	84	50.0	N/A
N/A	S5P1	27	81	28.4	1: spouse
S6	N/A	17	51	96.1	N/A
N/A	S6P1	16	48	102.1	4: friend
S7	N/A	27	81	91.4	N/A
N/A	S7P1	28	84	40.5	4: friend

[a]N/A: not applicable.

As noted in Table 5, some participants and peers deviated from the expected 28 days of participation in the study. Those with <28 days stopped participating in the study after informing the research team (S13P1 in *study A* and S6 and S6P1 in *study B*) or stopped participating without notification. On the other hand, those with >28 days were due to, in study A, the smartphone operating system at times cutting off the service that kept the daily counter in our mobile app in response to users' settings of the battery-saving profile. Consequently, some dyads went beyond 28 days. Finally, in *study B,* S1 and S1P1 continued contributing assessments before noticing the study had ended (which are displayed on the home screen of the mobile app), at which time they completed the exit survey and completed the study.

Additionally, in Table 5, the number of triggered surveys in *study A* is smaller than the expected value (8 times the number of days) due to a failure in the mobile app that did not count notifications that expired after 30 min. On the other hand, when the user interrupted the completion of a survey (eg, by opening another app), upon returning to the survey, a new record was mistakenly added to the triggered list, resulting in a number larger than the expected value (8 times the number of days). Both issues were later resolved in the platform.

For each participant and peer, we normalized the EMA/PeerMA assessments to 0 to 1 based on the highest and lowest assessment given by each person. Table 6 shows the median, mean, and SD of all assessments for stress, fatigue, and anxiety in study A. The last 3 columns show the calculated value and computed well-being as defined in the *Methods* section. In Table 6, lower values in the "Median" and "Mean" columns represent more desirable states, whereas higher values represent less desirable states. Table 7 shows the dataset for *study B*. In addition to the computed well-being, this study shows the actual *reported well-being*, including the median, mean, and SD.

Table 6. Study A: Summary of the ecological momentary assessment and peer-ceived momentary assessment values. Each row shows the median, mean, and SD for the corresponding participant or peer calculated from all the assessments issued by that person.

Participant ID	Peer ID	Stress (0-1)		Fatigue (0-1)		Anxiety (0-1)		Computed well-being (0-1)	
		Median	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)
S1	N/A[a]	0.35	0.41 (0.27)	0.22	0.27 (0.23)	0.30	0.37 (0.27)	0.69	0.65 (0.21)
N/A	S1P1	0.70	0.62 (0.27)	0.62	0.65 (0.27)	0.70	0.57 (0.29)	0.35	0.39 (0.27)
S2	N/A	0.17	0.28 (0.29)	0.29	0.31 (0.26)	0.19	0.31 (0.33)	0.65	0.70 (0.23)
N/A	S2P1	0.72	0.62 (0.30)	0.58	0.59 (0.28)	0.37	0.34 (0.23)	0.42	0.48 (0.23)
S3	N/A	0.61	0.51 (0.30)	0.61	0.62 (0.27)	0.37	0.39 (0.24)	0.40	0.49 (0.22)
N/A	S3P1	0.52	0.49 (0.27)	0.39	0.40 (0.21)	0.48	0.50 (0.27)	0.55	0.54 (0.22)
S4	N/A	0.18	0.26 (0.26)	0.30	0.36 (0.25)	0.35	0.43 (0.32)	0.71	0.65 (0.17)
N/A	S4P1	0.65	0.60 (0.30)	0.65	0.50 (0.36)	0.79	0.63 (0.33)	0.35	0.42 (0.28)
S5	N/A	0.11	0.23 (0.28)	0.35	0.36 (0.24)	0.00	0.10 (0.23)	0.82	0.77 (0.17)
N/A	S5P1	0.64	0.58 (0.32)	0.60	0.58 (0.33)	0.12	0.19 (0.28)	0.53	0.55 (0.27)
S6	N/A	0.37	0.44 (0.28)	0.32	0.34 (0.23)	0.40	0.43 (0.22)	0.74	0.60 (0.19)
N/A	S6P1	0.49	0.49 (0.23)	0.38	0.45 (0.32)	0.46	0.43 (0.24)	0.55	0.55 (0.24)
S7	N/A	0.13	0.17 (0.20)	0.36	0.37 (0.25)	0.12	0.15 (0.20)	0.80	0.77 (0.16)
N/A	S7P1	0.55	0.58 (0.22)	0.64	0.64 (0.25)	0.68	0.60 (0.23)	0.36	0.39 (0.22)
N/A	S7P2	0.31	0.34 (0.22)	0.65	0.57 (0.29)	0.34	0.39 (0.29)	0.60	0.57 (0.21)
S8	N/A	0.08	0.18 (0.25)	0.52	0.48 (0.25)	0.07	0.14 (0.22)	0.75	0.73 (0.18)
N/A	S8P1	0.70	0.68 (0.19)	0.62	0.65 (0.18)	0.67	0.57 (0.23)	0.34	0.37 (0.17)
N/A	S8P2	0.43	0.41 (0.25)	0.58	0.55 (0.27)	0.47	0.47 (0.25)	0.50	0.52 (0.24)
S9	N/A	0.00	0.16 (0.26)	0.38	0.43 (0.30)	0.26	0.27 (0.29)	0.74	0.71 (0.19)
N/A	S9P1	0.42	0.43 (0.31)	0.67	0.61 (0.29)	0.68	0.56 (0.25)	0.42	0.47 (0.21)
N/A	S9P2	0.50	0.55 (0.27)	0.51	0.55 (0.24)	0.45	0.58 (0.34)	0.45	0.44 (0.20)
S10	N/A	0.58	0.61 (0.32)	0.58	0.57 (0.27)	0.68	0.61 (0.28)	0.37	0.40 (0.12)
N/A	S10P1	0.36	0.38 (0.28)	0.34	0.34 (0.20)	0.34	0.44 (0.28)	0.63	0.62 (0.22)
N/A	S10P2	0.19	0.39 (0.43)	0.77	0.61 (0.39)	0.42	0.48 (0.46)	0.52	0.51 (0.35)
S11	N/A	0.48	0.52 (0.23)	0.18	0.29 (0.28)	0.41	0.42 (0.22)	0.62	0.59 (0.18)
N/A	S11P1	0.49	0.49 (0.23)	0.40	0.45 (0.31)	0.52	0.51 (0.25)	0.52	0.52 (0.24)
N/A	S11P2	0.48	0.42 (0.34)	0.63	0.57 (0.37)	0.50	0.50 (0.24)	0.52	0.50 (0.28)
S12	N/A	0.36	0.46 (0.29)	0.60	0.55 (0.27)	0.07	0.19 (0.28)	0.63	0.60 (0.17)
N/A	S12P1	0.67	0.56 (0.33)	0.77	0.65 (0.30)	0.57	0.49 (0.31)	0.31	0.43 (0.29)
N/A	S12P2	0.10	0.24 (0.32)	0.53	0.52 (0.25)	0.00	0.19 (0.29)	0.74	0.68 (0.20)
S13	N/A	0.38	0.40 (0.30)	0.03	0.16 (0.24)	0.30	0.31 (0.25)	0.74	0.71 (0.18)
N/A	S13P1	0.43	0.47 (0.36)	0.45	0.50 (0.30)	0.41	0.38 (0.33)	0.56	0.55 (0.28)
N/A	S13P2	0.51	0.55 (0.27)	0.42	0.43 (0.26)	0.61	0.62 (0.22)	0.48	0.47 (0.24)

[a]N/A: not applicable.

Table 7. Study B: Summary of ecological momentary assessment and peer-ceived momentary assessment values. Each row shows the median, mean, and SD for the corresponding participant or peer calculated from all the assessments issued by that person.

Participant ID	Peer ID	Stress (0-1)		Fatigue (0-1)		Anxiety (0-1)		Reported well-being (0-1)		Computed well-being (0-1)	
		Median	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)
S1	N/A[a]	0.49	0.36 (0.55)	0.58	0.37 (0.60)	0.58	0.66 (0.32)	0.76	0.72 (0.24)	0.39	0.52 (0.30)
N/A	S1P1	0.44	0.44 (0.21)	0.50	0.47 (0.23)	0.57	0.54 (0.21)	0.41	0.41 (0.21)	0.51	0.49 (0.21)
S2	N/A	0.35	0.34 (0.29)	0.42	0.44 (0.27)	0.42	0.43 (0.25)	0.73	0.63 (0.28)	0.66	0.63 (0.28)
N/A	S2P1	0.76	0.69 (0.24)	0.59	0.57 (0.20)	0.66	0.64 (0.25)	0.61	0.56 (0.29)	0.26	0.35 (0.22)
S3	N/A	0.50	0.49 (0.27)	0.60	0.58 (0.27)	0.39	0.42 (0.23)	0.48	0.47 (0.26)	0.37	0.43 (0.28)
N/A	S3P1	0.47	0.43 (0.24)	0.58	0.53 (0.32)	0.59	0.59 (0.23)	0.79	0.76 (0.20)	0.39	0.43 (0.29)
S4	N/A	0.50	0.53 (0.25)	0.56	0.59 (0.25)	0.42	0.44 (0.31)	0.64	0.56 (0.24)	0.56	0.52 (0.25)
N/A	S4P1	0.63	0.66 (0.28)	0.75	0.67 (0.25)	0.69	0.59 (0.32)	0.65	0.62 (0.28)	0.31	0.39 (0.30)
S5	N/A	0.22	0.27 (0.26)	0.15	0.24 (0.27)	0.33	0.34 (0.33)	0.78	0.66 (0.32)	0.67	0.67 (0.27)
N/A	S5P1	0.44	0.40 (0.29)	0.31	0.41 (0.32)	0.43	0.43 (0.30)	0.63	0.54 (0.29)	0.59	0.56 (0.29)
S6	N/A	0.54	0.56 (0.27)	0.53	0.54 (0.24)	0.55	0.59 (0.29)	0.48	0.51 (0.23)	0.45	0.39 (0.26)
N/A	S6P1	0.49	0.47 (0.28)	0.51	0.50 (0.29)	0.62	0.59 (0.32)	0.55	0.49 (0.32)	0.47	0.48 (0.29)
S7	N/A	0.00	0.21 (0.33)	0.54	0.56 (0.20)	0.26	0.35 (0.20)	0.62	0.59 (0.30)	0.81	0.66 (0.27)
N/A	S7P1	0.36	0.35 (0.34)	0.55	0.56 (0.28)	0.37	0.41 (0.36)	0.60	0.53 (0.26)	0.66	0.59 (0.30)

[a]N/A: not applicable.

Visual Exploration of Daily Dynamics

As this study primarily focused on assessing the feasibility of the method, we started the data analysis with the least complex visualization of raw datasets. We wanted to plot the values reported by the participants and their peers and understand the magnitude of agreement/disagreement in their ratings in time. We also imputed the missing PeerMA values using a spline function of order 4.

To illustrate, sample plots from *study A* are presented in Figure 2. For the dyad <S6, S6P1> shown in the 4 plots on the top, despite the differences in magnitude and time shifts, there are notable similarities for states such as stress and anxiety as well as nonnegligible overlapping for fatigue. For the triad <S8, S8P1, S8P2> shown in the 4 bottom plots, there is a pattern for stress and anxiety, where S8 started to report low values although the peers continued to report higher values with periods of mutual agreement. One can see that in this triad, there seemed to be higher agreement for fatigue. Consequently, either the peers were unable to truthfully report stress and anxiety or the participant did not truthfully report stress and anxiety (intentionally or otherwise). Both are hypotheses that can be explored further with focused experiments combining EMA and PeerMA with other qualitative methods.

Figure 3, similar to Figure 2, presents sample plots from *study B*. In the 5 plots on the top, corresponding to the dyad <S3-S3P1>, we observed high agreement for stress and fatigue and low agreement for anxiety. Additionally, in this particular case, the assessments for reported well-being differed significantly, whereas the computed well-being showed high agreement (in fact, their median was statistically equal). However, this could be because of either the participant or the peer answering the question of well-being on assigning higher importance to other aspects not inherently reflected as stress, fatigue, and anxiety (which are the sole variables used in our pragmatic computation of well-being, as defined in the *Methods* section).

Finally, the 5 plots at the bottom of Figure 3, for the dyad <S6, S6P1>, show very high agreement in all the states, and their median was statistically equal in all the cases. Assuming their observations genuinely reflect the actual state of the participant, with approximately 3 weeks of agreement in which there were high and low episodes, this dyad seems to have a shared

understanding of the concepts and excellent sensing skills by the peer. Such a dyad raises the confidence of a researcher to include them in a longitudinal study in which periods of disagreement between EMA and PeerMA motivate the use of other qualitative methods (such as the Day Reconstruction Method [61]) to study the root causes of such disagreements.

Figure 2. Ecological momentary assessment/peer-ceived momentary assessment. Plots from Study A. The x-axis represents days in the study, and the y-axis represents the magnitude of the normalized assessments for stress, fatigue, anxiety, and computed well-being.
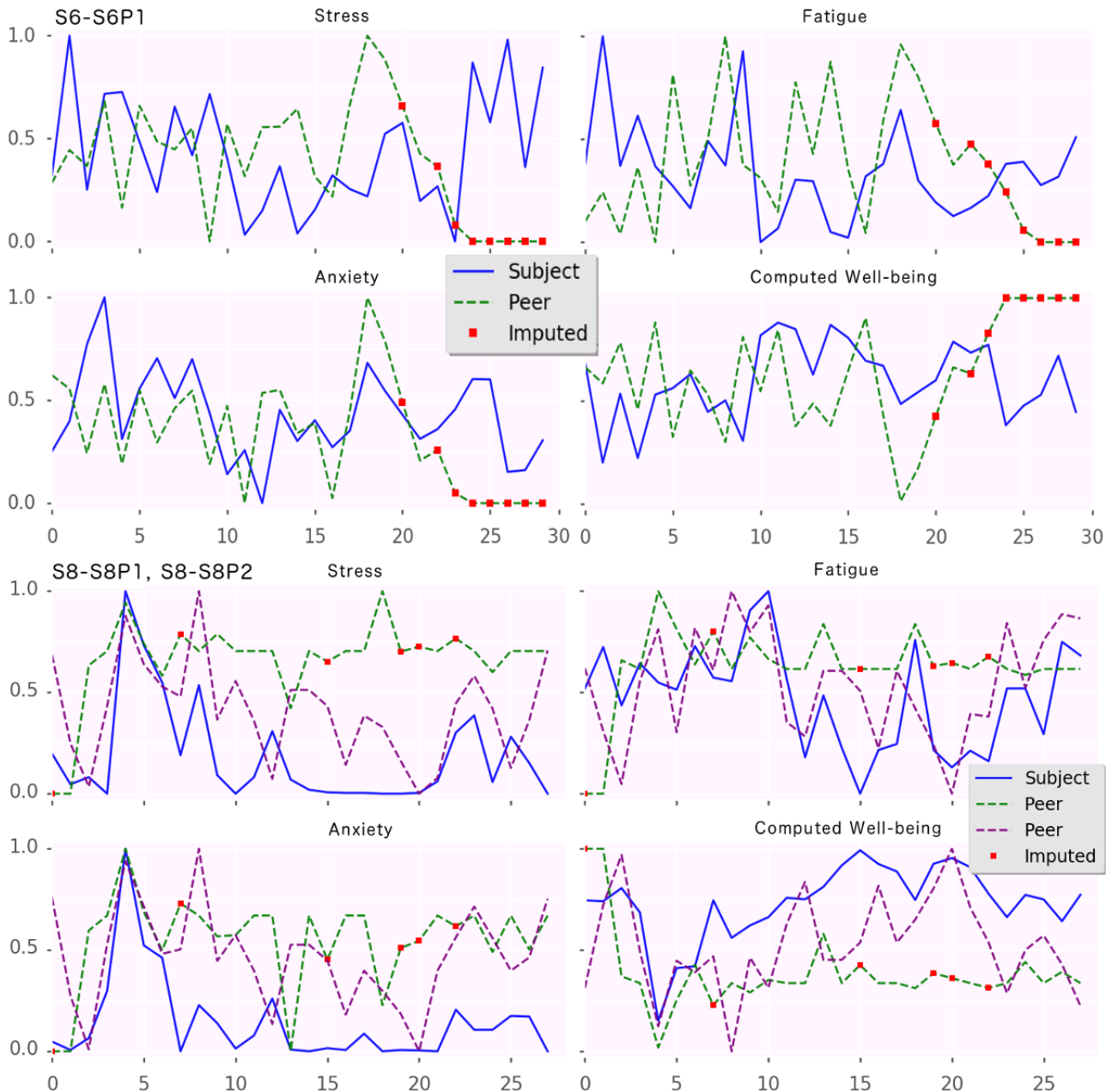
Figure 3. Ecological momentary assessment/peer-ceived momentary assessment. Plots from Study B. The x-axis represents days in the study, and the y-axis represents the magnitude of the normalized assessments for stress, fatigue, anxiety, and computed well-being.
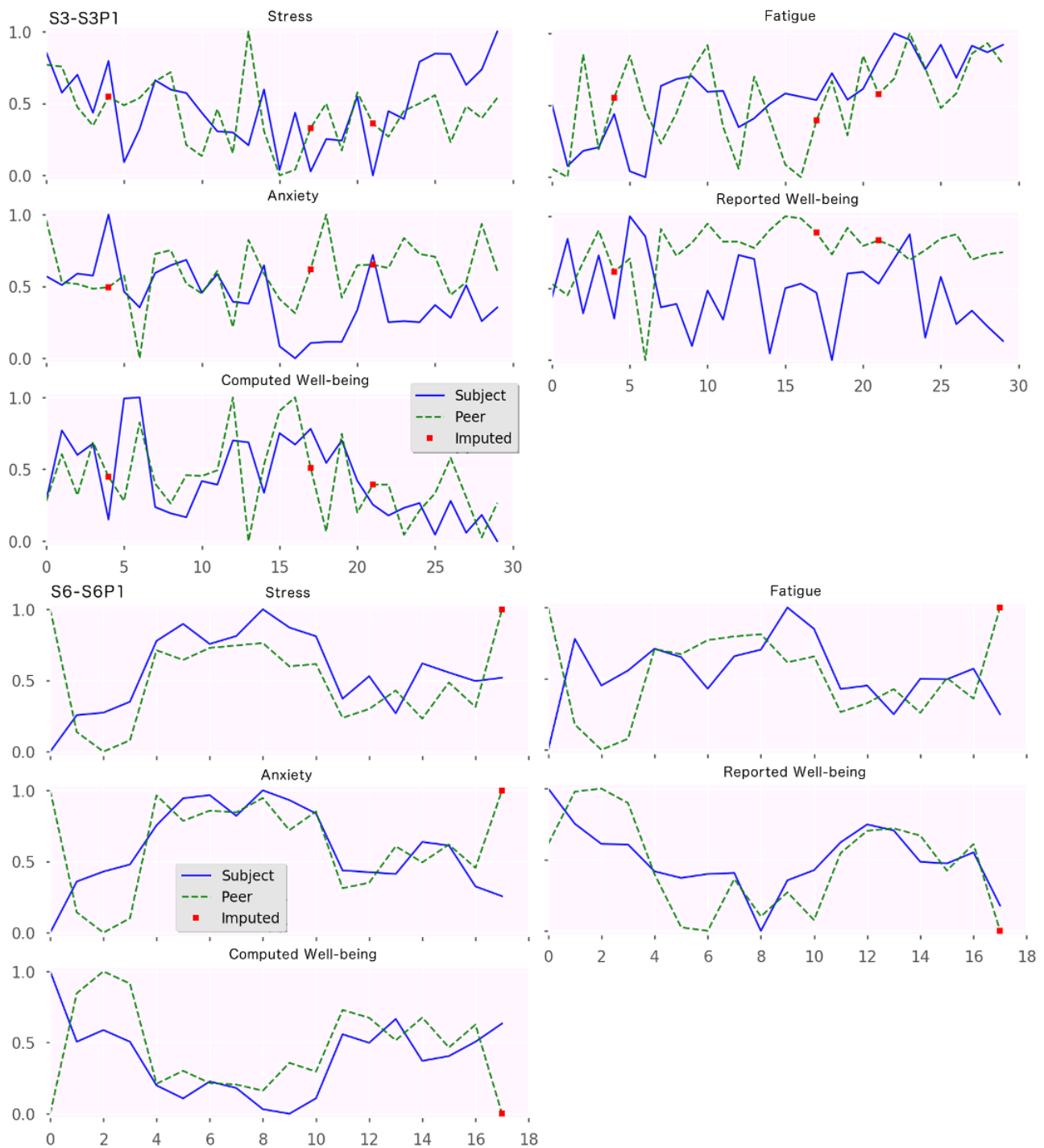
Mean Directional Accuracy of EMA/PeerMA

Various techniques can be used to quantify the daily agreement between the EMAs and PeerMAs. For instance, residual analysis such as mean absolute percent error (MAPE) or more robust alternatives such as mean arctangent absolute percentage error (MAAPE) [62]. In our case, the robustness of MAAPE is derived from the fact that MAPE produces undefined values when the assessment of a participant is equal to 0. These techniques can quantify the overall difference between the EMA and the PeerMA values over time. However, for the particular application of PeerMA, we expect to see periods of disagreement

as they represent the differences between the self-assessments and the observer assessments that are worth identifying and analyzing during field applications.

Therefore, we analyzed the daily agreement between the EMAs and the PeerMAs as follows. As a first approach, we reported the mean directional accuracy (MDA), which measures the agreement (ie, we report *a match*) in the direction of change of momentary assessments between participants and peers. MDA considers only the direction of the change (eg, upward or downward) and not its magnitude. For instance, let us suppose that a participant reported $(t_0, 0.4)$, $(t_1, 0.3)$, and $(t_2, 0.7)$ and a

peer reported (t_0, 0.2), (t_1, 0.4), and (t_2, 0.6); then, between t_0 and t_1, there is a mismatch as the participant's report decreased (negative direction) whereas the peer's report increased (positive direction). From t_1 to t_2, however, there is a match as both participants' and peers' reports increased (0.3 to 0.7 for the participant and 0.4 to 0.6 for the peer).

In the 2 studies, the number of assessments between participants and peers differed every day. Hence it was not possible to calculate the MDA for each individual EMA/PeerMA. Thus, we calculated the daily average for stress, fatigue, anxiety, and well-being for each participant and peer using all the assessments given that day. We then counted the number of days with *matches* and divided it by the number of days for

which both the participant and the peer issued at least one assessment. Figure 4 shows the accumulated MDA results for both studies. *Same day* is the percentage of days that participants and peers agreed in the directional change of their assessments the same day. As can be seen from the figure, the result is close to chance. However, it increased to approximately 73% in *study A* and to 79% in *study B* on counting matches occurring the same day or 1 day after. Naturally, MDA can be calculated with higher granularity down to every consecutive pair of EMA/PeerMA values. Overall, MDA shows that PeerMA is promising for identifying variations in mental or physical health early on (eg, accurately assessing changes in the individuals' day-to-day states).

Figure 4. Mean Directional Accuracy. "Same Day" is the average percentage of days that participants and peers agreed in the directional change of their assessments the same day (close to chance). "+1 Day" is the average percentage of days that participants and peers agreed in the directional change of their assessments the same day or the day after.
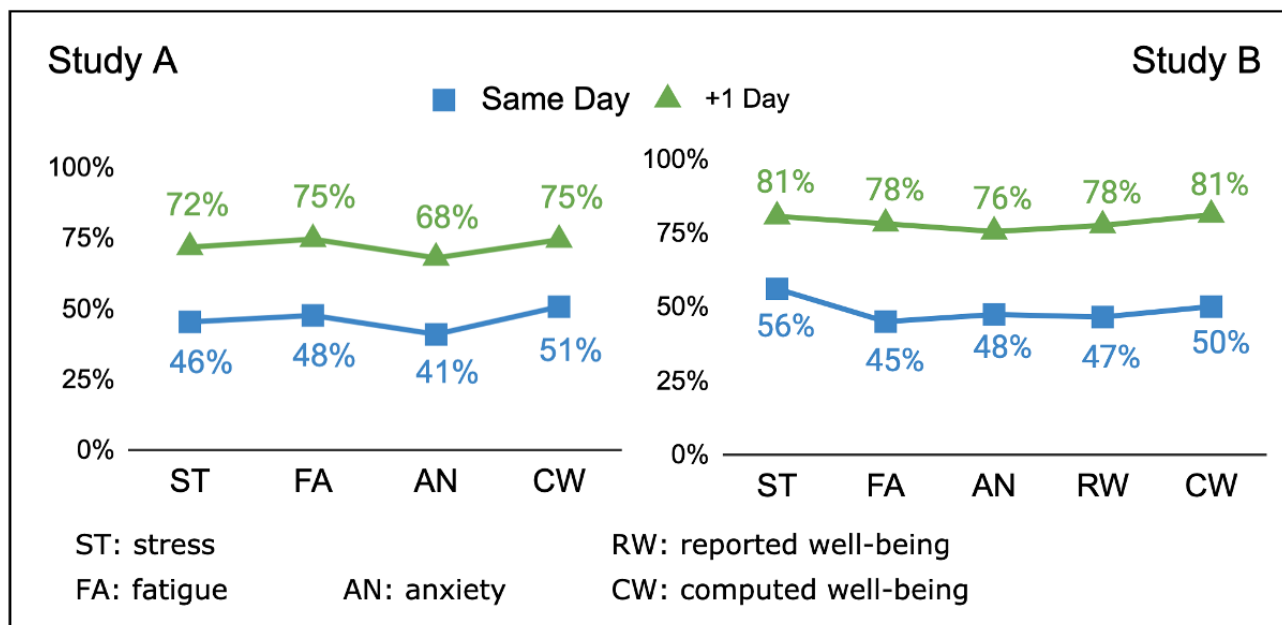
EMA and PeerMA Correlations

EMA/PeerMA Spearman Correlations

To further investigate the values of EMA and PeerMA, we conducted a correlation analysis for EMA/PeerMA values in both studies. By focusing on the correlation, we did not assume that the EMA and the PeerMA measure the same constructs; we investigated it later in this paper. Therefore, in this study, we applied the Spearman rank correlation method because (1) participant and peer assessments are not independent, they both refer to a state of the participant and (2) the Shapiro-Wilk, D'Agostino K^2, and Anderson-Darling tests indicate that not

all individuals' and peers' assessments were normally distributed.

For *study A*, the Spearman rank correlation coefficients for each of the states are presented in Table 8. Each row presents the correlation coefficient r_s and P value for a participant-peer dyad. We also show the correlation between the computed well-being of each participant and peer. In Table 9, we present a summary of the values from Table 8 classified in 6 correlation groups. In this sample, 40% of the correlations are weakly positive, followed by 36% being weakly negative and 20% being moderately positive.

Table 8. Study A: ecological momentary assessment/peer-ceived momentary assessment Spearman correlations calculated throughout the study. Each row shows the correlation between the participants' and peers' assessments.

Participant ID	Peer ID	Stress		Fatigue		Anxiety		Computed well-being	
		r_s	P value	r_s	P value	r_s	P value	r_s	P value
S1	S1P1	0.44	.02	0.28	.15	0.57	.002	0.58	.001
S2	S2P1	0.50	.003	0.26	.15	0.24	.17	0.63	<.001
S3	S3P1	–0.18	.36	–0.16	.40	–0.09	.64	–0.15	.45
S4	S4P1	0.06	.76	–0.10	.62	–0.09	.64	0.09	.67
S5	S5P1	–0.10	.62	–0.23	.24	–0.35	.07	–0.19	.33
S6	S6P1	–0.23	.23	–0.11	.57	0.27	.16	–0.08	.66
S7	S7P1	0.19	.33	–0.01	.96	–0.31	.11	–0.13	.49
S7	S7P2	–0.01	.96	0.28	.14	0.19	.32	0.36	.05
S8	S8P1	0.11	.57	0.08	.68	0.22	.25	0.26	.18
S8	S8P2	0.32	.10	0.59	<.001	0.33	.08	0.63	<.001
S9	S9P1	–0.25	.16	–0.05	.80	–0.28	.10	–0.37	.03
S9	S9P2	0.02	.91	0.29	.10	–0.09	.60	–0.16	.38
S10	S10P1	0.29	.13	–0.18	.35	0.04	.84	0.35	.06
S10	S10P2	0.14	.48	0.05	.80	–0.15	.45	–0.14	.46
S11	S11P1	0.39	.03	0.39	.03	–0.06	.75	0.28	.13
S11	S11P2	0.07	.70	0.03	.89	–0.06	.75	0.15	.42
S12	S12P1	0.32	.10	0.27	.16	0.21	.28	0.56	.002
S12	S12P2	0.07	.73	–0.24	.22	–0.02	.94	0.07	.71
S13	S13P1	0.41	.02	–0.19	.30	0.41	.02	0.27	.14
S13	S13P2	0.36	.05	–0.34	.06	0.43	.02	0.27	.15

Table 9. Study A: summary of ecological momentary assessment/peer-ceived momentary assessment Spearman correlations.

Correlation strength	Stress, n (%)	Fatigue, n (%)	Anxiety, n (%)	Computed well-being, n (%)	Total, n (%)
Highly positive[a]	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Moderately positive[b]	5 (25)	2 (10)	3 (15)	6 (30)	16 (20)
Weakly positive[c]	10 (50)	8 (40)	7 (35)	7 (35)	32 (40)
Weakly negative[d]	5 (25)	9 (45)	9 (45)	6 (30)	29 (36)
Moderately negative[e]	0 (0)	1 (5)	1 (5)	1 (5)	3 (4)
Highly negative[f]	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

[a](0.67 to 1.00): Values of the spearman correlation inside this interval are considered highly positive.

[b](0.34 to 0.66): Values of the spearman correlation inside this interval are considered moderately positive.

[c](0.00 to 0.33): Values of the spearman correlation inside this interval are considered weakly positive.

[d](–0.33 to 0.00): Values of the spearman correlation inside this interval are considered weakly negative.

[e](–0.66 to –0.34): Values of the spearman correlation inside this interval are considered moderately negative.

[f](–1.00 to –0.67): Values of the spearman correlation inside this interval are considered highly negative.

For *study B*, the Spearman rank correlation coefficients for each of the states are presented in Table 10. In this case, we reported the correlations between the *reported well-being* of each participant and peer. Correspondingly, Table 11 summarizes the values from Table 10. In this study, 43% of the correlations were weakly positive, followed by a tie of 23% between weakly negative and moderately positive.

XSL•FO

RenderX

Table 10. Study B: ecological momentary assessment/peer-ceived momentary assessment Spearman correlations calculated throughout the study. Each row shows the correlation between the participants' and peer's assessments calculated throughout the study.

Participant ID	Peer ID	Stress		Fatigue		Anxiety		Reported well-being		Computed well-being	
		r_s	P value	r_s	P value	r_s	P value	r_s	P value	r_s	P value
S1	S1P1	–0.39	.02	–0.36	.03	–0.33	.05	0.18	.30	–0.30	.08
S2	S2P1	–0.31	.10	0.33	.08	0.26	.18	0.40	.03	0.13	.51
S3	S3P1	0.37	.04	0.44	.01	0.04	.84	–0.08	.68	0.44	.01
S4	S4P1	0.24	.21	0.16	.42	0.10	.60	–0.17	.38	0.37	.048
S5	S5P1	0.04	.84	0.03	.88	0.06	.76	0.28	.13	–0.01	.95
S6	S6P1	0.44	.07	–0.06	.81	0.24	.34	0.83	<.001	0.26	.30
S7	S7P1	0.41	.03	0.20	.29	–0.25	.19	0.58	<.001	–0.58	.001

Table 11. Study B: summary of ecological momentary assessment/peer-ceived momentary assessment Spearman correlations.

Correlation strength	Stress	Fatigue	Anxiety	Reported well-being	Computed well-being	Total, n (%)
Highly positive[a]	0 (0)	0 (0)	0 (0)	1 (14)	0 (0)	1 (3)
Moderately positive[b]	3 (43)	1 (14)	0 (0)	2 (29)	2 (29)	8 (23)
Weakly positive[c]	2 (29)	4 (57)	5 (71%)	2 (29)	2 (29)	15 (43)
Weakly negative[d]	1 (14)	1 (14)	2 (29%)	2 (29)	2 (29)	8 (23)
Moderately negative[e]	1 (14)	1 (14)	0 (0)	0 (0)	1 (14)	3 (9)
Highly negative[f]	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

[a](0.67 to 1.00): Values of the spearman correlation inside this interval are considered highly positive.

[b](0.34 to 0.66): Values of the spearman correlation inside this interval are considered moderately positive.

[c](0.00 to 0.33): Values of the spearman correlation inside this interval are considered weakly positive.

[d](–0.33 to 0.00): Values of the spearman correlation inside this interval are considered weakly negative.

[e](–0.66 to –0.34): Values of the spearman correlation inside this interval are considered moderately negative.

[f](–1.00 to –0.67): Values of the spearman correlation inside this interval are considered highly negative.

EMA/PeerMA Subcorrelations Including Entry Survey Results

We conducted 3 more correlation analyses relevant to these studies, including the entry survey reports (GERT, PSS, and SDS) as collected within the studies. We again chose the Spearman ranked correlation method because of the small number of samples—17 for *study A* (after removing 3 participant-peer pairs who did not complete the entry surveys) and 7 for *study B*. This small sample size did not permit an accurate assessment of the underlying distribution of the data; hence, we did not assume that the samples were normally distributed. We report results relevant for the feasibility study conducted, even when they were not statistically significant.

EMA/PeerMA Values Versus Entry Surveys

The correlation between participants' and peers' SDS, PSS, and self-considered stressed score (from Table 4), and the r_s coefficient from the Spearman rank correlation for each state (from Table 8 and Table 10) was assessed. This was important to identify direct relationships between individual characteristics (obtained in the entry surveys) and the observed agreement between participants and peers. In *study A*, we found that the higher the participants' SDS scores, the lower the r_s correlation

coefficient for fatigue (Table 8; $r=-0.52$; $P=.03$). The result also holds true for *study B*, although it was not statistically significant (Table 10; $r=-0.40$; $P=.38$). Moreover, the more stressed the peers considered themselves ($r=0.45$; $P=.07$), the higher the correlation was between EMA/PeerMA and anxiety (Table 8). The result also holds true in *study B* (Table 10; $r=0.82$; $P=.02$).

Medians of Within-EMA/PeerMA Value

For both participants and peers, we derived the correlation between the median of the following pairs of states: stress-fatigue, stress-anxiety, stress–computed well-being, fatigue-anxiety, fatigue–computed well-being, and anxiety–computed well-being. *Study B* included the combinations with reported well-being in addition to the aforementioned pairs of states. This is relevant to understand whether participants and peers treat and report some of the states alike (ie, with the same semantics). We found no correlation between stress-fatigue or fatigue-anxiety for neither participants nor peers in both studies. Nevertheless, we observed a positive and close to statistically significant correlation in stress-anxiety for participants and peers in both studies. In *study A*, for participants: $r=0.73$; $P=.001$ and for peers: $r=0.44$; $P=.08$. In *study B*, for participants: $r=0.64$; $P=.12$ and for peers: $r=0.95$;

$P=.001$. Moreover, in *study B*, we did not observe a consistent or statistically significant correlation between reported and computed well-being for participants ($r=0.36$; $P=.42$) or peers ($r=-0.39$; $P=.38$). This result suggests that participants and peers may have considered other important variables beyond the assessments of stress, fatigue, and anxiety when they answered the question about well-being.

Median of EMA/PeerMA Values

The correlation between the participants' median of stress/fatigue/anxiety/well-being and the peers' median of the perceived state was assessed. This is important to understand whether there is high or low agreement in the assessments of the states at the sample level. We used the median because, at the individual level, the assessments are not independent and are not normally distributed. In *study A*, we observed a negative correlation for stress ($r=-0.39$; $P=.11$), a positive correlation for fatigue ($r=0.36$; $P=.15$), and a weakly negative correlation for computed well-being ($r=-0.14$; $P=.58$), although none of them was statistically significant. In *study B*, we observed a positive correlation for stress ($r=0.54$; $P=.20$), anxiety ($r=0.52$; $P=.22$), and computed well-being ($r=0.39$; $P=.38$), although, again, none of them was statistically significant.

EMA/PeerMA: Statistical Agreement

After the correlations evaluated within the previous sections, we focus on the statistical agreement of the EMA/PeerMA assessments by which we assume that the EMA and PeerMA measure the same constructs. To quantify the overall agreement between EMA and PeerMA, we applied the Wilcoxon signed-ranked test to determine whether the medians of the 2 sets (EMAs from participants and PeerMAs from peers) are statistically equal. The justification for choosing this test is that (1) participants' and peers' assessments are not independent; (2) the participant and peers' assessments are paired; and (3) in our datasets, not all individual and peer assessments are normally distributed. The null hypothesis, H_0, of the Wilcoxon signed-ranked test is that the 2 samples have the same distribution. Thus, failing to reject H_0 suggests that participants' and peers' assessments are statistically equal.

For *study A*, the dyads <participant, peer> for which at least one of the states is statistically equal (ie, $P>.05$) are presented in Table 12. For this particular study, 40% (8/20) of the peers reported daily stress assessments, which are statistically equal to those reported by the participant. The values obtained were 30% (6/20) for fatigue, 55% (11/20) for anxiety, and 35% (7/20) for computed well-being.

The results for *study B* are presented in Table 13. In this case, 29% (2/7) of peers reported daily stress assessments, which are statistically equal to those reported by the participant. The values were 43% (3/7) for fatigue and anxiety and 71% (5/7) for both reported and computed well-being.

Table 12. Study A: Wilcoxon signed-ranked significance tests for ecological momentary assessment/peer-ceived momentary assessment ($P=.05$).

Peer ID	*P* value (stress)	*P* value (fatigue)	*P* value (anxiety)	*P* value (computed well-being)
S1P1	.001	<.001	<.001	.34
S2P1	<.001	<.001	.58	.002
S3P1	.74	.004	.21	<.001
S4P1	<.001	.19	.07	.002
S5P1	<.001	.03	.09	<.001
S6P1	.96	.63	.09	.61
S7P2	.003	.002	.001	.18
S8P2	<.001	.13	<.001	<.001
S9P1	<.001	.02	<.001	.23
S9P2	<.001	.05	.001	<.001
S10P1	<.001	<.001	.02	.28
S10P2	.36	.01	.29	.72
S11P1	.37	.003	.40	<.001
S11P2	.77	<.001	.10	<.001
S12P1	.22	.01	<.001	<.001
S12P2	.03	.82	.94	<.001
S13P1	.06	<.001	.12	.002
S13P2	.13	.43	.80	.19

Table 13. Study B: Wilcoxon signed-ranked significance tests for ecological momentary assessment/peer-ceived momentary assessment (P=.05).

Peer ID	P value (stress)	P value (fatigue)	P value (anxiety)	P value (reported well-being)	P value (computed well-being)
S1P1	.003	.03	.01	.07	.42
S2P1	.001	.02	.002	.30	<.001
S3P1	.21	.18	.01	<.001	.50
S4P1	.01	.29	.07	.64	.05
S5P1	.02	.02	.35	.03	.13
S6P1	.06	.50	.56	.45	.06
S7P1	<.001	.03	.01	.07	.01

Operational and Human Factors

Users' Reflections on the mQoL Peers App

Positive Aspects

In general, users found the app easy to use and liked the brief surveys. Some users said that the app helped them to be more aware of their emotions. We quote some of their comments, *study A*: (subject) "Thanks to it I tried to be more aware of my stress level during the day" and (subject) "It makes me think about my attitude and feelings at the moment." For *study B*: (subject) "I liked that surveys were short and easy to understand" and (subject) "I liked the use of scale and colors I wish I can see the results." The users acknowledged the minimal obtrusiveness we aimed for.

Negative Aspects

For *study A*: (subject) "App is not very fun, the exercise quickly becomes monotonous" and (peer) "Same 4 questions 1 month extremely repetitive and annoying." For *study B*: (subject) "The app needs more questions and a dashboard" and (peer) "Short questions could be together in one page." The users acknowledged what we already knew, that the study length and frequency of self-reports are important for the quality of the data collected. In the *Discussion* section, we present implications stemming from these observations as well as future work areas related to the findings.

Users' Reflections on the Studies

Participants' Reflections

We asked participants how they felt about reflecting on their own emotional states. For *study A*, some participants confirmed a known risk of reactivity in *EMA*: (subject) "Sometimes the app has accentuated my stress" or (subject) "Answering the anxiety question often made me feel anxious." Other participants reported enriching experiences:

> I had the impression of being more aware of my anxiety during the study, before, I did not pay special attention to it. [Subject]

> It made me take a step back on myself. For example, when I was stressed, I told myself that I had to calm down. When I had high fatigue, I said to myself that I had to sleep better to recover, not look at my phone before sleeping. [Subject]

For *study B*, the experiences were mostly positive, for example:

> It allowed me to anchor myself and see what is causing me to feel stressed. [Subject]

> Loved it. It made me more aware of my emotional states than ever before. [Subject]

The higher sampling frequency in *study A* (8 per day vs 3 per day in *study B*) could explain the unintended side effects experienced by some of its participants.

Peers' Reflections

We asked peers how they felt about reflecting on someone else's emotional states during the study. In both studies, some peers said that the task was challenging to complete at times. Others said that participating in the study allowed them to learn more about emotional states. For example, for *study A*: (peer) "Sometimes it was hard because we hadn't talked for hours" and (peer) "If I did not see her, I had no idea what her emotional state was." On the contrary, peers reporting more enriching experiences said:

> I have the feeling to take stress problems more seriously, not like everyone is stressed and it is normal, but to understand that stress can block life of some people. [Peer]

> It made me think if she is doing well and this experience made me write to her more often. [Peer]

For *study B* these peers stated:

> It was a little harder than I had expected. I typically use facial expression and tone to determine my friends emotional states. On days that I don't see her, I'd have to rely on how she texts. [Peer]

> It allowed me to learn about his mood every day and know him better and any problems going on. [Peer]

Implications of Technology Choices

This section summarizes the results related to the technology choices in the 2 studies and how they may influence the methods' feasibility. In the 2 studies, we assumed that both EMA and PeerMA have the same technological requirements: (1) the mechanism to trigger the questions to participants at desired moments and (2) the channel to trigger the questions to the user and collect the answers reliably. As explained in the *Tools* section, we implemented EMA and PeerMA using our own *mQoL Lab* platform, detailed in the study by Berrocal et al [51].

To trigger the questions, van Berkel et al [63] provide a comprehensive overview of the multiple options researchers can choose from. We used uniformly randomized signal-contingent triggers throughout the day (waking hours). In *study A*, the *mQoL Peers* app was responsible for scheduling and triggering the signals in the participants' smartphones to present an EMA. These signals triggered a push notification to the peers' smartphones to initiate a corresponding PeerMA. This approach had 2 disadvantages: (1) it used computing resources of the users' smartphones and (2) the users could (accidentally or intentionally) change the system settings causing negative consequences for the study. In *study B*, we reduced those risks and obtained favorable results by scheduling and triggering the signals from a server component, also via push notifications, simultaneously to both participants and peers.

Regarding the channel itself, smartphones are commonly used for mobile human studies as they are often close to the owner [64]. However, cross-platform compatibility may be an obstacle as the smartphone type would need to be used as inclusion criteria, potentially biasing the study results. We had this experience ourselves. Namely, for *study A*, our platform was compatible only with Android. For *study B*, we improved our platform to support iOS-based devices either by the participant, the peer, or both. Nevertheless, in *study B*, 2 senior candidates were excluded during the study recruitment phase as they did not use the smartphone often and their preferred channel was their tablet computer. Researchers designing studies may consider this audience as well and design their study accordingly.

Suggestions From Users About Technology and Methods

The following are recommendations made by users (participants and peers) in our studies. First, they wanted to have some kind of dashboard to see their previous assessments and track how many they completed each day. This may have positive effects on response compliance; however, it may imply higher reactivity to the study itself, where a momentary, ecologically valid EMA may be influenced by the number or content of the past EMAs (depending on the dashboard design). The participants also wanted greater freedom to select the moods or states that they felt confident about reporting at a given moment. Such a level of freedom is possible, but it should be done carefully to salvage the collection of ecologically valid EMAs and PeerMAs, assuring the joint understanding of participants/peers of the state being assessed (eg, snacking on foods throughout a day) as well as to ensure that the data being collected directly relates to the main goal of the study. Additionally, such a participant-driven EMA/PeerMA study design may result in bias stemming from collecting only the states the participants want to report.

One peer in *study A* (whose relationship with the subject was "a friend from the university") suggested that the participants' assessment should start with a question: "Is your peer next to you to answer a short survey?" This is a valid observation that can be explored in future studies as a way to study the validity of the PeerMA answers. In our studies, peers were able to express their confidence levels for each individual PeerMA.

They may have selected low confidence if they had not been in contact (either physical or virtual) with the participant in the recent past. Future research could include software assessment of proximity between participants and peers (eg, using Bluetooth, using Wi-Fi, using Geo Fences, or leveraging the social network apps' usage patterns) as a way to reduce the burden for peers being requested to assess the state of the individual when the probability of obtaining a reliable assessment is low.

Discussion

Structure

In summary, our first research aim was to explore the feasibility of applying the PeerMA method to involve peers to assess phenomena such as mental states in healthy individuals. The second aim was to determine operational aspects and human factors that need to be taken into account to most effectively use the PeerMA method. We reflect upon the overall results to answer these questions.

Feasibility of Using PeerMA

In this section, to address the first research aim, we summarize the experience from both studies related to the tangible contributions from participants and peers using both the EMA and the PeerMA methods.

Regarding the users' recruitment and participation, we conclude the following: we first noticed that some participants had difficulties finding a peer. In *study A*, we initially had 18 participants; 5 participants who attempted to enroll and agreed to try and find a peer stopped later, indicating that they were unable to find a peer to participate. Hence, we ended up with 13 participants with peers in *study A*. In *study B*, we also started with 18 participants, with 8 cases in which the participant enrolled in the study but never enrolled a peer. Two of them informed us about the situation and six stopped without further contact. The three participants who completed the full study but did not manage to involve a peer were excluded from the analysis.

Once enrolled, participant retention was high. Although recruiting participants with peers required more effort than recruiting participants for EMA-only studies, we successfully completed two field studies of 4 weeks' duration in two geographically distant locations, as described here. In both studies, we observed that almost all participants who enrolled with a peer from the beginning were able to continue in the study till the end, on average, 29 (SD 1.7) days for participants and 25 (SD 5.9) days for peers in *study A* and 27 (SD 4.5) days for participants and 27 (SD 4.8) days for peers in study B. However, we noticed that some individuals had a low EMA/PeerMA response rate, as shown in Table 5, especially in *study A* (<10%). We believe this was at least partially due to a high number of daily EMA/PeerMA notifications (8 per day) that came along usual participants and peers' obligations during the day. We limited the number of notifications in *study B* to 3 per day, which resulted in a better EMA/PeerMAs response rate (Table 5; only 1 peer had a response rate as low as 30%).

When it comes to the overall agreement between EMA and PeerMA assessments, the evidence for the feasibility of the PeerMA method is as follows. Although not conclusive or generalizable, we observed a strong, close to statistically significant correlation between participant and peer's assessments of stress and anxiety (for *study A*, participants: $r=0.73$*; P=*.001 and peers: $r=0.44$*; P=*.08 and for *study B*, participants $r=0.64$*; P=*.12 and peers $r=.95$*; P=*.001; details in the *EMA and PeerMA Correlations* section). In both studies, the more stressed the peers described themselves in the entry survey (*study A*: $r=0.45$*; P=*.07 and *study B*: $r=0.82$*; P=*.02), the higher the correlation between EMA and PeerMA for anxiety (Table 8 and Table 10, respectively; details in the *EMA and PeerMA Correlations* section). Collectively, this suggests that stressed individuals tend to be more effective in detecting anxiety in their peers (already studied by Harrigan et al [65]) and that anxiety could be a proxy for the level of stress of a person as stress and anxiety were positively correlated in these particular samples.

Additionally, we recall the overall percentage of peers who reported daily assessments statistically equal to those reported by their participants. In *study A*, the results were 40% for stress, 30% for fatigue, 55% for anxiety, and 35% for computed well-being. In *study B*, the results were 29% for stress, 43% for fatigue and anxiety, and 71% for computed and reported well-being (details in the *EMA and PeerMA Statistical Agreement* section). In summary, in both studies, participants and peers achieved higher agreement in their assessments of anxiety and fatigue and lower agreement in their assessment of stress. We reason for this result as follows.

Anxiety is a complex state and highly involuntary [65], and it is presumably difficult to hide it as specific individuals try to do with stress. Manifestations of anxiety are anchored to a particular event or situation. Close peers who are aware of the events or situations can use that knowledge to estimate the state of the person [58]. Alternatively, a high agreement could also result from a case in which participants and peers interpret stress and anxiety as the same or very similar state and report it accordingly.

In addition, we consider that detecting fatigue in peers is less challenging as people tend to talk about it openly. Nevertheless, one person may be highly fatigued after 1 night of poor sleep, whereas another person may not reach that same level after 2 or 3 nights of poor sleep. For this purpose, psychometric models at the individual level help make comparisons among the assessments [58].

Finally, stress has a social component affected by stereotypes [66] (eg, work or school performance, fear of public speaking), and the social (external) manifestation of stress may produce different physiological (internal) reactions and external observable behaviors among individuals. It is ultimately the physiological stress that causes the underlying threat to a person's health. Potentially due to the social complexity of the stress phenomena, our observed results do not lead to strong conclusions for employing PeerMA when assessing stress. Nevertheless, the combined use of EMA and PeerMA within a specific study, to complement other data collection methods (eg, psychophysiology), represents a possible way to study peer-based stress assessment further.

In summary, there are challenges and open questions regarding interoceptive awareness (ability to consciously sense the inner state of the body) to consider when using EMA to study emotional and physical states [67]. PeerMA is constrained by similar, but not identical, limitations as peers report about a state that occurs in the individual being observed [58]. Nevertheless, a frequent and careful pairing of self-assessments and others' assessments (as shown in Figures 2 and 3), although subjective, may lead to new sources of information that we consider valuable to study how a person's emotional and physical states unfolded over time. The results of our studies suggest that applying the PeerMA method to the study of complex phenomena, such as mental states in individuals, is feasible and can open up new perspectives to examine the relationships between self-assessments and others' perceptions that are not possible to obtain from studies based on surveys or EMAs alone.

Human Factors and Operational Aspects of PeerMA

We address the second research aim by summarizing operational aspects and human factors derived from the experience of using the PeerMA method in the two studies. We discuss the implications of certain technological choices and offer recommendations for researchers who wish to include the method in their studies.

Recruiting and Retaining Participants

To begin with, recruiting participants for studies is a known challenge [68]. Hence, we know that recruiting participants for paired studies that combine PeerMA with EMA is challenging as well. In both studies, we observed that several participants who initially applied to join the study did not actually enroll after the web-based meeting with the researcher (in which they learned that bringing a peer was a requirement for these studies). In *study A*, 5 participants who actually enrolled and agreed to try and find a peer later stopped participating, indicating that they were unable to find a peer willing to participate. In *study B*, we observed 8 cases in which the participant enrolled in the study but never enrolled a peer. Two of them informed us about the situation, whereas the remaining six simply stopped without further contact.

We found that one recommendation is to start applying the PeerMA method in cohorts for which reaching out to a peer becomes less complicated, and more motivating and valuable for both the participants and the peers. For instance, at the time of this writing, we are conducting one study with adult patients of the Stanford Medical Center recovering from a liver transplant. In this case, the patients answer EMAs, whereas their support person answers PeerMAs. Recruiting peers for this particular group was more straightforward because of the anticipated clinical value of such a study.

Another recommendation, which relates to the technological choices influencing the feasibility of the methods, is to include gamification techniques (eg, scoring points, winning prizes, and solving puzzles) as part of the study dynamics, which can

provide an incentive for users to contribute data along the study duration and complete the study [69,70].

PeerMA: Methodological Opportunities

On the contrary, based on the experience during these two exploratory studies, we found it worth exploring how the EMA and PeerMA can be combined during a study to provide accurate, timely information about the observed participant state (and its changes). One suggestion is to modulate the administration of EMA and PeerMA based on prior knowledge about the participants' state of interest and individual just-in-time answers. This would imitate the computerized adaptive tests, in which the questions are tailored to the past answers of the individual. In our case, the question of well-being (or a similarly discriminating question) could always be first. From our study design and results, we know that well-being encompasses stress, fatigue, or anxiety, and potentially, other states. Then, the next question would be chosen based on the answer (high or low well-being); for low well-being, relevant question(s) for the current state (eg, stress, fatigue, or anxiety) would be triggered to the participant and the peer. For high well-being, the survey may be completed. This approach may reduce the participants' burden of monotonously answering questions when there may not be new information to provide. In other cases, the peers could be asked only to validate the response of the participant (eg, "Have you taken your medication?"). Nevertheless, some participants may feel their privacy is at stake when their peers answer a PeerMA every time they answer an EMA. Participants could have the option to decide—in real time—not to send the PeerMA to peers if they want to regain control over their privacy.

Another variation for the PeerMA design is to give users the options to slightly customize the time windows when they feel available and willing to answer certain types of questions. Some of them may be more engaged if they can choose the type of signal as well as the time window when they are better prepared to take an EMA or PeerMA (in a similar way as they make other daily choices like taking a cup of coffee or making a personal phone call). On the other hand, this could be detrimental to the research if participants are preparing to give a specific answer, knowing that the questions are being triggered at specific times, or if they choose to not answer questions at times when they feel more stressed, causing data loss and result bias. Additionally, some studies with PeerMA may allow users to take both the roles of participants and peers simultaneously, which allows them to report on each other's states, which may increase engagement.

Overall, as an exploratory method, there are yet many opportunities to design studies leveraging PeerMA. The main methodological question relates to what information or measure about the user state could be collected in reliable and minimally obtrusive ways from the participant and his/her peers. If chosen properly, we believe that such information collected from participants and peers simultaneously could enable further understanding of the observed state.

Limitations and Future Work

One limitation of these studies is the lack of ground truth of the assessed conditions (stress, fatigue, or anxiety). Despite the data presented in this preliminary analysis, we were unable to determine whether the EMA or the PeerMA were closer to the actual state of the participants. The limitation is inherent to any self-report, EMA-based study. To further examine the reliability of PeerMA, more research is needed to incorporate more modalities, such as heart rate variability for physiological stress.

Another limitation of our work is the small sample size of participants and peers. As indicated earlier, a larger sample is necessary to further investigate the reliability of the assessments as well as the effects introduced by sample characteristics such as the amount of time peers interact with the participants during a day, type of relationship between participant and peer, or coresidence as reported by Neumann et al [8], among others.

Additionally, research is needed to examine whether PeerMA affects the usual behavior of the participant during a study. Namely, if the participant, knowing that he/she is being observed, explicitly changes his/her state and behavior when interacting with the peer. Similarly, further research is needed to include the state of the peer at the moment of answering a PeerMA, for example, to understand the possible effects or biases related to the ego depletion theory [71].

Finally, our future work includes a case by case examination of how peers' reported level of confidence (ie, low, moderate, or high), as well as other socioeconomic characteristics, influence the results derived from the EMA and PeerMA agreement.

Conclusions

We presented results from two user studies conducted in the participants' natural daily life environments, evaluating the first version of a platform implementing the PeerMA method deployed on users' personal smartphones. The studies showed encouraging results from a total of 20 participants and 27 peers contributing multiple daily assessments for approximately 4 weeks each. In the studies, we collected empirical evidence regarding the feasibility of the method. We discussed the methodological and human aspects related to the application of the PeerMA method to study real-life phenomena, including those related to mental health. We demonstrated that users accepted the method and provided valuable feedback. We identified and discussed improvement opportunities that could lead to higher user engagement as well as more elaborate methodological options for researchers to explore when leveraging PeerMA in their studies. We discussed technical aspects to consider for a reliable, technology agnostic, and minimally obtrusive implementation of the PeerMA method.

We believe that the PeerMA method evaluated in this study opens a new perspective to study an individual's state based on frequent and possibly paired observations from trusted peers beyond the information traditionally obtained with EMA. As an independent observation, it has value for applications in clinical settings to evaluate the severity of and support treatment of mental disorders such as OCD or addictions. However, more research is needed to guarantee reliable utilization with sufficient

control to manage potential emergent biases stemming from either the participants or the peers or the momentary context in which PeerMA is triggered.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. Ann Behav Med 1994(3):199-2020 [FREE Full text] [doi: 10.1093/abm/16.3.199]
2. Csikszentmihalyi M, Larson R, Prescott S. The ecology of adolescent activity and experience. J Youth Adolesc 1977 Sep;6(3):281-294. [doi: 10.1007/BF02138940] [Medline: 24408457]
3. Reed LC, Csikszentmihalyi M. The experience sampling method BT. In: Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi. Netherlands, Dordrecht: Springer; 2014.
4. Paulhus DL, Vazire S. The self-report method. In: Robins RW, Fraley RC, Krueger RF, editors. Handbook of Research Methods in Personality Psychology. New York, USA: The Guilford Press; 2007.
5. Vazire S. Informant reports: a cheap, fast, and easy method for personality assessment. J Res Pers 2006 Oct;40(5):472-481 [FREE Full text] [doi: 10.1016/j.jrp.2005.03.003]
6. Kazdin AE. Observer effects: Reactivity of direct observation. New Direct Method Soc Behav Sci 1982;14:5-9 [FREE Full text]
7. Mayo NE, Figueiredo S, Ahmed S, Bartlett SJ. Montreal accord on patient-reported outcomes (PROs) use series - paper 2: terminology proposed to measure what matters in health. J Clin Epidemiol 2017 Sep;89:119-124. [doi: 10.1016/j.jclinepi.2017.04.013] [Medline: 28433673]
8. Neumann PJ, Araki SS, Gutterman EM. The use of proxy respondents in studies of older adults: lessons, challenges, and opportunities. J Am Geriatr Soc 2000 Dec;48(12):1646-1654. [doi: 10.1111/j.1532-5415.2000.tb03877.x] [Medline: 11129756]
9. Kromm W, Gadinger MC, Schneider S. Peer ratings of chronic stress: can spouses and friends provide reliable and valid assessments of a target person's level of chronic stress? Stress Health 2010 Jan 21;26(4):292-303 [FREE Full text] [doi: 10.1002/smi.1297]
10. Berrocal A, Wac K. Peer-Vasive Computing: Leveraging Peers to Enhance the Accuracy of Self-reports in Mobile Human Studies. In: Adjunct Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2018 ACM International Symposium on Wearable Computers. 2018 Presented at: Ubicomp/ISWC'18; October 8-12, 2018; Singapore URL: https://doi.org/10.1145/3267305.3267542 [doi: 10.1145/3267305.3267542]
11. McElroy SL, Phillips KA, Keck PE. Obsessive compulsive spectrum disorder. J Clin Psychiatry 1994 Oct;55(Suppl):33-51; discussion 52. [Medline: 7961531]
12. Balsis S, Cooper LD, Oltmanns TF. Are informant reports of personality more internally consistent than self reports of personality? Assessment 2015 Aug;22(4):399-404 [FREE Full text] [doi: 10.1177/1073191114556100] [Medline: 25376588]
13. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York, USA: Springer; 1984.
14. Sharpe M, Wilks D. Fatigue. BMJ (Clinical research ed 2020:1756-1833 [FREE Full text] [doi: 10.1136/bmj.325.7362.480]
15. Jansen KL, Fortenberry KT, Clark MS. Anxiety and Its Measurement. New York, USA: Springer; 2013.
16. Huppert FA. Psychological well-being: evidence regarding its causes and consequences. Appl Psychol: Health Well-Being 2009;1(2):164 [FREE Full text] [doi: 10.1111/j.1758-0854.2009.01008.x]
17. Versluis A, Verkuil B, Lane RD, Hagemann D, Thayer JF, Brosschot JF. Ecological momentary assessment of emotional awareness: preliminary evaluation of psychometric properties. Curr Psychol 2018 Nov 24:- [FREE Full text] [doi: 10.1007/s12144-018-0074-6]
18. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
19. MacKerron G, Mourato S. Happiness is greater in natural environments. Global Environ Chang 2013 Oct;23(5):992-1000 [FREE Full text] [doi: 10.1016/j.gloenvcha.2013.03.010]
20. Runyan JD, Steinke EG. Virtues, ecological momentary assessment/intervention and smartphone technology. Front Psychol 2015;6:481 [FREE Full text] [doi: 10.3389/fpsyg.2015.00481] [Medline: 25999869]
21. Santangelo PS, Ebner-Priemer UW, Trull TJ. Experience Sampling Methods in Clinical Psychology. Number July. 2019. URL: https://doi.org/10.1093/oxfordhb/9780199793549.013.0011 [accessed 2020-06-23]

22.   Trull TJ, Ebner-Priemer UW. Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. Psychol Assess 2009 Dec;21(4):457-462 [FREE Full text] [doi: 10.1037/a0017653] [Medline: 19947780]

23.   Yang YS, Ryu GW, Choi M. Methodological strategies for ecological momentary assessment to evaluate mood and stress in adult patients using mobile phones: systematic review. JMIR Mhealth Uhealth 2019 Apr 1;7(4):e11215 [FREE Full text] [doi: 10.2196/11215] [Medline: 30932866]

24.   Beal DJ. ESM 2.0: state of the art and future potential of experience sampling methods in organizational research. Annu Rev Organ Psychol Organ Behav 2015 Apr 10;2(1):383-407 [FREE Full text] [doi: 10.1146/annurev-orgpsych-032414-111335]

25.   Walker M, Consolvo S. Using the experience sampling method to evaluate ubicomp applications. Pervasive Comput 2003;2(2):24-31 [FREE Full text] [doi: 10.1109/mprv.2003.1203750]

26.   Gosling SD, John OP, Craik KH, Robins RW. Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. J Pers Soc Psychol 1998 May;74(5):1337-1349. [doi: 10.1037//0022-3514.74.5.1337] [Medline: 9599447]

27.   Paul TC, McCrae RR. The revised NEO personality inventory (NEO- PI-R). In: The Sage Handbook of Personality Theory and Assessment. Thousand Oaks, CA: Sage Publications; 2008.

28.   Watson D, Hubbard B, Wiese D. Self-other agreement in personality and affectivity: the role of acquaintanceship, trait visibility, and assumed similarity. J Pers Soc Psychol 2000 Mar;78(3):546-558. [doi: 10.1037//0022-3514.78.3.546] [Medline: 10743880]

29.   Watson D, Clark LA. The Panas-X: Manual for the Positive and Negative Affect Schedule - Expanded Form. Department of Psychological & Brain Sciences. 1994. URL: http://www2.psychology.uiowa.edu/faculty/Clark/PANAS-X.pdf [accessed 2020-06-23]

30.   Chmielewsk MS, Morgan TA. Five-Factor Model of Personality. In: Encyclopedia of Behavioral Medicine. New York, NY: Springer; 2013:803-804.

31.   Drum DJ, Brownson C, Burton A, Smith SE. New data on the nature of suicidal crises in college students: shifting the paradigm. Prof Psychol-Res Pr 2009;40(3):213-222 [FREE Full text] [doi: 10.1037/a0014465]

32.   Eisenberg D, Hunt J, Speer N, Zivin K. Mental health service utilization among college students in the United States. J Nerv Ment Dis 2011 May;199(5):301-308. [doi: 10.1097/NMD.0b013e3182175123] [Medline: 21543948]

33.   Eisenberg D, Hunt J, Speer N. Help seeking for mental health on college campuses: review of evidence and next steps for research and practice. Harv Rev Psychiatry 2012;20(4):222-232. [doi: 10.3109/10673229.2012.712839] [Medline: 22894731]

34.   Cusick CP, Brooks CA, Whiteneck GG. The use of proxies in community integration research. Arch Phys Med Rehabil 2001 Aug;82(8):1018-1024. [doi: 10.1053/apmr.2001.25098] [Medline: 11494179]

35.   Duncan PW, Lai SM, Tyler D, Perera S, Reker DM, Studenski S. Evaluation of proxy responses to the stroke impact scale. Stroke 2002 Nov;33(11):2593-2599. [doi: 10.1161/01.str.0000034395.06874.3e] [Medline: 12411648]

36.   Sneeuw KC, Aaronson NK, de Haan RJ, Limburg M. Assessing quality of life after stroke. The value and limitations of proxy ratings. Stroke 1997 Aug;28(8):1541-1549. [doi: 10.1161/01.str.28.8.1541] [Medline: 9259746]

37.   Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 8;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]

38.   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 7-11, 2014; Osaka, Japan URL: https://doi.org/10.1145/2632048.2632054 [doi: 10.1145/2632048.2632054]

39.   Ciman M, Wac K. Individuals' stress assessment using human-smartphone interaction analysis. IEEE Trans Affective Comput 2018 Jan 1;9(1):51-65 [FREE Full text] [doi: 10.1109/taffc.2016.2592504]

40.   Gjoreski M, Gjoreski H, Lustrek M, Gams M. Preface. In: International Conference on Intelligent Environments. 2015 Presented at: IE'15; July 15-17, 2015; Prague, Czech Republic URL: https://doi.org/10.1109/IE.2015.27 [doi: 10.1109/ie.2015.27]

41.   Hovsepian K, Al'absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. cStress: Towards a Gold Standard for Continuous Stress Assessment in the Mobile Environment. 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 URL: https://doi.org/10.1145/2750858.2807526 [doi: 10.1145/2750858.2807526]

42.   Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. J Med Internet Res 2018 Jun 8;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]

43.   Sano A, Picard RW. Stress Recognition using Wearable Sensors and Mobile Phones. In: Humaine Association Conference on Affective Computing and Intelligent Interaction. 2013 Presented at: ICII'13; September 2-5, 2013; Geneva, Switzerland URL: https://doi.org/10.1109/ACII.2013.117 [doi: 10.1109/acii.2013.117]

44. Suhara Y, Xu Y, Pentland AS. DeepMood: Forecasting Depressed Mood Based on Self-Reported Histories via Recurrent Neural Networks. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3-7, 2017; Perth, Australia URL: https://doi.org/10.1145/3038912.3052676 [doi: 10.1145/3038912.3052676]

45. Wang R, Aung MS, Andullah S, Brian R, Campbell AT, Choudhury T, et al. Crosscheck: Toward Passive Sensing and 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 URL: https://doi.org/10.1145/2971648.2971740 [doi: 10.1145/2971648.2971740]

46. Harari GM, Müller SR, Aung MS, Rentfrow PJ. Smartphone sensing methods for studying behavior in everyday life. Curr Opin Behav Sci 2017 Dec;18:83-90 [FREE Full text] [doi: 10.1016/j.cobeha.2017.07.018]

47. Gresham G, Hendifar AE, Spiegel B, Neeman E, Tuli R, Rimel BJ, et al. Wearable activity monitors to assess performance status and predict clinical outcomes in advanced cancer patients. NPJ Digit Med 2018;1:27 [FREE Full text] [doi: 10.1038/s41746-018-0032-6] [Medline: 31304309]

48. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. Perspect Psychol Sci 2016 Nov;11(6):838-854 [FREE Full text] [doi: 10.1177/1745691616650285] [Medline: 27899727]

49. Gamba J, Rashed M, Razaghpanah A, Tapiado J, Vallina-Rodriguez N. An analysis of pre-installed android software. Arxiv 2019:- epub ahead of print(1905.02713) [FREE Full text]

50. Lazar A, Koehlern C, Tanenbaum J, Nguyen DH. Why We Use and Abandon Smart Devices. 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 URL: https://doi.org/10.1145/2750858.2804288 [doi: 10.1145/2750858.2804288]

51. Berrocal A, Manea V, DeMasi A, Wac K., Vlad, Alexandre, Katarzyna. mqol-lab: Step-by-step creation of a flexible platform to conduct studies using interactive, mobile, wearable and ubiquitous devices. Proceedings of the 17th International Conference on Mobile Systems and Pervasive Computing, MobiSPC. 2020 Aug. URL: https://www.qualityoflifetechnologies.com/r-publications/mqol-lab-step-by-step-creation-of-a-flexible-platform-to-conduct-studies-using-interactive-mobile-wearable-and-ubiquitous-devices/

52. DeMasi A, Ciman M, Gustarini M, Wac K. mQoL Smart Lab: Quality of Life Living Lab for Interdisciplinary Experiments. 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. [doi: 10.1145/2968219.2971593]

53. QoL Technologies Lab. 2019. URL: https://www.qualityoflifetechnologies.com/ [accessed 2019-11-15]

54. Schlegel K, Scherer KR. Introducing a short version of the Geneva emotion recognition test (GERT-S): psychometric properties and construct validation. Behav Res Methods 2016 Dec;48(4):1383-1392. [doi: 10.3758/s13428-015-0646-4] [Medline: 26416137]

55. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]

56. Crowne DP, Marlowe DA. A new scale of social desirability independent of psychopathology. J Consult Psychol 1960 Aug;24:349-354. [doi: 10.1037/h0047358] [Medline: 13813058]

57. Rosenzveig A, Kuspinar A, Daskalopoulou SS, Mayo NE. Toward patient-centered care: a systematic review of how to ask questions that matter to patients. Medicine (Baltimore) 2014 Nov;93(22):e120 [FREE Full text] [doi: 10.1097/MD.0000000000000120] [Medline: 25396331]

58. Uher J. Quantitative data from rating scales: an epistemological and methodological enquiry. Front Psychol 2018;9:2599 [FREE Full text] [doi: 10.3389/fpsyg.2018.02599] [Medline: 30622493]

59. Bandelow B, Michaelis S. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin Neurosci 2015 Sep;17(3):327-335 [FREE Full text] [Medline: 26487813]

60. Buysse DJ. Sleep health: can we define it? Does it matter? Sleep 2014 Jan 1;37(1):9-17 [FREE Full text] [doi: 10.5665/sleep.3298] [Medline: 24470692]

61. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: the day reconstruction method. Science 2004 Dec 3;306(5702):1776-1780 [FREE Full text] [doi: 10.1126/science.1103572] [Medline: 15576620]

62. Kim S, Kim H. A new metric of absolute percentage error for intermittent demand forecasts. Int J Forecast 2016 Jul;32(3):669-679 [FREE Full text] [doi: 10.1016/j.ijforecast.2015.12.003]

63. van Berkel N, Ferreira D, Kostakos V. The experience sampling method on mobile devices. ACM Comput Surv 2018 Jan 12;50(6):1-40 [FREE Full text] [doi: 10.1145/3123988]

64. Dey AK, Wak K, Ferreira D, Tassini K, Hong J, Rojas J. Getting Closer: an Empirical Investigation of the Proximity of User to Their Smart Phones. In: Proceedings of the 13th International Conference on Ubiquitous Computing. 2011 Presented at: UbiComp'11; September 11-17, 2011; Beijing, China URL: https://doi.org/10.1145/2030112.2030135 [doi: 10.1145/2030112.2030135]

65. Harrigan JA, Lucic KS, Bailyn L, Zarnowiecki S, Rosenthal R. Judging others' anxiety. J Appl Social Pyschol 1992 Jun;22(11):855-873 [FREE Full text] [doi: 10.1111/j.1559-1816.1992.tb00929.x]

66.   Robinson MD, Johnson JT. Is it emotion or is it stress? Gender stereotypes and the perception of subjective experience. Sex Roles 1997 Feb;36(3-4):235-258 [FREE Full text] [doi: 10.1007/bf02766270]

67.   Price CJ, Hooven C. Interoceptive awareness skills for emotion regulation: theory and approach of mindful awareness in body-oriented therapy (MABT). Front Psychol 2018;9:798 [FREE Full text] [doi: 10.3389/fpsyg.2018.00798] [Medline: 29892247]

68.   Manea V, Hansen MS, Elbeyi SE, Wac K. Towards personalizing participation in health studies. In: Proceedings of the 4th International Workshop on Multimedia for Personal Health & Health Care. 2019 Presented at: MM'19; October 21, 2019; Nice, France. [doi: 10.1145/3347444.3356241]

69.   Taylor S, Ferguson C, Peng F, Schoeneich M, Picard RW. Use of in-game rewards to motivate daily self-report compliance: randomized controlled trial. J Med Internet Res 2019 Jan 3;21(1):e11683 [FREE Full text] [doi: 10.2196/11683] [Medline: 30609986]

70.   van Berkel N, Goncalves J, Hosio S, Kostakos V. Gamification of mobile experience sampling improves data quality and quantity. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-21 [FREE Full text] [doi: 10.1145/3130972]

71.   Muraven M. Ego Depletion: Theory and Evidence. Oxford, UK: Oxford University Press; 2012.

Abbreviations

EMA: ecological momentary assessment
ESM: Experience Sampling Method
GERT: Geneva Emotion Recognition Test
MAAPE: mean arctangent absolute percentage error
MAPE: mean absolute percent error
MDA: mean directional accuracy
OCD: obsessive-compulsive disorder
PeerMA: peer-ceived momentary assessment
PSS: Perceived Stress Scale
SDS: Social Desirability Scale
UNIGE: University of Geneva

<u>Original Paper</u>

Sensory-Discriminative Three-Dimensional Body Pain Mobile App Measures Versus Traditional Pain Measurement With a Visual Analog Scale: Validation Study

Niko Kaciroti[1,2,3*], PhD; Marcos Fabio DosSantos[4,5*], DDS, MS; Brenda Moura[2,5*], DDS, MS; Emily Light Bellile[6], MSc; Thiago Dias Nascimento[1], DDS, MSci; Eric Maslowski[7], BSc; Theodora E Danciu[8], PhD; Adam Donnell[9], PhD; Alexandre F DaSilva[1,2], DMed

[1]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, United States

[2]Headache & Orofacial Pain Effort (H.O.P.E.), Department of Biologic and Materials Sciences, School of Dentistry, University of Michigan, Ann Arbor, MI, United States

[3]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, United States

[4]Instituto de Ciências Biomédicas (ICB), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

[5]Programa de Pós-Graduação em Medicina (Radiologia), Universidade Federal do Rio de Janeiro (UFRJ)., Rio de Janeiro, Brazil

[6]Department of Public Health, University of Michigan, Ann Arbor, MI, United States

[7]Moxytech Inc, Ann Arbor, MI, United States

[8]Department of Periodontics and Medicine School of Dentistry, University of Michigan, Ann Arbor, Michigan, MI, United States

[9]Orthodontics, Department of Developmental Biology, Harvard University, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Alexandre F DaSilva, DMed
Headache & Orofacial Pain Effort (H.O.P.E.)
Department of Biologic and Materials Sciences
School of Dentistry, University of Michigan
1011 N. University Ave., Room 1014A,
Ann Arbor, MI, 48109-1078
United States
Phone: 1 (734) 615 3807
Email: adasilva@umich.edu

Abstract

Background: To quantify pain severity in patients and the efficacy treatments, researchers and clinicians apply tools such as the traditional visual analog scale (VAS) that leads to inaccurate interpretation of the main sensory pain.

Objective: This study aimed to validate the pain measurements of a neuroscience-based 3D body pain mobile app called GeoPain.

Methods: Patients with temporomandibular disorder (TMD) were assessed using GeoPain measures in comparison to VAS and positive and negative affect schedule (PANAS), pain and mood scales, respectively. Principal component analysis (PCA), scatter score analysis, Pearson methods, and effect size were used to determine the correlation between GeoPain and VAS measures.

Results: The PCA resulted in two main orthogonal components: first principal component (PC1) and second principal component (PC2). PC1 comprises a combination score of all GeoPain measures, which had a high internal consistency and clustered together in TMD pain. PC2 included VAS and PANAS. All loading coefficients for GeoPain measures in PC1 were above 0.70, with low loadings for VAS and PANAS. Meanwhile, PC2 was dominated by a VAS and PANAS coefficient >0.4. Repeated measure analysis revealed a strong correlation between the VAS and mood scores from PANAS over time, which might be related to the subjectivity of the VAS measure, whereas sensory-discriminative GeoPain measures, not VAS, demonstrated an association between chronicity and TMD pain in locations spread away from the most commonly reported area or pain epicenter ($P=.01$). Analysis using VAS did not detect an association at baseline between TMD and chronic pain. The long-term reliability (lag >1 day) was consistently high for the pain area and intensity number summation (PAINS) with lag autocorrelations averaging between 0.7 and 0.8, and greater than the autocorrelations for VAS averaging between 0.3 and 0.6. The combination of higher reliability

XSL•FO
RenderX

for PAINS and its objectivity, displayed by the lack of association with PANAS as compared with VAS, indicated that PAINS has better sensitivity and reliability for measuring treatment effect over time for sensory-discriminative pain. The effect sizes for PAINS were larger than those for VAS, consequently requiring smaller sample sizes to assess the analgesic efficacy of treatment if PAINS was used versus VAS. The PAINS effect size was 0.51 SD for both facial sides and 0.60 SD for the right side versus 0.35 SD for VAS. Therefore, the sample size required to detect such effect sizes with 80% power would be n=125 per group for VAS, but as low as n=44 per group for PAINS, which is almost a third of the sample size needed by VAS.

Conclusions: GeoPain demonstrates precision and reliability as a 3D mobile interface for measuring and analyzing sensory-discriminative aspects of subregional pain in terms of its severity and response to treatment, without being influenced by mood variations from patients.

Introduction

Background

The true complexity of a given pain area and intensity and pain progression are never precisely documented at individual, longitudinal, and epidemiologic levels. Hence, crucial information in pain investigation and treatment is inexorably lost. This is especially true when patients have chronic and overlapping pain conditions, where pain is refractory to treatment and usually transverse or co-exist in multiple locations and dermatomes with different intensities. Research studies from the past 10 years have identified a large overlap between a number of chronic pain conditions, including temporomandibular disorder (TMD), migraine, neuropathic pain, fibromyalgia, irritable bowel syndrome, and so forth [1]. In fact, the presence of multiple large body areas with pain is associated with worse pain experience and prognosis [2-5]. Nonetheless, patients are frequently asked to represent their overall pain by giving a single number (numerical rating scale, NRS), verbally state its level (verbal rating scale, VRS), or similarly mark on a 10 cm line their pain level (visual analog scale, VAS). Despite being considered the gold standard in clinical trials for pain, traditional measurement tools have huge disadvantages, including low precision and higher rates of incorrect responses [6,7]. Perhaps, the most important aspect is the assumption that overall pain is a unidimensional experience that can be measured with a single-item scale [8,9]. As per ClinicalTrials.gov data, research trials for pain killers have increased at an average rate of greater than 20% in the past 15 years. Despite the rise in costs and boost in investments, medical assessments of patients' pain and the response to current and potential novel analgesic therapies are mostly based on subjective tools, such as grading of pain on a numerical scale. More objective approaches are needed to precisely measure and track in real time, more single and overlapping pain conditions, their evolution, and the therapeutic approaches best suited for treating pain in an individual and in large populations. The bottom line is that sensory pain assessment allows clinicians and scientists to monitor the longitudinal severity of the pain disorder and to quantify analgesic treatment effects [10], independent of the emotional and cognitive impact. One missing disconnect of the clinical assessment of pain from the years of pain neuroimaging research is that the sensory-discriminative aspect of pain is not only processed by the brain for its intensity but also for its accurate account of region and even subregional pain location and area. Together with other pain-related brain structures, the primary somatosensory cortex (S1) ultimately processes pain based on the homuncular noxious ascending inputs from multiple anatomic subregions [11,12].

To address the inaccuracy issues related to the main sensory pain scales traditionally used in the clinical and research practice, a collaborative effort from pain neuroimaging and 3D experts at the University of Michigan has developed a 3D body mobile in-house and optimized it for multiple pain disorders to better match the objectivity of pain neuroimaging and neuromodulation studies [13]. GeoPain (licensed by MoxyTech Inc) is a free or customized mobile app for tracking, analyzing, and communicating pain on multiple mobile platforms. Please, refer to Multimedia Appendix 1 (flowchart of the patients included in the clinical trial [13]) and to Multimedia Appendix 2 (demographic data of patients included in the clinical trial [13]). It is a reproducible and quantifiable 3D navigation system of grids that generates clinical and research tools for single and overlapping pain disorders by providing detailed sensory-discriminative measurements for the full body and by subregions. The personalized interface allows the patient to quickly delineate the intensity and area of pain on diverse rotating 3D body models (different gender and age avatars) by simply touching and zooming on the screen to where it hurts using a touch device such as touch-screen desktops, mobile phones, or tablets. The AI-enabled time-stamped technology precisely and quantitatively records and communicates their pain(s) and associated symptoms, which better mirrors the way the brain decodes pain severity across the body. For instance, neuroimaging studies have reported that the level of endogenous μ-opioid activation in episodic and chronic migraine patients is highly affected by the pain area and intensity number summation (PAINS) [14], one of the main sensory pain measurements provided by GeoPain [14-16]. However, no correlations were found with μ-opioid receptor binding based on the attack pain intensity or area separately, or the traditional VAS score. In our high-definition neuromodulation study targeting the unilateral primary motor cortex, a significant pain difference between sham and active groups was detected 1 month before the traditional VAS by accurately looking at subregional pain area,

intensity, and both sensory-discriminative measures combined (PAINS) [13].

Objectives

Following these multiple neuroimaging and neuromodulation studies, we aimed to specifically validate sensory-discriminative GeoPain measurements from our TMD neuromodulation trial [13] and better understand their reliability to assess sensory pain impact and sample size of patients needed compared with the traditional VAS score.

Methods

Study Design

The study was a randomized, placebo-controlled, blinded clinical trial of high-definition transcranial direct current stimulation (HD-tDCS) of the motor cortex in 24 female patients with chronic TMD. The data obtained has been used to assess the validity, reliability, and utility of GeoPain measures in comparison to the VAS. The results of the clinical trial, which are not the aim of this study, have already been published (Trial Registration: Clinicaltrials.gov NCT02247063) [13].

After being assigned to the active or sham HD-tDCS group, participants presented during week 1 for a baseline visit. The protocol included 5 days of stimulation, a 1-week follow-up, and 1-month follow-up. The pain measures VAS (rated from 0-10), short form of the McGill Pain Questionnaire, and GeoPain app (initially released as PainTrek) were performed, thereby allowing tracking of the effectiveness of the treatment. Subjects' measures derived from GeoPain included PAINS were collected. The positive and negative affect schedule (PANAS) was used to assess mood changes. The University of Michigan Institutional Review Board approved the study (HUM00070766) and written informed consent was obtained from all participants.

GeoPain Technology

GeoPain is a free stand-alone and embedded mobile app developed in collaboration with the Headache and Orofacial Pain Effort (HOPE) at the University of Michigan and is currently licensed by the spinoff MoxyTech Inc. This pain app is available for free on Google Play and the Apple App Store. GeoPain provides a 3D body map based on a squared grid system with vertical and horizontal coordinates using anatomical landmarks. Each quadrangle, measuring approximately 1.6 cm $\times$ 1.6 cm, frames well-detailed 3D body regions, such as trunk, extremities (arms and hands, legs and feet), and craniofacial and intra oral areas for the patient to express his or her exact global and sectional pain location and intensity, as well as signs and symptoms (Figure 1).

At each session, using the app on an iPad (Apple Inc), participants drew their pain in multiple shades of red on the touch-sensitive screen ranging from pink (mild pain), red (moderate pain), and dark red (severe pain). It is also categorized with *mild*, *moderate,* and *severe* to provide some guidance for their selection and follows the accepted grouping mentioned next. Average pain is the average score of all cells that are marked as painful, with a scale of 1-3 (mild=1, moderate=2, severe=3). Pain area was the percent of the area of the head and neck region that was experiencing pain, with a scale of 0%-100% of all cells. The general size of the cells is about right for the adult male model; however, those sizes will scale depending on the body type chosen but remain accurate relative to anatomical landmarks. Finally, PAINS was the cumulative score for the cells. On the GeoPain version for this particular trial, there were a total of 2026 cells across the body. The head and neck have 382, and just the head has 322. The cells of the body are broken into 54 regions; 27 if you do not distinguish between left and right sides of the body. The average pain is the normal average for a particular region or full body. GeoPain takes all cells in a specified range, adds their intensities together, and then divides that by the number of entries. For the three pain measures, the analysis was performed for the entire head and neck area, or unilaterally, to understand how sensory-discriminative pain measures changed ipsilateral or contralateral to the putative primary cortex (M1) stimulation.

Figure 1. GeoPain tracks and measures pain in diverse single or group 3D-body models (gender and age group) and across multiple body locations. Whether the patient has TMD, migraine, fibromyalgia, or dental pain, the app can measure if a particular medication or clinical procedure is effective for each localized or spread pain condition. TMD: temporomandibular disorder.

Statistical Methods

Principal component analysis (PCA) was used to determine the correlation structure among the variables: PAINS, average pains, maximum pain, area of pain, VAS, PANAS. To assess the test-retest reliability of the PAINS and VAS measures over time, the autocorrelation between neighboring time points (ranging from 1 to 40 days) was performed. These lag autocorrelations were estimated using Pearson methods. Smoothing splines were used to model the time trend of such autocorrelations for both PAINS and VAS. All statistical analyses were performed using SAS version 9.4 (SAS Institute Inc). Cohen's effect size for both the PAINS and VAS were calculated for the delta change at 1 month from baseline to assess the longer term (1-month post-treatment) sensitivity of each measure for evaluating the treatment efficacy.
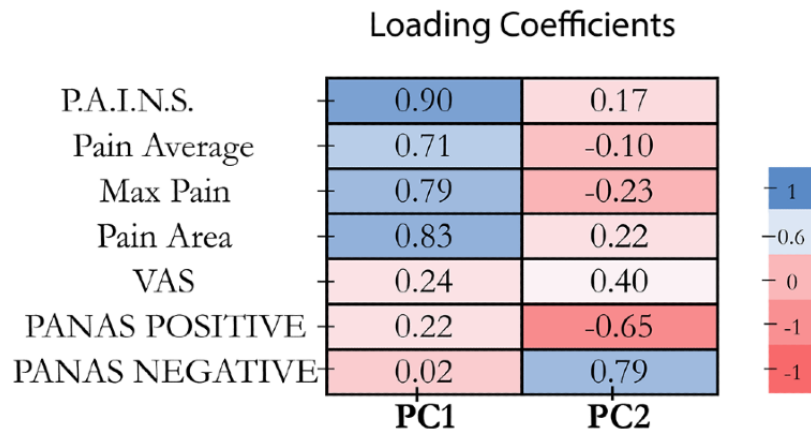
Results

Principal Component Analysis: GeoPain Versus Visual Analog Scale

PCA resulted in two main orthogonal components. The first principal component (PC1) comprises a combination score of all PAINS measures which had a high internal consistency and clustered together in TMD pain (Figure 2). The second principal component (PC2) included VAS and PANAS. The loading coefficients for each variable in PC1 and PC2 are shown in Figure 3. All loading coefficients for PAINS measures in PC1 were above 0.70, with low loadings for VAS and PANAS. The loading coefficient for PC2 was dominated by VAS and PANAS all >0.4. The biplot shown in Figure 2 shows that PC1 dominated by the GeoPain measures added is approximately orthogonal to PANAS, whereas PC2 comprises VAS along with PANAS.

Figure 2. PCA graph showing the most representative correlation among the measures at the baseline visit. PAINS: pain area and intensity number summation; PANAS: positive and negative affect schedule; PCA: principal component analysis; PC1: first principal component; PC2: second principal component; VAS: visual analog scale.
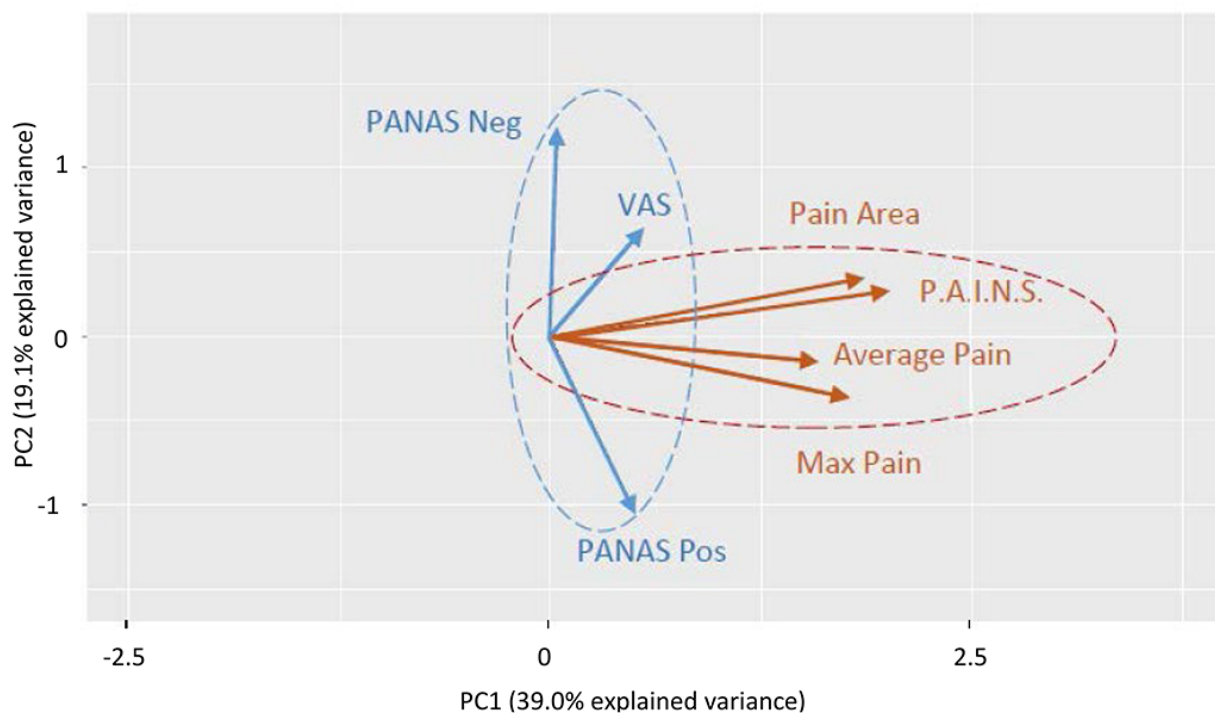
Figure 3. PC1 shows the correlation among GeoPain measures, and PC2 shows a more relevant correlation between VAS and PANAS on the baseline visit. The darker colors indicate higher loading coefficients. PAINS: pain area and intensity number summation; PANAS: positive and negative affect schedule; PC1: first principal component; PC2: second principal component; VAS: visual analog scale.

Loading Coefficients

	PC1	PC2
P.A.I.N.S.	0.90	0.17
Pain Average	0.71	-0.10
Max Pain	0.79	-0.23
Pain Area	0.83	0.22
VAS	0.24	0.40
PANAS POSITIVE	0.22	-0.65
PANAS NEGATIVE	0.02	0.79

Pain Epicenter: Correlating Duration of Pain with GeoPain and Visual Analog Scale Measures

GeoPain measures demonstrated an association between chronicity and pain in locations further away from the most commonly reported area, or pain epicenter, in the chronic TMD cohort (P=.01). We found a significant relationship between pain duration and further spread from pain epicenter cells using scatter score analysis (Figure 4). There was no correlation between VAS and pain duration in our patients.

Figure 4. GeoPain measures demonstrated an association between chronicity and pain in locations farther away from the pain epicenter (P=.01). Analysis using VAS detected neither an association at baseline for pain area nor chronicity. TMD: temporomandibular disorder; VAS: visual analog scale.
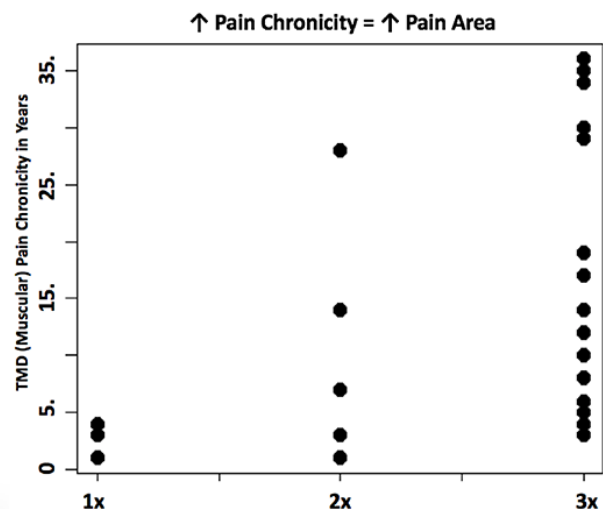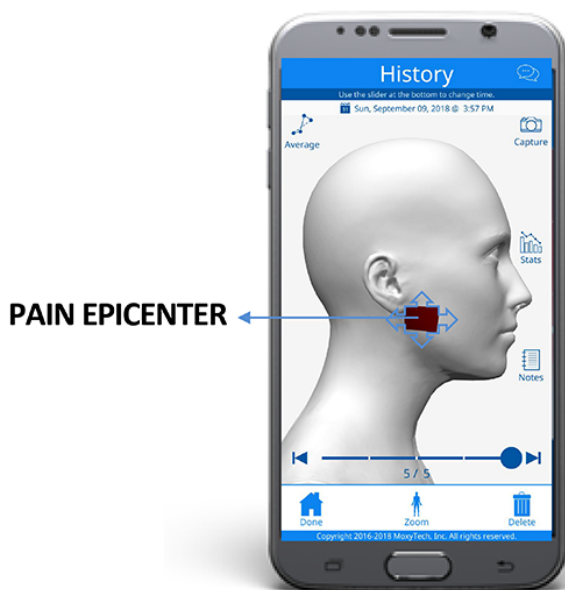

Clinical Trial: Delta Change and Effect Sizes Based on Pain Area and Intensity Number Summation and Visual Analog Scale

GeoPain (PAINS) showed consistent test-retest reliability compared with VAS, demonstrated by higher lag-auto correlations across the entire clinical trial data. Figure 5 displays the within-measure autocorrelation between neighboring time points from 1 to 40 days for PAINS and VAS. The short-term 1-day reliability for PAINS and VAS was similar and high, with an autocorrelation of around 0.7. However, longer term reliability (>1 day) is consistently high for PAINS averaging between 0.7 and 0.8, and higher than the autocorrelations for VAS averaging between 0.3 and 0.6. The combination of higher reliability for PAINS and its objectivity displayed by the lack of association with PANAS as compared with VAS suggests that PAINS demonstrates a potential better sensitivity for measuring treatment effect over time for sensory-discriminative pain.

Figure 5. Correlation over time between PAINS and VAS scores. The short-term 1-day reliability for PAINS (red) and VAS (blue) scores were similar. However, longer term reliability is consistently high for PAINS and higher than that for VAS on average. PAINS: pain area and intensity number summation; VAS: visual analog scale.
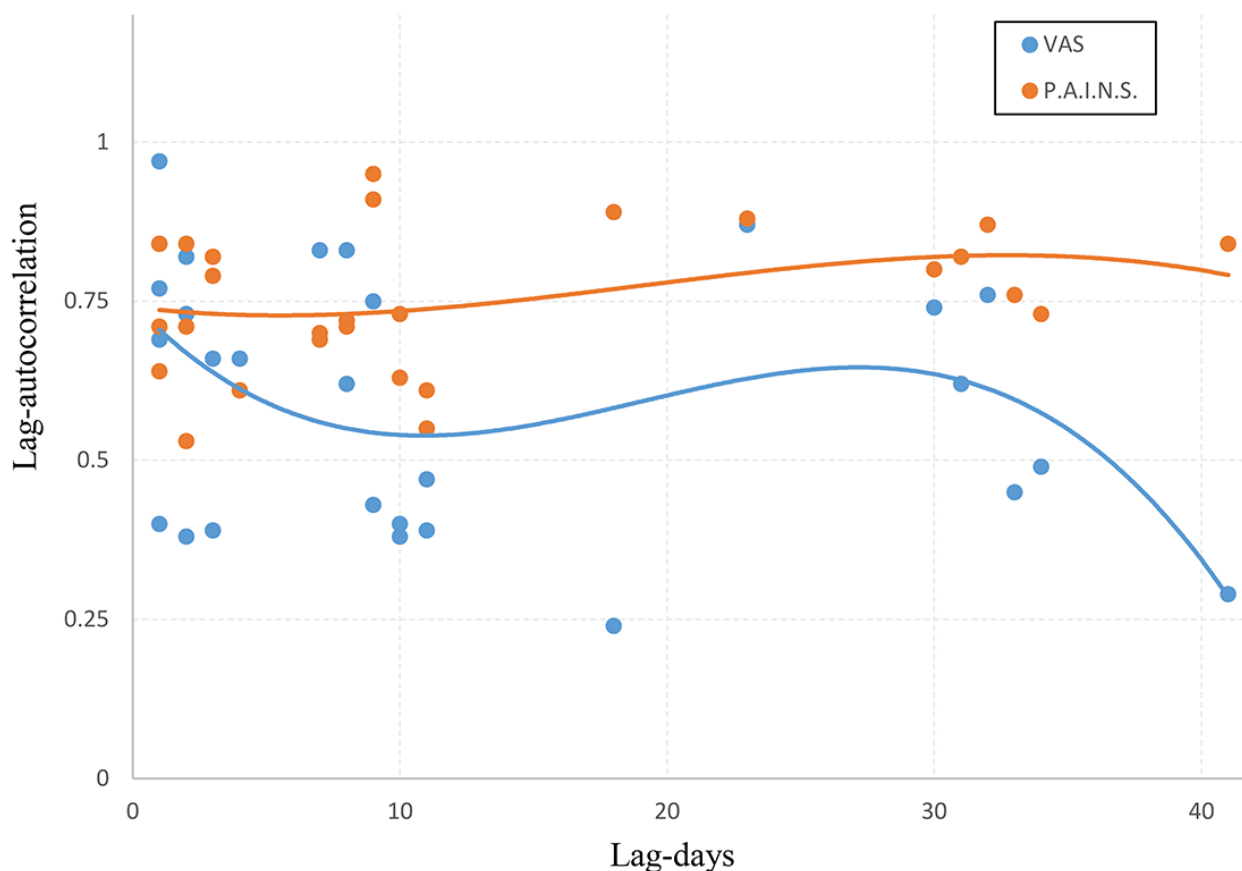
The GeoPain measure is a more specific sensory-discriminative measure of pain compared with the VAS, resulting in higher statistical power. For example, we were able to detect significant differences by treatment during the 5-day trial period using GeoPain but not when VAS was used. Higher differences were also observed using GeoPain during the 1-month follow-up after the trial. On the basis of the study [13], to achieve a statistical power of 80% when using VAS versus GeoPain a sample size twice or thrice as high for 1 month comparisons from the traditional 0 to 10 pain measurement method is required (Figure 6).

Figure 6. Graphs showing the power between the GeoPain × VAS. This figure shows the box plot for delta change by treatment and the corresponding effect sizes based on PAINS and VAS. The effect sizes for PAINS are larger than those for VAS, consequently requiring smaller sample sizes to assess the utility of treatment if PAINS is used versus VAS. Specifically, the effect sizes for PAINS were 0.51 SD for both regions and 0.60 SD for right side versus 0.31 SD for VAS. Thus, the sample size required to detect such effect sizes with 80% power would be n=145 per group for VAS, but for PAINS n=62 per group if both sides were used, n=44 per group for the right side, and n=82 per group for the left. PAINS: pain area and intensity number summation; VAS: visual analog scale.
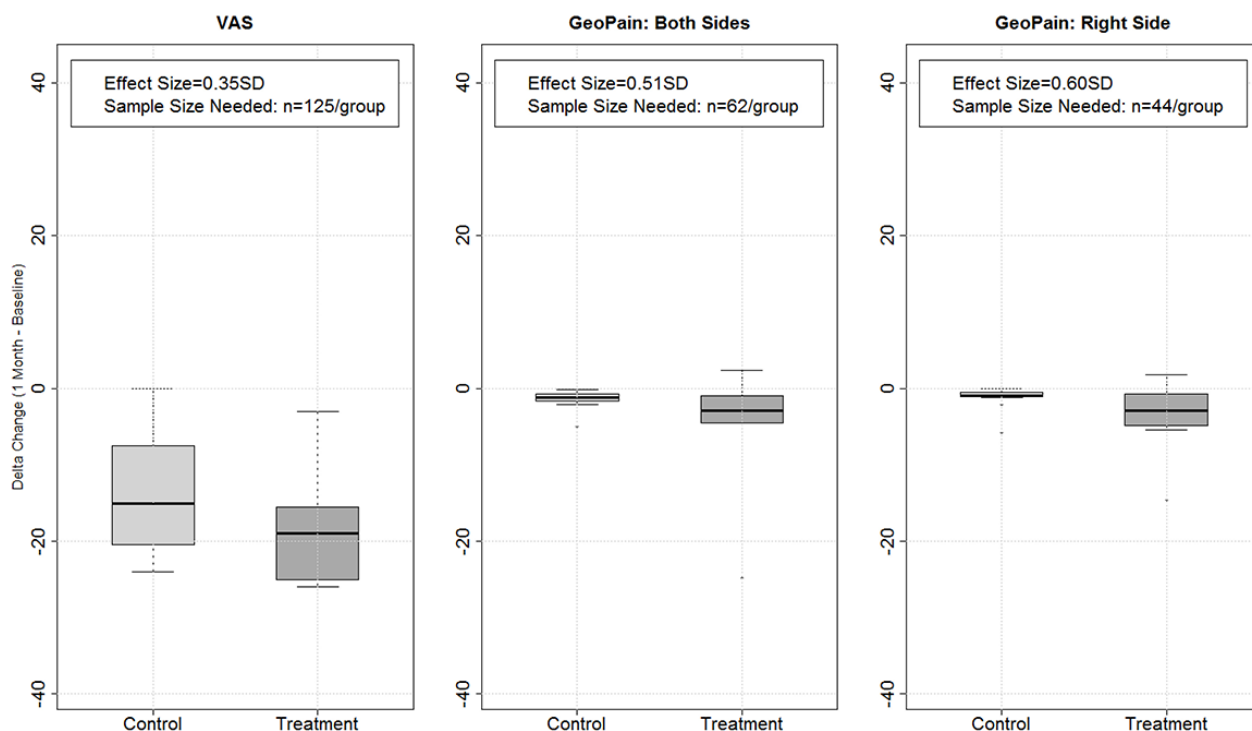
Discussion

Principal Findings

This study specifically assessed and validated the effectiveness of 3D–body pain measurements from a free mobile appl, called GeoPain, compared with VAS in a clinical trial with chronic TMD patients following sham or active 5-daily M1 HD-tDCS neuromodulation sessions. The results showed that the sensory-discriminative GeoPain measures, which included global and subregional pain area, intensity (average and maximum pain), and their combination (PAINS), were consistent and clustered together in TMD pain at baseline. On the contrary, VAS scores were bidirectionally correlated with swinging in patients' PANAS. In addition, the more chronic the pain in years, the larger the pain area spread from its epicenter. During the clinical trial (1 to 40 days), the long-term reliability was also steadily high for PAINS and low for VAS. Subsequently, the effect size for PAINS was larger than that for VAS, and as a result, half to a third of sample sizes of patients are needed to evaluate a particular pain-relief therapy if PAINS is used compared with VAS.

Patients' Positive Mood Influences Visual Analog Scale Pain Scores

Previous studies have demonstrated that unidimensional numerical scales such as the NRS or VAS provide a false impression of being sensitive and reliable measurements of pain performed in millimeters or numerals [17-19], and also lead to inaccuracies and biases. In addition, a large body of literature

has reported that patients' pain experience is actively related to extensive factors, such as mood changes and affective states [20-23]. Although all our sensory-discriminative measures from GeoPain grouped together, there was a separate assembly composed of VAS pain measurement and mood in our results. Over the course of the study, we noticed a bidirectional association between PANAS positive mood scores and VAS (P=.009), but not with PAINS (P=.14). Meanwhile, the higher the VAS pain levels reported, the worse the PANAS positive mood scores, and the more positive the patient felt, the lower was his, her, or their VAS. As humans are vastly susceptible to diurnal and seasonal mood variations with work, sleep, and daylength across multiple cultures [24], these mood swings create a large potential for disparities in VAS scores reported by even the same patient along the trial. These disparities became obvious in our study when we analyzed autocorrelations among neighboring time points of VAS scores that showed consistent lower test-retest long-term reliability compared with GeoPain's PAINS. Hence, varying degrees of altered mood in patients could lead to inaccuracy in the VAS-based results from one time point to another in clinical pain trials. It is not a surprise that many medications that improve or stabilize mood, such as antidepressants (eg, tricyclic antidepressant), are widely considered effective for the treatment of chronic pain conditions; and perhaps their success in clinical pain trials is *in part* explained by their additional indirect effect on VAS scoring. On the other hand, the minimal association of GeoPain's measurements with mood suggests that they have better

sensitivity for evaluating treatment effect over time for sensory-discriminative pain.

Worsening in Pain Area Is Linked to Chronification, Not Visual Analog Scale Score

In addition to pain intensity, a crucial component of pain processing in the peripheral and central nervous system is area extension of the pain. In our study, the spread of pain from its epicenter in our TMD patient group was significantly correlated with the years of their pain suffering, not their overall intensity on the VAS score. This is frequently seen in pain or more specifically in migraine neuroimaging studies that show neuroplasticity at the functional, structural, and molecular levels. Hence, the impact of pain chronification is linked to its area extension, which might be associated with the progression of central sensitization. For instance, molecular studies with positron emission tomography in chronic patients in vivo have shown that there is a dysfunction in the endogenous mu-opioid system that is highly related to years of pain or more specifically in migraine suffering and PAINS score [14-16]. This is arguably one of the main analgesic systems in our brains.

To address the clinical and research conundrum above, multiple groups have developed questionnaires with 2D body map tools and required patients to delineate the pain area, as a cross, checkmark, or score by counting large body regions affected. Attempts to analyze such recorded data are still too subjective and serve the purpose of general assumptions of patients' clinical pain complexity. Some studies have addressed the complexity of pain evaluation in the clinical setting [25-28]. For instance, one study indicated that the complexity of chronic pain in a biopsychosocial context includes not only physical but also mental and social outcomes [25]. Another study explored the accuracy of the questionnaire painDETECT to detect neuropathic components of orofacial pain when compared with a reference standard of clinical diagnosis. According to the results of that study, painDETECT, as well as other generic screening tools, must be adapted and revalidated specifically for orofacial pain patients [26]. Our results reinforce the need for more detailed and intuitive scoring of 3D pain area and intensity combined, even within body subregions such as GeoPain's PAINS, which provides a better assessment of the sensory-discriminative pain severity status quo in real time.

Show Me Where Your Pain Area and Intensity Number Summation Are, and I Will Tell You Where Your Treatment Is Working

In the era of precision medicine, pain treatments have become more focused and personalized. Consequently, more accurate maps to assess pain and its response to specific therapies are needed. Our clinical trial was a well-fitted example; we used high-definition neuromodulation of the patients' unilateral motor cortex, purposely in the craniofacial homuncular region. The results demonstrated that compared with VAS, PAINS provided a much higher power of analysis with 2 or 3 times less number of patients needed, depending on the craniofacial subregion analyzed both sides, ipsilateral or contralateral side to the cortical neuromodulation. As reported previously [13], the sensitivity of the GeoPain score was able to detect differences between active and sham groups in the first week of treatment, instead of only after 1 month with VAS.

The motor and sensory homuncular representations from our bodies are extremely accurate, especially from the trigeminal nervous system [12]. The cells and coordinates of GeoPain's grid system are based on multiple neuroanatomical landmarks, facilitating the translation back from the app to the patient's own body of the subregion affected by dictating the pain location in medical terms, by therapeutically targeting the region (eg, trigger point injection), by challenging it with quantitative sensory testing, with precise homuncular matching at individual and group levels [11]. Further pixilation of the grid is possible, but it loses the neurophysiological and clinical meaning. On the other hand, tracking pain changes via the traditional single VAS scoring in the overall body or even for large regions would not be sensitive enough to detect those changes within the subregions and in reasonable time. Figuratively, it would be as if driving using a map with only the country or state borders depicted, but without any local road descriptions or coordinates. This deficient GPS is even more inefficient when managing chronic pain patients suffering from multiple pain disorders, such as concurrent migraine and fibromyalgia, undergoing different pain therapies or not (eg, monoclonal antibodies that target calcitonin gene-related peptide and pregabalin). The extension of the patient's pain comorbidities could cloud the evaluation of the severity of each pain disorder (eg, head and full body pain) and the precise analgesic effect and subregional site of action from a new pain therapy in a clinical trial.

Conclusions

GeoPain measures exhibited great precision for capturing severity at baseline and treatment effect over time of sensory-discriminative pain compared with the traditional VAS score, which was highly modulated by mood. Further studies with different pain disorders and treatments are needed to confirm our validation results that were based on our previous clinical translational studies. Nonetheless, GeoPain is a valid, consistent, and reliable 3D mobile app for tracking, analyzing, and communicating pain at the individual and group levels.

XSL•FO

RenderX

Authors' Contributions

AFD, MD, BM, EB, AD, and NK developed and designed the study. AFD, TN, EM, MD, and AD acquired the data. AFD, MD, BM, EB, AD, and NK analyzed and interpreted the data. AFD, MD, TD, BM, EB, AD, and NK drafted the manuscript. NK conducted the statistical analysis. All authors had full access to all the data in the study. AFD is a guarantor and takes responsibility for the integrity of the data and the accuracy of the data analysis. NK, MD, and BM coshare the first authorship.

Conflicts of Interest

AFD is the cocreator of GeoPain (previously named PainTrek) and also the cofounder of MoxyTech Inc that licensed the technology from the University of Michigan.

Multimedia Appendix 1
Flowchart of the patients included in the clinical trial.
[PNG File , 67 KB - mhealth_v8i8e17754_app1.png]

Multimedia Appendix 2
Demographic data of the patients included in the clinical trial.
[PNG File , 213 KB - mhealth_v8i8e17754_app2.png]

References

1. Maixner W, Fillingim RB, Williams DA, Smith SB, Slade GD. Overlapping chronic pain conditions: implications for diagnosis and classification. J Pain 2016 Sep;17(9 Suppl):T93-107 [FREE Full text] [doi: 10.1016/j.jpain.2016.06.002] [Medline: 27586833]

2. Maixner W, Diatchenko L, Dubner R, Fillingim RB, Greenspan JD, Knott C, et al. Orofacial pain prospective evaluation and risk assessment study-the OPPERA study. J Pain 2011 Nov;12(11 Suppl):T4-11.e1 [FREE Full text] [doi: 10.1016/j.jpain.2011.08.002] [Medline: 22074751]

3. Slade GD, Sanders AE, Bair E, Brownstein N, Dampier D, Knott C, et al. Preclinical episodes of orofacial pain symptoms and their association with health care behaviors in the OPPERA prospective cohort study. Pain 2013 May;154(5):750-760 [FREE Full text] [doi: 10.1016/j.pain.2013.01.014] [Medline: 23531476]

4. Larsson B, Gerdle B, Björk J, Grimby-Ekman A. Pain sensitivity and its relation to spreading on the body, intensity, frequency, and duration of pain: a cross-sectional population-based study (SWEPAIN). Clin J Pain 2017 Jul;33(7):579-587. [doi: 10.1097/AJP.0000000000000441] [Medline: 27648588]

5. Dave AJ, Selzer F, Losina E, Usiskin I, Collins JE, Lee YC, et al. The association of pre-operative body pain diagram scores with pain outcomes following total knee arthroplasty. Osteoarthritis Cartilage 2017 May;25(5):667-675 [FREE Full text] [doi: 10.1016/j.joca.2016.12.013] [Medline: 27986621]

6. Pathak A, Sharma S, Jensen MP. The utility and validity of pain intensity rating scales for use in developing countries. Pain Rep 2018;3(5):e672 [FREE Full text] [doi: 10.1097/PR9.0000000000000672] [Medline: 30534623]

7. Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, European Palliative Care Research Collaborative (EPCRC). Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manage 2011 Jun;41(6):1073-1093 [FREE Full text] [doi: 10.1016/j.jpainsymman.2010.08.016] [Medline: 21621130]

8. Katz J, Melzack R. Measurement of pain. Surg Clin North Am 1999 Apr;79(2):231-252. [doi: 10.1016/s0039-6109(05)70381-9] [Medline: 10352653]

9. Melzack R. The McGill pain questionnaire: major properties and scoring methods. Pain 1975 Sep;1(3):277-299. [doi: 10.1016/0304-3959(75)90044-5] [Medline: 1235985]

10. Fillingim RB, Loeser JD, Baron R, Edwards RR. Assessment of chronic pain: domains, methods, and mechanisms. J Pain 2016 Sep;17(9 Suppl):T10-T20 [FREE Full text] [doi: 10.1016/j.jpain.2015.08.010] [Medline: 27586827]

11. DaSilva AF, Becerra L, Makris N, Strassman AM, Gonzalez RG, Geatrakis N, et al. Somatotopic activation in the human trigeminal pain pathway. J Neurosci 2002 Sep 15;22(18):8183-8192 [FREE Full text] [doi: 10.1523/JNEUROSCI.22-18-08183.2002] [Medline: 12223572]

12. Tal Z, Geva R, Amedi A. Positive and negative somatotopic BOLD responses in contralateral versus ipsilateral penfield homunculus. Cereb Cortex 2017 Feb 1;27(2):962-980 [FREE Full text] [doi: 10.1093/cercor/bhx024] [Medline: 28168279]

13. Donnell A, Nascimento TD, Lawrence M, Gupta V, Zieba T, Truong DQ, et al. High-definition and non-invasive brain modulation of pain and motor dysfunction in chronic TMD. Brain Stimul 2015;8(6):1085-1092 [FREE Full text] [doi: 10.1016/j.brs.2015.06.008] [Medline: 26226938]

14. Jassar H, Nascimento TD, Kaciroti N, DosSantos MF, Danciu T, Koeppe RA, et al. Impact of chronic migraine attacks and their severity on the endogenous μ-opioid neurotransmission in the limbic system. Neuroimage Clin 2019;23:101905 [FREE Full text] [doi: 10.1016/j.nicl.2019.101905] [Medline: 31279240]

XSL•FO
RenderX

15. DaSilva AF, Nascimento TD, DosSantos MF, Lucas S, van Holsbeec K, DeBoer M, et al. Association of μ-opioid activation in the prefrontal cortex with spontaneous migraine attacks-brief report I. Ann Clin Transl Neurol 2014 Jun 1;1(6):439-444 [FREE Full text] [doi: 10.1002/acn3.65] [Medline: 25072055]

16. Nascimento TD, DosSantos MF, Lucas S, van Holsbeeck H, DeBoer M, Maslowski E, et al. μ-opioid activation in the midbrain during migraine allodynia-brief report II. Ann Clin Transl Neurol 2014 Jun;1(6):445-450 [FREE Full text] [doi: 10.1002/acn3.66] [Medline: 25328905]

17. Cahill CM, McClellan KA, Morinville A, Hoffert C, Hubatsch D, O'Donnell D, et al. Immunohistochemical distribution of delta opioid receptors in the rat central nervous system: evidence for somatodendritic labeling and antigen-specific cellular compartmentalization. J Comp Neurol 2001 Nov 5;440(1):65-84. [doi: 10.1002/cne.1370] [Medline: 11745608]

18. Svensson E. Concordance between ratings using different scales for the same variable. Stat Med 2000 Dec 30;19(24):3483-3496. [doi: 10.1002/1097-0258(20001230)19:24<3483::aid-sim786>3.0.co;2-a] [Medline: 11122509]

19. Peters ML, Sorbi MJ, Kruise DA, Kerssens JJ, Verhaak PF, Bensing JM. Electronic diary assessment of pain, disability and psychological adaptation in patients differing in duration of pain. Pain 2000 Feb;84(2-3):181-192. [doi: 10.1016/s0304-3959(99)00206-7] [Medline: 10666523]

20. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013 Mar 8;15(3):e51 [FREE Full text] [doi: 10.2196/jmir.2350] [Medline: 23475457]

21. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. J Clin Nurs 2005 Aug;14(7):798-804. [doi: 10.1111/j.1365-2702.2005.01121.x] [Medline: 16000093]

22. Eich E, Reeves JL, Jaeger B, Graff-Radford SB. Memory for pain: relation between past and present pain intensity. Pain 1985 Dec;23(4):375-380. [doi: 10.1016/0304-3959(85)90007-7] [Medline: 4088698]

23. Gendreau M, Hufford MR, Stone AA. Measuring clinical pain in chronic widespread pain: selected methodological issues. Best Pract Res Clin Rheumatol 2003 Aug;17(4):575-592. [doi: 10.1016/s1521-6942(03)00031-7] [Medline: 12849713]

24. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science 2011 Sep 30;333(6051):1878-1881 [FREE Full text] [doi: 10.1126/science.1202775] [Medline: 21960633]

25. Kaiser U, Kopkow C, Deckert S, Neustadt K, Jacobi L, Cameron P, et al. Developing a core outcome domain set to assessing effectiveness of interdisciplinary multimodal pain therapy: the VAPAIN consensus statement on core outcome domains. Pain 2018 Apr;159(4):673-683. [doi: 10.1097/j.pain.0000000000001129] [Medline: 29300277]

26. Jafree DJ, Zakrzewska JM, Bhatia S, Nova CV. Accuracy of the painDETECT screening questionnaire for detection of neuropathic components in hospital-based patients with orofacial pain: a prospective cohort study. J Headache Pain 2018 Nov 6;19(1):103 [FREE Full text] [doi: 10.1186/s10194-018-0932-5] [Medline: 30400770]

27. García-Azorin D, Yamani N, Messina LM, Peeters I, Ferrili M, Ovchinnikov D, European Headache Federation School of Advanced Studies (EHF-SAS). A PRISMA-compliant systematic review of the endpoints employed to evaluate symptomatic treatments for primary headaches. J Headache Pain 2018 Sep 21;19(1):90 [FREE Full text] [doi: 10.1186/s10194-018-0920-9] [Medline: 30242571]

28. Steiner TJ, Jensen R, Katsarava Z, Linde M, MacGregor EA, Osipova V, et al. Aids to management of headache disorders in primary care (2nd edition) : on behalf of the European headache federation and lifting the burden: the global campaign against headache. J Headache Pain 2019 May 21;20(1):57 [FREE Full text] [doi: 10.1186/s10194-018-0899-2] [Medline: 31113373]

Abbreviations

HD-tDCS: high-definition transcranial direct current stimulation
M1: putative primary cortex
NRS: numerical rating scale
PAINS: pain area and intensity number summation
PANAS: positive and negative affect schedule
PC1: first principal component
PC2: second principal component
PCA: principal component analysis
TMD: temporomandibular disorder
VAS: visual analog scale

XSL•FO
RenderX

XSL•FO
RenderX

Original Paper

Comparing a Mobile Phone Automated System With a Paper and Email Data Collection System: Substudy Within a Randomized Controlled Trial

Diana M Bond[1], PhD; Jeremy Hammond[2], PhD; Antonia W Shand[3,4], MB ChB; Natasha Nassar[1,3], PhD

[1]Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[2]Strategic Ventures, University of Sydney, Sydney, Australia

[3]Children's Hospital at Westmead Clinical School, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[4]Department for Maternal Fetal Medicine, Royal Hospital for Women, Sydney, Australia

Corresponding Author:
Diana M Bond, PhD
Sydney School of Public Health
Faculty of Medicine and Health
University of Sydney
Level 2
Charles Perkins Centre D17
Sydney, 2006
Australia
Phone: 61 2 9036 7006
Email: diana.bond@sydney.edu.au

Abstract

Background: Traditional data collection methods using paper and email are increasingly being replaced by data collection using mobile phones, although there is limited evidence evaluating the impact of mobile phone technology as part of an automated research management system on data collection and health outcomes.

Objective: The aim of this study is to compare a web-based mobile phone automated system (MPAS) with a more traditional delivery and data collection system combining paper and email data collection (PEDC) in a cohort of breastfeeding women.

Methods: We conducted a substudy of a randomized controlled trial in Sydney, Australia, which included women with uncomplicated term births who intended to breastfeed. Women were recruited within 72 hours of giving birth. A quasi-randomized number of women were recruited using the PEDC system, and the remainder were recruited using the MPAS. The outcomes assessed included the effectiveness of data collection, impact on study outcomes, response rate, acceptability, and cost analysis between the MPAS and PEDC methods.

Results: Women were recruited between April 2015 and December 2016. The analysis included 555 women: 471 using the MPAS and 84 using the PEDC. There were no differences in clinical outcomes between the 2 groups. At the end of the 8-week treatment phase, the MPAS group showed an increased response rate compared with the PEDC group (56% vs 37%; $P<.001$), which was also seen at the 2-, 6-, and 12-month follow-ups. At the 2-month follow-up, the MPAS participants also showed an increased rate of self-reported treatment compliance (70% vs 56%; $P<.001$) and a higher recommendation rate for future use (95% vs 64%; $P<.001$) as compared with the PEDC group. The cost analysis between the 2 groups was comparable.

Conclusions: MPAS is an effective and acceptable method for improving the overall management, treatment compliance, and methodological quality of clinical research to ensure the validity and reliability of findings.

XSL•FO
RenderX

Introduction

Background

Participant engagement and response is a vital aspect of any clinical research study. Many research studies are costly, labor intensive, and potentially compromised because of the difficulties associated with patient compliance, engagement, incomplete data collection, and inadequate follow-up [1-3]. The method and type of data collection system utilized to recruit participants and collect data throughout the study is important to ensure the quality, reliability, and validity of data collection. In addition, it must be cost-effective and acceptable to participants, funding organizations, and researchers [4-6].

Paper-based data collection in research studies is gradually being replaced or used in conjunction with electronic data collection systems [7], primarily in the form of emails containing links to web-based surveys. Comparison of these two methods has been well documented [8-11].

In recent years, mobile phone technology has been increasingly used to promote health-related behavioral change and self-management of care via the use of apps and automated SMS text messages. Studies have shown effective changes in psychological and physical symptoms [12-14] as well as specific pregnancy and breastfeeding outcomes [15,16] by sending individually tailored text messages to participants. However, a Cochrane review specifically looking at mobile phone apps as a method of data delivery for self-administered questionnaires found that none of the included trials in the review reported data accuracy or response rates [17]. Furthermore, a review of studies utilizing mobile phones for data collection showed that they were based on very small sample sizes, collected intermittent data (as opposed to daily), or had limited longitudinal data collection (maximum 9 months) [18-21]. There is also limited assessment of mobile phone technology as part of a web-based automated system, integrating randomization, SMS delivery, and electronic data collection into a streamlined data management system. Although previous studies have compared traditional paper-based data collection with data collection using mobile phones [22,23], there is limited evidence assessing the effectiveness of a combination of paper or email-based methods in comparison with mobile phones as part of an automated data collection management system. In addition, longitudinal data collection using mobile phone technology has not been assessed, particularly in maternal and infant health, despite adults of reproductive age currently being the largest users of mobile phones [24].

Objectives

The primary aims of this study were to compare a web-based research management system utilizing mobile phone technology with a traditional delivery and data collection system using a combination of paper- and email-based methods on clinical research outcomes and to assess the acceptability and effectiveness of use, including cost analysis.

Methods

Design

We conducted a prespecified substudy as part of the APProve (CAn Probiotics ImProve Breastfeeding Outcomes?) trial to compare a mobile phone automated system (MPAS) with a paper and email data collection (PEDC) system. APProve was a double-blind randomized controlled trial (RCT) evaluating the effectiveness of an oral probiotic versus a placebo for preventing mastitis in breastfeeding women. It was conducted between April 2015 and December 2016 in 3 maternity hospitals in Sydney, Australia. Detailed methods have been published previously [25]. Briefly, it involved the evaluation of a probiotic versus a placebo taken daily for 8 weeks for the prevention of mastitis, which was assessed using short daily and slightly longer weekly questionnaires during the first 8 weeks following birth and longer follow-up questionnaires at 2, 6, and 12 months.

The MPAS was a data delivery and collection system that combined treatment randomization, SMS delivery to participants, electronic data collection, and data management. It was developed by the study team with the aid of an eResearch (electronic research) company, which developed the system based on our prospective design specifications. The system integrated 2 established software services, SMS delivery and a web-based survey tool, which were then linked to a secure web-based data management system. The MPAS sent automated text messages to the participants' mobile phones with links to self-administered web-based surveys. Each survey link was embedded with the participant's unique identifier, enabling comparison across multiple surveys. A maximum of 2 automated reminders were integrated into the system if a participant did not respond after 3 days. The MPAS was pilot tested by 17 members of the research department, with feedback and suggestions integrated into the system before study commencement.

The PEDC included a combination of an 8-week calendar diary provided to participants at the time of trial entry and emailed links to weekly and follow-up surveys. The calendar diaries were identified with the participant study number at the time of treatment randomization, and the start date was manually entered. The A4-size calendar was preserved with a waterproof coating, allowing for daily entries by pen. Participants were encouraged to hang the calendar in a prominent place at home. PEDC users were supplied with a stamped, addressed envelope to post the calendar back to the trial coordinating center at the end of the treatment phase.

The study was approved by the Northern Sydney Local Health District Human Research Ethics Committee, approval number HREC/14/HAWKE/358, and registered with the Australian New Zealand Clinical Trials Registry, registration number ACTRN12615000923561. Written informed consent was obtained from all participants.

Participants and Study Procedures

Of the 639 women randomized to the APProve trial, 539 women were allocated to the MPAS and 100 women to the PEDC. A quasi-randomization process was applied for PEDC recruitment,

which was conducted on randomly preassigned days of the week and continued until 100 participants were recruited. Both groups of women were identified, approached, and consented to the study in the postnatal ward in the same way, but the treatment randomization process was slightly different.

For the women allocated to the MPAS group, a research assistant entered their details into the web-based data management system, which then automatically generated a unique participant identification number and treatment allocation. The randomization schedule was built into the system and generated using a computer random number generator with random block sizes. Randomization of participants using the PEDC was conducted using sealed, opaque envelopes, with the randomization schedule developed using a similar but separate process compared with the MPAS group.

Data Collection

Baseline sociodemographic, clinical, and birth characteristics collected in this study are shown in Table 1. All daily, weekly, and follow-up questionnaires were identical for the 2 groups.

Table 1. Characteristics of participants using the mobile phone automated system compared with the paper and email data collection system.

Participant characteristics	MPAS[a] (n=526)	PEDC[b] (n=94)	Statistics[c]		P value
			t value (df)	Chi-square (df)	
Maternal					
Maternal age (years), mean (SD)	33.4 (4.9)	33.5 (4.0)	0.06 (618)	N/A[d]	.95
Born in Australia, n (%)	256 (48.7)	56 (59.6)	N/A	3.8 (1)	.05
Ethnicity, n (%)			N/A	0.2 (2)	.92
Asian	110 (20.9)	19 (20.2)	N/A	N/A	N/A
White	365 (69.4)	67 (71.3)	N/A	N/A	N/A
Other	51 (9.7)	8 (8.5)	N/A	N/A	N/A
Tertiary education[e], n (%)	440 (83.7)	77 (81.9)	N/A	0.2 (1)	.68
Alcohol in pregnancy, n (%)	58 (11.0)	11 (11.7)	N/A	0.0 (1)	.85
First baby, n (%)	312 (59.3)	44 (46.8)	N/A	5.1 (1)	.02
Allocated to probiotic, n (%)	265 (50.4)	46 (48.9)	N/A	0.1 (1)	.80
Birth, infant, and postpartum					
Caesarean section, n (%)	163 (31.0)	25 (26.6)	N/A	0.7 (1)	.39
Birthweight (grams), mean (SD)	3421 (458.1)	3456 (451.6)	0.69 (618)	N/A	.49

[a]MPAS: mobile phone automated system.

[b]PEDC: paper and email data collection.

[c]Test statistics using Pearson chi-square test for categorical variables and 2-tailed, independent sample t test for continuous variables with their respective df are presented.

[d]N/A: not applicable.

[e]College, university, or vocational training after high school.

For the MPAS group, each study site was provided with an electronic tablet with internet connectivity to enable the research assistant to enter the participants' details, conduct treatment randomization, and enter baseline and hospital data directly into the web-based data management system. All research assistants were trained in the use of the MPAS and given individualized password-protected access to the website, which could be accessed by phone, tablet, or computer. Only deidentified data were entered into the database and linked to an individual study number generated automatically at randomization. The only paper-based data for this cohort included a signed patient information and consent form and a trial entry form containing the participants' contact details. Once randomized, the study number generated by the MPAS was written in the trial entry form to allow for reidentification, if required. An audit trail was integrated into the MPAS to log all SMS messages sent and surveys completed. Daily and weekly outcome data for the

APProve trial for the first 8 weeks (56 days) following birth were collected via self-completed questionnaires using automated weblinks sent directly via SMS to the participant's mobile phone. Before the follow-up questionnaires at 2, 6, and 12 months (63, 180, and 360 days), participants were sent an automated link asking for their preferred method of receiving the questionnaires, with SMS, email, or post as options. On the basis of the response, the MPAS would either send the participant an SMS link to the relevant survey or alert the trial coordinator by an automated email of the preference for an emailed or a postal questionnaire.

For the PEDC participants, baseline and hospital data were collected on paper-based data forms and then entered into the web-based system at the trial coordinating center. Once randomized to their allocated treatment, participants were given a calendar diary by the research assistant to record daily outcomes for 8 weeks. Weekly outcome data for the first 8

XSL•FO

RenderX

weeks and follow-up questionnaires at 2, 6, and 12 months were collected by an emailed weblink to a web-based survey sent by

the clinical trial coordinator (Figure 1).

Figure 1. Flow diagram comparing the mobile phone automated system with paper and email data collection. MPAS: mobile phone automated system; PEDC: paper and email data collection.
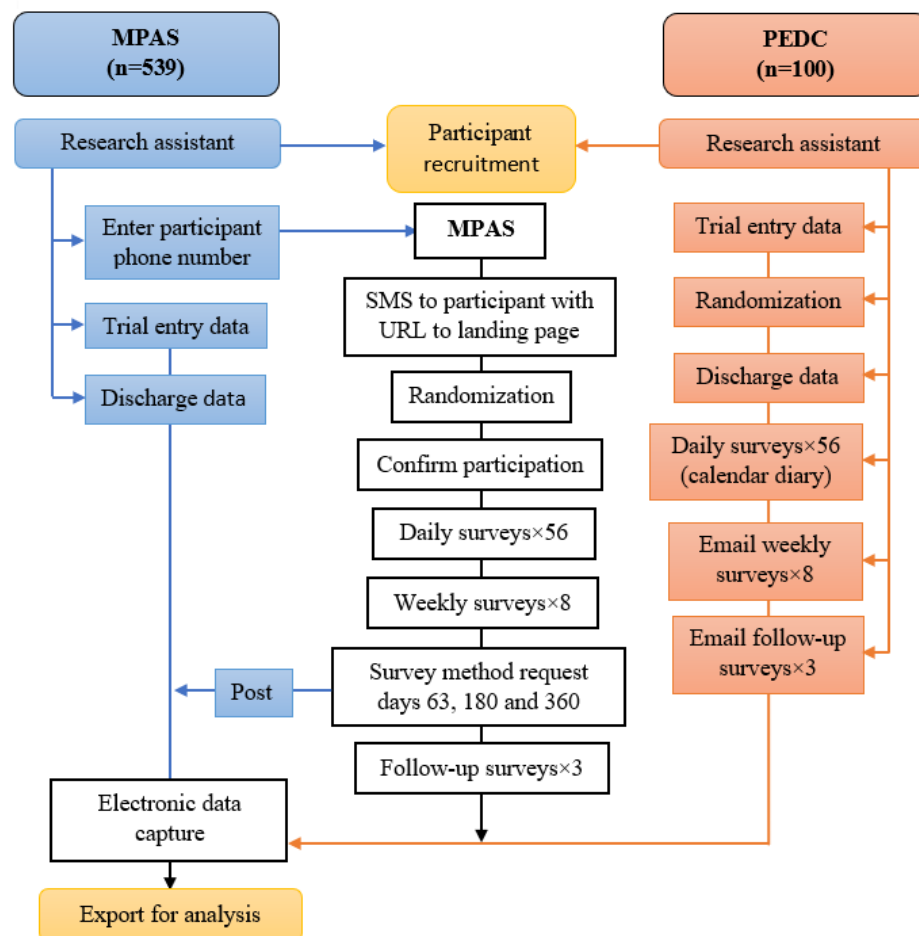
Outcomes

Outcomes evaluating participant acceptability, treatment compliance, and effectiveness of data collection comparing the MPAS with the PEDC were assessed in the 2-month follow-up questionnaire. Data were collected on the ease of participation in the trial and the ease of remembering to take the study treatment every day (both rated from 0 [very difficult] to 5 [very easy]), self-reported compliance with taking the allocated treatment (compliance was defined as having taken the product for ≥42 of 56 days, semicompliance as having taken the product for 15-41 of 56 days, and noncompliance as having taken the product for ≤14 of 56 days), whether the method of data collection was helpful in reminding the participant to take the treatment (ranked from 0 [not helpful at all] to 5 [very helpful]), recommendation of the allocated method of data collection for future studies, and the preference for how the participant wanted to receive the follow-up questionnaires (SMS, email, or post). The effectiveness of data collection was defined as the frequency of completing the questionnaires at all time points.

We also assessed whether the data collection method had any impact on the clinical trial outcomes. Clinical outcomes were collected during the daily, weekly, and 2-month surveys. They

included mastitis, maternal infection, and breastfeeding status up to 2 months after birth. The mastitis outcome measure was based on self-reported symptoms related to breast infection or a clinical diagnosis of mastitis by a care provider [26].

Satisfaction with using their assigned method of data collection (MPAS or PEDC) was assessed by using open-ended free text questions to elicit written comments pertaining to what the participants liked the most and the least about their assigned method of data collection and what suggestions could be provided for future use. In addition, satisfaction with the method of data collection was elicited from the MPAS users and responses ranked from 0 (did not like at all) to 5 (really liked it). This response was subgrouped into 2 categories: *satisfied* (4-5) and *less satisfied* (0-3).

The cost analysis of utilizing the MPAS compared with the PEDC was also performed. Costs included those associated with the initial development and ongoing usage of each system and personnel time associated with trial participant survey collection and follow-up. A web-based time tracking report was generated weekly to determine the average time required for creating and sending emails and manual data entry from paper survey collection.

Statistical Analysis

Baseline sociodemographic, clinical, and birth characteristics were compared between the 2 groups. Categorical data were summarized using percentages, and the differences in the characteristics between the 2 groups were assessed using a chi-square test. Continuous outcomes with a normal distribution were summarized using mean and SD, and the characteristics between the 2 groups were compared using t tests. Data with a nonnormal distribution were summarized using medians, and the groups were compared using nonparametric Wilcoxon tests. Satisfaction with the MPAS was analyzed by maternal sociodemographic characteristics and treatment compliance. Written responses were thematically assessed by 2 authors and an external researcher, who each independently coded the data, followed by group discussion. Common themes and relevant responses were identified, and frequency was quantified. Analyses were conducted using SPSS version 24 (IBM SPSS Statistics, 2016 IBM Corporation), and P value <.05 was used for statistical significance.

Results

Participant Characteristics

Of 620 women, 526 women were quasi-randomized to the MPAS group and 94 women to the PEDC group. There were no differences between the groups except that a higher percentage of women in the MPAS group gave birth to their first baby (P=.02; Table 1). After loss to follow-up of 10.5% (55/526) participants in the MPAS group and 11% (10/94) in the PEDC group, secondary outcomes were analyzed for 555 women. We found no difference in the trial outcomes between the 2 data collection groups (Table 2). There was also no difference in the ease of use between the MPAS and PEDC groups. However, a higher proportion of participants using the MPAS were compliant with taking the study treatment (331/471, 70.3% vs 47/84, 56%; P<.001), were more likely to rate their method of data collection as being a helpful reminder to record their symptoms (median 4.37 vs 2.63; P<.001), and were more likely to recommend their assigned method for future use (330/349, 94.6% vs 36/56, 64%; P<.001). There was little difference among the characteristics of the women who were lost to follow-up compared with those for whom we had follow-up data, except that at 2 months postpartum, the former were less likely to be tertiary educated (45/65, 69% vs 472/555, 85.0%; P=.001).

Table 2. Impact and acceptability of the mobile phone automated system compared with the paper and email data collection system.

Maternal outcomes	MPAS[a] (n=471)	PEDC[b] (n=84)	Statistics[c]		Odds ratio (95% CI)	P values
			t value (df)	Chi-square (df)		
Mastitis, n (%)	90 (19.1)	15 (17.9)	N/A[d]	0.1(1)	1.09 (0.59 to 1.99)	.79
Infections (other than mastitis), n (%)	77 (16.3)	20 (23.8)	N/A	2.8 (1)	0.63 (0.36 to 1.09)	.10
Any breastfeeding at 2 months, n (%)	443[e] (94.5)	77 (91.7)	N/A	0.1 (1)	1.55 (0.65 to 3.69)	.32
Exclusive breastfeeding at 2 months, n (%)	385[f] (82.3)	67 (79.8)	N/A	0.3 (1)	1.18 (0.66 to 2.11)	.58
Ease of participation (0-5, 5=very easy), mean (SD)	3.76 (1.31)	3.57 (1.40)	−1.02 (428)	N/A	0.19 (−0.56 to 0.18)	.31
Ease of remembering to take product (independent of method; 0-5, 5=very easy), mean (SD)	3.21 (1.43)	2.95 (1.50)	−1.3 (427)	N/A	0.21 (−0.66 to 0.14)	.21
Compliant with treatment, n (%)			N/A	15.8 (2)	N/A	<.001
Compliant (≥42 of 56 days)	331 (70.3)	47 (56.0)	N/A	N/A	N/A	N/A
Semicompliant (15-41 of 56 days)	87 (18.5)	14 (16.7)	N/A	N/A	N/A	N/A
Noncompliant (≤14 of 56 days)	53 (11.3)	23 (27.4)	N/A	N/A	N/A	N/A
Helpful reminder (data collection; 0-5, 5=very helpful), mean (SD)	4.37 (1.19)	2.63 (1.85)	−9.3 (403)	N/A	0.19 (−2.11 to −1.38)	<.001
Recommend for future, n (%)	330 (94.6)[g]	36 (64.3)[h]	N/A	50.8 (1)	0.19 (−2.11 to −1.38)	<.001

[a]MPAS: mobile phone automated system.

[b]PEDC: paper and email data collection.

[c]Test statistics using Pearson chi-square for categorical variables and 2-tailed, independent sample t test for continuous variables with their respective df are presented.

[d]N/A: not applicable.

[e]N=469.

[f]N=468.

[g]N=349.

[h]N=56.

Effectiveness and Satisfaction

The frequency with which women completed the daily and weekly questionnaires was consistently higher among the MPAS users, with a 56% average response rate over the 8-week treatment period compared with 37% (P<.001) among the PEDC users (Figure 2). There was a gradual decrease in the MPAS daily response rate over the course of the treatment phase from 70% in the first week to less than half the women completing the questionnaires by 8 weeks. Although the daily response rate from PEDC users was lower than MPAS users, there was a notable spike in the response rate among the PEDC users on the days the weekly questionnaires were sent by email (Figure 2). Response rates for the follow-up questionnaires showed a 12% higher rate of survey completion among the MPAS users at 2 months compared with the PEDC participants, with an 18% difference at 12 months (P<.05; Figure 2).

Figure 2. Effectiveness of data collection between the mobile phone automated system and the paper and email data collection. MPAS: mobile phone automated system; PEDC: paper and email data collection.
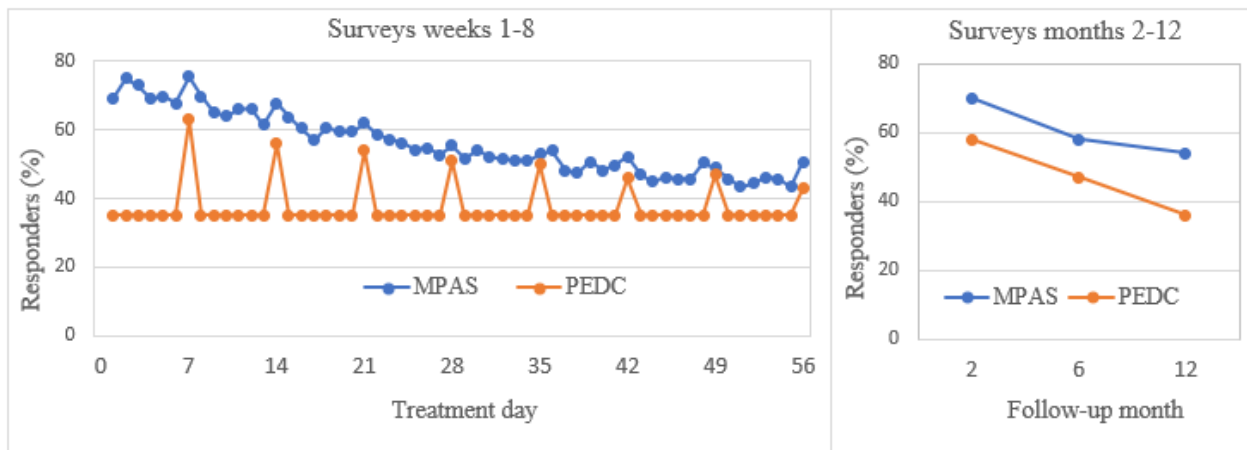
Among the MPAS users, satisfaction was high with a mean score of 4.49 out of 5 (SD 1.0). There was no difference in satisfaction scores among maternal characteristics. There was a difference in satisfaction related to compliance, with participants most compliant with treatment being the most satisfied with the use of the MPAS (*P*<.001; Figure 3). Nearly half of the participants preferred to receive the questionnaires by either SMS (135/289, 46.7%) or email (139/289, 48.0%) at 2 months; however, the preference for SMS increased to 60% for both the 6- and 12-month questionnaires (142/241,58.9% and 135/224,60.2%, respectively). Very few women opted to receive questionnaires by post (<5%).

Figure 3. Treatment compliance and satisfaction for the mobile phone automated system (n=555).
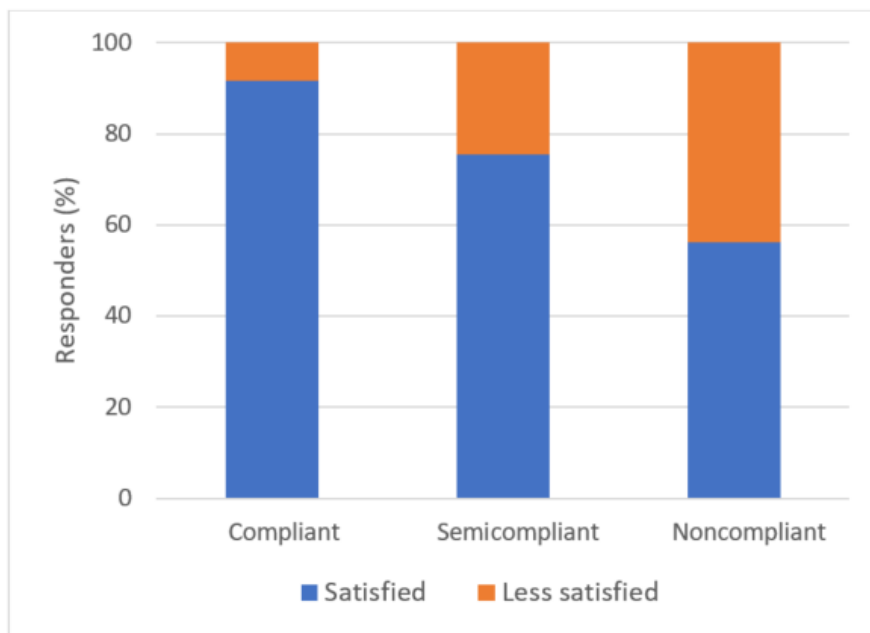
Compliant (treatment taken for ≥42 of 56 days)
Semicompliant (treatment taken for 15-41 of 56 days)
Noncompliant (treatment taken for ≤14 of 56 days)

Responses to open-ended questions in the 2-month questionnaires were received from 74.1% (349/471) MPAS participants and 67% (56/84) PEDC participants. The themes identified were related to the factors that the participants *liked most* and *liked least* about their method of data collection as outlined in Table 3. Most of the MPAS participants stated that the MPAS was easy, convenient, quick, accessible, and efficient to use. In particular, many commented that web-based questionnaires were easy to complete while breastfeeding.

Overall, less than 5% (16/349) of the MPAS participants stated that it was difficult to remember to complete the survey every day, compared with 25% (14/56) PEDC participants. Approximately, 1 in 5 participants in each group commented on the functionality of either the diary or the MPAS, such as difficulty with formatting, size restrictions, Wi-Fi accessibility, and inability to enter additional comments. Although 11 women in the MPAS group stated that they found the text messages

intrusive, 3 participants stated that they liked the fact that this   method was not intrusive.

Table 3. Qualitative analyses of the likes and dislikes of mobile phone automated system users compared with paper and email data collection system users.

Participant factors related to method of data collection	MPAS[a] (n=349), n (%)	PEDC[b] (n=56), n (%)
Liked the most		
Ease of use	325 (93.1)	7 (12.5)
Good reminder to take treatment	75 (21.5)	10 (17.8)
Liked the least		
Nothing	168 (48.1)	10 (17.8)
Time consuming	24 (6.9)	12 (21.4)
Functionality issues	77 (22.1)	10 (17.8)
Difficult to remember to complete survey	16 (4.6)	14 (25.0)

[a]MPAS: mobile phone automated system.

[b]PEDC: paper and email data collection.

Suggestions for future use by the MPAS participants included allowing users to select the time of day to receive the SMS and to opt in or out of reminder messages, limiting the number of questions on the questionnaire to minimize scrolling, diversifying the content of each SMS for improved interest, and improving the functionality to allow the questionnaires to be completed later if interrupted. Many of the PEDC participants recommended the use of SMS or a web-based app for data collection (Textbox 1).

Textbox 1. Participants' comments about the mobile phone automated system compared with the paper and email data collection system.

Mobile phone automated system

- "I found using my phone to complete the surveys great as I could do it easily when feeding my daughter."

- "It was great—something to look forward to everyday. It was easy and also a great reminder in case I had forgotten to take my daily Approve sachets."

- "So easy to remember and to complete the daily survey. I often completed the survey while out and about."

- "Most people have a smartphone on hand. Much easier than using a computer or a paper record. Ease of use—always with me. Could answer questions while breastfeeding my baby."

- "Sometimes it took a while to upload the questions"

- "Reminders were great but sometimes daily were a bit annoying"

- "Weekly questionnaires bit lengthy"

- "It would often change my response (touch feature too sensitive)"

- "Hard to see if the survey was completed if forgotten to complete the previous ones"

Paper and email data collection system

- "I liked to be a part of this study but it was not that easy to remember it to take every day...I missed sometimes."

- "Helped to keep on track. Encouraged me to have a morning routine that incorporated having breakfast at the similar time each morning."

- "The calendar was quick and easy. Can't imagine also having to write in a diary on a daily basis."

- "Now that everyone is on the phone maybe there could be a daily reminder on the participants' phone, creating an app or site so the data goes straight to the research office daily or weekly."

- "Filling out the manual form is troublesome."

- "Forgetting to fill in the daily diary even though it was clearly explained to me before I agreed to do the trial. I'm so sorry. I only found it the other day in a pile of paperwork. I do everything electronically."

- "Keeping track and filling as was not doing it every day so it was hard to remember after 15-20 days for that period, sorry."

- "The progress chart would be easier if online or an app so it could be filled in on a smartphone during feeds."

- "Probably use a different stock as it could be hard to write on."

Cost Analysis

Cost analysis between the 2 groups showed a comparable per-person cost, with the MPAS costing on average Aus $10 (US $7.21) more (Tables 4 and 5).

Table 4. Cost analysis for paper and email data collection

Paper and email data collection	Cost, Aus $ (US $)[a]
Diaries	
Printing	854.05 (615.65)
Labels for diaries	58.85 (42.42)
Stamps	150 (108.13)
Envelopes	40 (28.83)
Paper and printing (case report forms)	10 (7.21)
Emails[b,c]	5000 (3604.29)
Data collection forms	2500 (1802.14)
Reminder emails: 35% (33/94) return rate	500 (360.43)
Total×100 participants	9112.90 (6569.11)
Total cost per person	91.13 (65.69)

[a]All costs are calculated in Australian dollars (Aus $1=US $0.72).

[b]Labor is calculated at Aus $50 (US $36.04) per hour.

[c]Emails are calculated at 5 min per email.

Table 5. Cost analysis for mobile phone automated system.

Mobile phone automated system	Cost, Aus $ (US $)[a]
Tablets×3	1060 (764.11)
Intersect: data hosting	3300 (2378.83)
Intersect: app development	29,080 (20,962.50)
Intersect: trial Infrastructure	14,500 (10,452,40)
Web survey tool (Aus $780 per year×2)	1560 (1124.54)
SMS service (45000@Aus $0.069 per sms)	3105 (2238.26)
Mobile service number (Aus $25 per month×24)	600 (432.52)
Website hosting (Aus $25 per year×2)	50 (36.04)
Broadband (Aus $30 per month×24)	720 (519.02)
Reminder emails: 46.7% (246/526) return rate)[b,c]	450 (324.39)
Total×529 participants	54,425 (39,232.70)
Total cost per person	102.88 (74.16)

[a]All costs are calculated in Australian dollars (Aus $1=US $0.72).

[b]Labor is calculated at Aus $50 (US $36.04) per hour.

[c]Emails are calculated at 5 min per email.

Discussion

Principal Findings

This study demonstrates that an MPAS is an effective and acceptable tool for improving study delivery and data collection within a randomized trial as compared with a more traditional system. We have shown that the mobile phone system improved treatment compliance and response rates, demonstrated greater user satisfaction, is comparable in cost to PEDC, and does not impact study outcomes.

Comparison With Prior Work

Our study supports previous studies which showed that SMS messaging could improve treatment adherence and was acceptable to participants [16,19,27]. Despite concerns about long-term attrition in previous studies [28], the MPAS results showed that even with a decrease in response rates over time, the response rates were consistently higher than the PEDC rates over the same period, possibly because of better engagement among the users. Although the response rate of the PEDC participants showed that 37% (35/94) of the participants returned

a completed questionnaire, it is likely that some of the days may have been retrospectively completed, compromising the accuracy of the data. The peak completion rate of the PEDC questionnaires was on the day the weekly questionnaires were emailed to the participants, suggesting that emailed links are a more effective method of data collection compared with paper-based data collection, although they are more time consuming for the trial coordinator compared with automated SMS links. Despite no difference in clinical outcome measures between the 2 groups, the increased response rates to the daily surveys provided rich data regarding breastfeeding habits, confirming the feasibility of using an MPAS as a means of improving the reliability of outcome data in breastfeeding research [23].

The daily questionnaires of the MPAS appeared to have a secondary effect of improving treatment compliance by serving as a daily reminder, which in turn increased engagement with the system, resulting in a higher rate of satisfaction. Anecdotally, satisfaction among the research assistants was also high, with the majority saying that the MPAS was easy to use and less time-consuming for randomization and data entry as compared with paper forms. Moreover, the MPAS minimized the use of paper.

Despite previous research showing a 55% reduction in cost upon using electronic data collection compared with paper data collection [10], our study indicates that the cost per person is comparable between PEDC and MPAS. This is largely because of the differences in electronic data capture between the 2 studies, with the earlier study collecting, monitoring, and entering data directly into a web-based database, whereas the major expenditure to our study was the development of a research management system that integrated randomization, automated SMS, and data collection. It is important to note that once the trial infrastructure and data hosting was installed and initiated, there was potential to significantly scale up the number of participants and the duration of the study without an incremental increase in cost, whereas an increase in PEDC participants would constitute a supplemental increase in labor costs. An additional 18 PEDC participants in our study would have balanced the costs between the 2 groups. Furthermore, the scope for contact and engagement with participants with the MPAS is greater compared with paper and email methods of data collection. For example, the PEDC participants each received a minimum of 11 emails. Conversely, the MPAS participants received an average of 61 automated text messages, including welcome texts, daily SMS, and reminder messages. If the same number of texts were sent by email by a clinical trial coordinator, the cost would have increased to an additional Aus $200 (US $144.17) per participant (Aus $292 [US $210.49] PEDC vs Aus $102 [US $73.53] MPAS).

There is very little data to evaluate the use of SMS as a consolidated research management tool. We found many benefits of using MPAS in the multicenter APProve trial, including a centralized system to manage randomization, data collection across all stages of the trial, automated reminders and alerts, reduced paper transfer of sensitive patient information between sites, reduced potential for transcription error [11,29,30], and improved reliability of daily data collection

associated with reduced risk of recall bias [23]. Reducing the burden and time of data collection on the research assistant was significant, along with issues associated with patient confidentiality and storage of physical case report forms [23,29]. The advantage of integrating the MPAS via a web-based platform ensured access across mobile phone platforms and enabled accessibility to a large and diverse population, especially for those living in rural, remote, or disadvantaged areas or where mobility is restricted [31,32]. In addition, staff sick leave and absences were less of an issue because of the automated nature of the system, leading to increased flexibility of the research team, which is important when managing research studies on small budgets in small teams.

Strengths and Limitations

The main strength of our study was embedding the assessment of the MPAS versus PEDC as a substudy in an RCT with quasi-randomization to treatment group showing little difference between study groups. Most studies comparing paper-based data collection and electronic data collection had very small sample sizes, 20 to 116 participants [20,33], whereas we were able to show an effective difference with a statistically robust sample size. Furthermore, daily data collection for 8 weeks and comparison of responses at 3 strategic time points over the course of 1 year was instrumental in the accurate assessment of outcomes and minimizing errors in recall bias [34]. The inclusion of data accuracy and response rates fills a gap in the literature as addressed by a relevant Cochrane review [17]. Furthermore, the method of data collection for both groups allowed for objectivity of responses without *gratitude bias*, as is often seen in questionnaires of a face-to-face nature [35,36].

One of the limitations of the study was the difference in sample size between the 2 groups. As this was a substudy of an RCT, it was not powered for this secondary outcome. Random sampling was performed to ensure that the MPAS did not adversely affect the primary outcome. Although baseline maternal characteristics show that more women in the MPAS group gave birth to their first baby, possibly because the paper diary appeared more overwhelming for first-time mothers, there were no differences between maternal health and breastfeeding outcomes. In addition, self-reported compliance can be perceived as subjective and prone to bias, but as compliance was measured by the same method in both groups, the bias would be nondifferential. There were also issues with the interface and usability for completing the questionnaires via the web for both MPAS and PEDC participants. However, we were able to resolve many of the issues and make slight modifications to the software over time. This did not negatively impact the response rates. A final limitation was that no assessment of participant time was included in the cost analysis. This was not included as it was not anticipated that there would have been a discernible difference in time cost between the 2 groups. Posting the diaries and logging on to the computer for the weekly questionnaires may have elicited more time from the PEDC participants, but this would have been negligible.

Conclusions

Despite the increasing growth of web-based clinical trial management systems, there has been little or no evaluation of

these systems against traditional methods of trial management systems. Since the commencement of our trial, there have been improvements in the quality and availability of electronic data collection systems. For example, REDCap (Research Electronic Data Capture) is a secure web application for building and managing web-based surveys and databases, specifically for research studies and operations [37]. The system offers an easy-to-use and secure method of flexible yet robust data collection, which is free to researchers affiliated with universities. Using such a system would have decreased the costs associated with the development of the web-based survey tool we utilized as well as eliminated many of the functionality issues we experienced to reduce future research costs.

Future research should focus on how to maximize the effect of mobile phone technology, such as implementing strategies to improve long-term engagement with participants by simplifying questionnaires, optimizing the number of text messages, and personalizing the content and timing of messages.

Although we evaluated MPAS in a perinatal population, the use of mobile phone technology provides the opportunity to facilitate and improve the quality and effectiveness of clinical research studies; enhance patient interaction; and improve clinical research across a wide range of methodologies, disciplines, and health care settings. Integration and evaluation of mobile phone research management systems that are cost-effective, efficient, and acceptable to both researchers and patients is essential, given the increasing use of mobile phone technology [24] and high costs of undertaking research. We have shown that the use of an integrated MPAS is an effective and acceptable method for improving the overall management, treatment compliance, and methodological quality of a randomized clinical trial to ensure validity and reliability of findings, in addition to being cost-effective.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003 Apr;24(2):182-199. [doi: 10.1016/s0197-2456(02)00320-3] [Medline: 12689739]

2. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clin Trials 2004;1(4):368-376. [doi: 10.1191/1740774504cn032oa] [Medline: 16279275]

3. Jüni P, Altman D, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. Br Med J 2001 Jul 7;323(7303):42-46 [FREE Full text] [doi: 10.1136/bmj.323.7303.42] [Medline: 11440947]

4. Sibbald B, Roland M. Understanding controlled trials. Why are randomised controlled trials important? Br Med J 1998 Jan 17;316(7126):201 [FREE Full text] [doi: 10.1136/bmj.316.7126.201] [Medline: 9468688]

5. Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. Am J Prev Med 2007 Aug;33(2):155-161. [doi: 10.1016/j.amepre.2007.04.007] [Medline: 17673104]

6. Whitford H, Donnan P, Symon A, Kellett G, Monteith-Hodge E, Rauchhaus P, et al. Evaluating the reliability, validity, acceptability, and practicality of SMS text messaging as a tool to collect research data: results from the feeding your baby project. J Am Med Inform Assoc 2012;19(5):744-749 [FREE Full text] [doi: 10.1136/amiajnl-2011-000785] [Medline: 22539081]

7. Nahm ML, Pieper CF, Cunningham MM. Quantifying data quality for clinical trials using electronic data capture. PLoS One 2008 Aug 25;3(8):e3049 [FREE Full text] [doi: 10.1371/journal.pone.0003049] [Medline: 18725958]

8. Fitzgerald D, Hockey R, Jones M, Mishra G, Waller M, Dobson A. Use of online or paper surveys by Australian women: longitudinal study of users, devices, and cohort retention. J Med Internet Res 2019 Mar 14;21(3):e10672 [FREE Full text] [doi: 10.2196/10672] [Medline: 30869647]

9. Chen L, Chapman JL, Yee BJ, Wong KK, Grunstein RR, Marshall NS, et al. Agreement between electronic and paper Epworth sleepiness scale responses in obstructive sleep apnoea: secondary analysis of a randomised controlled trial undertaken in a specialised tertiary care clinic. BMJ Open 2018 Mar 8;8(3):e019255 [FREE Full text] [doi: 10.1136/bmjopen-2017-019255] [Medline: 29523562]

10. Pavlović I, Kern T, Miklavcic D. Comparison of paper-based and electronic data collection process in clinical trials: costs simulation study. Contemp Clin Trials 2009 Jul;30(4):300-316. [doi: 10.1016/j.cct.2009.03.008] [Medline: 19345286]

11. Boyer K, Olson J, Calantone R, Jackson E. Print versus electronic surveys: a comparison of two data collection methodologies. J Oper Manage 2002 Feb 21;20(4):357-373. [doi: 10.1016/s0272-6963(02)00004-9]

12. Heron K, Smyth J. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]

13. 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. Br Med J 2018 May 17;361:k1959 [FREE Full text] [doi: 10.1136/bmj.k1959] [Medline: 29773539]

14. Redfern J. Smart health and innovation: facilitating health-related behaviour change. Proc Nutr Soc 2017 Aug;76(3):328-332. [doi: 10.1017/S0029665117001094] [Medline: 28857019]

15. Lau Y, Htun TP, Tam WS, Klainin-Yobas P. Efficacy of e-technologies in improving breastfeeding outcomes among perinatal women: a meta-analysis. Matern Child Nutr 2016 Jul;12(3):381-401 [FREE Full text] [doi: 10.1111/mcn.12202] [Medline: 26194599]

16. Poorman E, Gazmararian J, Parker R, Yang B, Elon L. Use of text messaging for maternal and infant health: a systematic review of the literature. Matern Child Health J 2015 May;19(5):969-989. [doi: 10.1007/s10995-014-1595-8] [Medline: 25081242]

17. Marcano Belisario JS, Jamsek J, Huckvale K, O'Donoghue J, Morrison CP, Car J. Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. Cochrane Database Syst Rev 2015 Jul 27(7):MR000042. [doi: 10.1002/14651858.MR000042.pub2] [Medline: 26212714]

18. Jimoh F, Lund E, Harvey L, Frost C, Lay W, Roe M, et al. Comparing diet and exercise monitoring using smartphone app and paper diary: a two-phase intervention study. JMIR Mhealth Uhealth 2018 Jan 15;6(1):e17 [FREE Full text] [doi: 10.2196/mhealth.7702] [Medline: 29335239]

19. Mougalian SS, Epstein LN, Jhaveri AP, Han G, Abu-Khalaf M, Hofstatter EW, et al. Bidirectional text messaging to monitor endocrine therapy adherence and patient-reported outcomes in breast cancer. JCO Clin Cancer Inform 2017 Nov;1:1-10 [FREE Full text] [doi: 10.1200/CCI.17.00015] [Medline: 30657377]

20. Christie A, Dagfinrud H, Dale O, Schulz T, Hagen KB. Collection of patient-reported outcomes;--text messages on mobile phones provide valid scores and high response rates. BMC Med Res Methodol 2014 Apr 16;14:52 [FREE Full text] [doi: 10.1186/1471-2288-14-52] [Medline: 24735061]

21. Lim MS, Sacks-Davis R, Aitken CK, Hocking JS, Hellard ME. Randomised controlled trial of paper, online and SMS diaries for collecting sexual behaviour information from young people. J Epidemiol Community Health 2010 Oct;64(10):885-889. [doi: 10.1136/jech.2008.085316] [Medline: 19767322]

22. Price M, Kuhn E, Hoffman JE, Ruzek J, Acierno R. Comparison of the PTSD checklist (PCL) administered via a mobile device relative to a paper form. J Trauma Stress 2015 Oct;28(5):480-483. [doi: 10.1002/jts.22037] [Medline: 26375277]

23. Bruun S, Buhl S, Husby S, Jacobsen LN, Michaelsen KF, Sørensen J, et al. Breastfeeding, infant formula, and introduction to complementary foods-comparing data obtained by questionnaires and health visitors' reports to weekly short message service text messages. Breastfeed Med 2017 Nov;12(9):554-560. [doi: 10.1089/bfm.2017.0054] [Medline: 28832183]

24. Chaffey D. Mobile Marketing Statistics Compilation. Smart Insights. 2019. URL: https://www.smartinsights.com/mobile-marketing/mobile-marketing-analytics/mobile-marketing-statistics/ [accessed 2019-02-06] [WebCite Cache ID 75z4Idit4]

25. Bond D, Morris J, Nassar N. Study protocol: evaluation of the probiotic Lactobacillus Fermentum CECT5716 for the prevention of mastitis in breastfeeding women: a randomised controlled trial. BMC Pregnancy Childbirth 2017 May 19;17(1):148 [FREE Full text] [doi: 10.1186/s12884-017-1330-8] [Medline: 28525984]

26. Amir L, Lumley J, Garland S. A failed RCT to determine if antibiotics prevent mastitis: cracked nipples colonized with Staphylococcus aureus: a randomized treatment trial [ISRCTN65289389]. BMC Pregnancy Childbirth 2004 Sep 16;4(1):19 [FREE Full text] [doi: 10.1186/1471-2393-4-19] [Medline: 15369597]

27. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]

28. 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]

29. Cole E, Pisano ED, Clary GJ, Zeng D, Koomen M, Kuzmiak CM, et al. A comparative study of mobile electronic data entry systems for clinical trials data collection. Int J Med Inform 2006;75(10-11):722-729. [doi: 10.1016/j.ijmedinf.2005.10.007] [Medline: 16386460]

30.   Jones S, Murphy F, Edwards M, James J. Doing things differently: advantages and disadvantages of web questionnaires. Nurse Res 2008;15(4):15-26. [doi: 10.7748/nr2008.07.15.4.15.c6658] [Medline: 18700656]

31.   Bensley RJ, Hovis A, Horton KD, Loyo JJ, Bensley KM, Phillips D, et al. Accessibility and preferred use of online web applications among WIC participants with internet access. J Nutr Educ Behav 2014;46(3 Suppl):S87-S92. [doi: 10.1016/j.jneb.2014.02.007] [Medline: 24810002]

32.   Vangeepuram N, Mayer V, Fei K, Hanlen-Rosado E, Andrade C, Wright S, et al. Smartphone ownership and perspectives on health apps among a vulnerable population in East Harlem, New York. Mhealth 2018;4:31 [FREE Full text] [doi: 10.21037/mhealth.2018.07.02] [Medline: 30221166]

33.   Dale O, Hagen KB. Despite technical problems personal digital assistants outperform pen and paper when collecting patient diary data. J Clin Epidemiol 2007 Jan;60(1):8-17. [doi: 10.1016/j.jclinepi.2006.04.005] [Medline: 17161749]

34.   Fadnes LT, Taube A, Tylleskär T. How to identify information bias due to self-reporting in epidemiological research. Int J Epidemiol 2009 Jan;7(2):1-21. [doi: 10.5580/1818]

35.   Lumley J. Assessing satisfaction with childbirth. Birth 1985;12(3):141-145. [doi: 10.1111/j.1523-536x.1985.tb00952.x] [Medline: 3850745]

36.   van Teijlingen ER, Hundley V, Rennie A, Graham W, Fitzmaurice A. Maternity satisfaction studies and their limitations: 'what is, must still be best'. Birth 2003 Jun;30(2):75-82. [doi: 10.1046/j.1523-536x.2003.00224.x] [Medline: 12752163]

37.   Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

Abbreviations

APProve: CAn Probiotics ImProve Breastfeeding Outcomes?
MPAS: mobile phone automated system
PEDC: paper and email data collection
RCT: randomized controlled trial

Original Paper

Evaluating a Theoretically Informed and Cocreated Mobile Health Educational Intervention for First-Time Hearing Aid Users: Qualitative Interview Study

David W Maidment[1,2,3], BSc, MSc, PhD; Rachel Heyes[2,3], BMedSci; Rachel Gomez[2,4], BSc; Neil S Coulson[5], MA, MSc, PhD; Heather Wharrad[6], BSc, PhD; Melanie A Ferguson[2,3,4,7], BSc, MSc, PhD

[1]School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, United Kingdom

[2]National Institute for Health Research, Nottingham Biomedical Research Centre, Nottingham, United Kingdom

[3]Hearing Sciences, Division of Clinical Neuroscience, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[4]Nottingham University Hospitals National Health Service Trust, Nottingham, United Kingdom

[5]Division of Rehabilitation, Aging and Wellbeing, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[6]School of Health Sciences, University of Nottingham, Nottingham, United Kingdom

[7]National Acoustic Laboratories, Sydney, Australia

Corresponding Author:
David W Maidment, BSc, MSc, PhD
School of Sport, Exercise and Health Sciences
Loughborough University
Epinal Way
Loughborough, LE11 3TU
United Kingdom
Phone: 44 1509225439
Email: D.W.Maidment@lboro.ac.uk

Abstract

Background: Adults living with hearing loss have highly variable knowledge of hearing aids, resulting in suboptimal use or nonuse. This issue can be addressed by the provision of high-quality educational resources.

Objective: This study aims to assess the everyday experiences of first-time hearing aid users when using a newly developed, theoretically informed cocreated mobile health (mHealth) educational intervention called m2Hear. This intervention aims to deliver greater opportunities for individualization and interactivity compared with our previously developed multimedia intervention, C2Hear.

Methods: A total of 16 first-time hearing aid users trialed m2Hear for a period of 10-weeks in their everyday lives, after which individual semistructured interviews were completed. The data were analyzed using an established deductive thematic analysis procedure underpinned by the Capability, Opportunity, Motivation-Behavior model. The model stipulates that to engage in a target behavior, an individual must have physical and psychological capability, physical and social opportunity, and automatic and reflective motivation.

Results: Capability—m2Hear was viewed as a concise and comprehensive resource, suitable for a range of digital literacy skills. It was stated that m2Hear could be conveniently reused to provide useful reminders that facilitate knowledge of hearing aids and communication. Opportunity—m2Hear was simple and straightforward to use, enabling greater individualization and independence. The availability of m2Hear via mobile technologies also improved accessibility. Motivation—m2Hear provided greater support and reassurance, improving confidence and empowering users to self-manage their hearing loss.

Conclusions: Overall, this qualitative study suggests that m2Hear supports first-time hearing aid users to successfully self-manage their hearing loss postfitting. Furthermore, this study demonstrates the utility of employing a combined theoretical and ecologically valid approach in the development of mHealth educational resources to meet the individual self-management needs of adults living with hearing loss.

Trial Registration: ClinicalTrials.gov NCT03136718; https://clinicaltrials.gov/ct2/show/NCT03136718

XSL•FO
RenderX

Introduction

Background

Chronic health conditions, including hearing loss, are a leading cause of morbidity and mortality worldwide [1]. Empowering patients with long-term conditions to manage their own health can improve outcomes as well as quality of life [2]. Nevertheless, supporting patients to successfully self-manage their health is complex and multifaceted [3]. In the context of hearing loss, hearing aids are the primary clinical intervention strategy. As such, self-management support in aural rehabilitation should facilitate knowledge of hearing loss and hearing aids, hearing aid handling and communication skills, monitoring the development of and solving new problems, psychosocial well-being, and collaborative decision making with hearing health care professionals [4-6].

However, knowledge about hearing loss and hearing aids is often poor. First-time users, for instance, experience difficulties using their hearing aids because they struggle to remember all of the information given to them by their audiologist at the time of fitting [7,8]. Similarly, hearing aid handling skills in existing hearing aid users are highly variable, ranging from poor to excellent [9]. As a result, hearing aids are often used suboptimally or not at all, with estimates of nonuse varying from 3% to 24% [10]. Unmanaged hearing loss results in persistent psychosocial difficulties that can lead to social withdrawal and isolation for both individuals and their frequent communication partners [11,12]. Although audiological counseling post hearing aid fitting aims to address suboptimal use and nonuse of hearing aids, information in clinical settings is typically delivered verbally. Consequently, most of the information provided to patients is forgotten or retained incorrectly [13-15]. However, this difficulty can be overcome through the provision of supplementary educational support.

Our Original Multimedia Intervention: C2Hear

We previously developed a home-delivered educational intervention for first-time hearing aid users based on the concept of reusable learning objects (RLOs) [16,17], which aimed to improve hearing aid use. RLOs are interactive *chunks* of multimedia learning that contain highly visual components, such as animations, video clips, and patient testimonials. The RLOs that we have developed cover a range of topics, prioritized by hearing health care professionals, and include both practical (eg, *how to insert hearing aids*; *hearing aid care*) and psychosocial (eg, *what to expect when wearing hearing aids*; *communication tactics*) aspects of the adult aural rehabilitation process [16]. A registered randomized controlled trial (RCT) involving 203 first-time hearing aid users demonstrated that, in comparison to standard care, the RLOs were clinically effective; users of the RLOs demonstrated superior practical hearing aid handling skills, better knowledge of hearing aids and communication, and greater hearing aid use in those who did not wear hearing aids all of the time [17]. Additionally, in a further clinically registered RCT, the RLOs were also shown

to significantly improve self-efficacy for hearing aids [18]. Following the original RCT, the RLOs were refined based on participant feedback and are now called C2Hear [19,20]. In November 2015, C2Hear was made freely available via YouTube [19] and currently averages around 7000 views per month, with over 250,000 views globally. In 2019, a dedicated website for people living with hearing loss, frequent communication partners, and health care professionals was launched [20].

Although C2Hear has been shown to provide a range of benefits, there are several shortcomings. C2Hear was originally developed for a DVD-based platform because research at the time of development suggested that this format would be most accessible to the first-time hearing aid user age group (70-74 years) [21]. However, a DVD mode of delivery limited opportunities for individualization and interactivity. Additionally, the RLOs were between 5 and 8 min in duration, with some participants in the original RCT reporting that the RLOs were too long. Finally, the *one-size-fits-all* approach made it difficult to locate personally relevant information with ease [17]. Mobile health (mHealth) interventions, defined as health practices that are supported by mobile devices [22], could address these limitations. Specifically, mHealth interventions delivered via smartphone technologies (eg, smartphones, tablets, wearables) have been shown in other chronic health care domains, such as diabetes and asthma, to provide a platform that is both accessible and engaging, promoting greater self-management [23,24]. Consequently, we redeveloped C2Hear into an mHealth intervention (ie, m2Hear) that aims to meet the specific informational needs of first-time hearing aid users [25,26].

Development of an mHealth Intervention: m2Hear

It has been recognized for some time that, to be effective, health-related behavior change interventions should be underpinned by the appropriate behavioral theory [27]. Popular theories applied within the field of audiology include the Health Belief Model [28], Theory of Planned Behaviour [29], and Transtheoretical Model [30]. However, these models have been widely criticized because they often fail to explain variations in complex human behavior [31]. Coulson et al [31] subsequently argued that the Capability, Opportunity, Motivation-Behavior (COM-B) model, a contemporary supratheory of behavior change, is better suited to understanding and describing behavior. The COM-B model proposes that for an individual to engage in a specific health-related behavior (B), they must have physical and psychological capability (C), physical and social opportunity (O), and automatic and reflective motivation (M). A more detailed understanding of capability, opportunity, and motivation can be further derived from the theoretical domains framework (TDF), which consists of a number of different constructs (Multimedia Appendix 1) that are necessary to bring about behavior change [32]. In combination, the COM-B model and TDF can be used to identify the essential components (or *active ingredients*) that should be included in an intervention to facilitate the target behavior.

XSL·FO

RenderX

Consequently, the redevelopment of C2Hear into m2Hear was theoretically grounded, whereby the COM-B model and TDF were used to identify the components of the original C2Hear RLOs that facilitate the intended target behavior (ie, hearing aid use) [33]. Overall, we found that all RLOs consisted of multiple TDF components associated with capability (ie, knowledge, skills, memory and decision processes, behavioral regulation). However, different RLOs covered a broad range of domains relating to opportunity and motivation. For example, the *communication tactics* RLO included a high proportion of content related to opportunity, such as social influences and environmental context, whereas the *adapting to wearing your hearing aids* RLO contained content specific to motivation, such as beliefs about consequences, intentions, and goals (Figure 1).

Further to this theoretical underpinning, we employed an ecologically valid approach, whereby the original C2Hear RLOs were repurposed into 42 shorter mobile-enhanced RLOs (mRLOs). Each mRLO was designed to be a small *chunk* of learning, each with a mean duration of about 1 min (range 20 seconds to 1 min and 56 seconds). To ensure that m2Hear met the needs of the end user, a think-aloud technique was used to label each mRLO. This technique is an established observational method [34], which has been widely used in health research to develop and evaluate digital interventions [35]. We completed the think-aloud interviews with existing hearing aid users who watched the mRLOs and concurrently described their views on the content in their own words. Using an established inductive

thematic analysis procedure [36], each mRLO was then labeled in accordance with data generated from participants. For example, the *what to expect when wearing hearing aids* RLO was divided into the following 2 mRLOs: (1) *What can I expect when wearing hearing aids for the first time?* and (2) *How do I get used to wearing my hearing aids?* These short mRLOs, along with the option to select the appropriate earmold coupling (open fit or custom earmold) and 5 higher-level categories corresponding to the likely need along the patient's journey post hearing aid fitting (eg, *using your hearing aids*; *looking after your hearing aids*), aimed to provide individualized learning opportunities, whereby the user could decide what they wanted to view according to their own needs and preferences.

In addition to individualization, the mHealth intervention also enabled users to actively engage in a range of optional learning activities and quizzes to further enhance an individual's learning potential. For example, a *drag-and-drop* activity was developed that accompanied the mRLOs *How do I put my hearing aids in?* and *How do I take my hearing aids out?* This activity required users to place images in the correct order to reinforce the mRLO learning objectives, namely, how to correctly insert and remove the earmold and hearing aid. Both individualized and interactive elements were incorporated into the design of the mobile platform for delivery of the intervention, a process that was iterative and followed a user-centered and participatory design approach. The final m2Hear intervention is a freely available web-based intervention [37]; see also Figure 2 for screenshots.

Figure 1. Radar chart showing the proportion of time the theoretical domains framework (TDF) factors were included in the communication tactics and adapting to wearing your hearing aids C2Hear reusable learning objects (RLOs). Percentages are plotted for the 14 TDF factors on individual axes. Concentric grid lines connecting axes increase in 20% increments, from 0% (center point) to 100% (outer edge). Each data point has been connected to form the black shaded area.
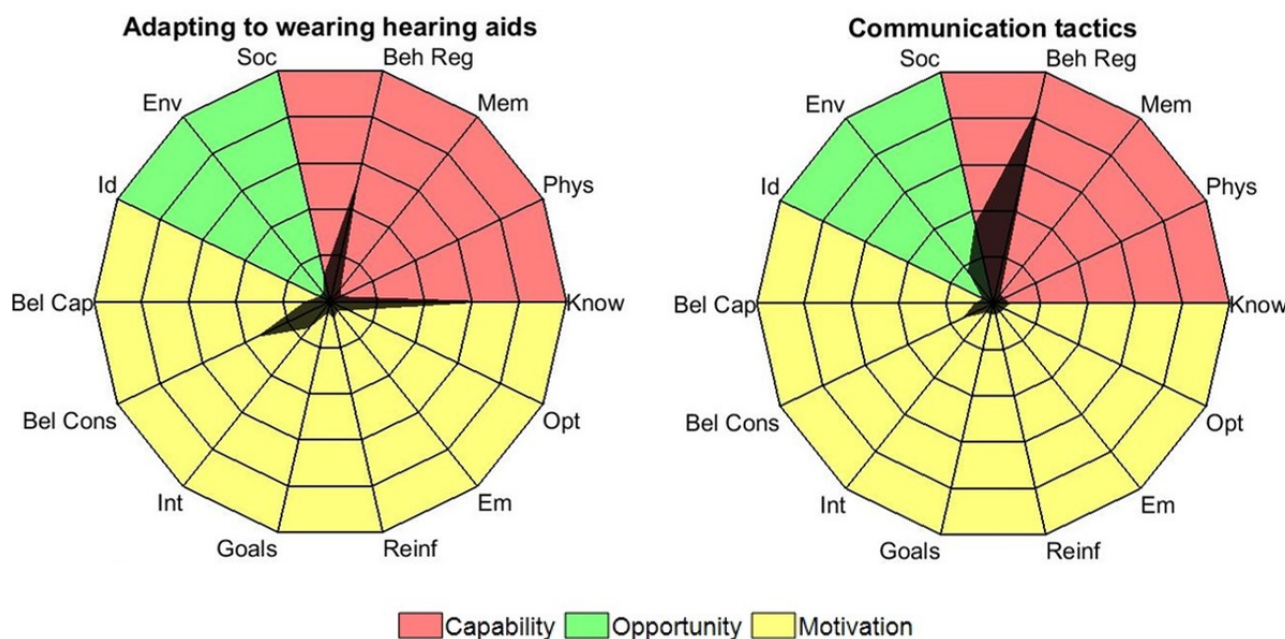
Figure 2. Screenshots of the m2Hear intervention, which is a freely available web-based intervention.
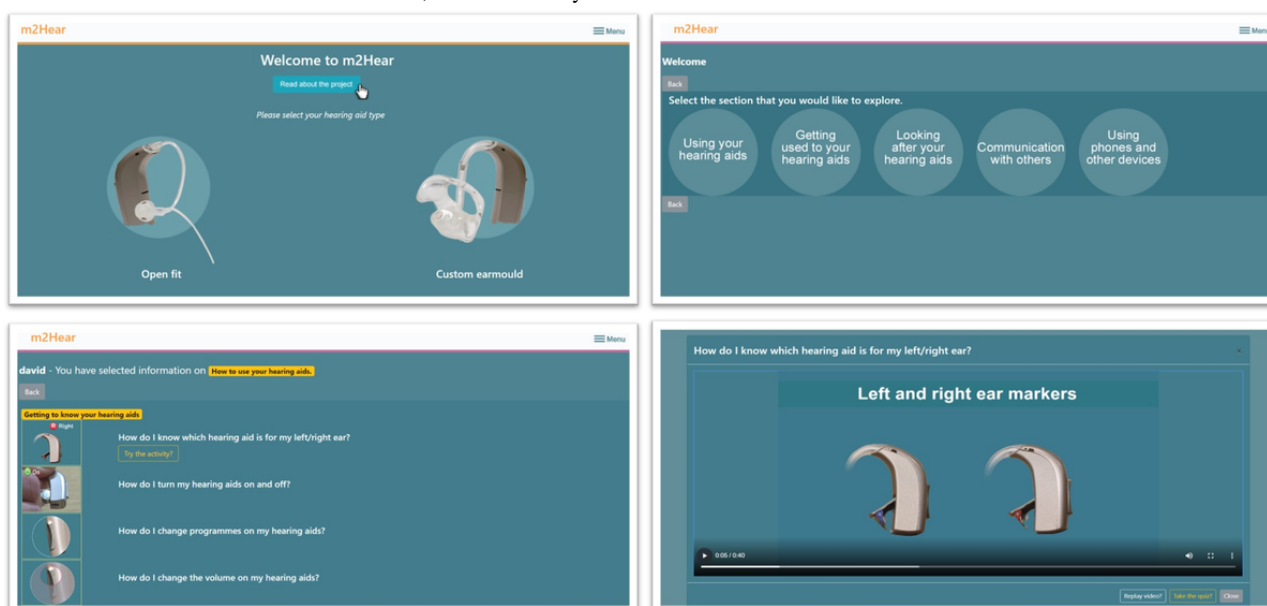
Study Aims

Following the development of m2Hear, a clinically registered study [38] assessed the feasibility of the intervention in naïve first-time hearing aid users using a mixed methods approach. In this paper, we aimed to present the results of the qualitative evaluation of this study. Specific aims were as follows:

1. Gain an in-depth insight into the views of first-time hearing aid users toward the barriers and facilitators of using m2Hear in everyday life underpinned by the COM-B model and TDF.
2. Compare barriers and facilitators between m2Hear and C2Hear to assess whether the individualized and interactive elements incorporated in the newly developed m2Hear intervention result in greater patient benefits.

Methods

Participants

A total of 16 participants were recruited from the Nottingham Adult Audiology Service, Nottingham University Hospitals National Health Service (NHS) Trust. A purposive sampling strategy was used, whereby participants met the following inclusion criteria: (1) adults aged ≥18 years, (2) adults who had never worn hearing aids, (3) adults who were familiar with smartphone technologies, and (4) adults who had a good understanding of the English language to understand the mRLO content. Exclusion criteria included those who were unable to use m2Hear unassisted due to cognitive decline or dementia, determined via a self- or familial report. The demographic information of the sample is provided in Table 1. Overall, participants presented with mild-to-moderate hearing loss and self-reported as *competent* users of digital technologies [21].

Table 1. Demographic information of participants who trialed either m2Hear alone or both m2Hear and C2Hear.

Demographic characteristics	m2Hear only	m2Hear and C2Hear	Overall
Gender, n (%)			
Female	3 (30)	3 (50)	6 (37)
Male	7 (70)	3 (50)	10 (63)
Age (years)			
Mean (SD)	67.70 (11.01)	70.67 (17.53)	68.81 (13.32)
Range	46-81	39-85	39-85
Better ear pure-tone average (0.25-4 kHz; dB HL)			
Mean (SD)	26.90 (5.55)	33.50 (8.34)	28.79 (6.85)
Range	17-34	25-45	17-45
Self-reported digital technology competency, n (%)			
Beginner	0 (0)	2 (33)	2 (12)
Competent	10 (100)	4 (67)	14 (88)

XSL·FO
RenderX

Procedure

Potentially eligible participants were invited to participate in the study during their hearing assessment appointment at the Nottingham Adult Audiology Service. Interested patients attended an initial study session at the National Institute for Health Research (NIHR) Nottingham Biomedical Research Centre (BRC). This session was completed shortly after participants had received their hearing aids (mean 5.06, SD 4.28 days postfitting). During this session, participants were shown how to use m2Hear and were asked to use it in their everyday lives as and when required. They were encouraged to use as much of the content that they felt was relevant to them. A subgroup of participants (n=6) was also asked to use the original C2Hear intervention via YouTube, which they were asked to compare to their use of m2Hear.

Following a period of independent use (mean 10.82, SD 0.70 weeks), participants attended a second study session at the NIHR Nottingham BRC. All participants were interviewed by the lead author (DM). The interview schedules were flexible due to the semistructured design of the interviews, although the core content remained the same (Multimedia Appendices 2 and 3). The interviews were conducted face-to-face in a quiet room and lasted for approximately 1 hour. Each interview was audio-recorded and transcribed verbatim.

All participants were each paid a nominal inconvenience allowance and travel expenses. The study was approved by the NHS Health Research Authority, East of England–Cambridgeshire and Hertfordshire Research Ethics Committee, and Nottingham University Hospitals NHS Trust Research and Innovation Department.

Data Analysis

NVivo 10 software (QSR International) was used to organize and support the analysis of the semistructured interview data.

Anonymized identification codes were assigned to each participant (eg, M060, M061, etc). The data were analyzed using the thematic analysis procedure by Braun and Clarke [36], which consists of specific analytical phases: data familiarization, generating initial codes, searching for themes, reviewing themes, and defining and naming themes. The analysis was deductive (or theoretical), as themes were derived from the components of the TDF, which each link to a specific aspect of the COM-B model [32]. In this study, themes were considered in relation to the barriers and facilitators that impacted the use of m2Hear in everyday life. The process of combining or redefining the codes led to the generation of initial themes. Overarching themes stemmed from the TDF, while inductive subthemes were devised by the first author. Subthemes were defined as something important about the data that captured important information about the research aims and represented repeated patterns of response or meaning that were prevalent (ie, reported by several participants) across the entire data set [36]. For rigor, coding comparison [39] was undertaken to ensure that the interpretation of the data was not limited to the perspective of the first author. Specifically, a second author (RH) independently coded 6 (37%) of the transcripts and formulated potential themes. Any discrepancies were discussed, and an agreement was made regarding which codes should be applied. The themes were refined and defined through re-analysis of the data and discussions among coauthors (DM and RH).

Results

Theoretical domains that were and were not coded against the TDF are shown in Table 2. A summary of the subthemes classified according to the TDF is also provided in Table 3. Each domain was mapped onto the corresponding determinants of the target behavior (ie, use of m2Hear) in accordance with the COM-B model (Multimedia Appendix 1).

Table 2. Theoretical domains that were coded against the theoretical domains framework and those that were not.

COM-B[a], TDF[b]	Coded?
Capability	
Knowledge	No
Physical skills	Yes
Memory, attention, and decision processes	Yes
Behavioral regulation	No
Opportunity	
Social influences	Yes
Environmental context	Yes
Motivation	
Social and professional role and identity	No
Beliefs about capabilities	Yes
Beliefs about consequences	No
Intentions	No
Goals	Yes
Reinforcement	No
Emotion	No
Optimism	No

[a]COM-B: Capability, Opportunity, Motivation-Behavior.
[b]TDF: theoretical domains framework.

Table 3. A summary of subthemes generated in relation to the Capability, Opportunity, Motivation-Behavior model and theoretical domains framework.

COM-B[a], TDF[b]	Subtheme
Capability	
Physical skills	Digital literacy skills
Memory, attention, and decision processes	Conciseness
Memory, attention, and decision processes	Repeated use to aid memory
Opportunity	
Social influences	Independence and autonomy
Social influences	Relevance to communication partners
Social influences	Requirement of audiological input
Environmental context	Convenience and portability
Environmental context	Ease of use and individualization
Motivation	
Beliefs about capabilities	Learning and understanding
Beliefs about capabilities	Reassurance and self-efficacy
Goals	Improving hearing aid handling and communication skills

[a]COM-B: Capability, Opportunity, Motivation-Behavior.
[b]TDF: theoretical domains framework.

Capability

Digital Literacy Skills

Having sufficient digital literacy skills was considered essential for users to successfully access and use m2Hear. For example,

M067 commented, "I could use it, and then my Dad could use it, so…there's no sort of barriers in it. It's open to everyone to use." Some thought that m2Hear might be difficult to use for older people who lack digital skills. M061 remarked, "I think if you're computer-savvy…it could be a benefit...for other

people, I'm just thinking that the older age group, I think, will struggle." As a potential solution, friends or family could support the use of m2Hear:

> Somebody else who hasn't got that confidence or the capability or the technology to do it, needs to talk to somebody…If they've got that reference, they've got a daughter or a son...they can be assisted by their own family. [M065]

Nevertheless, given the ubiquitous nature of smartphone technologies, digital literacy was considered to be less of a concern:

> The only people who won't have [m2Hear] available to them all the time are those who haven't got a smartphone, a tablet or a computer. I would say they're going to be non-existent, very shortly. I mean, even Granny's got a smartphone now. [M065]

Conciseness

The short mRLOs enabled participants to locate personally relevant information with ease:

> I think the value is that you can just go to the subjects you want. For bits that you're not interested in, well, what's the point of going to that? What you want is information, and to be able to go straight to that bit of information, I think, is a valuable part of [m2Hear]. [M064]

The succinct mRLOs also encouraged access and re-engagement, "They're concise and logical, and that's why I went back to them a couple of times" (M065). Furthermore, the mRLOs were a good use of time and effort: M062 said, "It meant best use of my time. I didn't have to spend ages getting bored and wondering where it was, I could clearly see what it was I needed." The concise content also held participants' interest, which improved information retention:

> It's better because you only have a certain amount of attention span, and I think because they are only like a minute, that's about right…there's so much more information in longer ones, whereas if it's a little bit of information it makes it easier to take in. [M067]

Taken together, the mRLOs were concise and subsequently viewed favorably.

Repeated Use to Aid Memory

Although participants reported that they initially decided to view all mRLOs within the first week, they would regularly reuse m2Hear throughout the 10-week study period to improve their knowledge and understanding of specific topics. For example, participants consistently commented that they would rewatch the mRLOs for troubleshooting advice, particularly when they experienced a problem with their hearing aid or communication:

> I'd click on the "communicate with others" one, and "looking after my hearing aid". I'd click on those two if I'd had a bad day about something, so I'd click on there just to have a look, and then make sure I was

> looking after them properly. Just to refresh what I should do with them. [M067]

It was consistently reported that m2Hear supplemented and expanded upon the information provided by the audiologist during their hearing aid fitting appointment, as well as providing useful reminders of the information they had been given:

> Although the audiologist had probably given me all the right information, you don't store it. You don't remember it because there's so much coming in…Because I'd got [m2Hear], I could go back, and I could go and check on something. [M062]

Participants commented that it was often difficult to retain all the information provided during audiological appointments and that re-using m2Hear during the study period provided a means to refresh their memory, without feeling overwhelmed or confused:

> You can only take in so much at any one time…It's nice to have a reference to go back to it, just to check it on your own…The information is there if you need it. That's what a [m2Hear] is all about. [M065]

In addition, the quizzes and interactive features also ensured that the content of the mRLOs were successfully retained and remembered, "Yes, it was like a little test to work on my memory. Even though I was watching the [mRLO] I was probably not watching it, whereas a quiz it was making me remember stuff" (M067). Nevertheless, some participants stated that they did not think it necessary for them to reuse m2Hear once they had viewed all the mRLOs, as they understood the information and were confident using their hearing aids:

> Well, in the first instance of course it was very good because it told me about the hearing aid. After a while, no, I don't think there's much more you need to know. [M068]

Opportunity

Independence and Autonomy

One of the main reported advantages of m2Hear was that it enabled participants to self-manage their hearing loss. This was summarized by M060:

> The thing is that everyone is responsible for their own wellbeing. You shouldn't rely on other people. For me, I don't feel that I should be relying on other people. Other people are there in a desperate need, which is why [m2Hear] is good. As I keep saying, it's a prop.

Participants stated that managing their hearing loss independently was crucial to their self-esteem:

> I haven't got to rely on anyone else. I haven't got to ask someone else; I don't like asking people…I'd rather just be able to do it myself...If I'm constantly asking people, I just feel helpless, and I don't like that. [M067]

In addition, participants were often reluctant to telephone and arrange an appointment with an audiologist due to limited health

care resources and stated that m2Hear empowered them to be more self-sufficient:

> I think [m2Hear] is very useful. I mean, in audiology, they tell you things and that, but the limited time. Yes, it just seems a useful reference instead of having to keep phoning someone up or go back... You're straight there for as many times as you want it. [M064]

Relevance to Communication Partners

m2Hear was not only relevant to people living with hearing loss, but also to frequent communication partners. Indeed, several participants stated that they shared m2Hear with friends and family members to improve their communication. For example, M060 said:

> I showed my wife...to make sure that she's aware of the subtleties of not talking to me in a different room. That if she's talking to me in a different room then we're back to square one.

The added value of m2Hear was that it provided an objective voice that was external to the participant's immediate social network:

> I think because it's produced by somebody professionally, rather than me just giving my opinion or my interpretation...therefore that gave it more weight as far as he [husband] was concerned. [M062]

However, some participants preferred to use m2Hear alone, as they did not think it would be suitable or relevant for communication partners that do not have a hearing loss, such as M069, who said, "Well, I didn't think it was particularly important, you know, really. They were…appropriate to me, but not to anybody else."

Requirement of Audiological Input

Although m2Hear facilitated self-management of hearing loss, there were occasions when it was considered necessary to seek help and advice from an audiologist. Changes in the overall hearing level, for instance, would prompt participants to arrange an audiological reassessment; "If my hearing changes I would ring audiology," remarked M060. In addition, participants reported that they would arrange an appointment if their hearing aid was faulty or causing discomfort:

> I suppose if I was having specific problems with my hearing aids and thought there was something malfunctioning or something like that, then I would ring up the [audiology department] and speak to somebody. [M062]

M069 also commented that they felt it was permissible to contact audiology if they experienced difficulties they were unable to resolve themselves using m2Hear: "I think if something goes wrong, you need advice about that type of thing, then…it's nice to know that you've got somebody to talk to." On this basis, m2Hear is a useful "tool" (M062) that can be used to supplement the provision of face-to-face appointments with an audiologist.

Convenience and Portability

Whether participants opted to access m2Hear from a handheld device (eg, smartphone, tablet) was based on immediacy and convenience. For example, M066 said, "I had a smartphone…that's what I use all the time…It's just easier to pick up and use." Similarly, M069 commented:

> I've got a laptop, but I don't use it very often. The tablet is much more convenient...it's easier to handle. It's smaller, and also, to be able to access the Internet, it's almost immediate. Whereas, the laptop isn't.

Accessing m2Hear from smartphone technologies was advantageous because of portability: "It's transferrable isn't it, between devices…it's portable, everywhere, any device. It makes it interesting" (M067). Furthermore, m2Hear could be used in multiple listening situations whenever desired. For example, comparing m2Hear with written information, M066 commented, "It's more convenient to use, wherever you are. You can just get your phone out, whereas you might not have the booklet with you." Nevertheless, some participants preferred to use a laptop or desktop computer to access m2Hear, citing that a larger screen size was necessary to optimize visual acuity: "My eyesight is going as well. So, whilst the phone, I've got an iPhone, is great, I'm finding if I need to look at something that I feel I get a bigger picture" (M061). Taken together, while the device used to access m2Hear varied across participants, this was often based on personal preference and convenience. Nevertheless, smartphone technologies have improved accessibility.

Ease of Use and Individualization

All participants commented that m2Hear was simple and straightforward to use, "It wasn't loads of diving off into other areas and that, it was…nice and simple. There wasn't anything complicated" (M067). It was also well-organized and structured, ensuring that information could be easily located: M062 said, "It's very easy to get to the section you need to look at or you want to look at…you can see quite clearly which elements you need to go to. That was good." In relation, the questions accompanying each individual mRLO facilitated navigation:

> They seemed relevant to the [mRLO] you'd just watched. I think the questions really weren't difficult questions. I presume the questions were designed in mind of, "Have you got the hang of what the [mRLO] was on about?" I think, yes, the questions tease that answer out quite well. [M064]

Participants also reported that the organization of m2Hear enabled greater personalization:

> It means that you haven't got to start right from the beginning again...I could just go straight to the one that I want. That's all I want to go to, and it doesn't hold you back, it just makes it easier to access it. [M067]

For these reasons, participants reported that m2Hear was preferable to written information: M065 summarized, "In the written information you've got to do a whole lot of searching.

[m2Hear] leads you by the nose, to be quite honest, and everything is logical about it, and it just takes a click."

Motivation

Improved Learning and Understanding

Participants commented that the mRLO content was comprehensive, useful, and relevant, which they believed facilitated their knowledge of hearing aids and communication. For example, M060 said, "I can't think of any topics that were actually left out...It's got everything in there that I needed to know, and it's there for me to look back on if I forget." However, some participants stated that they would have liked additional information to have been included, such as the physiology of the ear and hearing loss, "I'd like a bit more technical depth to it...I'd like a little bit more depth about how the hearing system works" (M063). In addition to the mRLOs, optional quizzes and interactivities were also perceived to improve participants' learning and understanding. For example, remarking on one of the activities, M060 said:

> It gives you a bit of food for thought, so you think about it again. It's no good learning something by rote, as it were. It's understanding it. So next time you come across that you've got an understanding.

However, some stated that the quizzes were not beneficial, as they understood the content of the mRLO:

> In my case, I think I pretty much understood everything on the [mRLOs]. The quizzes for me were slightly irrelevant in terms of my understanding, my learning. I didn't need them. [M060]

Therefore, although some participants expressed ambivalence, most stated that they used m2Hear because they believed that both the mRLOs and interactive components would improve their knowledge of hearing aids and communication, resulting in more successful outcomes.

Reassurance and Self-Efficacy

Several participants reported that they regularly referred back to m2Hear as a means of support and reassurance that they were handling their hearing aids correctly:

> I think it's a kind of reassurance thing...You watch it and you think, "Well, yes, I am." I think it's just useful as a reminder, having forgotten something or so on. Yes, I think it's a useful tool, isn't it? Yes, it's a good backup, I think. [M064]

Subsequently, m2Hear improved both self-efficacy for hearing aids and coping with hearing loss:

> I can't reinforce how useful I feel that [m2Hear] is. As I say, I've gone back to recap on different things...It's just really given me the confidence...I feel that I can cope with any situation with my hearing aids now. [M060]

In addition, m2Hear reduced feelings of loneliness and despair, with M060 saying:

> I think when I was looking at it I was not only soaking in the information and stuff, I was thinking, well, this

> is something that's relevant to a lot of people...I'm not on my own, and I've got something to help me.

Therefore, using m2Hear facilitated an optimistic outlook, empowering participants to self-manage their hearing loss.

Improving Hearing Aid Handling and Communication Skills

The most commonly cited reason participants used m2Hear was to improve their hearing aid handling and communication skills. m2Hear helped participants to manage their expectations and encouraged them to persevere using their hearing aids:

> You expect to be able to put [hearing aids] on like you put glasses on. I did think it would work in the same way. I thought it would be instant and it's not. [m2Hear] is a useful reminder...this isn't going to be straightforward and you're going to have to work at it, but the benefits will be worthwhile. [M062]

Moreover, m2Hear facilitated hearing aid use, especially when experiencing communication difficulties that might otherwise result in social withdrawal. For example, M067 commented that the mRLOs:

> made a difference, they helped me. I mean I wouldn't have said to people, "I can't hear you," I probably would have just switched off, and sort of just not bothered probably. It did help me; it gave me a bit of a boot.

As such, m2Hear had a positive impact on hearing aid use and adherence as well as communication during the 10-week evaluation.

Differences Between m2Hear and C2Hear

In relation to the themes identified, differences between the barriers and facilitators that impacted the use of m2Hear compared with those that impacted the use of C2Hear were also reported, whereby m2Hear was consistently viewed more positively.

Digital Literacy

m2Hear was considered more appropriate for any level of digital literacy compared with C2Hear when accessed via YouTube. For example, M075 stated, "[m2Hear] is presented so well that I think most people, no matter how poor their understanding is, they'd still get on well with that." In comparison, the C2Hear YouTube channel was viewed as more difficult to use if an individual had poor digital literacy skills. When using YouTube, M078 remarked:

> I was terrified in case I'd make a mistake and I ruin my computer, you see. Because I'm not very computer-savvy. So, I just looked at [m2Hear]

Conciseness

A key difference reported between m2Hear and C2Hear was the difference in length of the mRLOs. The shorter mRLOs were preferred, as participants felt that they could retain the information more easily. For example, M075 commented, "You can assimilate the knowledge more easily and you retain it perhaps better." A further advantage of the shorter mRLOs was

that it was easier to locate personally relevant information: "m2Hear was covering a topic which was concise and to the point, so you could jump straight to the information that you wanted" (M074). Conversely, participants commented that the C2Hear RLOs were lengthy in duration: "I lost interest in the [C2Hear RLOs]. I found them a bit distracting because sometimes they told you a bit too much" (M076). Nevertheless, the longer RLOs might be preferable for individuals who like to access all the information in one sitting:

> Sometimes the length of the [RLOs] could be nice. When you have got all the information, it's not so bad to watch it all from beginning to end. [M076]

Convenience and Portability

m2Hear was more accessible than C2Hear, as it was more convenient to use with smartphone technologies. For example, M074 remarked:

> [m2Hear] was smarter on my iPhone to use, the interface was more workable to me on the iPhone, whereas the C2Hear...because it was on YouTube, it kept jumping up with other clips of other hearing research which was frustrating.

A further benefit of m2Hear being easily accessed via smartphone technologies was that it could be used whenever and wherever required:

> It's easier just to use [m2Hear] when you're out and about and you just want to sit down and have a look at it and just go through it, just to remember things more, rather than wait till I get home and just look on the laptop. [M076]

Conversely, participants reported that their use of C2Hear was restricted to when they had more time available due to the length of the RLOs: M076 said, "I used to do [C2Hear] when I'd got more time, so I could watch all of them, not dip in and out of it quite so much". Taken together, convenience and portability improved the accessibility of m2Hear compared with C2Hear.

Ease of Use and Individualization

An important difference between m2Hear and C2Hear was that m2Hear offered the opportunity for greater personalization. In addition, it was easier to find personally relevant information in m2Hear compared with C2Hear: "[m2Hear] is so easy, and it's so user-friendly, and it's so clear" (M075). C2Hear was less personalized and difficult to navigate. For example, M071 remarked, "In C2Hear you've got to go back, find various bits, and not quite start again, but it's much more difficult to go back." Nevertheless, C2Hear would have been acceptable if they had not been able to compare it to m2Hear: M071 stated, "If you hadn't shown me [m2Hear], I'd probably have been perfectly happy with [C2Hear]." Therefore, although participants were satisfied with C2Hear, they preferred m2Hear overall.

Learning and Understanding

Participants consistently commented that m2Hear was more interactive than C2Hear, which reinforced the knowledge gained from each mRLO and would appeal to different learning preferences. For example, M076 remarked, "It's a different way

of learning, a different way of putting information across." In comparison, C2Hear was considered less interactive, with some participants failing to notice that there was the opportunity to take quizzes at the end of each RLO: "I didn't spot [the quizzes] on [C2Hear]. Probably because…the [RLO] went on too long" (M076).

Reassurance and Self-Efficacy

Compared with C2Hear, m2Hear improved confidence and enabled greater independent use of hearing aids. M074 commented:

> [m2Hear] increases my confidence for looking after my hearing aids myself without needing to go and get help...It gives me reassurance that I'm doing the right thing and that I'm not going to break them.

Participants also reported that the m2Hear interactivities provided further reassurance: "They made you think and made you realise what there was and how easy it was to get the information out" (M075). Thus, in comparison to C2Hear, m2Hear improved self-efficacy for hearing aids and communication, facilitating greater self-management of hearing loss.

Discussion

Principal Findings

The functionality of mHealth technologies has been shown to enable greater individualization and interactivity in multiple chronic health care domains, which has the potential to improve self-management [23,24]. In this study, we assessed the barriers and facilitators of using a newly developed mHealth educational intervention, m2Hear, designed specifically for first-time hearing aid users. We assessed the views of first-time hearing aid users toward m2Hear when used in everyday life. To gain an in-depth insight into potential barriers and facilitators, participants' experiences were evaluated within the context of the TDF and COM-B model [40,41], which are discussed as follows.

Capability

With regard to capability, digital literacy skills were identified as important for the usability and adherence of m2Hear. For m2Hear, participants commented that any level of skill would be sufficient in this area, given that it was relatively straightforward to use and navigate, whereas C2Hear accessed via YouTube required a high level of digital literacy. It is likely that the ease of use of m2Hear is attributable to the iterative, user-centered, and participatory design approach that was employed during the development of m2Hear [25,26]. Such an approach, which has been shown to improve usability, acceptability, and adherence of interventions [42], likely ensured that m2Hear met the specific needs of the end user. This is encouraging given that the proportion of older adults (≥55 years) who use smartphone technologies continues to increase exponentially in the typical first-time hearing aid user age group [43,44]. Thus, mobile technologies should be considered an acceptable and accessible mode of delivery for educational support throughout the hearing aid user's journey, as digital literacy skills are becoming less of a barrier in this population.

A further theme related to capability was that m2Hear was reused throughout the 10-week study period because it provided useful reminders that expanded upon the information provided by the audiologist during the hearing aid fitting appointment. This was further facilitated by the concise duration of the mRLOs, which enabled participants to easily locate and revisit the desired information with ease. In support of these findings, existing research in the area of multimedia learning recommends dividing content into shorter learning segments to reduce cognitive load (or memory capacity), improving knowledge acquisition and long-term retention [45]. On this basis, mHealth interventions have the potential to improve the likelihood that first-time hearing aid users will acquire the necessary knowledge and skills to successfully self-manage their hearing loss (eg, improve hearing aid use and social participation).

Opportunity

One of the most pertinent social factors identified in this study was whether participants shared m2Hear with their family and friends (ie, frequent communication partners). Although some participants felt it necessary to share m2Hear with others to improve mutual communication, others did not, citing that their hearing loss was not a concern for others who did not experience hearing difficulties. This latter finding should be addressed, given that frequent communication partners play a pivotal role in hearing loss management and communication [46]. Furthermore, hearing loss in older adults can result in continued communication difficulties, leading to social isolation and withdrawal for both the person living with hearing loss and their communication partners, termed third-party hearing disability [47].

Meeting the informational needs of communication partners has been shown to be highly beneficial [12,47-49]. For example, Barker et al [12] suggest that information and support should be offered to both individuals and their communication partners to align coping strategies and improve outcomes for both parties. Consequently, we have redeveloped and tailored the original *communication tactics* C2Hear RLO into an mRLO suitable for communication partners [26]. Specifically, we have altered the wording so that it is more generic for *others*, such as family members and the general public. Interactive components have also been incorporated, including simulated hearing losses in the presence and absence of background noise. This mRLO for communication partners and the general public is available on the web [20]. The quality, usability, relevance, and impact of the repurposed mRLO have subsequently been examined using think-aloud techniques with dyads comprising adults with hearing loss and their communication partners [26]. We found that these dyads led to greater inclusivity; the mRLO enabled greater joint working and joint responsibility, whereby both parties became jointly aware of factors that prevented and facilitated optimal communication. As a result, mHealth educational interventions that incorporate greater individualization and interactivity have the potential to improve outcomes for adults living with hearing loss and their frequent communication partners. This is also highlighted in recent recommendations published by the UK National Institute for Health and Care Excellence, which states that, in addition to the person with hearing loss, information about hearing loss and how it can be managed should also be given to family members and caregivers [50].

Another theme related to opportunity included environmental factors that promoted greater use of m2Hear such as convenience and portability. These findings reflect a key benefit of mHealth technologies, as they can be used when comfort and convenience are paramount [51]. Furthermore, in the context of learning, the perceived convenience of mobile technologies has also been shown to have a positive impact on attitudes and intentions toward using an educational intervention [52]. Other environmental factors identified included the ease of finding personally relevant information due to improved organization and navigation. This is likely attributable to the extensive iterative usability testing that was employed during the development of m2Hear [25,26]. The ability to discover relevant information independently as well as to control the pace of learning via well-indexed content has been shown to enhance learning potential [53,54]. As such, these findings lend further support for the notion that a user-centered and participatory design approach should be utilized when developing mHealth interventions so that they meet the specific educational needs of the end user.

Motivation

Participants consistently reported that they were motivated to use m2Hear because it improved their knowledge of specific topics relating to hearing aids and communication. Supplemental interactivities further improved perceived learning and understanding, presumably through active engagement with learning materials [54,55]. Participants were also motivated to use m2Hear because it provided reassurance and increased their confidence (or self-efficacy) to use hearing aids and communicate successfully. Self-efficacy refers to a domain-specific construct associated with particular tasks, abilities, skills, or actions that are needed to achieve a certain behavior, including health-related behaviors [56]. Perceived self-efficacy is being increasingly recognized as playing a key role in the audiological rehabilitation process [57-59]. Previous studies have shown that individuals with higher levels of self-efficacy are more likely to obtain hearing aids and become successful users [60-62]. In addition, self-efficacy has also been shown to predict hearing outcomes, including satisfaction [63], and has been shown to be a modifiable factor that could be targeted to improve hearing loss self-management [6]. We have shown that C2Hear significantly increases self-efficacy for hearing aids and readiness to act, with large clinical effect sizes, compared with a printed booklet. This was shown to be highly efficacious even when delivered at the hearing assessment appointment, thus priming patients before the provision of hearing aids [18].

Study Limitations and Future Research

There are several caveats to the design of this study that could be addressed in future research. For example, a purposive sampling strategy was employed, whereby participants were recruited based on prespecified inclusion and exclusion criteria, such as familiarity with smartphone technologies. This was necessary to ensure that participants would be able to use and access m2Hear throughout the home-based evaluation. As a

result, it is perhaps unsurprising that most participants self-reported as *competent* users of digital technologies, which arguably limits the generalizability of the study findings to individuals with lower levels of self-perceived competency. Future studies could address this limitation by enhancing the representativeness of the sample in terms of digital literacy skills as well as other demographic and clinical characteristics such as age, gender, and hearing loss severity.

A further consideration is that this study used a formalized, deductive (or theory-driven) thematic analysis approach, whereby themes were underpinned by the COM-B model and TDF. The application of theories and models from health psychology in audiological rehabilitation research continues to rise [11,64-68]. However, popular models frequently used in the field of audiology (eg, the Health Belief Model [28], Theory of Planned Behaviour [29], Transtheoretical Model [30]) have been widely criticized because they fail to reliably account for variations in complex human behavior [31]. As a result, Coulson et al [31] suggested that the use of unreliable models to explain and predict hearing health behaviors should be replaced by more contemporary behavior change science, namely, the COM-B model. As such, this study adds to a growing body of literature that has utilized the COM-B model to inform adult aural rehabilitation practices [26,67,69].

Clinical Implications

As we have argued from the outset, a key advantage of mHealth interventions is that they enable the individual to tailor the information they need as well as increase user interaction, resulting in a more patient-centered approach. Patient-centered care is widely accepted as a fundamental practice that supports an individual to be an active participant in the management of their health [70]. Critically, involving patients in their own care can result in empowerment, conceptualized as a process that enhances feelings of autonomy, control, self-efficacy, and coping [71]. The concept of empowerment was also conveyed in this study, whereby participants reported that m2Hear improved their confidence to take control and participate more

fully in the management of their hearing health. It is likely that this finding stems from a combination of factors afforded by using m2Hear, including increased knowledge of hearing aids and optimal communication strategies. In support, in their qualitative assessment of patients' perspectives of empowerment, Small et al [71] found that improved knowledge and understanding is a pertinent factor necessary to empower patients to manage long-term health conditions. Additionally, identified themes surrounding reassurance and self-efficacy suggest that using m2Hear not only benefitted psychological capability but also reflective motivation for hearing loss self-management. This suggests that m2Hear fulfills 3 cornerstones of successful hearing loss self-management: (1) enabling the acquisition of knowledge, (2) prompting actions (eg, practicing hearing aid insertion), and (3) the adoption of a positive psychological stance (ie, self-efficacy). On this basis, we advocate the widespread implementation of mHealth educational resources in adult aural rehabilitation given that they have substantial potential to facilitate patient-centered care and improve hearing health outcomes.

Summary and Conclusions

This qualitative study provides an in-depth assessment of an mHealth educational intervention used by first-time hearing aid users in their everyday lives. Underpinned by a contemporary model of health behavior change, the COM-B model, we identified key factors that influenced intervention use. Specifically, m2Hear was viewed as a concise and comprehensive resource that is simple and straightforward to use and enables greater individualization and independence. In addition, m2Hear provides greater support and reassurance, improves confidence, and empowers users to self-manage hearing loss. On this basis, this study suggests that m2Hear can be used to supplement existing aural rehabilitation practices to support successful self-management in first-time hearing aid users. Furthermore, this study demonstrates the utility of employing theoretical and ecologically valid approaches in the development of mHealth educational resources to meet the individual needs of the end user.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions of the theoretical domains framework (TDF).
[DOCX File , 21 KB - mhealth_v8i8e17193_app1.docx]

XSL•FO

RenderX

Multimedia Appendix 2
Semistructured interview schedule for participants evaluating m2Hear only (n=10).
[DOCX File , 21 KB - mhealth_v8i8e17193_app2.docx]

Multimedia Appendix 3
Semistructured interview schedule for participants evaluating both m2Hear and C2Hear (n=6).
[DOCX File , 20 KB - mhealth_v8i8e17193_app3.docx]

References

1. Integrated chronic disease prevention and control. World Health Organization. URL: https://www.who.int/chp/about/integrated_cd/en/ [accessed 2020-01-09]
2. McGowan PT. Self-management education and support in chronic disease management. Prim Care 2012 Jun;39(2):307-325. [doi: 10.1016/j.pop.2012.03.005] [Medline: 22608868]
3. Taylor SJ, Pinnock H, Epiphaniou E, Pearce G, Parke HL, Schwappach A, et al. A rapid synthesis of the evidence on interventions supporting self-management for people with long-term conditions: PRISMS - Practical systematic Review of Self-Management Support for long-term conditions. Southampton (UK): NIHR Journals Library 2014 Dec;2(53). [doi: 10.3310/hsdr02530] [Medline: 25642548]
4. Barker F, Mackenzie E, Elliott L, Jones S, de Lusignan S. Interventions to improve hearing aid use in adult auditory rehabilitation. Cochrane Database Syst Rev 2016 Aug 18(8):CD010342 [FREE Full text] [doi: 10.1002/14651858.CD010342.pub3] [Medline: 27537242]
5. Convery E, Meyer C, Keidser G, Hickson L. Assessing hearing loss self-management in older adults. Int J Audiol 2018 Apr;57(4):313-320. [doi: 10.1080/14992027.2017.1390268] [Medline: 29081257]
6. Convery E, Hickson L, Meyer C, Keidser G. Predictors of hearing loss self-management in older adults. Disabil Rehabil 2019 Aug;41(17):2026-2035. [doi: 10.1080/09638288.2018.1457091] [Medline: 29587551]
7. Ferguson M, Brandreth M, Brassington W, Wharrad H. Information retention and overload in first-time hearing aid users: an interactive multimedia educational solution. Am J Audiol 2015 Sep;24(3):329-332. [doi: 10.1044/2015_AJA-14-0088] [Medline: 26649541]
8. Reese J, Smith S. Recall of hearing aid orientation content by first-time hearing aid users. Semin Hear 2006 Nov;27(4):337-344. [doi: 10.1055/s-2006-954862]
9. Desjardins JL, Doherty KA. Do experienced hearing aid users know how to use their hearing AIDS correctly? Am J Audiol 2009 Jun;18(1):69-76. [doi: 10.1044/1059-0889(2009/08-0022)] [Medline: 19380509]
10. Ferguson M, Kitterick P, Chong L, Edmondson-Jones M, Barker F, Hoare D. Hearing aids for mild to moderate hearing loss in adults. Cochrane Database Syst Rev 2017 Sep 25;9:CD012023 [FREE Full text] [doi: 10.1002/14651858.CD012023.pub2] [Medline: 28944461]
11. Heffernan E, Coulson NS, Henshaw H, Barry JG, Ferguson MA. Understanding the psychosocial experiences of adults with mild-moderate hearing loss: an application of Leventhal's self-regulatory model. Int J Audiol 2016 Jul;55(Suppl 3):S3-12 [FREE Full text] [doi: 10.3109/14992027.2015.1117663] [Medline: 26754550]
12. Barker AB, Leighton P, Ferguson MA. Coping together with hearing loss: a qualitative meta-synthesis of the psychosocial experiences of people with hearing loss and their communication partners. Int J Audiol 2017 May;56(5):297-305. [doi: 10.1080/14992027.2017.1286695] [Medline: 28599604]
13. Kessels RPC. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222 [FREE Full text] [doi: 10.1258/jrsm.96.5.219] [Medline: 12724430]
14. Latorre-Postigo JM, Ros-Segura L, Navarro-Bravo B, Ricarte-Trives JJ, Serrano-Selva JP, López-Torres-Hidalgo J. Older adults' memory for medical information, effect of number and mode of presentation: an experimental study. Patient Educ Couns 2017 Jan;100(1):160-166. [doi: 10.1016/j.pec.2016.08.001] [Medline: 27567498]
15. Nair EL, Cienkowski KM. The impact of health literacy on patient understanding of counseling and education materials. Int J Audiol 2010 Feb;49(2):71-75. [doi: 10.3109/14992020903280161] [Medline: 20151879]
16. Ferguson M, Leighton P, Brandreth M, Wharrad H. Development of a multimedia educational programme for first-time hearing aid users: a participatory design. Int J Audiol 2018 Aug;57(8):600-609. [doi: 10.1080/14992027.2018.1457803] [Medline: 29718733]
17. Ferguson M, Brandreth M, Brassington W, Leighton P, Wharrad H. A randomized controlled trial to evaluate the benefits of a multimedia educational program for first-time hearing aid users. Ear Hear 2016;37(2):123-136 [FREE Full text] [doi: 10.1097/AUD.0000000000000237] [Medline: 26565785]
18. Gomez R, Ferguson M. Improving self-efficacy for hearing aid self-management: the early delivery of a multimedia-based education programme in first-time hearing aid users. Int J Audiol 2020 Apr;59(4):272-281. [doi: 10.1080/14992027.2019.1677953] [Medline: 31651206]
19. C2Hear Online. YouTube. URL: https://www.youtube.com/c2hearonline [accessed 2020-07-01]
20. C2Hear Online. URL: https://c2hearonline.com/ [accessed 2020-07-01]

XSL•FO
RenderX

21. Henshaw H, Clark DP, Kang S, Ferguson MA. Computer skills and internet use in adults aged 50-74 years: influence of hearing difficulties. J Med Internet Res 2012 Aug 24;14(4):e113 [FREE Full text] [doi: 10.2196/jmir.2036] [Medline: 22954484]

22. World Health Organization. Mhealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on Ehealth. Geneva, Switzerland: World Health Organization; 2011. URL: https://apps.who.int/iris/handle/10665/44607/ [accessed 2020-01-09]

23. 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]

24. 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]

25. Ferguson M, Maidment D, Henshaw H, Heffernan E. Evidence-based interventions for adult aural rehabilitation: that was then, this is now. Semin Hear 2019 Feb;40(1):68-84 [FREE Full text] [doi: 10.1055/s-0038-1676784] [Medline: 30728650]

26. Ferguson M, Maidment D, Henshaw H, Gomez R. Knowledge is power: improving outcomes for patients, partners, and professionals in the digital age. Perspect ASHA SIGs 2019 Feb 26;4(1):140-148. [doi: 10.1044/2018_pers-sig7-2018-0006]

27. Medical Research Council. Developing and evaluating complex interventions: new guidance. London, United Kingdom. 2006. URL: https://mrc.ukri.org/documents/pdf/developing-and-evaluating-complex-interventions/ [accessed 2020-01-09]

28. Rosenstock IM. Why people use health services. Milbank Q 1966 Jul;44(3):94. [doi: 10.2307/3348967]

29. Fishbein M, Ajzen I. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading, MA: Addison-Wesley; 1977.

30. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [doi: 10.1037//0022-006x.51.3.390] [Medline: 6863699]

31. Coulson NS, Ferguson MA, Henshaw H, Heffernan E. Applying theories of health behaviour and change to hearing health research: time for a new approach. Int J Audiol 2016 Jul;55(Suppl 3):S99-104. [doi: 10.3109/14992027.2016.1161851] [Medline: 27138716]

32. 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]

33. 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]

34. Fonteyn ME, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. Qual Health Res 1993 Nov 1;3(4):430-441. [doi: 10.1177/104973239300300403]

35. Todhunter F. Using concurrent think-aloud and protocol analysis to explore student nurses' social learning information communication technology knowledge and skill development. Nurse Educ Today 2015 Jun;35(6):815-822. [doi: 10.1016/j.nedt.2015.01.010] [Medline: 25910633]

36. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

37. m2Hear. University of Nottingham. 2019. URL: https://www.nottingham.ac.uk/helm/dev-test/m2hear/ [accessed 2020-07-01]

38. mHealth Technologies for Hearing Aid Users (m2Hear). Clinical Trials. URL: https://clinicaltrials.gov/ct2/show/NCT03136718 [accessed 2020-07-01]

39. Yardley L. Demonstrating validity in qualitative psychology. In: Smith JA, editor. Qualitative Psychology: A Practical Guide to Research Methods. London, UK: Sage Publications; 2008.

40. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]

41. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. London, UK: Silverback; 2014.

42. Scariot CA, Heemann A, Padovani S. Understanding the collaborative-participatory design. Work 2012;41(Suppl 1):2701-2705. [doi: 10.3233/WOR-2012-0656-2701] [Medline: 22317129]

43. Mobile Consumer Survey 2019: Plateauing at the Peak: the State of the Smartphone. Deloitte UK. 2019. URL: https://www.deloitte.co.uk/mobileuk [accessed 2020-07-01]

44. Mobile Fact Sheet. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-07-01]

45. Ibrahim M, Antonenko PD, Greenwood CM, Wheeler D. Effects of segmenting, signalling, and weeding on learning from educational video. Learn Media Technol 2011 Jun 28;37(3):220-235. [doi: 10.1080/17439884.2011.585993]

46. Kamil RJ, Lin FR. The effects of hearing impairment in older adults on communication partners: a systematic review. J Am Acad Audiol 2015 Feb;26(2):155-182. [doi: 10.3766/jaaa.26.2.6] [Medline: 25690776]

47. Scarinci N, Worrall L, Hickson L. Factors associated with third-party disability in spouses of older people with hearing impairment. Ear Hear 2012;33(6):698-708. [doi: 10.1097/AUD.0b013e31825aab39] [Medline: 22710661]

48. Ekberg K, Meyer C, Scarinci N, Grenness C, Hickson L. Family member involvement in audiology appointments with older people with hearing impairment. Int J Audiol 2015 Feb;54(2):70-76. [doi: 10.3109/14992027.2014.948218] [Medline: 25141941]

49. Stark P, Hickson L. Outcomes of hearing aid fitting for older people with hearing impairment and their significant others. Int J Audiol 2004;43(7):390-398. [doi: 10.1080/14992020400050050] [Medline: 15515638]

50. Hearing loss in adults: assessment and management. National Institute for HealthCare Excellence. 2018 Jun 21. URL: https://www.nice.org.uk/guidance/ng98 [accessed 2020-01-09]

51. 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]

52. Chang C, Yan C, Tseng J. Perceived convenience in an extended technology acceptance model: mobile technology and English learning for college students. Australas J Educ Technol 2012 Jul 26;28(5):809-826. [doi: 10.14742/ajet.818]

53. Bennett GG, Glasgow RE. The delivery of public health interventions via the internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]

54. Zhang D, Zhou L, Briggs RO, Nunamaker JF. Instructional video in e-learning: assessing the impact of interactive video on learning effectiveness. Inf Manag 2006 Jan;43(1):15-27. [doi: 10.1016/j.im.2005.01.004]

55. Evans C, Gibbons NJ. The interactivity effect in multimedia learning. Comput Educ 2007 Dec;49(4):1147-1160. [doi: 10.1016/j.compedu.2006.01.008]

56. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191]

57. Jennings MB, Cheesman MF, Laplante-Lévesque A. Psychometric properties of the self-efficacy for situational communication management questionnaire (SESMQ). Ear Hear 2014;35(2):221-229. [doi: 10.1097/01.aud.0000441081.64281.b9] [Medline: 24556967]

58. West RL, Smith SL. Development of a hearing aid self-efficacy questionnaire. Int J Audiol 2007 Dec;46(12):759-771. [doi: 10.1080/14992020701545898] [Medline: 18049965]

59. Smith SL, Pichora-Fuller KM, Watts KL, la More C. Development of the listening self-efficacy questionnaire (LSEQ). Int J Audiol 2011 Jun;50(6):417-425. [doi: 10.3109/14992027.2011.553205] [Medline: 21470067]

60. Hickson L, Meyer C, Lovelock K, Lampert M, Khan A. Factors associated with success with hearing aids in older adults. Int J Audiol 2014 Feb;53(Suppl 1):S18-S27. [doi: 10.3109/14992027.2013.860488] [Medline: 24447233]

61. Meyer C, Hickson L, Lovelock K, Lampert M, Khan A. An investigation of factors that influence help-seeking for hearing impairment in older adults. Int J Audiol 2014 Feb;53(Suppl 1):S3-17. [doi: 10.3109/14992027.2013.839888] [Medline: 24405446]

62. Meyer C, Hickson L, Fletcher A. Identifying the barriers and facilitators to optimal hearing aid self-efficacy. Int J Audiol 2014 Feb;53(Suppl 1):S28-S37. [doi: 10.3109/14992027.2013.832420] [Medline: 24447235]

63. Ferguson MA, Woolley A, Munro KJ. The impact of self-efficacy, expectations, and readiness on hearing aid outcomes. Int J Audiol 2016 Jul;55(Suppl 3):S34-S41. [doi: 10.1080/14992027.2016.1177214] [Medline: 27266542]

64. Laplante-Lévesque A, Hickson L, Worrall L. Stages of change in adults with acquired hearing impairment seeking help for the first time: application of the transtheoretical model in audiologic rehabilitation. Ear Hear 2013;34(4):447-457. [doi: 10.1097/AUD.0b013e3182772c49] [Medline: 23364333]

65. Saunders GH, Frederick MT, Silverman S, Papesh M. Application of the health belief model: development of the hearing beliefs questionnaire (HBQ) and its associations with hearing health behaviors. Int J Audiol 2013 Aug;52(8):558-567. [doi: 10.3109/14992027.2013.791030] [Medline: 23682849]

66. Meister H, Grugel L, Meis M. Intention to use hearing aids: a survey based on the theory of planned behavior. Patient Prefer Adherence 2014;8:1265-1275 [FREE Full text] [doi: 10.2147/PPA.S65008] [Medline: 25258520]

67. Barker F, Atkins L, de Lusignan S. Applying the COM-B behaviour model and behaviour change wheel to develop an intervention to improve hearing-aid use in adult auditory rehabilitation. Int J Audiol 2016 Jul;55(Suppl 3):S90-S98. [doi: 10.3109/14992027.2015.1120894] [Medline: 27420547]

68. Ferguson MA, Coulson NS, Henshaw H, Heffernan E. Application of health behaviour theory to hearing healthcare research: the state of play and beyond. Int J Audiol 2016 Jul;55(Suppl 3):S1-S2. [doi: 10.1080/14992027.2016.1206215] [Medline: 27420545]

69. Maidment DW, Ali YH, Ferguson MA. Applying the COM-B model to assess the usability of smartphone-connected listening devices in adults with hearing loss. J Am Acad Audiol 2019 May;30(5):417-430. [doi: 10.3766/jaaa.18061] [Medline: 31044693]

70. Pulvirenti M, McMillan J, Lawn S. Empowerment, patient centred care and self-management. Health Expect 2014 Jun;17(3):303-310 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00757.x] [Medline: 22212306]

71. Small N, Bower P, Chew-Graham CA, Whalley D, Protheroe J. Patient empowerment in long-term conditions: development and preliminary testing of a new measure. BMC Health Serv Res 2013 Jul 8;13:263 [FREE Full text] [doi: 10.1186/1472-6963-13-263] [Medline: 23835131]

Abbreviations

BRC: Biomedical Research Centre
COM-B: Capability, Opportunity, Motivation-Behavior
mRLO: mobile-enhanced reusable learning object
NHS: National Health Service
NIHR: National Institute for Health Research
RCT: randomized controlled trial
RLO: reusable learning object
TDF: theoretical domains framework

XSL·FO
RenderX

Original Paper

Influence of Personality on mHealth Use in Patients with Diabetes: Prospective Pilot Study

Jingyuan Su[1], MSc; Michelle Dugas[2], PhD; Xitong Guo[1], PhD; Guodong (Gordon) Gao[2], PhD

[1]eHealth Research Institute, School of Management, Harbin Institute of Technology, Harbin, China

[2]Center for Health Information & Decision Systems, Department of Decision, Operations, and Information Technologies, Robert H Smith School of Business, University of Maryland, College Park, MD, United States

Corresponding Author:
Xitong Guo, PhD
eHealth Research Institute
School of Management
Harbin Institute of Technology
92 West Dazhi Street
Nangang District
Harbin,
China
Phone: 86 451 86414022
Email: xitongguo@hit.edu.cn

Abstract

Background: Mobile technology for health (mHealth) interventions are increasingly being used to help improve self-management among patients with diabetes; however, these interventions have not been adopted by a large number of patients and often have high dropout rates. Patient personality characteristics may play a critical role in app adoption and active utilization, but few studies have focused on addressing this question.

Objective: This study aims to address a gap in understanding of the relationship between personality traits and mHealth treatment for patients with diabetes. We tested the role of the five-factor model of personality traits (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) in mHealth adoption preference and active utilization.

Methods: We developed an mHealth app (DiaSocial) aimed to encourage diabetes self-management. We recruited 98 patients with diabetes—each patient freely chose whether to receive the standard care or the mHealth app intervention. Patient demographic information and patient personality characteristics were assessed at baseline. App usage data were collected to measure user utilization of the app. Patient health outcomes were assessed with lab measures of glycated hemoglobin (HbA_{1c} level). Logistic regression models and linear regression were employed to explore factors predicting the relationship between mHealth use (adoption and active utilization) and changes in health outcome.

Results: Of 98 study participants, 46 (47%) downloaded and used the app. Relatively younger patients with diabetes were 9% more likely to try and use the app ($P=.02$, odds ratio [OR] 0.91, 95% CI 0.85-0.98) than older patients with diabetes were. Extraversion was negatively associated with adoption of the mHealth app ($P=.04$, OR 0.71, 95% CI 0.51-0.98), and openness to experience was positively associated with adoption of the app ($P=.03$, OR 1.73, 95% CI 1.07-2.80). Gender ($P=.43$, OR 0.66, 95% CI 0.23-1.88), education (senior: $P=.99$, OR 1.00, 95% CI 0.32-3.11; higher: $P=.21$, OR 2.51, 95% CI 0.59-10.66), and baseline HbA_{1c} level ($P=.36$, OR 0.79, 95% CI 0.47-1.31) were not associated with app adoption. Among those who adopted the app, a low education level (senior versus primary $P=.003$; higher versus primary $P=.03$) and a high level of openness to experience ($P=.048$, OR 2.01, 95% CI 1.01-4.00) were associated with active app utilization. Active users showed a significantly greater decrease in HbA_{1c} level than other users ($\Delta HbA_{1c}=-0.64$, $P=.05$).

Conclusions: This is one of the first studies to investigate how different personality traits influence the adoption and active utilization of an mHealth app among patients with diabetes. The research findings suggest that personality is a factor that should be considered when trying to identify patients who would benefit the most from apps for diabetes management.

Introduction

Diabetes Self-Management

Globally, type 2 diabetes is a common chronic disease, and its incidence is rapidly increasing in countries such as China [1,2]. Treatment for type 2 diabetes is largely self-managed; patients are responsible for engaging in health-promoting behavior on a day-to-day basis [3]. Health-promoting behaviors include dietary control, physical activity, and blood glucose monitoring, and these health-promoting behaviors are often incorporated as essential components of treatment programs in order to keep blood glucose levels within target ranges and to prevent long-term complications [4]. Although long-term self-management and lifestyle behaviors are critical for controlling diabetes, these skills prove difficult for many patients to develop [3].

Role of mHealth

Mobile technology for health (mHealth) interventions can benefit chronic disease management by delivering real-time monitoring and reminders to a large number of people, enabling the delivery of tailored support and providing low-cost, remote health care services [5]. For diabetes, mHealth interventions are increasingly being used to assist patients with lifestyle changes and health-promoting behaviors [6]. Some recent studies and reviews [7-9] have shown that mHealth smartphone interventions for diabetes self-management have reduced glycated hemoglobin (HbA_{1c}) levels and have significantly facilitated self-management for patients; however, the effectiveness of mHealth in improving diabetes outcomes critically depends on voluntary patient engagement. Most studies, in contrast, rely upon randomized trials where patients are exogenously assigned to the mHealth interventions [10]. In many studies that are designed as such, the patients who are required to use the mHealth interventions showed low utilization of the health apps and often had high dropout rates [5,11]. Therefore, within the literature to date, there is a limited understanding of how willing diabetic patients are to adopt mHealth interventions. Within this question, there is also the matter of individual differences that may make patients more or less likely to engage with a given intervention. In order to make use of the full potential of mHealth, researchers, technology companies, and clinicians have been exploring ways of attracting users and increasing mHealth use [12], with the understanding that personality characteristics may be an important consideration when building or recommending different mHealth products [13]. Based on this, we drew on insights from personality research to explore individual differences that could explain heterogeneity in the adoption and active utilization of mHealth to improve diabetes self-management.

Personality Traits and mHealth Intervention

The five-factor model of personality (also referred to as the big five personality traits) offers a comprehensive framework for examining distinct personal characteristics and their influences. Within the big five, *agreeableness* encompasses the traits of courtesy, cooperation, trust, and tolerance. *Conscientiousness* represents tendencies such as being self-disciplined, organized, and persistent in goal-directed behavior. *Extraversion* is frequently associated with being sociable, gregarious, and optimistic. *Neuroticism* (or emotional instability) is characterized by insecurity, anxiousness, and hostility. *Openness to experience* represents one's curiosity and willingness to explore new ideas [14]. Taken together, the big five capture the essence of one's personality.

Past research has recognized that the personal characteristics of users are essential factors in predicting technology adoption and continued utilization of different kinds of apps [15-17]. For example, previous studies have found that more conscientious individuals may be less likely to use leisure apps, but would prefer communication and business apps [18,19]. Openness to experience can be used to predict greater smartphone use and may be positively correlated with the use of social apps [16,20]. Extraverts are more likely to use social networking and instant messaging apps than they are to use apps related to books, references, and education [20]. Individuals with high agreeableness are less likely to use apps related to communication, browser usage, productivity, and gaming [18,19]. Neurotic individuals are likely to adopt apps in general, due to their fastidious and meticulous nature as well as their interest in creative activities [19].

Previous studies [21-25] have also examined the effects of the five personality traits on patient intent to use mHealth apps. Openness to experience had a positive influence on acceptance of an mHealth app for hypertension [21] and was correlated with better adherence to a self-care app for cancer [22]. High conscientiousness was significantly associated with the total number of points earned on an mHealth weight loss app [23]. Gender played a moderating role in the relationship between two specific personality traits, extraversion and emotional stability, and in the behavioral intention to use mHealth apps, in general [24]. Personality traits have been found to influence self-care behavior and glycemic control in patients with type 2 diabetes [25]; however, little is known about how the big five personality traits influence adoption preference and active utilization of diabetic self-management apps. Our exploratory study was designed to help fill this knowledge gap.

Study Objectives

Using data from a 3-month period, this study explored three relationships: (1) the relationship between the big five personality traits and mHealth adoption, (2) the relationship between the big five personality traits and active utilization of the app, and (3) the effect of mHealth app usage on health outcomes in patients with diabetes. Findings related to the association between personal characteristics and mHealth use in the context of diabetes self-care will contribute to further understanding of the mechanisms underlying high dropout rates

and could inform the development of tailored interventions that will engage and improve patient self-management.

Methods

Ethical Considerations

The pilot study was approved by the ethics committees at the University of Maryland (1331093-1) and Harbin Institute of Technology School of Management. Informed written consent was obtained from each individual prior to their participation, and patients were provided with complete assurance that all information would be kept confidential. Participation was completely voluntary.

Recruitment

Patients were recruited through their affiliation with the Endocrinology Department at the Fifth Hospital, Daqing City, China. Daqing has played a key role in diabetic research, as it was the setting for the first batch of studies on diabetes in China, which were well-recognized in journals such as the Lancet [26]. Inclusion criteria for this study were a diagnosis of type 2 diabetes and more than 1 year since that diagnosis. Exclusion criteria were individuals who were blind, deaf, had a diagnosis of serious mental illness (active psychosis, bipolar disorder, schizophrenia, borderline personality disorder, active alcohol addiction, or other), and who lacked a stable residence.

Procedure

Eligible patients were identified from electronic medical records and were then notified of the time, address, and content of an upcoming orientation session by phone call.

Patients who were interested in participating attended an orientation session, which included an educational presentation on diabetes by a physician and a detailed description of the study by the research team. In the latter portion of the session, staff explained that patients would choose which treatment they wanted to receive—either the app treatment or standard care—if they chose to participate. The staff then presented the app to explain its functions and why and how it could help the patients. Patients were informed that all participants in both groups would receive the same perks including free HbA_{1c} tests, a portioned dinner plate as a gift, an educational handbook about diabetes self-management, and physician guidance for their health management. They were also informed that, as an additional incentive, those in either group who decreased their HbA_{1c} level during the 3-month period would receive a gift worth 50 RMB (approximately US$ 7). At the end of the orientation session, the prospective participants were entirely free to choose which group to join, and if they chose to participate, they signed a statement to indicate that they were providing informed consent.

Participants attended a second session on the morning following the first session. In the second session, all participants underwent baseline measurement, including HbA_{1c} laboratory tests, and BMI measurements, and completed a questionnaire. Then, participants who chose the mHealth intervention were sent instructions on how to use the app.

Over the course of 1 month (November 2018), we conducted seven orientation sessions and recruited 98 patients. Participant exclusions and categories are summarized in Figure 1. The study duration was 6 months, but the analyses in this paper reflect midpoint findings (3 months).

Figure 1. CONSORT flow diagram.
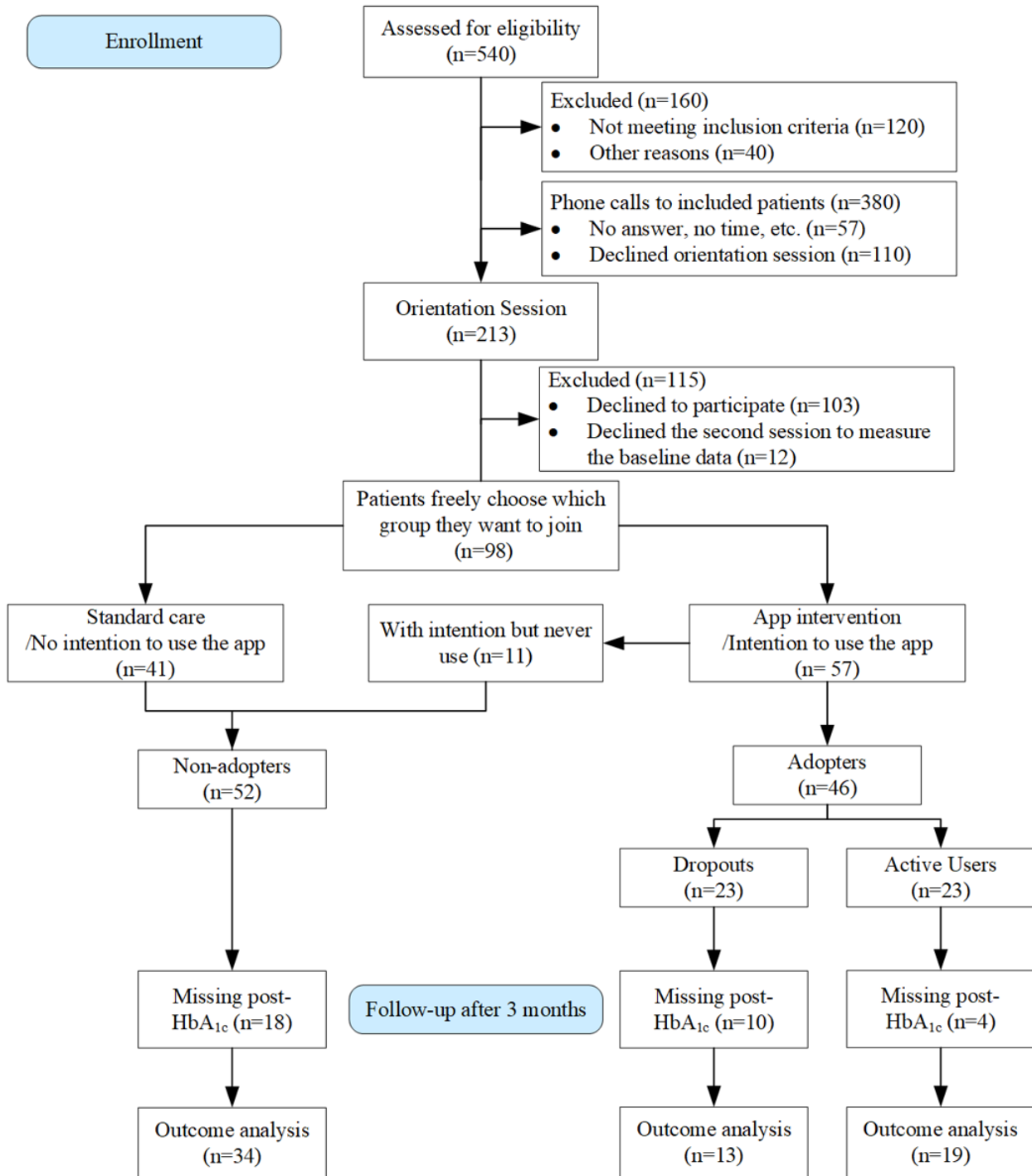
mHealth Intervention

Participants who indicated intention to use the app were given access to the health management app (DiaSocial) designed by our research team; screen captured images from the app are shown in Figure 2. The app was available in both iPhone and Android versions. App components consisted of educational resources, tracking features, and feedback. Participants had continuous access to the app throughout the intervention period.

Educational information included basic self-management strategies as well as guidance for app use. Participants who used the app were instructed to use it daily and record their progress towards diabetes self-care goals such as exercise, a healthy diet, managing glucose levels, and medication adherence. Physical activity was objectively measured using the participant's smartphone pedometer or was manually entered by participants into the app. Other records were manually entered.

XSL•FO
RenderX

Figure 2. Screen capture images from the DiaSocial app.

The app provided feedback through graphical displays of logged behavior. Two types of graphical feedback were provided. The first comprised separate graphs of glucose levels, sleep, medication adherence, diet, weight, and moderate to vigorous intensity physical activity. An information breakdown was provided in the form of daily, weekly, and monthly line charts. The second mode of graphical feedback was a red arrow displayed on the analysis graphs designed to alert participants (1) when blood glucose level prior to meals (breakfast, lunch, or dinner) was greater than 7 mmol/L, (2) when blood glucose level after meals (breakfast, lunch, or dinner) was greater than 10 mmol/L, (3) when BMI was greater than 25 kg/m^2.

The app was also gamified such that participants earned scores within the app for use and for reaching target levels of glucose, exercise, nutrition, and medication adherence. The app ranked all the participants according to their daily total scores. At the same time, app users were also assigned to different treatment arms, such as team competition and individual competition, but these experimental conditions did not work out as planned and all analyses were collapsed across the various treatment arms.

HbA$_{1c}$ Laboratory Test

Glycated hemoglobin (HbA$_{1c}$ level) was chosen as our key clinical outcome because higher HbA$_{1c}$ levels have been associated with more complications and poorer health outcomes [27]. HbA$_{1c}$ level provides a picture of average blood glucose level over a period of months. Patients had blood drawn within 7 days of the study's launch to assess baseline HbA$_{1c}$ level and 90 days later to determine postintervention HbA$_{1c}$ (post HbA$_{1c}$) levels. Given the 1-month recruitment span, there was variability in the dates on which HbA$_{1c}$ levels were measured. HbA$_{1c}$ values of participants were measured at Daqing Hospital as part of the study, and researchers collected these lab values for analysis.

Questionnaire Design

The baseline questionnaire was prepared in two languages, English and Chinese, using a semantic translation technique. The content validity of this instrument was assessed by an endocrinologist and also by type 2 diabetic patients to ensure that the questions were appropriate and relevant in the current Chinese setting and culture. The questionnaire items included demographic information (age, gender, education, BMI, and time since diagnosis), measures of health behaviors and attitudes, and a personality inventory.

Personality was assessed with the previously developed Chinese version of the Ten-Item Personality Inventory [28], a widely used, brief measure of the big five personality traits; it has two items per dimension, with each item consisting of a pair of adjectives. Half of the items represent the positive pole of the given dimension and the other half represent the negative pole. Participants rated each item on a 7-point Likert scale, ranging from 1 (disagree strongly) to 7 (agree strongly). A total score for each dimension was generated by transforming the scores of the negative pole items (reverse-scored items), then computing the averaged dimension score. Higher scores indicated a higher level of that personality dimension.

Statistical Analysis

All analyses were performed with Stata MP (version 15.1; StataCorp LLC) software. Given our research questions, we had three outcome variables of interest: adoption preference (nonadopters=0, adopters=1), app active utilization (dropouts=0, active users=1), and health outcome (ΔHbA$_{1c}$= post HbA$_{1c}$ level – baseline HbA$_{1c}$ level).

We conducted binary logistic regression analyses to determine the association of personality traits with mHealth adoption preference and active utilization. To address our first research question, all patients who agreed to participate in the study were included. To address the second question, the subset of patients who actually used the mHealth app were used in the analysis. An analysis of variance was used for numerical variables and chi-square tests were used for categorical variables to compare the differences between the three types of participants after three months: active users, dropouts, and nonadopters. We also conducted multiple linear regression to explore the relationship between the amount of app usage (number of days per week) and ΔHbA$_{1c}$ for adopters. Analyses included several covariates as well as the five personality traits of interest.

Results

Participant Classification and Summary

Based on app usage (average days per week) for each participant (see Multimedia Appendix 1), we defined participant categories as shown in Table 1. Of the patients who agreed to participate (N=98), 58% (57/98) initially indicated their willingness to download and use the app; however, 11 patients never used the app. Nonadopters, therefore, included patients who were not willing to use the app and patients who never used the app even if they initially indicated a willingness to do so. Adopters included dropouts (those who initially used the app, but then became inactive later on) and active users. As defined in Table 1, if an app user did not continue to use the app past the first two months, we regarded that user as a dropout. Users who kept using the app until the third month, were defined as active users. Active users were further divided into 2 groups: low-frequency users (whose weekly average app usage was less than 3.5 days) and high-frequency users (whose weekly average app usage was greater than 3.5 days). Detailed demographics for participant subcategories are shown in Multimedia Appendix 2.

Table 1. Definitions of participant categories.

Categories and subcategories	n	Definition
Participants	**98**	**All patients willing to participate in the project.**
Intention to use[a]	57	All patients who indicated willingness to use the app at the beginning of the project.
Nonadopters	**52**	
No intention to use	41	Patients who were not willing to use the app at the beginning of the project.
With intention but never used	11	Patients who indicated willingness to use the app, but then did not use it at all.
Adopters	**46**	
Dropouts	23	Patients who used the app, but did not continue use the app in the third month.
Active users	23	Patients who continued to use the app into the third month.
Low-frequency	14	Patients whose average weekly app usage was less than 3.5 days
High-frequency	9	Patients whose average weekly app usage was more than 3.5 days

[a]Includes *Nonadopters: With intention but never used*, and *Adopters*.

App Adoption Preference Among Diabetic Patients and Association With Personality Traits

Table 2 presented the results of the binary logistic regression analyses associating app adoption with sociodemographic and personality traits among diabetic patients. Relatively younger patients with diabetes (mean 56.07, SD 9.06 years) had 9% higher odds of trying the app (P=.02, odds ratio [OR] 0.91, 95% CI 0.85-0.98) compared to the odds of older patients with diabetes (mean 62.04, SD 8.04 years). Diabetic patients who were less extraverted had 29% higher odds of trying the app (P=.04, OR 0.71, 95% CI 0.51-0.98) compared to the odds of those who were more extraverted. At the same time, diabetic patients who were more open to experience had 73% higher odds of adopting use of the app (P=.03, OR 1.73, 95% CI 1.07-2.80) compared to the odds of those who were less openness. Gender (P=.43, OR 0.66, 95% CI 0.23-1.88), education (senior: P=.99, OR 1.00, 95% CI 0.32-3.11; higher: P=.21, OR 2.51, 95% CI 0.59-10.66), and baseline HbA_{1c} level (P=.36, OR 0.79, 95% CI 0.47-1.31) were not associated with app adoption. Ordered logistic regression for all participants, which had 5 ordered levels ranging from no intention to use to high-frequency users, also provided similar results and can be found in Multimedia Appendix 3.

Table 2. Logistic regression for personality traits associated with adoption preference.

Variables	Participants (N=98)		Logistic regression model			
	Nonadopters (n=52)	Adopters (n=46)	β	OR[a]	95% CI	*P* value
Age (years), mean (SD)	62.04 (8.04)	56.07 (9.06)	–0.09	0.91	0.85-0.98	.02
Gender, n (%)						
Female	23 (44)	14 (30)	–0.42	0.66	0.23-1.88	.43
Male	29 (56)	32 (70)	N/A[b]	N/A	N/A	N/A
Education, n (%)						
Primary/junior	23 (44)	10 (22)	N/A	N/A	N/A	N/A
Senior/vocational	24 (46)	19 (41)	0.001	1.00	0.32-3.11	.99
Higher/university	5 (10)	17 (37)	0.92	2.51	0.59-10.66	.21
BMI (kg/m^2), mean (SD)	25.95 (3.15)	25.64 (3.27)	–0.13	0.87	0.74-1.04	.13
Time since diagnosis (years), mean (SD)	9.52 (8.72)	9.02 (7.02)	0.04	1.04	0.97-1.11	.30
Baseline HbA$_{1c}$ (%), mean (SD)	7.17 (1.05)	7.14 (0.96)	–0.24	0.79	0.47-1.31	.36
Personality traits, mean (SD)						
Extraversion	4.97 (1.52)	4.01 (1.81)	–0.34	0.71	0.51-0.98	.04
Agreeableness	5.14 (1.12)	5.23 (1.15)	–0.10	0.91	0.55-1.50	.70
Conscientiousness	4.42 (1.19)	4.86 (1.32)	0.11	1.11	0.74-1.67	.61
Emotional stability	4.29 (1.41)	4.51 (1.52)	0.18	1.20	0.82-1.77	.35
Openness	3.99 (1.06)	4.72 (1.30)	0.55	1.73	1.07-2.80	.03
Constant			8.51			.08
Chi-square (*df*)			34.1 (12)			<.001

[a]OR: odds ratio.

[b]N/A: not applicable.

Relationship Between Personality Traits and App Usage Among Diabetic Patients

Table 3 presented the relationship between personality traits and app usage. The results showed that a lower education level—specifically, participants who completed primary school compared to those who had a vocational (senior) or university (higher) education—was associated with greater odds of active app utilization (senior versus primary *P*=.003 and higher versus primary *P*=.03). Higher openness (*P*=.048, OR 2.01, 95% CI 1.01-4.00) was also significantly associated with greater active utilization. We found similar results by using the amount of app usage (number of days per week) as an outcome variable to conduct ordinary least squares linear regression in Multimedia Appendix 4, and by ordered logistic regression in Multimedia Appendix 5.

Table 3. Logistic regression for personality traits associated with active utilization.

Variables	Adopters (n=46)		Logistic regression model			
	Dropouts (n=23)	Active users (n=23)	β	OR[a]	95% CI	*P* value
Age (years), mean (SD)	54.96 (8.76)	57.17 (9.80)	0.04	1.04	0.93-1.17	.45
Gender, n (%)						
Female	7 (30)	7 (30)	0.34	1.40	0.17-11.41	.75
Male	16 (70)	16 (70)	N/A[b]	N/A	N/A	N/A
Education, n (%)						
Primary/junior	1 (4)	9 (39)	N/A	N/A	N/A	N/A
Senior/vocational	14 (61)	5 (22)	–4.43	0.01	0.001-0.22	.003
Higher/university	8 (35)	9 (39)	–3.14	0.04	0.003-0.60	.03
BMI (kg/m^2), mean (SD)	25.96 (3.08)	25.31 (3.49)	–0.12	0.88	0.66-1.19	.41
Time since diagnosis (years), mean (SD)	8.65 (7.07)	9.39 (7.11)	0.04	1.04	0.91-1.18	.57
Baseline HbA$_{1c}$ (%), mean (SD)	7.03 (0.97)	7.25 (0.96)	0.27	1.31	0.50-3.40	.58
Personality traits, mean (SD)						
Extraversion	4.33 (1.90)	3.70 (1.70)	–0.37	0.69	0.38-1.27	.23
Agreeableness	5.20 (1.04)	5.26 (1.28)	0.56	1.75	0.68-4.52	.25
Conscientiousness	4.78 (1.41)	4.94 (1.26)	0.06	1.05	0.46-2.40	.92
Emotional stability	4.74 (1.44)	4.28 (1.59)	–0.17	0.85	0.44-1.63	.62
Openness	4.46 (1.16)	4.98 (1.41)	0.70	2.01	1.01-4.00	.048
Constant			–2.71			.77
Chi-square (*df*)			23.5 (12)			.02

[a]OR: odds ratio.

[b]N/A: not applicable.

mHealth App Usage and Patient Health Outcomes

Because some patients did not have their final HbA$_{1c}$ level measured, Table 4 reports descriptive statistics and differences among categories for retained participants after 3 months. Active users showed a greater decrease in HbA$_{1c}$ level (ΔHbA$_{1c}$=–0.64, *P*=.05) than those shown by users in the other groups. We also observed an unexpected decrease in HbA$_{1c}$ level for nonadopters and increase in HbA$_{1c}$ level for dropouts. We observe similar connections between extraversion and openness and health outcomes as we did for overall app adoption—less extraverted participants and those who were more open to experience were more likely to experience a decrease in HbA$_{1c}$ level.

Next, we used linear regression to test the relationship between the amount of app usage (number of days per week) and ΔHbA$_{1c}$ among adopters (Table 5). We found that if app users increased their usage by ten days, their HbA$_{1c}$ level decreased 0.20 points (β=–0.02, *P*=.02) under controlled covariables. More detailed comparisons among all categories are shown in Multimedia Appendix 6.

XSL•FO

RenderX

Table 4. Descriptive statistics after 3 months and the differences among categories (N=66).

Variables	Nonadopters (n=34)	Dropouts (n=13)	Active users (n=19)	$F_{2,63}$[a] or chi-square (df)[b]	P value
Age (years), mean (SD)	63.4 (7.96)	53.23 (9.33)	58.11 (9.98)	6.8[a]	.002
Gender, n (%)				**0.7 (2)[b]**	**.70**
Female	14 (41)	4 (31)	6 (32)		
Male	20 (59)	9 (69)	13 (68)		
Education, n (%)				**17.1 (4)[b]**	**.002**
Primary/junior	15 (44)	1 (8)	6 (32)		
Senior/vocational	17 (50)	6 (46)	4 (21)		
Higher/university	2 (6)	6 (46)	9 (47)		
BMI (kg/m^2), mean (SD)	26.21 (3.30)	26.72 (3.47)	25.06 (3.68)	1.1[a]	.36
Time since diagnosis (years), mean (SD)	9.32 (8.19)	8.08 (7.71)	8.90 (6.95)	0.1[a]	.89
Personality traits, mean (SD)					
Extraversion	4.47 (1.41)	4.81 (1.38)	3.66 (1.67)	3.8[a]	.03
Agreeableness	5.03 (1.67)	5.00 (0.89)	5.39 (1.21)	0.8[a]	.48
Conscientiousness	4.53 (1.22)	4.54 (1.41)	4.89 (1.22)	0.6[a]	.57
Emotional stability	4.26 (1.49)	4.62 (1.49)	4.36 (1.67)	0.2[a]	.78
Openness	4.04 (1.13)	4.69 (0.88)	5.03 (1.43)	4.5[a]	.01
Baseline HbA$_{1c}$ (%), mean (SD)	7.22 (1.07)	6.93 (0.89)	7.14 (0.89)	0.5[a]	.64
Post HbA$_{1c}$ (%), mean (SD)	6.97 (1.22)	7.01 (0.81)	6.49 (0.95)	1.7[a]	.18
ΔHbA$_{1c}$ (%), mean (SD)	–0.25 (0.82)	0.08 (0.78)	–0.64 (0.86)	3.1[a]	.050

[a]F test was used for numerical variables.

[b]Chi-square was used for categorical variables.

Table 5. Model estimates predicting ΔHbA_{1c} for adopters (n=32).

Variables	β	SE	t test	P value
Age	0.03	0.02	1.69	.11
Female vs male	–0.68	0.47	–1.44	.17
Education				
Senior vs Primary	0.52	0.49	1.07	.30
Higher vs Primary	0.57	0.46	1.24	.23
BMI	0.03	0.05	0.65	.52
Time since diagnosis	–0.01	0.02	–0.31	.76
Baseline HbA_{1c}	–0.32	0.17	–1.87	.08
Personality traits				
Extraversion	–0.02	0.12	–0.17	.87
Agreeableness	0.09	0.20	0.44	.66
Conscientiousness	–0.23	0.16	–1.46	.16
Emotional stability	–0.10	0.12	–0.86	.40
Openness	0.23	0.14	1.62	.12
Days of app usage	–0.02	0.01	–2.58	.02
Constant	–0.24	3.20	–0.07	.94
Active utilization	0.65			
$F_{13,18}$	2.5			.03

Discussion

Principal Results

This study suggested a relationship between individual characteristics and mHealth app use (adoption and active utilization) in patients with diabetes. We found that patients with diabetes who were relatively younger, less extraverted, and more open to experience were more likely to adopt use of the mHealth self-management app. In addition, education level and openness to experience were associated with active utilization of the app. Finally, active users were also associated with better clinical outcomes than those of dropouts.

Comparison With Prior Work

Our finding that relatively younger patients were more likely than older patients to try to adopt the app was in agreement with previous studies [29,30] which found that older adults were less likely to adopt new technology. Older people reported that they do not go online for various reasons including cost, lack of skills, lack of interest, and concerns about information security; however, once older adults did get online, they were just as enthusiastic as younger users [31]. This was also consistent with our own findings, as age was negatively associated with initial adoption, but not with actual use once adopted. In contrast to the outcomes suggested by prior research [32,33], we unexpectedly found that gender, education, and baseline HbA_{1c} were not associated with app adoption.

We also found that mHealth app users with lower education levels may have been less likely to drop out. Low education levels have been associated with inadequate health literacy

[34,35], which, in turn, has been associated with lower diabetes-related knowledge and lower engagement in mobile- and web-delivered self-care interventions [36]. This finding may therefore seem surprising; however, Paasche-Orlow et al [37] argued that patients with low health literacy may have difficulty acquiring self-management skills, but once these skills are acquired, they may follow directions more readily than those with higher literacy [37]. In line with this reasoning, our mHealth app provided diabetes self-management educational resources, which may have enabled patients with low education to learn more about diabetes management skills and may have encouraged app usage more than it did for those who already possessed diabetes self-management knowledge [38].

Our results also suggested that extraverted patients with diabetes were less likely to adopt the app, despite the fact that it incorporated and emphasized social features. This may possibly be explained by the fact that the study focused on motivating long-term and continuous self-management, which may not fulfill the social desires of extraverts. Compared to extraverts who prefer to meet friends or participate in social activities to get health support, introverted people may be more likely to use the mHealth app instead of a social network app for this purpose [19]. In fact, a previous study [39] found a similar pattern in which higher extraversion in students was associated with a preference for face-to-face mental health services over eHealth services.

The positive association between openness to experience and patient engagement in mHealth was consistent with the results of previous research [21] on technology adoption. High openness to experience is reflected in curiosity and novelty-seeking—open

individuals are willing to try new and different things, are willing to actively seek out new and varied experiences, and value change [15]. At the same time, some research has also found that openness plays an important role in promoting healthy behavior and lowering mortality for patients with chronic disease [40,41]. While mechanisms underlying these findings are not well understood, they are consistent with the notion that openness should be associated with the adoption of novel self-management tools [21] and associated with better health behavior adherence [22]. This means that people more open to experiences may be more likely to adopt and actively use an mHealth app for diabetes management.

The analysis of our findings showed that conscientiousness, agreeableness, and emotional stability did not have a significant relationship with acceptance of the mHealth app. Of particular note, conscientiousness, which reflects self-discipline, did not have a significant role in active app use, which was inconsistent with results of a previous study focusing on a weight loss app [23]. This unexpected finding underscored the need for future research, but differences could be attributed to different population characteristics. Perhaps, our study's sample of older participants had less prior experience with mHealth apps; therefore, open-mindedness was an especially important predictor of app adoption and active use. There have also been studies [42] suggesting that conscientious patients were more likely to fulfill tasks in accordance with existing strategies, whereas using new technologies could be viewed as more time-consuming and complex than the traditional methods they already use.

Lastly, our pilot study offered support that mHealth improved clinical outcomes of patients with diabetes. This reduction was on the magnitude of a 0.6-point reduction in HbA_{1c} level, which was sizable considering the relatively short intervention period; however, we also noticed that the HbA_{1c} level of participants who dropped out increased while the HbA_{1c} level of nonadopters who never used the app decreased, even though there was no significant difference between the two groups. This may be related to other differences between dropouts and nonadopters; patients with good daily self-management behavior may be less interested in adopting the mHealth app. Meanwhile, patients who had poor self-management behavior may have started using the app to improve their health, yet may not have continuously used the app or followed its guidance, and their HbA_{1c} level may have increased or remained constant. Because this study was not randomized, readers are cautioned against interpreting the results as causal.

Implications

To the best of our knowledge, this is the first empirical paper to investigate how different personality traits influence the adoption and the active utilization of an mHealth app

intervention for patients with diabetes. Our research extends the literature on the adoption of mHealth apps and active utilization to improve diabetes management by exploring the role of personality traits. This work has implications for app designers and practitioners who can leverage this knowledge to target individuals who are most likely to succeed. For instance, patients who are less extraverted and more open to experience may find an app-based intervention the most appealing and effective. With the development of machine learning models, then, an app designer may be able to predict mHealth app user personality traits using easily accessible data and in real time [43,44]. Consequently, designers and practitioners can create or administer personalized mHealth services in a way that enhances patient engagement with helpful apps and which ultimately improves their lifestyle and health.

Limitations

Caution must be shown in generalizing the findings of this work because it has several limitations, but these limitations also provide opportunities for future research. First, the study used a nonrandomized experimental design in which only those who indicated their willingness to use the mHealth app were placed in the intervention group. While this allowed us to answer research questions about mHealth adoption preference, it made assessment of app efficacy more challenging. In future, a hybrid preference–randomized controlled trial [45] would enable examination of both questions. Second, although all participants, no matter which intervention they chose, were told that those who decrease their HbA_{1c} level during the 3-month period would receive a gift as an incentive, we cannot exclude the possibility that this incentive may have influenced the efficacy of the mHealth app intervention. Building on the encouraging findings from this pilot, future studies may use more rigorous research designs to address this potential impact. Third, as a pilot study, the sample was relatively small in size and came from one hospital in one city. Increasing and diversifying the sample size would provide more confidence that these results are generalizable.

Conclusions

Although there has been some research about how personality traits impact the use of new technology, there is relatively little that focuses on understanding the impact of individual characteristics, including personality traits, on mHealth app adoption and active utilization among diabetic patients. Our pilot study has made a strong start in addressing this gap by extending the mHealth literature. The study revealed that diabetic patients who are relatively young, introverted, and open to experience were interested in and willing to use the app. Moreover, active use of the app was associated with greater improvements in blood glucose level control. These research findings may have practical effects on the future development of mobile health apps for patients with diabetes.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The weekly days of app usage for each participant with the intention to use.
[DOCX File , 40 KB - mhealth_v8i8e17709_app1.docx]

Multimedia Appendix 2
Detailed descriptive statistics.
[DOCX File , 23 KB - mhealth_v8i8e17709_app2.docx]

Multimedia Appendix 3
Ordered logistic regression for all participants.
[DOCX File , 21 KB - mhealth_v8i8e17709_app3.docx]

Multimedia Appendix 4
Model estimates predicting the days of app usage.
[DOCX File , 20 KB - mhealth_v8i8e17709_app4.docx]

Multimedia Appendix 5
Ordered logistic regression for adopters.
[DOCX File , 20 KB - mhealth_v8i8e17709_app5.docx]

Multimedia Appendix 6
The impact of app usage in the changes in HbA_{1c}.
[DOCX File , 22 KB - mhealth_v8i8e17709_app6.docx]

References

1. Wang L, Gao P, Zhang M, Huang Z, Zhang D, Deng Q, et al. Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013. JAMA 2017 Jun 27;317(24):2515-2523 [FREE Full text] [doi: 10.1001/jama.2017.7596] [Medline: 28655017]
2. Jia W, Weng J, Zhu D, Ji L, Lu J, Zhou Z, Chinese Diabetes Society. Standards of medical care for type 2 diabetes in China 2019. Diabetes Metab Res Rev 2019 Sep;35(6):e3158. [doi: 10.1002/dmrr.3158] [Medline: 30908791]
3. Ahola AJ, Groop P. Barriers to self-management of diabetes. Diabet Med 2013 Apr;30(4):413-420. [doi: 10.1111/dme.12105] [Medline: 23278342]
4. Funnell MM, Anderson RM. Empowerment and self-management of diabetes. Clin Diabetes 2004 Jul 01;22(3):123-127. [doi: 10.2337/diaclin.22.3.123]
5. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
6. Quinn CC, Butler EC, Swasey KK, Shardell MD, Terrin MD, Barr EA, et al. Mobile diabetes intervention study of patient engagement and impact on blood glucose: mixed methods analysis. JMIR Mhealth Uhealth 2018 Feb 02;6(2):e31 [FREE Full text] [doi: 10.2196/mhealth.9265] [Medline: 29396389]
7. 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]
8. Fu H, McMahon SK, Gross CR, Adam TJ, Wyman JF. Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: a systematic review. Diabetes Res Clin Pract 2017 Sep;131:70-81. [doi: 10.1016/j.diabres.2017.06.016] [Medline: 28692830]
9. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016 Nov 18;11(11):e0166718 [FREE Full text] [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]

10. 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]

11. 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]

12. Patient Adoption of mHealth. In: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf. 100 IMS Drive, Parsippany, NJ 07054, USA: IMS Institute for Healthcare Informatics; 2015.

13. Dugas M, Crowley K, Gao GG, Xu T, Agarwal R, Kruglanski AW, et al. Individual differences in regulatory mode moderate the effectiveness of a pilot mHealth trial for diabetes management among older veterans. PLoS ONE 2018 Mar 7;13(3):e0192807 [FREE Full text] [doi: 10.1371/journal.pone.0192807] [Medline: 29513683]

14. Digman JM. Personality structure: emergence of the five-factor model. Annu Rev Psychol 1990 Jan;41(1):417-440 [FREE Full text] [doi: 10.1146/annurev.ps.41.020190.002221]

15. Devaraj S, Easley RF, Crant JM. How does personality matter? relating the five-factor model to technology acceptance and use. Information Systems Research 2008 Mar;19(1):93-105. [doi: 10.1287/isre.1070.0153]

16. Correa T, Hinsley AW, de Zúñiga HG. Who interacts on the web?: the intersection of users' personality and social media use. Comput Human Behav 2010 Mar;26(2):247-253. [doi: 10.1016/j.chb.2009.09.003]

17. McElroy, Hendrickson, Townsend, DeMarie. Dispositional factors in internet use: personality versus cognitive style. MIS Quarterly 2007;31(4):809-820. [doi: 10.2307/25148821]

18. Stachl C, Hilbert S, Au J, Buschek D, De Luca A, Bischl B, et al. Personality traits predict smartphone usage. Eur J Pers 2017 Aug 02;31(6):701-722 [FREE Full text] [doi: 10.1002/per.2113]

19. Xu R, Frey RM, Fleisch E, Ilic A. Understanding the impact of personality traits on mobile app adoption – insights from a large-scale field study. Computers in Human Behavior 2016 Sep;62:244-256. [doi: 10.1016/j.chb.2016.04.011]

20. Kim Y, Briley DA, Ocepek MG. Differential innovation of smartphone and application use by sociodemographics and personality. Comput Human Behav 2015 Mar;44:141-147. [doi: 10.1016/j.chb.2014.11.059]

21. Breil B, Kremer L, Hennemann S, Apolinário-Hagen J. Acceptance of mHealth apps for self-management among people with hypertension. Stud Health Technol Inform 2019 Sep 03;267:282-288. [doi: 10.3233/SHTI190839] [Medline: 31483283]

22. Mikolasek M, Witt CM, Barth J. Adherence to a mindfulness and relaxation self-care app for cancer patients: mixed-methods feasibility study. JMIR Mhealth Uhealth 2018 Dec 06;6(12):e11271 [FREE Full text] [doi: 10.2196/11271] [Medline: 30522990]

23. Hales S, Turner-McGrievy GM, Wilcox S, Davis RE, Fahim A, Huhns M, et al. Trading pounds for points: Engagement and weight loss in a mobile health intervention. Digit Health 2017 Apr 24;3:1-9 [FREE Full text] [doi: 10.1177/2055207617702252]

24. Nunes A, Limpo T, Castro S. Individual factors that influence the acceptance of mobile health apps: the role of age, gender, and personality traits. In: Bamidis PD, Ziefle M, Maciaszek L, editors. Information and Communication Technologies for Ageing Well and e-Health. Cham: Springer; Mar 22, 2019:167-179.

25. Skinner TC, Bruce DG, Davis TME, Davis WA. Personality traits, self-care behaviours and glycaemic control in type 2 diabetes: the Fremantle diabetes study phase ii. Diabet Med 2013 Nov 18;31(4):487-492. [doi: 10.1111/dme.12339] [Medline: 24147848]

26. Li G, Zhang P, Wang J, Gregg EW, Yang W, Gong Q, et al. The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. Lancet 2008 May 24;371(9626):1783-1789. [doi: 10.1016/S0140-6736(08)60766-7] [Medline: 18502303]

27. Ye Q, Fu J. Paediatric type 2 diabetes in China-pandemic, progression, and potential solutions. Pediatr Diabetes 2017 Mar 22;19(1):27-35. [doi: 10.1111/pedi.12517] [Medline: 28326652]

28. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the big-five personality domains. J Res Pers 2003 Dec;37(6):504-528. [doi: 10.1016/S0092-6566(03)00046-1]

29. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]

30. Andone I, Błaszkiewicz K, Eibes M, Trendafilov B, Montag C, Markowetz A. How age and gender affect smartphone usage. 2016 Presented at: UbiComp '16: The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing:Adjunct; September 12-16; Heidelberg Germany p. 9-12 URL: https://dl.acm.org/doi/pdf/10.1145/2968219.2971451 [doi: 10.1145/2968219.2971451]

31. Fox S. Older Americans and the Internet. In: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2004/PIP_Seniors_Online_2004.pdf.pdf#page=9&zoom=auto,-14,733. WASHINGTON, D.C: Pew Internet and American Life Project; Mar 25, 2004.

32. 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]

33.   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 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]

34.   van DHI, Wang J, Droomers M, Spreeuwenberg P, Rademakers J, Uiters E. The relationship between health, education, and health literacy: results from the Dutch Adult Literacy and Life Skills Survey. J Health Commun 2013;18 Suppl 1:172-184 [FREE Full text] [doi: 10.1080/10810730.2013.825668] [Medline: 24093354]

35.   Schillinger D, Barton LR, Karter AJ, Wang F, Adler N. Does literacy mediate the relationship between education and health outcomes? a study of a low-income population with diabetes. Public Health Rep 2016 Aug 02;121(3):245-254. [doi: 10.1177/003335490612100305] [Medline: 16640146]

36.   Chen Q, Carbone ET. Functionality,implementation, impact, and the role of health literacy in mobile phone apps for gestational diabetes: scoping review. JMIR Diabetes 2017 Oct 04;2(2):e25 [FREE Full text] [doi: 10.2196/diabetes.8045] [Medline: 30291088]

37.   Paasche-Orlow MK, Schillinger D, Greene SM, Wagner EH. How health care systems can begin to address the challenge of limited literacy. J Gen Intern Med 2006 Aug;21(8):884-887 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00544.x] [Medline: 16881952]

38.   von Wagner C, Steptoe A, Wolf MS, Wardle J. Health literacy and health actions: a review and a framework from health psychology. Health Educ Behav 2008 Aug 26;36(5):860-877. [doi: 10.1177/1090198108322819] [Medline: 18728119]

39.   March S, Day J, Ritchie G, Rowe A, Gough J, Hall T, et al. Attitudes toward e-mental health services in a community sample of adults: online survey. J Med Internet Res 2018 Feb 19;20(2):e59 [FREE Full text] [doi: 10.2196/jmir.9109] [Medline: 29459357]

40.   Jackson JJ, Hill PL, Payne BR, Roberts BW, Stine-Morrow EAL. Can an old dog learn (and want to experience) new tricks? cognitive training increases openness to experience in older adults. Psychology and Aging 2012;27(2):286-292. [doi: 10.1037/a0025918] [Medline: 22251379]

41.   Jonassaint CR, Boyle SH, Williams RB, Mark DB, Siegler IC, Barefoot JC. Facets of openness predict mortality in patients with cardiac disease. Psychosom Med 2007;69(4):319-322. [doi: 10.1097/psy.0b013e318052e27d] [Medline: 17510289]

42.   Shambare N. Examining the Influence of Personality Traits on Intranet Portal Adoption by Faculty in Higher Education. In: Ph.D. Dissertation, Northcentral University. 789 East Eisenhower Parkway, P.O. Box 1346, Ann Arbor, MI 48106: ProQuest LLC; 2013:226.

43.   Azucar D, Marengo D, Settanni M. Predicting the big 5 personality traits from digital footprints on social media: a meta-analysis. Pers Individ Dif 2018 Apr;124:150-159. [doi: 10.1016/j.paid.2017.12.018]

44.   Mønsted B, Mollgaard A, Mathiesen J. Phone-based metric as a predictor for basic personality traits. Journal of Research in Personality 2018 Jun;74:16-22. [doi: 10.1016/j.jrp.2017.12.004]

45.   Almeida FA, Pardo KA, Seidel RW, Davy BM, You W, Wall SS, et al. Design and methods of "diaBEAT-it!": A hybrid preference/randomized control trial design using the RE-AIM framework. Contemporary Clinical Trials 2014 Jul;38(2):383-396. [doi: 10.1016/j.cct.2014.06.010] [Medline: 24956325]

Abbreviations

HbA$_{1c}$: glycated hemoglobin
ΔHbA$_{1c}$: change in HbA$_{1c}$ level
mHealth: mobile technology for health
OR: odds ratio
RMB: Chinese yuan renminbi

Original Paper

User Perceptions of Virtual Hospital Apps in China: Systematic Search

Yuan Wang[1], PhD; Yuqiao Liu[1], BSc; Yancui Shi[1], PhD; Yanjun Yu[1], BSc; Jucheng Yang[1], PhD

College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin, China

Corresponding Author:
Jucheng Yang, PhD
College of Artificial Intelligence
Tianjin University of Science and Technology
Room 316, Building 8
No 9, 13th Avenue, Binhai New District
Tianjin
China
Phone: 86 15602118875
Email: jcyang@tust.edu.cn

Abstract

Background: Virtual hospital apps are mobile apps that offer functionalities of online consultation, medical guidance, health community forums, referrals, outpatient appointments or virtual hospital-to-home care services. With an increasing number of online medical and health care consulting services, virtual hospital apps have made health care more accessible and fairer for all, especially in China. However, they have occurred without control or regulation. User evaluation can provide directions to help apps optimize identification, lower risks, and guarantee service quality.

Objective: We aimed to conduct a systematic search for virtual hospital apps in China. To get a global view, virtual hospital apps were assessed and characterized by means of quantitative analysis. To get a local view, we conducted a content feedback analysis to explore user requirements, expectations, and preferences.

Methods: A search was conducted of the most popular Apple and Android app stores in China. We characterized and verified virtual hospital apps and grouped apps according to quantification analysis. We then crawled apps and paid attention to corresponding reviews to incorporate users' involvement, and then performed aspect-based content labeling and analysis using an inductive approach.

Results: A total of 239 apps were identified in the virtual hospital app markets in China, and 2686 informative corresponding reviews were analyzed. The evidence showed that usefulness and ease of use were vital facts for engagement. Users were likely to trust a consulting service with a high number of downloads. Furthermore, users expected frequently used apps with more optimization to improve virtual service. We characterized apps according to 4 key features: (1) app functionalities, including online doctor consultation, in-app purchases, tailored education, and community forums; (2) security and privacy, including user data management and user privacy; (3) health management, including health tracking, reminders, and notifications; and (4) technical aspects, including user interface and equipment connection.

Conclusions: Virtual hospitals relying on the mobile internet are growing rapidly. A large number of virtual hospital apps are available and accessible to a growing number of people. Evidence from this systematic search can help various types of virtual hospital models enhance virtual health care experiences, go beyond offline hospitals, and continuously meet the needs of individual end users.

XSL•FO
RenderX

Introduction

Public health focuses on improving and protecting health and well-being among large groups of people. However, the borders of time and space limit equality in countries' medical systems, especially in rural areas and for vulnerable groups [1,2]. With the rapid development of the internet, virtual hospitals can provide registration, remote medical advice, and online doctor consultations. Researchers suggest that 25% to 70% of health care solution seeking does not need a face-to-face appointment with a doctor [3]. Importantly, the Chinese government is vigorously promoting this new type of health service to support health care reform [4,5]. Virtual hospitals play an effective role in reinforcing traditional health care and allow for a flexible relationship between physicians and patients [6] in many countries around the world [7-9].

Currently, China is the largest and fastest-growing market for the mobile internet. The number of mobile internet users in China is 1.31 billion and the mobile data traffic reaches 110.1 billion GB [10]. Residents in China have grown accustomed to using mobile health services for health-related information and health solution retrieval [11,12]. Existing virtual hospital apps in China mainly focus on the functionalities of online consultation, medical guidance, health community forum development, referral, or outpatient appointments. Recently, some apps have started providing appointments for virtual hospital-to-home care services [13]. Due to fair access opportunity to high-quality doctors on virtual hospital apps, the majority of people are willing to pay more for the service [14]. Some researchers have studied online consultation from doctors' perspectives [9].

However, few researchers focus on user evaluation of Chinese virtual hospital apps. There is little research on how user engagement happens and why users abandon a virtual hospital app. To the best of our knowledge, user reviews and basic app data are accessible and valuable [15,16], but have not been fully explored, especially for virtual hospital apps. Hence, we conducted a systematic search using statistical analysis and aspect-based content analysis to review apps of this theme, determine user engagement, and discuss user needs, expectations, concerns, and preferences for virtual hospital apps. We hope the in-depth analysis can provide guidance on the further optimization of apps and their services.

Methods

Selection of App Markets

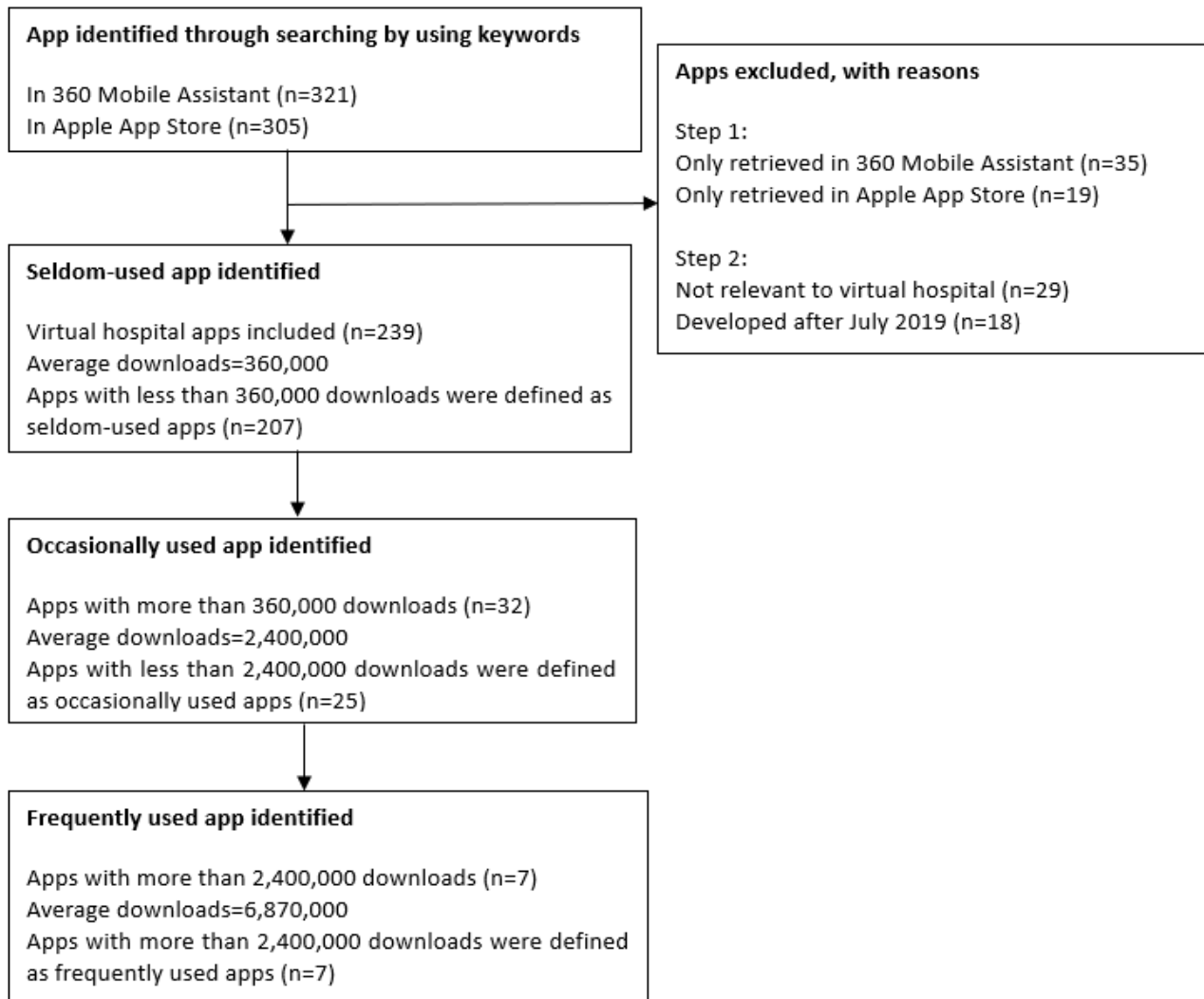According to the 2018-2019 China Mobile App Store Market Annual Monitoring Report from iiMedia Research, 360 Mobile Assistant was the most popular Android app market in China [17]. Hence, we selected 360 Mobile Assistant as the typical Chinese Android app market and the App Store as the typical iOS app market to retrieve virtual hospital apps.

Identification of Virtual Hospital Apps

On November 18, 2019, we crawled mobile apps in 360 Mobile Assistant (Android) and App Store (iOS). We used the following 7 Chinese keywords to conduct a preliminary search in app markets: see a doctor, medication guide, consultation, hospital registration, doctor, medication, medical treatment. We selected the apps that appeared in both Android markets and iOS markets for in-depth analysis. We found that many apps were short-lived, similar to mental health apps [18,19]. More than half of the studied apps were revised roughly every four months [18]. Due to these observations, we excluded the new apps that were first released after July 2019. For these apps, the information was insufficient and unstable for objective analysis. Three investigators screened the names and descriptions of these apps and then tested functionalities of the apps to distinguish between virtual hospital–related apps and unrelated apps. We excluded the apps that more than one investigator labeled as not relevant to the theme. In this way, we identified the included virtual hospital apps. We downloaded and crawled the qualified apps with the researchers' related information.

Apps for Analysis

All of the apps were sorted by number of downloads. For each included app, extracted data consisted of (1) flat user score, reflecting perceived satisfaction of a user; (2) the number of downloads, showing the measure of usage frequency and intention; (3) the number of positive, negative, and neutral ratings; (4) content of reviews; and (5) the last updated date of an app. Critically, due to the fact that Apple's App Store requires a minimum number of reviews before releasing average ratings (ie, small numbers are suppressed), we think that the score on the App Store is more referential. Thus, we used average ratings in App Store. While App Store does not provide the number of downloads, we used the number of downloads from 360 Mobile Assistant for analysis. We divided apps into three categories according to the number of downloads: (1) frequently used apps, (2) occasionally used apps, and (3) seldom-used apps. First, apps with fewer than the average number of downloads of all determined apps were defined as seldom-used apps. Second, apps with fewer than the average number of downloads of the remaining apps were defined as occasionally used apps. Other apps were defined as frequently used apps. Procedures are shown in Figure 1.

Figure 1. Flowchart of app classification selection processes.

App identified through searching by using keywords

In 360 Mobile Assistant (n=321)
In Apple App Store (n=305)

Apps excluded, with reasons

Step 1:
Only retrieved in 360 Mobile Assistant (n=35)
Only retrieved in Apple App Store (n=19)

Step 2:
Not relevant to virtual hospital (n=29)
Developed after July 2019 (n=18)

Seldom-used app identified

Virtual hospital apps included (n=239)
Average downloads=360,000
Apps with less than 360,000 downloads were defined as seldom-used apps (n=207)

Occasionally used app identified

Apps with more than 360,000 downloads (n=32)
Average downloads=2,400,000
Apps with less than 2,400,000 downloads were defined as occasionally used apps (n=25)

Frequently used app identified

Apps with more than 2,400,000 downloads (n=7)
Average downloads=6,870,000
Apps with more than 2,400,000 downloads were defined as frequently used apps (n=7)

Aspect-Based Content Labeling of Valuable User Reviews

According to the proposal put forward by Mendiola et al [19] and review data, we sum up the following four aspects that concern users: (1) app functionalities, including online doctor consultation, in-app purchases, tailored education, and community forums; (2) security and privacy, including user data management and user privacy; (3) health management, including health tracking, reminders, and notifications; and (4) technical aspects, including user interface and equipment connection.

For qualified apps, we selected and reviewed user-generated reviews after the last update of the 7 most popular apps. We found 4630 reviews. Before aspect-based content analysis, we filtered poor-quality or fake reviews from our data set based on the following rules: (1) review expresses only positive or negative emotions and without mentioning any specific aspects, such as "good!" or "very bad!" (n=1902); and (2) review is an apparent case of human manipulation, such as advertising reviews that have nothing to do with the corresponding app (n=42).

After screening, we got a clean review data set (n=2686), where each review was authentic and specific. On the clean review data set, we asked 3 investigators to use a binary system to assign "1" to a particular aspect in a review in which the user was satisfied or "0" to indicate that the user was dissatisfied with a particular aspect. Investigators independently conducted quality assessments and the final results were decided by majority vote. Examples are shown in Textbox 1 and Figure 2. Note that most of the reviews only included one or two aspects (2570 reviews with one aspect each, 116 reviews with two aspects each).

Textbox 1. Examples of reviews.

No. 1. "This app always takes mandatory access to my address book data and leaks out my personal privacy!"
No. 2. "The sound of the reminder to take my medicine is too low and unattractive and there is so much noise at home that I often can't hear it!"
No. 3. "The interface design is too complicated! I hope they make the design a little simpler. In addition, the connection between bracket and app is always unstable, easy to disconnect."
No. 4. "Every time the doctor can provide me with reasonable advice according to my physical condition and deliver medicine directly to the door. There are formal invoices and after-sales protection!"
No. 5. "I can neither upload my data to my Wechat, nor share it with my friends. I hope this feature is added!"

Figure 2. Aspect-based content labeling for the review examples in Textbox 1.

Aspect		No. 1	No. 2	No. 3	No. 4	No. 5
App functionalities						
	Online doctor consultation					
	In-app purchases				1	
	Tailored education				1	
	Community forums					
Security and privacy						
	User data management					0
	User privacy	0				0
Health management						
	Health tracking					
	Reminders and notifications		0			
Technical aspects						
	User interface			0		
	Equipment connection			0		

Results

App Characteristics

Table 1 summarizes the characteristics of included virtual hospital apps, including the statistical information of downloads and rating scores. According to the search strategy and data filtering, we took a total of 239 apps for analysis. The average rating score was 3.8 out of 5 and the average number of downloads was 36,464. We included a total of 92,366 rating scores, including 90,106 positive scores, 1512 negative scores, and 748 neutral scores. Note that people can give rating scores without reviews for apps. Thus, there was not a one-to-one

correspondence between content of reviews and the numbers of ratings. In addition, apps had an average of 386 scores, including 377 positive scores, 3 neutral scores, and 6 negative scores per app.

The results in Table 1 show that there was no linear relationship between the number of rating scores and the number of downloads. Frequently used apps only accounted for 2.9% (7/239) of the number of virtual hospital apps but accounted for 63.06% (5,496,000/8,715,000) of all downloads. There was not a big difference in average rating scores between the frequently used apps and the occasionally used apps, but the seldom-used apps scored notably lower than the other two. For

frequently used apps, negative rating scores accounted for the highest proportion of the total comparison theory, followed by seldom-used apps, and lastly occasionally used apps. People set higher requirements for popular virtual hospital apps (ie, frequently used apps).

Table 1. Characteristics of included virtual hospital apps.

Characteristic	Frequently used apps	Occasionally used apps	Seldom-used apps	All
Apps, n (%)	7 (2.93)	25 (10.46)	207 (86.61)	239 (100)
Downloads, n (%)	5,496,000 (63.06)	2,208,000 (25.34)	1,011,000 (11.60)	8,715,000 (100)
Positive ratings, n (%)	59,687 (97.17)	15,063 (98.49)	15,356 (98.15)	90,106 (97.55)
Neutral ratings, n (%)	577 (0.94)	83 (0.54)	88 (0.56)	748 (0.81)
Negative ratings, n (%)	1162 (1.89)	148 (0.97)	202 (1.29)	1512 (1.64)
Ratings, n (%)	61,426 (100)	15,294 (100)	15,646 (100)	92,366 (100)
Rating score, mean	4.8	4.6	3.7	3.8
Downloads, mean	785,142	88,320	4884	36,464
Positive ratings, mean	8526	602	74	377
Neutral ratings, mean	83	3	1	3
Negative ratings, mean	166	6	1	6
Number of ratings, mean	8775	611	75	386

Case Study of Selected Apps

We ranked apps according to their average rating scores and the number of downloads in the frequently used apps, occasionally used apps, and seldom-used apps categories. We took the top 3 and the last 3 apps in each category for in-depth analysis.

The results are shown in Table 2. For frequently used apps, the difference between average rating scores was not large, but the number of downloads and the number of positive, neutral, and negative ratings varied greatly. Although Dr Chunyu had the most downloads (17.98 million), it also had the most neutral rating scores and received a low average ranking score in this category. The majority of apps with more downloads had more positive reviews, neutral reviews, and negative reviews. There are also good virtual hospital apps in the occasionally used and seldom-used app categories, for example Good Doctor for People and Dr Ruyi, with average rating scores of 4.9. For seldom-used apps, apps with higher average rating scores had more downloads than apps with lower scores, except Dr Thumb. We observed that some apps in the seldom-used apps category specialized in types of virtual hospitals, such as Dr Doodle for pregnant women and newborns and Dr Warmth for video consultation.

Table 2. Characteristics of the typical apps in the frequently used app, occasionally used app, and seldom-used app categories (ordered by average rating scores).

App (app's name in Chinese Pinyin)	Score	Downloads (in millions)	Positive reviews, n (%)	Neutral reviews, n (%)	Negative reviews, n (%)
Frequently used apps					
Considerate Doctor (Tie Xin Yi Sheng)	4.9	5.54	9254 (99.92)	3 (0.03)	4 (0.04)
Online Family Doctor (Jia Ting Yi Sheng Zai Xian)	4.9	3.39	1088 (99.18)	5 (0.46)	4 (0.36)
Ping An Good Doctor (Ping An Hao Yi Sheng)	4.8	15.13	29,445 (96.80)	258 (0.85)	715 (2.35)
Enjoy Health (Yi Jia Jian Kang)	4.7	3.16	895 (95.93)	5 (0.54)	33 (3.54)
Dr Chunyu (Chun Yu Yi Sheng)	4.7	17.98	11,197 (95.66)	243 (2.08)	265 (2.26)
Online Good Doctor (Hao Dai Fu Zai Xian)	4.6	4.96	1944 (98.08)	10 (0.50)	28 (1.41)
Occasionally used apps					
Good Doctor for People (Ren Min Hao Yi Sheng)	4.9	2.16	816 (99.51)	0 (0.00)	4 (0.49)
9K Doctor (9k Yi Sheng)	4.9	1.01	401 (99.01)	0 (0.00)	4 (0.99)
Dr Lighthouse (Deng Ta Yi Sheng)	4.9	0.63	1137 (99.91)	0 (0.00)	1 (0.09)
39 Ask for Doctor (39 Wen Yi Sheng)	4.4	0.36	158 (86.81)	6 (3.30)	18 (9.89)
Female Private Doctor (Nv Xing Si Ren Yi Sheng)	4.2	1.28	298 (86.63)	17 (4.94)	29 (8.43)
Dr Almond (Xing Ren Yi Sheng)	3.8	0.40	21 (43.75)	3 (6.25)	24 (50)
Seldom-used apps					
Dr Ruyi (Ru Yi Yi Sheng)	4.9	0.23	738 (99.73)	0 (0.00)	2 (0.27)
Dr Doodle (Du Du Yi Sheng)	4.9	0.20	3163 (99.81)	2 (0.06)	4 (0.13)
Dr Half (Ban Ge Yi Sheng)	4.9	0.18	382 (99.48)	1 (0.26)	1 (0.26)
5U Family Doctor (5U Jia Ting Yi Sheng)	2.5	0.002	0 (0.00)	0 (0.00)	3 (100.00)
Dr Warmth (Wen Nuan Yi Sheng)	3.0	0.002	0 (0.00)	2 (100.00)	0 (0.00)
Dr Thumb (Mu Zhi Yi Sheng)	2.6	0.09	7 (46.67)	5 (33.33)	3 (20.00)

Aspect-Based Content Analysis of User Reviews

We analyzed authentic and specific reviews, defined in the section "Aspect-Based Content Labeling of Valuable User Reviews." The 4 themes were (1) app functionalities, (2) security and privacy, (3) health management, and (4) technical aspects. The themes and associated aspects are summarized in Table 3.

From Table 3, we observe statistical information regarding user concerns about virtual hospital apps. First, we compared the total number of reviews for each aspect. The majority of

reviews expressed opinions about online doctor consultation, health tracking, and user data management for virtual hospital apps in China, while fewer reviews discussed community forums, user privacy, and tailored education. Second, we determined the ratio of positive and negative evaluations. Many reviews praised online health services (948/600, 91.3% for online doctor consultation; 148/152, 97.4% for reminders and notifications). About 80% (32/40) of reviews showed positive opinions about community forums within virtual hospital apps, indicating that users like understanding, sharing, and discussing health-related issues in apps. However, it is imperative to improve security and privacy in virtual hospital apps, especially for user data management, which 86.0% (394/458) of reviews complained about. Details of typical user opinions about each aspect are described below.

Table 3. Themes and aspects of virtual hospital apps.

Aspect	Total, n (%)[a]	Positive, n (%)	Negative, n (%)
App functionalities			
Online doctor consultation	600 (22.34)	548 (91.33)	52 (8.67)
In-app purchases	358 (13.33)	188 (52.51)	170 (47.49)
Tailored education	80 (2.98)	44 (55.00)	36 (45.00)
Community forums	40 (1.48)	32 (80.00)	8 (20.00)
Security and privacy			
User data management	458 (17.05)	64 (13.97)	394 (86.03)
User privacy	66 (2.46)	2 (3.03)	64 (96.97)
Health management			
Health tracking	554 (20.63)	290 (52.35)	264 (47.65)
Reminders and notifications	152 (5.66)	148 (97.37)	4 (2.63)
Technical aspects			
User interface	286 (10.65)	190 (66.43)	96 (33.57)
Equipment connection	208 (7.74)	10 (4.81)	198 (95.19)

[a]Percentage is out of 2686 ratings.

Topic 1: App Functionalities

Online Doctor Consultation

Online doctor consultation, which has been adopted by all virtual hospital apps, is a functionality that provides convenient outpatient service online. This is the best feature for user engagement, with the highest number of positive reviews (n=548). For consulting, users preferred video chatting and shared more details with pictures of their condition. For example, a user said, "I am a patient with dermatosis. It is so painful that I can't describe how I feel, so I want to upload some videos or photos about my illness for help or advice." For consulting quality, most users were satisfied due to the truthfulness of doctors' information. Users particularly wanted to reach high-quality and certified doctors during web-based consultation services. A user said, "I hope apps can collect a large amount of information about doctors to provide valuable references for patients, including doctor's reputation and experience." With this information, regardless of time or location, patients can find and reach high-quality doctors from all over the country.

In-App Purchases

A total of 52.5% (188/358) of users mentioned that they felt it was convenient to buy drugs or other health care–related goods in apps. Users said that they were concerned about whether the drug provided was safe enough. Users further talked about their opinion that (1) apps should take actions to avoid counterfeit medications and (2) after-sales protection could make them feel at ease when buying drugs in virtual hospital apps. Related to online doctor consultation, users also wished that doctors who recommended medications were professional enough to judge disease symptoms and know side effects of drugs. Moreover, users wanted apps to show clear and accurate information about the terms and conditions of in-app purchases.

Tailored Education

Users desired apps to have the ability to tailor schemes for different populations by analyzing personal data. For patients with chronic diseases, users prefer personalized knowledge and consulting recommendations in apps based on their previous records of vital signs, laboratory test results, or medical records. For example, a patient with diabetes said, "This app scored for various histopathological indicators by my uploaded data, then returned my physical health condition to me and provided nutrition knowledge and counseling. I like it!"

Community Forums

Social features in virtual hospital apps were a significant influence on ratings. We found a total of 40 reviews about community forums, 80% (32/40) of which provided positive opinions. Users wished for apps to provide a forum to elaborate on particular health care topics, share arduous experiences, discuss their views, and even share treatments for diseases.

Patients with chronic conditions wanted to have doctors join discussions to share knowledge and experience, ask for advice on diagnosis and treatment for specific patient cases, and discuss issues about their disease and related topics. Pregnant women preferred an in-app feature to chat with other pregnant women, just like chatting with each other when lining the hospital corridors.

Topic 2: Security and Privacy

User Data Management

On the premise of ensuring data security, users wanted to get more health data administrative privileges. Reviews that discussed data management made up the highest number of negative reviews (n=394). Users wanted to synchronize personal health data between multiple apps or to migrate data to another app. Apps should provide intuitive and easy features to query, display, and download data. Meanwhile, apps with local data management should remind users to save data correctly. For example, a pregnant user said, "After re-installing the app, I lost all my data for five months of pregnancy, which made it impossible for me to continue my physical management." In this case, data retention was unsatisfied.

User Privacy

For online inquiries about diseases, users required that their data be protected and anonymized. Users complained about privacy settings within the highest ratio of negative reviews (64/66, 97%,). They needed authentication, authorization, and a privacy mechanism to protect their data. For general diseases, some users were embarrassed to consult with their own identity. They requested the ability to have a doctor consultation in an anonymous way and have an option to destroy their data immediately after consultation. Users expected to be notified via messages or emails if their accounts were logged in abnormally, such as logging in with new devices. Those who were willing to consider sharing data also emphasized data control. They talked about the risk of sharing their health data. Some of them preferred local databases that were disconnected from the internet or distributed databases over several servers.

Topic 3: Health Management

Health Tracking

Health tracking is the right way for users to be aware of changes in health status. In this way, it is essential to understand how apps collect user data and provide processed visualized data to users. Only 52.4% (290/554) of reviews were positive about health tracking in virtual hospital apps. These reviews mentioned that an automatic and effective way of health tracking in apps could help make doctor consultation better. For some patients with chronic conditions, app features that monitored daily blood pressure and heartbeat data were helpful. Patients with insomnia hoped that the app could access sleep data about sleep latency, sleep duration, sleep efficiency, and sleep quality assessment to provide to doctors during consultation.

Reminders and Notifications

Personalized reminders can reinforce behavioral changes. A total of 97.4% (148/152) reviews supported timely and accurate reminders and notifications. For negative reviews, users expressed challenges with limited alert customization, especially in terms of the alarm loudness. A user commented, "No matter what tone, the tone is too soft and doesn't ring long enough to wake me up." Users mentioned that customization with a ringtone was "an essential feature since it is how you will be notified." Sometimes, users hoped that notifications could not be closed forcibly. In addition, users said that apps should remind them of potential risks for improper usage or possible adverse effects along with the reminders and notifications.

Topic 4: Technical Aspects

User Interface

For user interface design, users thought simplicity was the most important design feature and argued for adaptation for vulnerable groups and multiple modes. An excellent user interface can have a powerful impact on the usability and user experience of an app. About 66.4% (190/286) of reviews expressed satisfaction with the user interface of virtual hospital apps in China.

Users appreciated clear, intuitive, and helpful design apps that were easy to use (ie, navigation was inherent). Reviewers complimented apps that had a "clean and simple format" and a "beautiful layout." On the contrary, reviews commented on "clunky" user interfaces. For example, apps should try to avoid lots of extra buttons and submenus. Elderly people in particular frequently listed inconvenience of use, which influenced their engagement. Moreover, the interface design should follow the same pattern. That is, all graphic elements (ie, typographies, icons, and buttons) should have a consistent appearance. The features of each component (ie, navigation menu, lists, and photo gallery) should be identified.

Besides these features, virtual hospital apps could pay more attention to health, for example, by providing a night mode for eyesight protection. We also found that monolingual apps sometimes made consultation difficult. These apps were difficult for foreigners to register for and use. Some users said that multilingual settings could benefit foreign user groups; even basic machine translation could help. These reviews demonstrate that medical service via virtual hospital apps in China could think about globalization to serve more people.

Equipment Connection

In China, health monitoring and wearable devices are becoming extremely popular [20], as they can record and transmit real-time, health-related data from a patient's home to any medical situation. However, this aspect was a top complaint along with user privacy. The percentage of reviews about equipment connection that were negative was 95.2% (198/208). Users often complained that (1) apps could not detect the device, (2) apps could not connect to the device for data transmission, (3) data transferred to the client were missing or inaccurate, (4) there was an unstable connection, (5) there was poor compatibility when connecting another device, (6) they were unable to unpair, or (7) there were external interruptions (eg, incoming calls or messages). Besides these issues, users mentioned that both Wi-Fi and Bluetooth could be supported or that the app could be made to work in flight mode.

Discussion

Overview

This study explored user preferences and requirements for virtual hospital apps to predict and analyze current models and future directions of the medical service market with mobile internet. First, we provided a statistical description of full data to compare distributions of rating scores and number of downloads with 3 different categories. Second, we conducted in-depth aspect-based content analysis to discover opinion details about interest in and barriers to participation. In this study, the app and review selection strategy benefited research, with better generalization and time-efficiency. Despite many works that have discussed virtual hospitals from the perspective of economics [21] or doctors [9], this paper is the first to provide an in-depth report and highlight situations from the users' perspective.

Principal Findings

End user engagement and follow-up are prerequisites for sustained promotion and effectiveness. In the field of virtual hospitals, frequently used apps with the most downloads received the highest average rating scores (4.8/5), while there was great variation in the average rating scores among seldom-used apps. Interestingly, the ratios between positive and negative reviews were similar in frequently used, occasionally used, and seldom-used apps, although the Matthew effect was clearly seen in the number of downloads. We noticed that frequently used apps suffered from more negative reviews (1162/61,426, 1.89%) with detailed user experience. Users tending to give high rating scores for those popular apps had higher expectations for conversational features, personalization, performance, and high-quality content and preferred putting forward the related proposals. Frequently used apps should take responsibility for possible updates and improvements.

We noticed that high-probability access to individualized expert consultation was the main reason for both new and old user engagement. First, health care service apps can greatly relieve unequally distributed medical resources in China. Medical resources are geographically unbalanced, especially for rural areas and for persons with a disability [2]. For virtual hospitals in apps, users can access a virtual hospital regardless of time or location. All that is needed is an internet-enabled phone. In positive reviews, users said that they liked to try virtual hospital apps because they were a considerable distance away from good medical resources. When patients reported that they were cured by a doctor's treatment in an app, they were more likely to pay for in-app purchases for further health services [14]. Users felt comfortable with famous, professional doctors who also worked for offline hospitals rather than those who were from unknown or ambiguous sources. Our findings suggest that virtual hospital apps should rely on actual medical conditions and qualified physicians and medicines. With the support of government policies [4,5], users in China were willing to trust that apps could invite high-quality doctors to trade their leisure time for online consultation. Second, most users were satisfied with the prompt response, consulting attitude, and advanced knowledge of doctors in apps. Users saved waiting time, reduced

transportation costs, and shared fair doctor navigation. Third, a multichannel communication strategy for points of patient engagement was welcomed. Besides video chatting, which is similar to an offline service, new consultation options in apps, such as reservation calls, voice messages, graphic messages, video messages, or plain text, could relieve patients' tension and give patients thinking time.

Compared with online doctor consultations, users were wary of health products in in-app purchases, such as medicines, supplements, probiotic food, and home health care devices. About 47.5% (170/358) of users raised questions because distribution channels for medicine and medical devices lacked trusted verification. It was hard for customers to protect their rights with in-app purchases. Stakeholders need to pay more attention to consumers' rights protection in the selling of health-related products.

Users cared about their health-related data. On the one hand, users sought convenient and flexible data usage, including data association, data synchronization, data transmission, and data sharing across multiple apps and external devices. Specifically, inferred from patients' comments, we found that some users wanted to switch apps and share their data with other communities or doctors. However, that is a big challenge now [22]. An agreement is needed to put forward sharing and canonical management of health data in China. On the other hand, users asked for precision and integrity for data management. Accurate and complete health consultation records are often a great help to detect or review changes in physical and psychological indicators. Users wanted to learn from their health data history.

With regard to interactivity and visual settings, users preferred broad individualization. Personalization can enhance convenience and user-friendliness. For tailored education, relevant and meaningful individualized health knowledge provides engagement, serendipity, and high value to users, while irrelevant or unindividualized health care information displays make users feel unengaged. This type of individualization has been shown to be vital for sustained engagement [23]. For user interface, about 66.4% (190/286) of comments showed positive attitudes to the current design. The others, including foreign app users, elderly persons, and visually impaired users, asked for alternative customization view modes for promoting large-scale, easy usage. The design of indicative icons and feedback functions required more consideration. For reminders and notifications, most users found such functions helpful for consultation appointments, abnormal account access, or medicine plans. Notifications allow users to tackle health problems and to schedule time for maintaining health [24,25]. Personalized reminders can better reinforce behavioral changes in app users.

Community forums based on peer support, education for health practice, health promotion, and emotional support can serve as an effective supplement to professional doctor consultation. It can relieve China's insufficient medical resources. A total of 1.48% (40/2686) of reviews mentioned that social health education is a powerful force in human health. People share their health experiences, knowledge, and practical support, and

XSL•FO

RenderX

others learn from that. Such client-to-client communication and question answering could meet the needs of finding people who have similar health issues. In reviews, users provided tips to promote peer support, for example, placing forum joining in a prominent position during patients' in-app journeys.

Equipment connection and user privacy remain a considerable challenge in virtual hospital apps. Satisfaction of these two aspects was relatively low, with a satisfaction rate of only 4.8% (10/208) and 3.0% (2/66), respectively. With fast growth of wearable devices [20], inexpensive health-related wearable devices provide everyone the possibility of constantly tracking personal health data. However, due to inaccurate data estimation and poor connection of devices, the user experience was negative. Inaccurate data estimation limits data usage, such as using data as an accurate and valid reference for doctor consultation. Data show that user privacy protection was not paid enough attention. Almost all comments about this issue were negative. Apps lacked explicit features to protect user accounts, especially for abnormal log-in attempts. Users suggested that email or a message alert could be an easy way to improve security. App makers need to consider how to balance user security and privacy with flexible data access.

Virtual hospital apps are convenient and suitable for two types of users: (1) users who are far away from good medical resources in time or space, especially for rural users and users with disabilities, and (2) users who require medical support and group support but are uncomfortable with face-to-face communication. However, users who demand high security and privacy need to practice caution during app selection and usage.

A practical finding from our analysis is that the Chinese market has begun to pay attention to privacy, security, personalization, and socialization in virtual hospital apps. We estimate that these will provide better solutions in future development.

Limitations

This research has some limitations. First, we explored only 2 app markets, although they are the most representative ones for Android and iOS operating systems in China. Second, we were not sure about the recall ratio for the whole mobile virtual hospital market because there were other popular ways to access virtual hospital services with mobile internet, such as temporal mobile online meetings, Wechat applets, and online chat groups. Except for platform operators, data of these tools were incomplete and inaccessible in most situations. Thus, we mainly focused on public app markets. Third, we limited content analysis on frequently used apps, as we thought that other data may suffer from severe bias due to the small number of helpful reviews and downloads. However, it is known that this can be a risk for sampling strategies. Finally, this research did not include opinions or usage experiences of medical workers, such as doctors and nurses.

Conclusions

Virtual hospital apps are growing in popularity with the high pace of mobile internet in China. User evaluation provides a valuable insight for app development and online medical service reform. This study provides mobile internet–based statistical data for medical service providers, domain policy makers, and potential related stakeholders, and it provides aspect-based content analysis to support a better understanding of virtual hospital app users.

Authors' Contributions

YW and YL designed the study. YL and YY were responsible for app search, analysis, and data collection. YW and YL analyzed and interpreted the data. YL drafted the manuscript. All authors contributed to the review and editing of the manuscript.

Conflicts of Interest

None declared.

References

1. Tang Y, Yang Y, Shao Y. Acceptance of Online Medical Websites: An Empirical Study in China. Int J Environ Res Public Health 2019 Mar 15;16(6) [FREE Full text] [doi: 10.3390/ijerph16060943] [Medline: 30884754]
2. Ma Y, Zhang L, Boswell M. Inequities in the allocation of medical resources in China's Township Health Centers. China Ag Economic Review 2016 Nov 07;8(4):637-646 [FREE Full text] [doi: 10.1108/caer-11-2015-0170]
3. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Med Care 2004 Mar;42(3):200-209. [doi: 10.1097/01.mlr.0000114908.90348.f9] [Medline: 15076819]
4. Li K. Report on the Work of the Government 2018. Secondary Report on the Work of the Government. 2018. URL: http://www.gov.cn/zhuanti/2018lh/2018zfgzbg/2018zfbgdzs.htm#book7/page1 [accessed 2020-07-28]
5. Li K. Report on the Work of the Government 2019. Secondary Report on the Work of the Government. 2019. URL: http://www.gov.cn/premier/2019-03/16/content_5374314.htm [accessed 2019-04-10]

6.   Brockes C, Schenkel JS, Buehler RN, Grätz K, Schmidt-Weitmann S. Medical online consultation service regarding maxillofacial surgery. J Craniomaxillofac Surg 2012 Oct;40(7):626-630. [doi: 10.1016/j.jcms.2012.03.018] [Medline: 22555239]

7.   Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]

8.   Online Doctor Consultation Market: Global Market Synopsis, Growth Factors, Industry Segmentation, Regional Analysis And Competitive Analysis 2017-2025.: Garner Insights; 2018 Mar. URL: http://garnerinsights.com/ Online-Doctor-Consultation-Market-Global-Market-Synopsis-Growth-Factors-Industry-Segmentation-Regional-Analysis-And-Competitive-Analysis-2017—2025 [accessed 2020-07-28]

9.   Li Y, Yan X, Song X. Provision of Paid Web-Based Medical Consultation in China: Cross-Sectional Analysis of Data From a Medical Consultation Website. J Med Internet Res 2019 Jun 03;21(6):e12126 [FREE Full text] [doi: 10.2196/12126] [Medline: 31162129]

10. Economic operation of the communications industry from January to November 2019.: Ministry of Industry and Information Technology of the People's Republic of China; 2019 Dec 18. URL: http://www.miit.gov.cn/n1146312/n1146904/n1648372/ c7572427/content.html [accessed 2020-07-28]

11. 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]

12. Harte R, Hall T, Glynn L, Rodríguez-Molinero A, Scharf T, Quinlan LR, et al. Enhancing Home Health Mobile Phone App Usability Through General Smartphone Training: Usability and Learnability Case Study. JMIR Hum Factors 2018 Dec 26;5(2):e18 [FREE Full text] [doi: 10.2196/humanfactors.7718] [Medline: 29699969]

13. McGillion M, Ouellette C, Good A, Bird M, Henry S, Clyne W, et al. Postoperative Remote Automated Monitoring and Virtual Hospital-to-Home Care System Following Cardiac and Major Vascular Surgery: User Testing Study. J Med Internet Res 2020 Mar 18;22(3):e15548 [FREE Full text] [doi: 10.2196/15548] [Medline: 32186521]

14. China's Paid Knowledge Industry Development White Paper 2016.: Analysys; 2016. URL: https://www.analysys.cn/article/ analysis/detail/1000416 [accessed 2020-07-28]

15. Park JYE, Li J, Howren A, Tsao NW, De Vera M. Mobile Phone Apps Targeting Medication Adherence: Quality Assessment and Content Analysis of User Reviews. JMIR Mhealth Uhealth 2019 Jan 31;7(1):e11919 [FREE Full text] [doi: 10.2196/11919] [Medline: 30702435]

16. Aji M, Gordon C, Peters D, Bartlett D, Calvo RA, Naqshbandi K, et al. Exploring User Needs and Preferences for Mobile Apps for Sleep Disturbance: Mixed Methods Study. JMIR Ment Health 2019 May 24;6(5):e13895 [FREE Full text] [doi: 10.2196/13895] [Medline: 31127714]

17. 2018-2019 China Mobile App Store Market Annual Monitoring Report.: iiMedia; 2018. URL: https://www.iimedia.cn/ c400/63240.html [accessed 2020-07-28]

18. Larsen ME, Nicholas J, Christensen H. Quantifying App Store Dynamics: Longitudinal Tracking of Mental Health Apps. JMIR Mhealth Uhealth 2016 Aug 09;4(3):e96 [FREE Full text] [doi: 10.2196/mhealth.6020] [Medline: 27507641]

19. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]

20. Llamas RT. Worldwide Wearables Market Shares, 2019: Apple Leads the Market in Volume and Value.: IDC; 2020. URL: https://www.idc.com/getdoc.jsp?containerId=US46092820 [accessed 2020-07-28]

21. Li J, Zhang Y, Ma L, Liu X. The Impact of the Internet on Health Consultation Market Concentration: An Econometric Analysis of Secondary Data. J Med Internet Res 2016 Oct 28;18(10):e276 [FREE Full text] [doi: 10.2196/jmir.6423] [Medline: 27793793]

22. Zhang P, Dong L, Chen H, Chai Y, Liu J. The Rise and Need for Mobile Apps for Maternal and Child Health Care in China: Survey Based on App Markets. JMIR Mhealth Uhealth 2018 Jun 08;6(6):e140 [FREE Full text] [doi: 10.2196/mhealth.9302] [Medline: 29884601]

23. Peters D, Calvo RA, Ryan RM. Designing for Motivation, Engagement and Wellbeing in Digital Experience. Front Psychol 2018;9:797 [FREE Full text] [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]

24. 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]

25. Tran N, Coffman JM, Sumino K, Cabana MD. Patient reminder systems and asthma medication adherence: a systematic review. J Asthma 2014 Jun;51(5):536-543. [doi: 10.3109/02770903.2014.888572] [Medline: 24506699]

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

Designing an Information and Communications Technology Tool With and for Victims of Violence and Their Case Managers in San Francisco: Human-Centered Design Study

Devika Patel[1], MS; Siavash Sarlati[2], MD; Patrick Martin-Tuite[3], MSc; Joshua Feler[4], MSc; Lara Chehab[1], MPH; Michael Texada[1]; Ruben Marquez[1], AA; F Julia Orellana[1]; Terrell L Henderson[1], BA; Adaobi Nwabuo[1], MBBS, MPH; Rebecca Plevin[1], MD; Rochelle Ami Dicker[5], MD; Catherine Juillard[5], MPH, MD; Amanda Sammann[1], MPH, MD

[1]Department of Surgery, University of California, San Francisco, CA, United States
[2]Department of Emergency Medicine, University of California, San Francisco, CA, United States
[3]University of California, San Francisco School of Medicine, San Francisco, CA, United States
[4]Yale University School of Medicine, New Haven, CT, United States
[5]Division of Surgical Critical Care, Department of Surgery, University of California, Los Angeles, CA, United States

Corresponding Author:
Devika Patel, MS
Department of Surgery
University of California, San Francisco
1001 Potrero Ave
San Francisco, CA, 94110
United States
Phone: 1 (628) 206 3764
Email: devika.patel@ucsf.edu

Abstract

Background: Violence is a public health problem. Hospital-based violence intervention programs such as the San Francisco Wraparound Project (WAP) have been shown to reduce future violent injury. The WAP model employs culturally competent case managers who recruit and enroll violently injured patients as clients. Client acceptance of the WAP intervention is variable, and program success depends on streamlined, timely communication and access to resources. High rates of smartphone usage in populations who are at risk for violent reinjury create an opportunity to design a tailored information and communications technology (ICT) tool to support hospital-based violence intervention programs.

Objective: Current evidence shows that ICT tools developed in the health care space may not be successful in engaging vulnerable populations. The goal of this study was to use human-centered design methodology to identify the unique communication needs of the clients and case managers at WAP to design a mobile ICT.

Methods: We conducted 15 semi-structured interviews with users: clients, their friends and families, case managers, and other stakeholders in violence intervention and prevention. We used a human-centered design and general inductive approach to thematic analysis to identify themes in the qualitative data, which were extrapolated to insight statements and then reframed into design opportunities. Wireframes of potential mobile ICT app screens were developed to depict these opportunities.

Results: Thematic analysis revealed four main insights that were characterized by the opposing needs of our users. (1) A successful relationship is both professional and personal. Clients need this around the clock, but case managers can only support this while on the clock. (2) Communications need to feel personal, but they do not always need to be personalized. (3) Healing is a journey of skill development and lifestyle changes that must be acknowledged, monitored, and rewarded. (4) Social networks need to provide peer support for healing rather than peer pressure to propagate violence. These insights resulted in the following associated design opportunities: (1) Maximize personal connection while controlling access, (2) allow case managers to personalize automated client interactions, (3) hold clients accountable to progress and reward achievements, and (4) build a connected, yet confidential community.

Conclusions: Human-centered design enabled us to identify unique insights and design opportunities that may inform the design of a novel and tailored mobile ICT tool for the WAP community.

XSL•FO
RenderX

KEYWORDS

Introduction

Violence is a public health problem that targets vulnerable populations in the United States [1,2]. Homicide is among the leading causes of death for young people aged 10-30 years, and for every fatal assault, members of this age group experience 90 non-fatal assault injuries [3]. Risk factors associated with urban violence are concentrated among vulnerable populations and include low income, unemployment, being of racial or ethnic minority background, low education levels, substance abuse, and neighborhood disorder [4-9]. Violence persists among these communities that experience social inequity, and individuals are caught in a cycle of constant perpetration or reinjury [6,10-14]. Previous exposure to violence is a strong predictor of future violence with data showing that up to 45% of patients injured by assault will be reinjured, and 20% will be killed within 5 years of an index injury [4,5,10,11].

The health care setting presents an opportunity to leverage a teachable moment to prevent violence by mitigating future risks [15]. Hospital-based violence intervention programs are cost-effective and successfully reduce violent reinjury among high-risk, assault-injured victims [16,17]. These programs employ culturally competent violence intervention specialists (case managers), who are often from the same neighborhoods and ethnic groups as their clients. These case managers help victims of violence (clients) navigate the emotional and logistical challenges following injury. Case managers use shared experiences, common culture and language, and a trauma-informed approach by recognizing the signs and symptoms of trauma and responding in a safe, respectful, and transparent way. As a result, these case managers are best equipped to foster the high level of trust and communication required to guide clients towards protective behaviors and services.

Information and communications technology (ICT) provides a promising mechanism to address future risks of violence, as mobile apps and web-based communication platforms are effective adjuncts in promoting behavior change and risk-factor modification in the health care setting [18-20]. Studies show high levels of smartphone penetration among marginalized and minority populations who access health care through the emergency department [21,22], and increased dependence on smartphones for internet use [23]. Ethnic minority groups are more likely to use smartphones to seek health information, employment, and educational content than non-smartphone users [24]. Despite these trends, few targeted digital interventions have been developed to address the issues faced by vulnerable and underserved populations, and even fewer have addressed the unique needs of victims of violence [25,26]. Digital health care tools are not culturally tailored and have not been purposefully designed to address the needs of vulnerable populations, leading to technology with poor usability, acceptability, and effectiveness in these groups [25,27,28]. As

such, ethnic minority groups, individuals at lower socioeconomic status levels, and other vulnerable groups are less likely to engage successfully with health care providers using available digital tools [27,29,30].

This study is a partnership with The San Francisco Wraparound Project (WAP), a hospital-based violence intervention program at a public safety-net hospital and level 1 trauma center in San Francisco. We use a human-centered design approach to understand user needs and expectations and develop ICT that is tailored to the vulnerable population served by the WAP in San Francisco. Human-centered design is a well-established methodology used for problem-solving and innovation, employing ethnographic research to develop a deep understanding of users' unmet needs, and iteratively prototypes solutions to address those needs [31]. The methodology is rooted in empathy to ensure that design solutions address real problems, avoid designer bias, and build trust with the community [31-33]. The purpose of this study is to develop a deep understanding of WAP client and case manager needs and to identify key design opportunities that can inform the development of a smartphone-enabled ICT tool that scales the unique impact of case managers in hospital-based violence intervention programs nationwide.

Methods

Study Design and Setting

This prospective observational study is a collaboration between the San Francisco Wraparound Project (WAP) and The Better Lab, a mixed-methods research center located at the Zuckerberg San Francisco General (ZSFG) Hospital and Trauma Center that specializes in human-centered design research techniques. Study activities were conducted at the hospital, which is an affiliate of the University of California, San Francisco (UCSF), and the only level one trauma center in the city. The UCSF institutional review board approved this study as an expedited human subjects research protocol, and verbal consent was approved for all activities.

The San Francisco Wraparound Project

The San Francisco WAP offers intensive case management addressing the psychosocial needs of victims of violence aged 10-30 years at the ZSFG hospital who are at high risk for reinjury. Over the first 10 years of WAP implementation, ZSFG saw a 42% relative reduction in assault-related trauma reinjuries, and for every 100 clients served, WAP confers a net benefit of 24 quality-adjusted life years over standard practices [16,34]. The WAP case managers use their professional expertise and community connections to develop direct pipelines to resources, which include jobs, educational programs, legal services, mental health services, relocation services, emotional outlets in the arts and community service projects, and camps and retreats. These assets are operationalized through face-to-face interactions, text messages, phone calls, and emails between clients and case

managers to support and direct clients towards the tasks necessary for them to "graduate" from WAP. A client graduates once their major needs are met by the program or independent means.

Human-Centered Design

Human-centered design is an approach to problem-solving that is rooted in ethnography and iterative prototyping in order to develop solutions that are tailored to the end-user group. This methodology has been used to effectively deliver context-specific experiences, services, and products to manage diseases such as diabetes, post-traumatic stress disorder, and patient falls [35-37], with evidence suggesting that focusing on the end user yields more effective and sustainable results

[38-40]. The human-centered design method leverages ethnographic interviews and observations to develop rich insights and prioritizes iterative and collaborative solution development with its users at each stage of the process. We chose this approach to develop ICT that is intentionally designed for victims of violence and maximizes the impact of WAP case managers. This study focuses on the first step of the human-centered design process, the 'inspiration' phase, to develop a thorough understanding of users' experiences through interviews, followed by a synthesis of themes that will inform the design of human-centered ICT (Figure 1). We will refer to all subjects involved in this study, including clients, case managers, and other violence prevention specialists as "users."

Figure 1. Human-Centered Design Process.
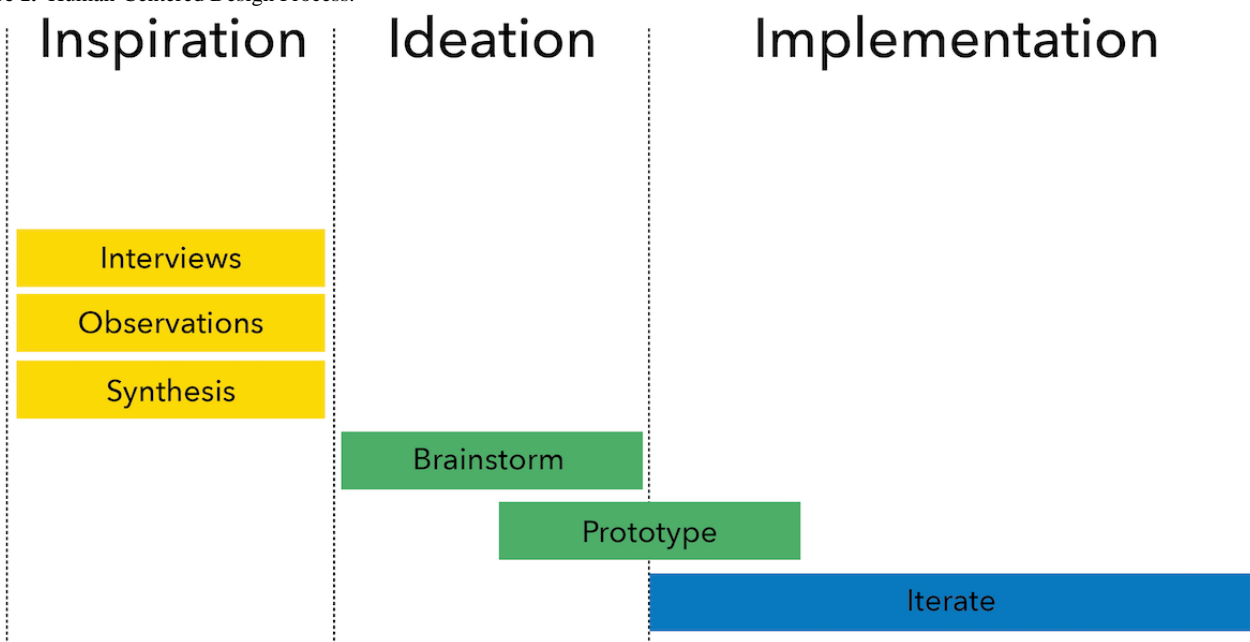
Qualitative Data Collection

Qualitative data were obtained through semi-structured user interviews to understand perspectives about WAP and the experience of being a victim of violence in the community. All interviews were anonymous, each lasting approximately 60 minutes at the WAP offices at ZSFG. Interviews were conducted in two phases. Phase one interviews were conducted with "violence intervention specialists" by a three-person team, including two design researchers from The Better Lab and one project manager from WAP. Interview subjects included WAP case managers and members of violence recovery and prevention organizations in the community. These interviews covered topics of daily workflow, communication activities, and barriers or challenges in their work. When appropriate and not violating the confidentiality of the clients, we asked our violence intervention specialists to describe experiences working with clients to illustrate challenges, barriers, and successes.

Phase two interviews were conducted with clients, ranging from 16 through 30 years old, and their friends and family. These interviews were performed by a four-person team, including two design researchers from The Better Lab, the client's case manager, and a project manager from WAP. Client interviews

covered topics including their personal background, violence history, experience with WAP, goals for the future, and thoughts on using technology to communicate with their case manager. Requests to audio record the interviews were declined due to the sensitive nature of the conversations. As a result, an additional design researcher was included during each interview to capture notes and quotes in real-time.

Interview Subject Selection and Recruitment

A total of 15 people participated in the study: 7 clients, 2 family members or friends of clients, 4 case managers, 1 program director from a partner violence prevention program, and 1 licensed clinical social worker specializing in crisis management. All participants were recruited through the WAP using purposeful sampling, a human-centered design approach to recruitment that draws a diverse cohort of subjects in order to represent all aspects of the victim of violence experience [41]. This approach to subject selection yields a heterogeneous sample of demographics and experiences with the WAP and violent injury community that can better inform the design of solutions. Table 1 describes each user. Participation was voluntary, and users from this second phase were compensated with $25 gift cards for their participation.

Table 1. Descriptions of Users.

Interviewee pseudonym	Description
Violence Prevention Specialists	
1	Longest-standing case manager with WAP who has a deep connection in the community.
2	Younger case manager from the neighborhoods WAP serves, seen as the role model for young men in the program.
3	Case manager with a background in grassroots community activism.
4	The only female-identifying case manager specializes in job placement.
5	Leader of a city initiative that conducts real-time violence mediation and street-level outreach to reduce street violence.
6	A licensed clinical social worker who works on crisis response by providing immediate emotional, mental, and logistical support following violent trauma.
Clients	
A	A client who graduated from WAP who is doing well and has not been reinjured.
B	The model client—has engaged with WAP to varying degrees for 5+ years without reinjury, proactive in seeking support when needed even after graduating, but also aims to support other clients and serve as a positive example.
C	A client who is a teenager recently returned to the Bay Area and re-engaging in WAP services—interviewed with a parent.
D	A client who is currently in WAP early in their work (ie, recently enrolled).
E	The parent/guardian of a client.
F	A client who is further along in WAP, but was reinjured by shooting.
G	A client who initially refused WAP services, but eventually enrolled.
H	A peer of a client who has not been a victim of violence.
I	A younger client (age early 20s), currently in the middle of the program.

Qualitative Data Analysis

Development of Insight Statements

The general inductive approach to thematic analysis was used to analyze interviews and to identify themes [42,43]. This approach was chosen to allow the emergence of themes that are closely related to the interview data [44]. First, all interview notes were read separately by two members of the research team. During this process, they identified specific themes that captured the core messages of the interviewees. Second, the researchers met to identify and describe common categories of themes. In this phase, descriptions of themes were developed and refined, and redundant themes were consolidated. Third, the researchers reviewed the interview notes a second time to extract key text or quotes associated with each thematic category. The data were reviewed a final time to ensure that a name, description, and supporting quotes for each thematic category were defined and consensus achieved.

Following the human-centered design method to qualitative analysis, researchers then extrapolated 'insight statements' from these themes. Insight statement development is an integral step in the human-centered design analysis process that involves re-reviewing the notes to understand themes in the context of the individual interviews and deduce unique human perspectives, motivations, or tensions from the thematic data. Insight statements ascribe meaning to the data and are then used to develop design opportunities [45].

Development of Design Opportunities

Design opportunities were developed from insight statements using the human-centered design approach to qualitative analysis. Design opportunities are action statements that provide direction to address tensions or challenges described in the insight. They guide the innovation process and solutions by framing and focusing the design effort [32,46]. Figure 2 depicts this process. The output of this process includes insight statements with supporting data and the associated design opportunity illustrated by a wireframe and list of features. These wireframes and features will be the first prototype to be iteratively tested with users in the next phase of this study.

Figure 2. Process for developing insights and opportunities using human-centered design.
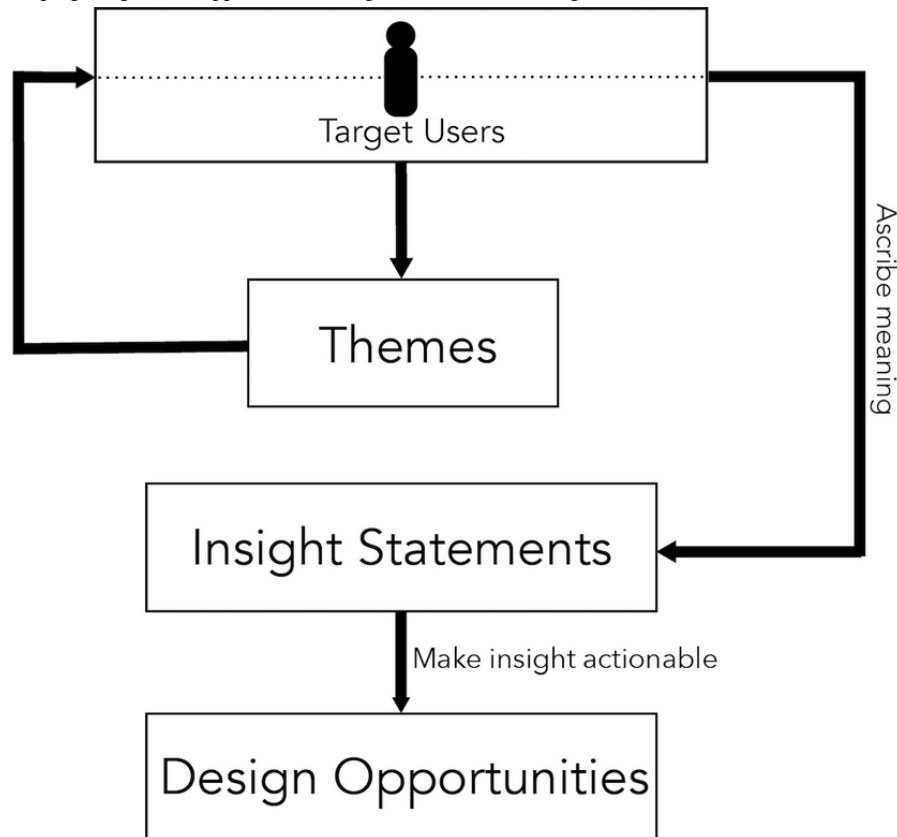
Results

Our analysis revealed four main insights that represent tensions or 'opposing needs' between the clients and case managers. These insights informed each subsequent design opportunity. This section presents the following for each insight: a description of the opposing needs of the case manager and client; an example of this tension from the interviews; the associated design opportunity; and an initial wireframe and list of features. Table 2 includes a summary of each insight with supporting quotations.

Insight 1: A Successful Relationship is Both Professional and Personal. Clients Need This Around the Clock, but Case Managers can Only Support This While on the Clock

Case managers are motivated by the steadfast goal of wanting to ensure success for their clients. However, maintaining both

a personal and professional relationship can be very disruptive to case managers' lives because it is challenging to set boundaries for clients who need support during nonbusiness hours. Clients are often in a vulnerable emotional, mental, and physical state and need a trusted confidant. The personal connection between case managers and clients is critical to WAP's success. Clients often consider the case managers to be a friend or a part of their family.

Example of Tension

To establish a trusting relationship, case managers often tell clients to call whenever there is an emergency. However, clients often call after hours with nonemergent issues, which leaves case managers feeling drained as they work around the clock. Clients value the deep bond they feel with their case managers and express a desire for continuous access in order to receive support at all hours across a range of issues.

Table 2. Supporting insight quotations from semi-structured interviews.

Insight and case manager quotations	Client quotations
Insight 1: A successful relationship is both professional and personal. Clients need this around the clock, but case managers can only support this while on the clock.	
Sometimes they share way too much. As workers, we sometimes need to look at the situation and get them help that they need… A lot of people who call just want someone to talk to. They trust the case managers because they're part of the community, but oftentimes they need a therapist… We explain that this is just our jobs. We explain that we need to talk to them in a certain way. [Subject 4]	[My case manager] does stuff he don't got to do… He is a big bro… I got to give him the same love he gives me. [Subject A]
We love you and want to support you, but this is our job. [Subject 4]	[My case manager] is my lifesaver, my #1 supporter... I've been through a lot with him. [Subject F]
I got a call at 2 am, which was just an excuse to call me... I'm a trusting ear. [Subject 4]	We're real tight. She's like family to me. [Subject G]
I've had clients call me on a Saturday. Some to pray. [Subject 1]	—
Insight 2: Communications need to feel personal, but they don't always need to be personalized.	
[The app needs to have] reminders. They will wait until the last hour, the last minute. [Subject 1]	Texting is cool, but I'd rather talk face to face. You can't tell what someone is going through on FB or text. [Subject C]
You need something to gather their attention... We need to give them something right now. [Subject 2]	Not really an email person, I would rather someone call... I also don't really check voicemails. [Subject A]
You've gotta create a plan. Everyone's looks different. [Subject 4]	[I] can talk to him like we're friends and text… he [case manager] also talks to my mom. [Subject I]
—	I feel good every time I talk to her. I can talk to her about anything... I know I'm safe when I talk to her. I know I don't have to worry about people knowing my business. [Subject G]
—	The communication was good, but it was kind of difficult because I took a night class... whenever [I] wanted to talk to [Subject 2], he would be off duty. [Subject I]
Insight 3: Healing is a journey of skill development and lifestyle changes that must be acknowledged, monitored, and rewarded.	
[After helping a client with housing], I need a way to hook you…something that can assure that I got you! [Subject 3]	The app should show accomplishments, something you can always return to and see… I did, and I can still do more. [Subject B]
I tell them, bring me a paycheck after thirty days. They gotta hold a job for thirty days. Then we'll get lunch. [Subject 4]	[This] is a complicated process…how can we make a roadmap? [Subject L]
We don't work for [the client], we work with [the client]. [Subject 3]	[Client referring to visual aspects of the app] "Once it's out of sight, you forget about it. [Subject D]
I will never give up on you. [Mistakes] are just a learning curve. [Subject 3]	—
Insight 4: Social networks need to provide peer support for healing, not peer pressure to propagate violence.	
Sometimes we'll monitor people through Facebook, see if they're posting suggestive videos [of violence]. [Subject 1]	The app should give a person a chance to help another person…to hear their stories and give advice. [Subject B]
If you have a permanent home, you're more likely to get hurt because there's a place to attach you to...social media doesn't help this. [Subject 3]	The app should maybe have a chatroom…start it anonymous with the option of revealing yourself. Some people feel comfortable typing, but don't feel comfortable sharing their feelings in person. [Subject G]
—	You're not the only person who got something… [being able to] chat with other people, that would be cool. [Subject I]
—	Nothing is safe on social media… [need] ability to make it private. [Subject D]

Design Opportunity 1: Maximize Personal Connection While Controlling Access

An ICT tool should be designed to encourage personal interactions while providing systems to differentiate nonemergent issues from emergencies in order to maintain professional boundaries for case managers. Figure 3 illustrates an initial prototype of this opportunity. The essential features include:

- Facilitate the role of the case manager as a trusted confidant by placing the onus of enforcing boundaries and deferring non-urgent client inquiries onto the system instead of on the individual case manager.

- Support clients at the greatest risk of reinjury by allowing case managers to negotiate and set client communication permissions on an individual basis.
- Minimize feelings of rejection by framing communication deferrals in terms that humanize the case managers (eg,

"Steve is out of the office right now and will be back on Monday. He is sleeping, is this an emergency?").
- Allow clients to escalate communications after hours as emergencies arise to ensure case manager response. As an example, the wireframe in Figure 3 depicts an example of an after-hours communication with a case manager.

Figure 3. Proposed depiction of a mobile application wireframe showing a conversation between a case manager and client. In this particular case, the client messages their case manager after hours. The mobile application allows the user to either escalate the message or wait until business hours.
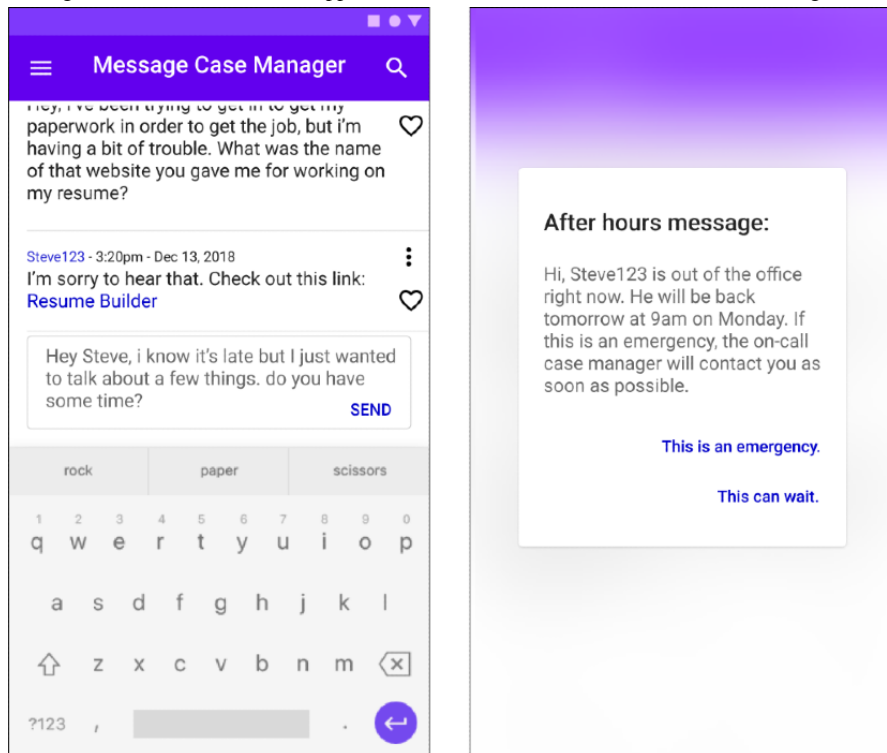
Insight 2: Communications Need to Feel Personal, but They do not Always Need to be Personalized

Case managers need a way to provide administrative and tactical communication to reduce the burden of continuous, individualized messaging. With a caseload of up to 10 clients, case managers need a way to attract and maintain their clients' attention that does not require them to craft each routine message or reminder manually. Clients want to communicate regularly and unpredictably with their case managers and value personal attention, positive feedback, and acknowledgment of their achievements. They prefer to meet their case managers face to face to discuss personal matters.

Example of Tension

Case managers spend significant time sending administrative messages such as appointment reminders and resume-building information. They want to focus on the high-value personal interactions and need a way to streamline the routine administrative tasks. However, clients cherish the personal relationship and fear that automating communications will lead

to an impersonal experience by diminishing their close relationship.

Design Opportunity 2: Allow Case Managers to Personalize Automated Client Interactions

An ICT tool must help case managers keep clients engaged through multiple communications and automate routine reminders that feel personal and are tailored to clients' unique needs. Figure 4 illustrates an initial prototype of this opportunity. The essential features include:

- Break down essential tasks such as building resumes, applying for schools, training, and jobs, searching for housing, and completing inquiries and applications for social services into sequential, achievable tasks. For example, the wireframe in Figure 4 focuses on the multiple and diverse tasks that are required to apply for a new job.
- Allow case managers and clients to set unique goals and deadlines for each task.
- Enable users to set automated reminders and suggest next steps after missed deadlines or successful task completion.

Figure 4. Proposed depiction of a mobile application wireframe showing a job application workflow. The client begins with a list of incomplete and complete tasks on the first screen. The screens Step 1 through Step 4 detail the step-by-step process of applying for a job, with options to ask questions, as well as a positive, affirming message to encourage the client to continue working towards their goals on the last screen.
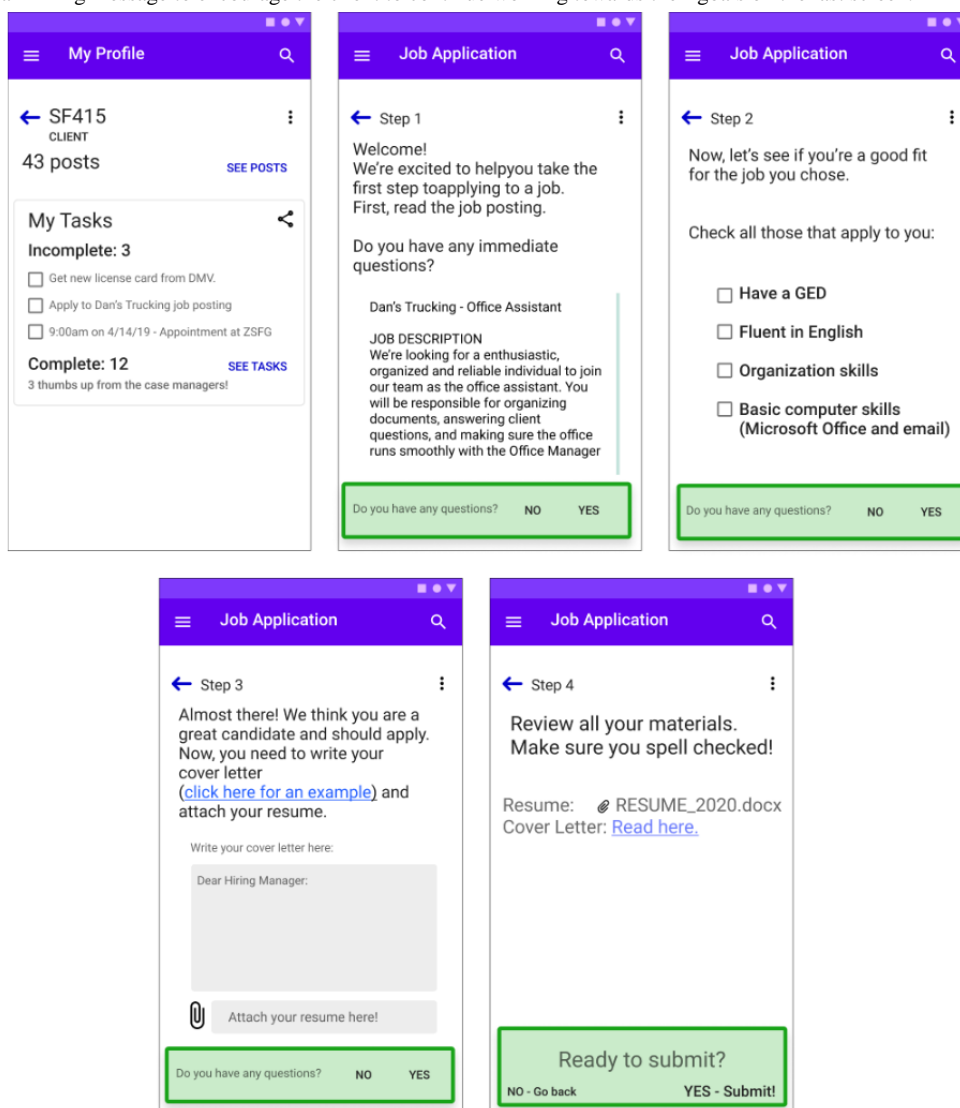
Insight 3: Healing is a Journey of Skill Development and Lifestyle Changes That Must be Acknowledged, Monitored, and Rewarded

Recovery after a violent injury extends far beyond physical recovery and includes significant lifestyle changes. Case managers work closely with clients to co-create a new future and support their achievements along the way. Case managers need a way to document, monitor, and share each client's unique journey in order to redefine expectations and set goals for their clients' future. They need to provide "tough love" by being stern on goals and expectations while offering compassion and encouragement to their clients. Clients need a clear roadmap for their recovery that includes feedback when they have gone off trajectory and celebration for the milestones they have successfully achieved.

Example of Tension

Case managers use their own systems to track a client's progress and send text messages or make reminder calls about upcoming appointments. When clients miss deadlines, case managers call

and text the client to discuss why it was missed and express their disappointment. Clients do not have tools to track their progress or access to their case manager's system, so they have no choice but to respond reactively to reminders without a clear picture of where they are in their journey.

Design Opportunity 3: Hold Clients Accountable for Progress and Reward Achievements

The ICT tool should incorporate a shared roadmap that clients and case managers can use to track progress. Figure 5 illustrates an initial prototype of this opportunity. The essential features include:

- Define and illustrate milestones throughout a client's recovery to measure success.
- Identify struggling clients early by implementing alerts to clients and case managers.
- Enable transparent and targeted interventions by allowing the case manager to explore each client's recent activity, historical trajectory, and future goals. As an example, the wireframe in Figure 5 focuses on how a case manager can

interact with and congratulate a client for a recent   accomplishment.

Figure 5. Proposed depiction of tracking a client's journey. Clients input their goals. Once a goal is marked as achieved, a photo of the achievement can be attached. These screens help a client visualize their successes and progress towards their goals.
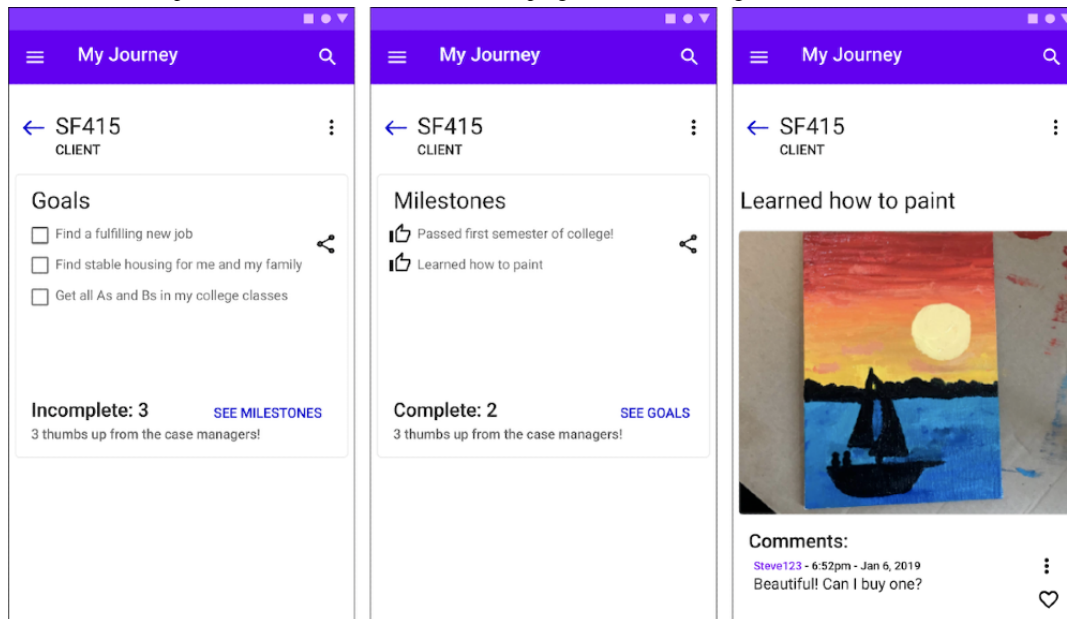
Insight 4: Social Networks Need to Provide Peer Support for Healing Rather Than Peer Pressure to Propagate Violence

Clients may be on opposing sides of violence within the community and need to know that there is a safe and confidential space to build community with peers who have experienced similar trauma. Clients want to share stories with others who have had similar experiences being a victim of violence. Case managers need a way to connect clients without compromising emotional, mental, or physical safety and security.

Example of Tension

Social media is a source of personal expression for clients, and some use it as their diary to document and articulate their emotions and thoughts. However, online platforms provide the opportunity to escalate interpersonal arguments and publicly pressure or publish threats of violence. Case managers do not have tools that allow them to create private, anonymized communities for their clients to connect, share, and learn from each other.

Design Opportunity 4: Build a Connected yet Confidential Community

The ICT should include the capability to create multiple community platforms that are curated by case managers where clients can engage anonymously. Figure 6 illustrates an initial prototype of this opportunity. The essential features include:

- Enable knowledge-sharing and community support by constructing a client social network. As an example, the wireframe below showcases a forum where clients can share their personal experiences.
- Mitigate violent communications by anonymizing identities, establishing clear rules for participation, enabling moderation by case managers, and providing training on internet safety.
- Stimulate useful conversations by structuring forums on topics of community interest so that clients can find the resources that they need.
- Ensure security with modern data-protection protocols such as HIPAA compliance.

Figure 6. Proposed depiction of a community forum and/or social network for clients to share resources and experiences with peers in an anonymous environment. The last screen depicts a reporting mechanism within the application to ensure safe interactions between clients within the community.
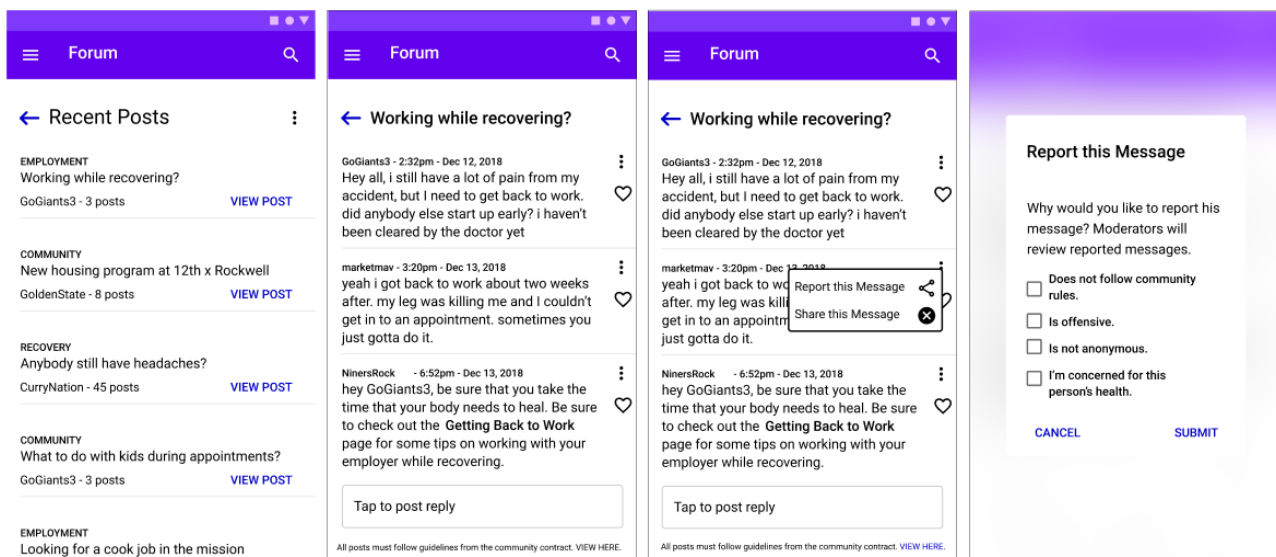
Discussion

This study used the human-centered design methodology to identify essential features for a smartphone-enabled ICT tool to support hospital-based violence intervention program clients and violence intervention specialists. The findings of this study identified opportunities and informed the initial wireframes that will be prototyped within the community of victims of violence in the next phase of the study. The human-centered design process was essential to this work because it allowed us to identify important opposing needs facing case managers and clients. Although case managers desired clear boundaries, automation of routine tasks, and caution in using social media, their clients wanted close relationships, personalized communication, and an online community. A unique ICT must be designed to address these opposing needs to support both stakeholder groups.

Processes and technology tools in health care tend to be designed in a silo and lack coordination with other aspects of care delivery [47]. The human-centered design method provides a structure that avoids the pitfalls common to traditional ICT development, such as lack of a deliberate process for innovation, all-in-one solutions that do not adequately meet users' needs, or solutions that are developed without user input [48,49]. For example, phase one of our interviews with case managers revealed that an ICT tool needed to focus primarily on efficiency to allow them to reach more clients. However, phase two of our interviews with clients revealed that the personal connection is what makes the violence intervention program successful. This tension between automation and personalization, among others described here, made it clear that the features of an effective ICT tool would require the artful synthesis of opposing needs that have not yet been addressed.

Thoughtful ICT is important in health care, but essential for the violence community because a poorly designed tool could compromise the safety of our clients and case managers. Our clients are a technologically savvy group with high digital literacy and significant social media activity. Although we identified an opportunity to build a social community for peer support, social networking in this population has the potential to amplify threats of violence in the form of what is known as "internet banging" [50]. Surveys of advocates for and victims of domestic violence and stalking show that technology can be used to threaten and perpetrate psychological and social attacks [51]. We identified these same concerns in our qualitative research, which reinforced the fact that an ICT tool that replicates current social media forums is both inadequate and potentially dangerous. A social networking tool on an ICT application for victims of violence needs features to create secured, anonymous, and moderated conversation forums.

Limitations

There are several limitations to our study and the human-centered design methodology. The features of the ICT tool we have described here are specific to our population and environment and represent early prototypes that will be iterated on in the next phase of the design process. The wireframes and their associated features are preliminary ideas developed by the design researchers in order to represent the design opportunity visually. The content of the wireframes is based on quotes and insights from our users. These wireframes have not been tested with users. Design and implementation specifications will be the focus of the next phase of our study. We anticipate that the features identified in this phase may change through our iterative process and may not be generalizable to other organizations. However, the broader insights and opportunities are likely transferable to other organizations interested in developing an ICT tool for victims of violence. Other settings that support vulnerable populations in health care may also find the human-centered design methodology to be an effective way to uncover previously unknown needs and potential tensions in the communities they aim to support through ICT.

Lastly, we used a purposeful sampling methodology in order to capture a diverse set of experiences in a time-efficient way, which is an essential part of the rapid iteration process at the

heart of human-centered design. However, this approach yields a relatively small sample size. For example, this study only included one peer and one parent/guardian representative of WAP clients. While this is not uncommon in human-centered design, we acknowledge that this small number of users may limit the generalizability of those findings to other communities and settings. Additionally, while interviewees were intentionally chosen based on the breadth and diversity of their experience in the community in order to provide information-rich insights and design opportunities, this approach to sampling carries a risk of selection bias. We attempted to mitigate this risk by asking our four experienced case managers to select the most diverse interview subjects.

Next Steps

This study identified key insights and design opportunities to develop a mobile ICT tool that meets the personal needs of WAP clients while enabling case managers to support their clients' goals effectively. The opportunities presented here will serve as core principles when developing, testing, and refining wireframes in the next phase of this work.

Conclusions

The success of hospital-based violence intervention programs is predicated on the value of intensive case management, ready access to services, and securing personal and professional development opportunities for the clients [52]. Meeting these goals requires that case managers effectively and efficiently provide life-changing resources. This study has contributed to the understanding of violence intervention program clients and case managers in the design and development of an ICT tool for case management. Importantly, our findings indicate that the use of human-centered design uncovered previously unknown tensions between the needs of clients and case managers. Designers of ICT tools for this population must consider how issues of safety, privacy, automation, personal connection, and efficiency impact their users.

Conflicts of Interest

AS, LZC, and DP are human-centered design consultants for The Empathy Studio, LLC. SS recently joined Anthem, Inc, as a Medical Director of Digital Technologies, where he is involved in digital health development that incorporates human-centered design. All other authors declare they have no competing interests.

References

1. Office of the Surgeon General. Surgeon General's Workshop on Violence and Public Health Report. Department of Health and Human Services & Department of Justice. 1986 May. URL: https://www.nlm.nih.gov/exhibition/confrontingviolence/materials/OB10998.pdf [accessed 2019-08-05] [WebCite Cache ID 72fpp7MoF]

2. Dalhberg L, Mercy, JA. The History of Violence as a Public Health Problem. AMA Journal of Ethics 2009;11(2):167-172. [doi: 10.1001/virtualmentor.2009.11.2.mhst1-0902]

3. CDC National Center for Injury Prevention Control. Web-based injury statistics query and reporting system (WISQARS). CDC. URL: https://www.cdc.gov/injury/wisqars/index.html [accessed 2019-01-10] [WebCite Cache ID 75KQORVC8]

4. Hankin A, Meagley B, Wei SC, Houry D. Prevalence of Exposure to Risk Factors for Violence among Young Adults Seen in an Inner-City Emergency Department. West J Emerg Med 2013 Aug;14(4):303-308 [FREE Full text] [doi: 10.5811/westjem.2013.2.14810] [Medline: 23930142]

5. Rivara FP, Shepherd JP, Farrington DP, Richmond PW, Cannon P. Victim as offender in youth violence. Ann Emerg Med 1995 Nov;26(5):609-614. [doi: 10.1016/s0196-0644(95)70013-7] [Medline: 7486371]

6. Green B, Horel T, Papachristos AV. Modeling Contagion Through Social Networks to Explain and Predict Gunshot Violence in Chicago, 2006 to 2014. JAMA Intern Med 2017 Mar 01;177(3):326-333. [doi: 10.1001/jamainternmed.2016.8245] [Medline: 28055070]

7. Newgard CD, Schmicker RH, Sopko G, Andrusiek D, Bialkowski W, Minei JP, Resuscitation Outcomes Consortium Investigators. Trauma in the neighborhood: a geospatial analysis and assessment of social determinants of major injury in North America. Am J Public Health 2011 Apr;101(4):669-677. [doi: 10.2105/AJPH.2010.300063] [Medline: 21389292]

8. Morenoff JD, Sampson RJ, Raudenbush SW. Neighborhood Inequality, Collective Efficacy, and the Spatial Dynamics of Urban Violence. Criminology 2001 Aug;39(3):517-558. [doi: 10.1111/j.1745-9125.2001.tb00932.x]

9. Rich JA, Grey CM. Pathways to recurrent trauma among young Black men: traumatic stress, substance use, and the "code of the street". Am J Public Health 2005 May;95(5):816-824. [doi: 10.2105/AJPH.2004.044560] [Medline: 15855457]

10. Goins WA, Thompson J, Simpkins C. Recurrent intentional injury. J Natl Med Assoc 1992 May;84(5):431-435. [Medline: 1495116]

11. Sims DW, Bivins BA, Obeid FN, Horst HM, Sorensen VJ, Fath JJ. Urban trauma: a chronic recurrent disease. J Trauma 1989 Jul;29(7):940-947. [Medline: 2746704]

12. Cunningham RM, Carter PM, Ranney M, Zimmerman MA, Blow FC, Booth BM, et al. Violent reinjury and mortality among youth seeking emergency department care for assault-related injury: a 2-year prospective cohort study. JAMA Pediatr 2015 Jan;169(1):63-70 [FREE Full text] [doi: 10.1001/jamapediatrics.2014.1900] [Medline: 25365147]

13. Turcotte Benedict F, Amanullah S, Linakis JG, Ranney M. Emergency Department Utilization Among Assault-Injured Youth: Implications for Youth Violence Screening. Pediatr Emerg Care 2017 Sep;33(9):607-612 [FREE Full text] [doi: 10.1097/PEC.0000000000000609] [Medline: 26785098]

14. Bingenheimer JB, Brennan RT, Earls FJ. Firearm violence exposure and serious violent behavior. Science 2005 May 27;308(5726):1323-1326 [FREE Full text] [doi: 10.1126/science.1110096] [Medline: 15919997]

15. Johnson SB, Bradshaw CP, Wright JL, Haynie DL, Simons-Morton BG, Cheng TL. Characterizing the teachable moment: is an emergency department visit a teachable moment for intervention among assault-injured youth and their parents? Pediatr Emerg Care 2007 Aug;23(8):553-559. [doi: 10.1097/PEC.0b013e31812c6687] [Medline: 17726415]

16. Juillard C, Smith R, Anaya N, Garcia A, Kahn JG, Dicker RA. Saving lives and saving money: hospital-based violence intervention is cost-effective. J Trauma Acute Care Surg 2015 Feb;78(2):252-257. [doi: 10.1097/TA.0000000000000527] [Medline: 25757108]

17. Strong BL, Shipper AG, Downton KD, Lane WG. The effects of health care-based violence intervention programs on injury recidivism and costs: A systematic review. J Trauma Acute Care Surg 2016 Nov;81(5):961-970. [doi: 10.1097/TA.0000000000001222] [Medline: 27537505]

18. Gonzalez VM, Dulin PL. Comparison of a smartphone app for alcohol use disorders with an Internet-based intervention plus bibliotherapy: A pilot study. Journal of Consulting and Clinical Psychology 2015 Apr;83(2):335-345. [doi: 10.1037/a0038620]

19. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes Self-Management Smartphone Application for Adults With Type 1 Diabetes: Randomized Controlled Trial. J Med Internet Res 2013 Nov 13;15(11):e235. [doi: 10.2196/jmir.2588]

20. Zhao J, Freeman B, Li M. Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]

21. Post LA, Vaca FE, Biroscak BJ, Dziura J, Brandt C, Bernstein SL, et al. The Prevalence and Characteristics of Emergency Medicine Patient Use of New Media. JMIR Mhealth Uhealth 2015 Jul 08;3(3):e72 [FREE Full text] [doi: 10.2196/mhealth.4438] [Medline: 26156096]

22. Ray M, Dayan PS, Pahalyants V, Chernick LS. Mobile Health Technology to Communicate Discharge and Follow-Up Information to Adolescents From the Emergency Department. Pediatr Emerg Care 2016 Dec;32(12):900-905. [doi: 10.1097/PEC.0000000000000970] [Medline: 27898635]

23. Pew Research Center. Internet/Broadband Fact Sheet. Pew Research Center. URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2019-08-05]

24. Anderson M. Racial and ethnic differences in how people use mobile technology. Pew Research Center. URL: http://www.pewresearch.org/fact-tank/2015/04/30/racial-and-ethnic-differences-in-how-people-use-mobile-technology/ [accessed 2019-08-05]

25. Stowell E, Lyson MC, Saksono H, Wurth RC, Jimison H, Pavel M, et al. Designing and Evaluating mHealth Interventions for Vulnerable Populations. CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems 2018;1:1-17. [doi: 10.1145/3173574.3173589]

26. Hicks ER. Digital Citizenship and Health Promotion Programs: The Power of Knowing. Health Promot Pract 2017 Jan;18(1):8-10. [doi: 10.1177/1524839916676263] [Medline: 27815394]

27. Vangeepuram N, Mayer V, Fei K, Hanlen-Rosado E, Andrade C, Wright S, et al. Smartphone ownership and perspectives on health apps among a vulnerable population in East Harlem, New York. Mhealth 2018;4(31). [doi: 10.21037/mhealth.2018.07.02] [Medline: 30221166]

28. Montague E, Perchonok J. Health and wellness technology use by historically underserved health consumers: systematic review. J Med Internet Res 2012;14(3):e78 [FREE Full text] [doi: 10.2196/jmir.2095] [Medline: 22652979]

29. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]

30. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]

31. Brown T. Design Thinking. Harvard Business Review. URL: https://new-ideo-com.s3.amazonaws.com/assets/files/pdfs/IDEO_HBR_DT_08.pdf [accessed 2020-08-19]

32. Ideo.org. The Field Guide to Human-Centered Design. URL: http://www.designkit.org/resources/1 [accessed 2019-08-05]

33. Ideo. Design Thinking. URL: https://designthinking.ideo.com/ [accessed 2019-08-05]

XSL•FO

RenderX

34. Juillard C, Cooperman L, Allen I, Pirracchio R, Henderson T, Marquez R, et al. A decade of hospital-based violence intervention: Benefits and shortcomings. J Trauma Acute Care Surg 2016 Dec;81(6):1156-1161. [doi: 10.1097/TA.0000000000001261] [Medline: 27653168]

35. LeRouge C, Wickramasinghe N. A review of user-centered design for diabetes-related consumer health informatics technologies. J Diabetes Sci Technol 2013 Jul;7(4):1039-1056 [FREE Full text] [Medline: 23911188]

36. Owen JE, Kuhn E, Jaworski BK, McGee-Vincent P, Juhasz K, Hoffman JE, et al. VA mobile apps for PTSD and related problems: public health resources for veterans and those who care for them. Mhealth 2018;4(28) [FREE Full text] [doi: 10.21037/mhealth.2018.05.07] [Medline: 30148141]

37. Harte R, Quinlan LR, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, et al. Human-Centered Design Study: Enhancing the Usability of a Mobile Phone App in an Integrated Falls Risk Detection System for Use by Older Adult Users. JMIR Mhealth Uhealth 2017 May 30;5(5):e71 [FREE Full text] [doi: 10.2196/mhealth.7046] [Medline: 28559227]

38. Altman M, Huang TTK, Breland JY. Design Thinking in Health Care. Prev Chronic Dis 2018 Sep 27;15:e117 [FREE Full text] [doi: 10.5888/pcd15.180128] [Medline: 30264690]

39. Compton-Phillips A, Mohta N. Care Redesign Survey: How Design Thinking Can Transform Health Care. NEJM Catalyst 2020 Jun 07 . [doi: 10.1056/cat.20.0130]

40. Kachirskaia I, Mate K, Neuwirth E. Human-Centered Design and Performance Improvement: Better Together. NEJM Catalyst 2018 Jun 28 [FREE Full text]

41. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]

42. Thomas DR. A General Inductive Approach for Analyzing Qualitative Evaluation Data. American Journal of Evaluation 2016 Jun 30;27(2):237-246. [doi: 10.1177/1098214005283748]

43. Frith H, Gleeson K. Clothing and Embodiment: Men Managing Body Image and Appearance. Psychology of Men & Masculinity 2004 Jan;5(1):40-48. [doi: 10.1037/1524-9220.5.1.40]

44. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

45. Dalton J. What is an Insight? The 5 Principles of Insight Definition. URL: https://thrivethinking.com/2016/03/28/what-is-insight-definition/ [accessed 2019-08-04]

46. Bridge Innovate. Turning Insights into Opportunities. URL: https://www.bridgeinnovate.com/blog/2017/12/19/turning-insights-into-opportunities [accessed 2020-08-19]

47. Kelly Y, Goodwin D, Wichmann L, Mendu ML. Breaking Down Health Care Silos. Harvard Business Review. 2019 Jul. URL: https://www.hbr.org/2019/07/breaking-down-health-care-silos/ [accessed 2019-08-04]

48. Blevins K, Azagury D, Wall J, Chandra V, Wynne E, Krummel T. How Good Ideas Die: Understanding Common Pitfalls of Medtech Innovation. In: Behrns KE, Gingles B, Sarr MG, editors. Medical Innovation: Concept to Commercialization. London: Academic Press; 2018:117-127.

49. Liappas N, Teriús-Padrón J, García-Betances R, Cabrera UM, Arredondo M. User-Centered Design in Defining and Developing Health Well-Being ICT Solutions. In: Chen F, García-Betances RI, Chen L, Cabrera Umpiérrez MF, Nugent C, editors. Smart Assisted Living. Computer Communications and Networks. Switzerland: Springer, Cham; 2019.

50. Patton DU, Eschmann RD, Elsaesser C, Bocanegra E. Sticks, stones and Facebook accounts: What violence outreach workers know about social media and urban-based gang violence in Chicago. Computers in Human Behavior 2016 Dec;65:591-600. [doi: 10.1016/j.chb.2016.05.052]

51. Woodlock D. The Abuse of Technology in Domestic Violence and Stalking. Violence Against Women 2017 Apr;23(5):584-602. [doi: 10.1177/1077801216646277] [Medline: 27178564]

52. Smith R, Dobbins S, Evans A, Balhotra K, Dicker RA. Hospital-based violence intervention: risk reduction resources that are essential for success. J Trauma Acute Care Surg 2013 Apr;74(4):976-980. [doi: 10.1097/TA.0b013e31828586c9] [Medline: 23511134]

Abbreviations

ICT: information and communication technology
UCSF: University of California at San Francisco
WAP: The Wraparound Project
ZSFG: Zuckerberg San Francisco General

XSL•FO

RenderX

<u>Original Paper</u>

Mobile App for Symptom Management and Associated Quality of Life During Systemic Treatment in Early Stage Breast Cancer: Nonrandomized Controlled Prospective Cohort Study

Cvetka Grašič Kuhar[1,2], MD, PhD; Tjaša Gortnar Cepeda[2], MD; Timotej Kovač[3], BSc; Matjaž Kukar[3], MSc, PhD; Nina Ružić Gorenjec[4], BSc, PhD

[1]Department of Medical Oncology, Institute of Oncology Ljubljana, Ljubljana, Slovenia

[2]Faculty of Medicine Ljubljana, University of Ljubljana, Ljubljana, Slovenia

[3]Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia

[4]Institute for Biostatistics and Medical Informatics, Faculty of Medicine Ljubljana, University of Ljubljana, Ljubljana, Slovenia

Corresponding Author:
Cvetka Grašič Kuhar, MD, PhD
Department of Medical Oncology
Institute of Oncology Ljubljana
Zaloška 2
Ljubljana, SI-1000
Slovenia
Phone: 386 1 5879 287
Fax: 386 1 5879 303
Email: cgrasic@onko-i.si

Abstract

Background: Providing patients with cancer who are undergoing systemic therapy with useful information about symptom management is essential to prevent unnecessary deterioration of quality of life.

Objective: The aim was to evaluate whether use of an app for symptom management was associated with any change in patient quality of life or use of health resources.

Methods: Outpatients with early stage breast cancer receiving systemic therapy were recruited at the Institute of Oncology in Ljubljana, Slovenia. Patients who received systemic therapy between December 2017 and March 2018 (control group) and between April 2018 and September 2018 (intervention group) were eligible. All patients received standard care, but only those in the intervention group were asked to use mPRO Mamma, an Android-based smartphone app, in addition. The app supported daily tracking of 50 symptoms, allowed users to grade their symptom severity (as mild, moderate, or severe), and also provided in-depth descriptions and recommendations based on reported symptom level. Patient-reported outcomes in both groups were assessed through the European Organisation for Research and Treatment of Cancer (EORTC) core (C-30) and breast cancer (BR-23) questionnaires, as well as a questionnaire about health resources use. The primary outcomes were the difference in the global quality of life between groups and the difference in summary score of the EORTC C-30 questionnaire between groups after 3 time periods (the first week of treatment, the first treatment cycle, and the entire treatment). The secondary outcome was the use of health resources (doctor visits and hospitalizations) in each time period. Other scales were used for exploratory analysis.

Results: The mean difference between the intervention group (n=46) and the control group (n=45) in global quality of life (adjusted for baseline and type of surgery) after the first week was 10.1 (95% CI 1.8 to 18.5, *P*=.02). The intervention group summary scores were significantly higher than those of the control group after the first week (adjusted mean difference: 8.9, 95% CI 3.1 to 14.7, *P*=.003) and at the end of treatment (adjusted mean difference: 10.6, 95% CI 3.9 to 17.3, *P*=.002). Use of health resources was not statistically significant between the groups in either the first week (*P*=.12) or the first treatment cycle (*P*=.13). Exploratory analysis findings demonstrated clinically important improvements (indicated by EORTC C-30 or BR-23 scale scores)—social, physical, role, and cognitive function were improved while pain, appetite loss, and systemic therapy side effects were reduced.

Conclusions: Use of the app enabled patients undergoing systemic therapy for early stage breast cancer to better cope with symptoms which was demonstrated by a better global quality of life and summary score after the first week and by a better

XSL•FO

RenderX

summary score at the end of treatment in the intervention group compared to those of the control group, but no change in the use of health resources was demonstrated.

KEYWORDS

breast cancer; systemic therapy; mobile application; patient-reported outcome; quality of life

Introduction

In the context of personalized treatment of a patient with cancer, the patient-reported outcome is gaining growing importance. This is because the physician's ability to evaluate the patient's subjective symptoms is not optimal. Physicians often underestimate the severity and frequency of symptoms experienced by patients during chemotherapy. Fromme et al [1] analyzed physician and patient reports on 8 symptoms in a metastatic prostate cancer chemotherapy trial and found that the physician reports had low sensitivity and specificity for detecting chemotherapy side effects [1]. Moreover, the agreement between multiple physicians was, at best, moderate [2]. Physician assessment of symptom presence and severity differed by one to two grades according to the Common Toxicity Criteria of Adverse Effects [3] which, the study [2] concluded, would have resulted in different treatment decisions (such as a difference between continuing chemotherapy, halting chemotherapy, or changing treatment dosage). To better assess patient safety and toxicity in clinical studies of treatments, the National Cancer Institute of the United States of America developed a patient-reported outcome version of the Common Toxicity Criteria of Adverse Effects (PRO-CTCAE), which consists of 78 adverse effects criteria that are appropriate for patient self-reporting. For each adverse effect, one or more descriptive measures such as the adverse effect severity, frequency, and interference with daily activity are assessed yielding 124 PRO-CTCAE items, each graded on a 5-point scale (from 0 to 4, where 0 indicates the adverse effect is absent, and 4 indicates the most disabling adverse effect) [4]. An appropriate subset of these items should be selected for each clinical trial; however, most patients are treated outside of clinical trials.

With the widespread use of smartphones, electronic collection of patient-reported outcomes is becoming possible. Several studies [5-12] have shown the feasibility and clinical utility of electronic capture and monitoring of patient-reported outcome for symptoms, functional status, and quality of life during systemic treatment of cancer. In advanced solid tumors, Basch et al [5] showed improved health-related quality of life measures and reduced emergency room admittance in patients who received weekly notifications and reported their symptoms (also weekly) [5]. Integration of the patient-reported outcome into routine care was even shown to prolong overall survival in patients with advanced solid tumors by 5 months [13]. With increased patient willingness to use technology, there are possibilities for expansion of patient-centered medicine [14-16]. At present, few apps are used for the self-management of symptoms [7,9-11,14]. In one study [10], higher cognitive function and a high symptom burden in patients were predictive

factors for whether they would use symptom self-management strategies.

The aim of our study was to assess whether the use of our innovative mobile app by patients with early stage breast cancer who were receiving systemic treatment was associated with any changes in their health-related quality of life (global quality of life, ability to function in daily life, symptom severity) and in the use of health resources. Our hypothesis was that the mobile app users would have better health-related quality of life outcomes, especially after the first week and after the first cycle of chemotherapy, and would use fewer health resources.

Methods

Study Design and Population

We conducted a prospective nonrandomized controlled study with a cohort consisting of outpatients who were being treated with systemic therapy (chemotherapy) for early stage breast cancer at the Institute of Oncology in Ljubljana. The study protocol was approved by the Ethical Commission at the Institute of Oncology and Commission of the Republic of Slovenia for Medical Ethics (0120-386/2017/3). The researchers followed the principles of the Declaration of Helsinki.

Patients were recruited after their first appointment with a medical oncologist. All study information was provided by oncologists, medical residents, and medical students working on the project. Patients were eligible if they were scheduled to receive neoadjuvant or adjuvant chemotherapy (and anti–human epidermal growth factor 2 targeted therapy, if indicated) for early stage breast cancer, their smartphone was Android-based, and they were willing to fill in paper and pencil questionnaires reporting their quality of life while receiving treatment. The patients were ineligible if their smartphone was Windows-based or an iPhone, they were unfamiliar with using a smartphone, or they were unable to communicate effectively in Slovenian. All participants provided informed consent prior to being enrolled in the study.

Study Group and Intervention

The control group consisted of patients enrolled between December 2017 and March 2018; the intervention group consisted of patients enrolled between April 2018 and September 2018 (since the mobile app was not finished until April). The control group received the standard information about chemotherapy side effects. The intervention group received both the standard information and access to the mPRO Mamma mobile app [17] which was accompanied by a detailed user instruction guide.

Study Procedures

Patients in both groups were asked to complete the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ C-30) and Quality of Life Questionnaire Breast Cancer Module (EORTC QLQ BR-23) [18,19] and a questionnaire about health resources use at 4 time points: before systemic therapy (baseline), after the first week, after the first cycle, and at the end of systemic therapy. Upon enrollment, each patient received questionnaires (4 sets) along with instructions to return completed questionnaires by mail in a timely manner. The patients who returned the baseline questionnaires and questionnaires from at least one other time point were included in the analysis.

Mobile App

For the purpose of this study, we developed an Android mobile app called mPRO Mamma [17]. It conforms to standard Google Material Design guidelines [20], but additionally features customized slider controls to visually indicate symptom severity levels (Figure 1). It allows for quick daily recording of symptoms and symptom severity (with reminder notifications) and can send encrypted reports to the patient's oncologist, if selected. The app was designed and revised several times in cooperation with experienced medical oncologists, medical students, and computer scientists. It has been implemented according to the best principles of Material Design multilingual support and currently supports Slovenian and English.

Figure 1. The mobile app user interface with customized slider controls allows for quick daily recording of symptoms and their severity levels.

We selected 50 symptoms that were relevant to the expected adverse effects of either cytotoxic regimens or targeted drugs administered to patients during the neoadjuvant or adjuvant breast cancer treatments from the 78 symptoms of PRO-CTCAE. Each symptom had a description, a severity scale, and advice for its management. Patients labeled and reported a symptom only if it occurred. To make our app clear and useful, symptom severity was graded on a 3-point scale as mild, moderate, or severe.

For each grade of symptom severity, a health professional's advice was presented. For mild or moderate ratings by the patient, suggestions that would allow them to alleviate the symptoms on their own were displayed. If symptoms were rated as severe, the app advised the patient to visit either their general practitioner or an emergency department. Patients reported their symptoms in the app on a daily basis and sent the symptom report to their treating oncologist as an encrypted email on the day before their next therapy cycle (every 2 or 3 weeks, depending on their chemotherapy schedule). Screenshots of the symptom entry page and daily report are presented in Multimedia Appendix 1, Multimedia Appendix 2, and Multimedia Appendix 3.

Outcome Measures

The EORTC QLQ C-30 and BR-23 consist of a global quality of life score, and several functional and symptom scales, which can be combined in a summary score. Our primary outcomes were the global quality of life score and summary score at three time points (after the first week, after the first cycle, and at the end of treatment). Other scales were used in our exploratory analysis. The EORTC QLQ scoring manual was followed [21]. All scales had values from 0 to 100, where 100 represented the best global quality of life, the best functioning, or the worst symptoms. The summary score, which ranged from 0 (worst) to 100 (best), was calculated from 13 out of 15 EORTC

QLQ-C30 scales (the global quality of life and the financial impact scale were excluded) in accordance with Giesinger et al [22] and instructions from the EORTC [21-23]. We used previously established thresholds [24] for important differences in quality of life—clinically meaningful difference was set to 10 points.

The secondary outcome was the use of health resources which was measured by self-reported number of visits to the doctor and number of hospitalizations in the first week and in the first cycle of therapy (questionnaires at the end of treatment were not included because they may not have been reliable due to the long time span between the start and the end of treatment).

Statistical Analysis

Sample Size Determination

A sample size sufficient to detect a difference between the intervention and control groups for changes of the global quality of life scores from baseline scores using an independent t test was determined using a power analysis. For a significance level of α=.05 and 80% power to reject the null hypothesis of equal means of groups when the population mean difference is equal to 13 yielded a sample size of 90 (45 per group); a standard deviation of 22 was used for both groups based on the results from a recent randomized trial [25]. It was subsequently determined that a more advanced analysis had to be conducted, as explained below.

Data Analysis

Categorical variables (tumor size, axillary nodes, histologic type, tumor grade, estrogen receptors, progesterone receptors, HER2, Ki-67 expression, tumor subtype, type of surgery, breast reconstruction, chemotherapy, type of chemotherapy, doctor visits during first week, doctor visits during the first cycle, hospitalization during the first week, and hospitalization during the first cycle) were summarized with frequencies and percentages. Numerical variables (age, number of chemotherapy cycles, global quality of life score, and summary score) were described with means and standard deviations (or medians and interquartile ranges if distributions were asymmetric). To compare the intervention and control groups at baseline, chi-square tests (or Fisher exact test, if more than 20% of expected frequencies were below 5) were used for categorical variables, and two-tailed independent t tests (or Mann-Whitney U tests, if distributions were asymmetric) were used for numerical variables.

It was required that the model account for differences between the groups in the type of surgery (a higher proportion of mastectomies, and consequently, reconstruction in the control group), as well as in quality of life and summary score at baseline. The primary outcomes of global quality of life score and summary score were analyzed using linear mixed-effects models [26]. Linear mixed-effects models allowed for the use of one model to represent all three time points, adjusting for covariates at baseline and where a random intercept (indexed by patient ID) accounted for repeat measurements of the same patient. In each model, the fixed effects were time (after the first week, after the first cycle, or at the end of treatment), group (intervention or control), interaction between group and time, value at baseline, and type of surgery (breast-conserving surgery or mastectomy).

In the exploratory analysis, we examined all 19 other scales of the QLQ-C30 and the QLQ-BR23 (excluding sexual enjoyment and upset by hair loss, which were both dependent upon answering affirmative to other questions, and excluding financial difficulties). The change from baseline between the groups at each of the three time points was analyzed separately using the Mann-Whitney U test because of either distribution asymmetry or the low number of possible values of the scale. All P values in the exploratory analysis served only as an aid to evaluate the differences between the groups. Nevertheless, with so many hypotheses tested, the P values were adjusted using Holm method to control the family-wise error rate.

An (adjusted) P<.05 was considered as statistically significant. All analyses were performed using R statistical software (version 3.4.3) [27].

Results

Patient Characteristics

The study flow chart is presented in Figure 2. We performed a CONSORT-EHEALTH checklist (Multimedia Appendix 4) according to pilot testing [28].

The demographic, clinical, and treatment data for the control (n=45) and the intervention (n=46) groups are presented in Table 1. Patients in both groups were similar with respect to age (t_{89}=0.86, P=.39), tumor size (Fisher exact test, P=.48), nodal involvement (Fisher exact test, P=.13), histology (Fisher exact test, P=.42), tumor grade (Fisher exact test, P=.38), hormone receptors (progesterone: χ^2_1=1.23, P=.27; estrogen: χ^2_1=2.33, P=.13), human epidermal growth factor 2 status (χ^2_1=1.76, P=.19), tumor subtype (Fisher exact test, P=.30), type of chemotherapy (Fisher exact test, P=.41), and number of chemotherapy cycles (Mann-Whitney U test, P=.73); however, there was a significant difference between the intervention and control groups with respect to type of surgery (χ^2_3=12.60, P=.006). Patients in the control group had undergone more mastectomies, and consequently, had also undergone reconstruction more often (Fisher exact test, P=.01). At baseline, patients in the control group had lower quality of life and summary scores than those of the intervention group (quality of life: Mann-Whitney U test, P=.06; summary: Mann-Whitney U test, P=.02). Baseline differences in quality of life and summary scores and for type of surgery (mastectomy or not) between the groups were, therefore, taken into account and adjusted for, within the model.

Figure 2. Consolidated Standard of Reporting Trials (CONSORT) diagram showing the flowchart of patients with breast cancer by group and reasons for exclusion from the study and the analysis.

XSL•FO

RenderX
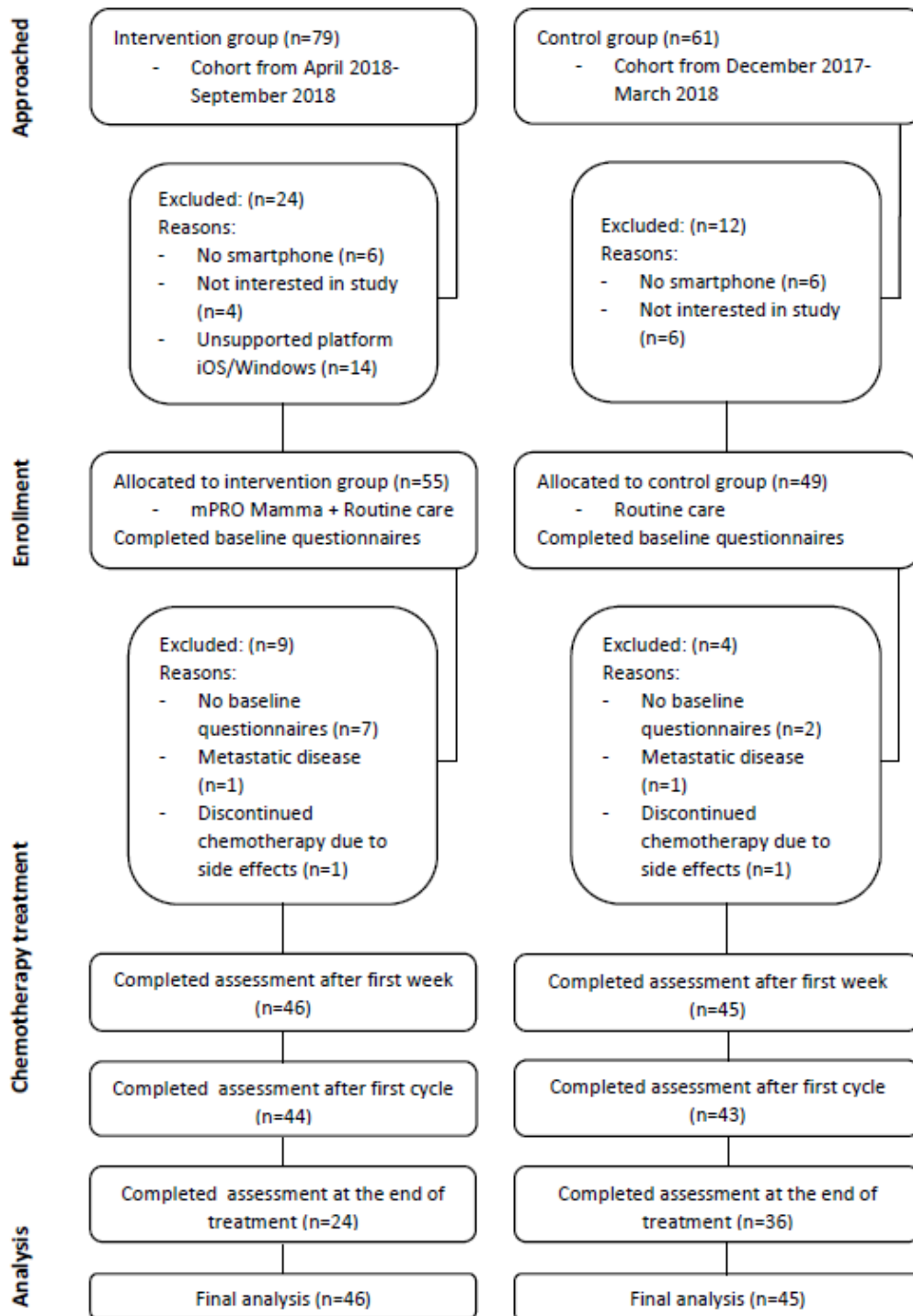
Table 1. Patient characteristics in the intervention and control group.

Characteristics	Groups			Comparison between groups	
	All patients (n=91)	Intervention (n=46)	Control (n=45)	Chi-square[a] or t test[b] (df)	P value
Age in years, mean (SD)	51.7 (9.5)	50.9 (9.3)	52.6 (9.6)	0.86[b] (89)	.39
Number of chemotherapy cycles, median (IQR)	6 (4-7)	6 (5-7)	6 (4-7)	N/A[c]	.73
Quality of life score at baseline, median (IQR)	66.7 (58.3-83.3)	83.3 (66.7-91.7)	66.7 (58.3-83.3)	N/A[c]	.06
Summary score at baseline, median (IQR)	90.9 (83.8-96.2)	93.6 (86.2-97.9)	89.1 (83.3-93.7)	N/A[c]	.02
Tumor size, n (%)				N/A[d]	.48
≤20 mm	43 (47)	24 (52)	19 (42)		
21-50 mm	39 (43)	19 (41)	20 (44)		
>50 mm	9 (10)	3 (7)	6 (13)		
Axillary nodes, n (%)				N/A[d]	.13
Negative nodes	41 (45)	23 (50)	18 (40)		
1-3 positive nodes	37 (41)	15 (33)	22 (49)		
4-6 positive nodes	9 (10)	7 (15)	2 (4)		
>6 positive nodes	4 (4)	1 (2)	3 (7)		
Histologic type, n (%)				N/A[d]	.42
Invasive ductal carcinoma	77 (85)	40 (87)	37 (82)		
Invasive lobular carcinoma	5 (6)	1 (2)	4 (9)		
Other	9 (10)	5 (11)	4 (9)		
Tumor grade, n (%)				N/A[d]	.38[e]
Grade I	2 (2)	0 (0)	2 (4)		
Grade II	30 (33)	14 (30)	16 (36)		
Grade III	58 (64)	31 (67)	27 (60)		
Unknown	1 (1)	1 (2)	0 (0)		
Estrogen receptors, n (%)				2.33[a] (1)	.13
Negative	18 (20)	12 (26)	6 (13)		
Positive	73 (80)	34 (74)	39 (87)		
Progesterone receptors, n (%)				1.23[a] (1)	.27
Negative	25 (28)	15 (33)	10 (22)		
Positive	66 (73)	31 (67)	35 (78)		
HER2[f] , n (%)				1.76[a] (1)	.19
Negative	65 (71)	30 (65)	35 (78)		
Positive	26 (29)	16 (35)	10 (22)		
Ki-67[g] expression, n (%)				1.69[a] (1)	.19
Negative (≤15 %)	21 (23)	8 (17)	13 (29)		
Positive (> 15 %)	70 (77)	38 (83)	32 (71)		
Tumor subtype, n (%)				N/A[d]	.30
Luminal A–like	9 (10)	4 (9)	5 (11)		
Luminal B HER2 negative	43 (47)	20 (44)	23 (51)		
Luminal B HER2 positive	17 (19)	8 (17)	9 (20)		

XSL•FO
RenderX

Characteristics	Groups			Comparison between groups	
	All patients (n=91)	Intervention (n=46)	Control (n=45)	Chi-square[a] or t test[b] (df)	P value
HER2 positive	8 (9)	7 (15)	1 (2)		
Triple negative	14 (15)	7 (15)	7 (16)		
Type of surgery, n (%)				12.60[a] (3)	.006
Breast conserving surgery + SNB[h]	36 (40)	26 (57)	10 (22)		
Breast conserving surgery + axillary lymphadenectomy	12 (13)	6 (13)	6 (13)		
Mastectomy + SNB	21 (23)	6 (13)	15 (33)		
Mastectomy + axillary dissection	22 (24)	8 (17)	14 (31)		
Breast reconstruction, n (%)				N/A[d]	.01
No	70 (77)	41 (89)	29 (64)		
Deep inferior flap	12 (13)	2 (4)	10 (22)		
Tissue expander	9 (10)	3 (7)	6 (13)		
Chemotherapy, n (%)				0.38[a] (1)	.54
Adjuvant chemotherapy	64 (69)	31 (67)	33 (73)		
Neoadjuvant chemotherapy	27 (30)	15 (33)	12 (27)		
Type of chemotherapy, n (%)				N/A[d]	.41
Anthracycline + taxanes	63 (69)	35 (76)	28 (62)		
Anthracyclines alone	17 (19)	6 (13)	11(24)		
Taxanes alone	8 (9)	3 (7)	5 (11)		
CMF[i]	3 (3)	2 (4)	1 (2)		

[a]Denotes a chi-square test value.

[b]Denotes a t test value.

[c]N/A: not applicable; in this instance, denotes the use of Mann-Whitney U test.

[d]N/A: not applicable; in this instance; denotes the use of Fisher exact test.

[e]Unknown values were excluded from calculations.

[f]HER2: human epidermal growth factor 2.

[g]Ki-67: proliferation marker.

[h]SNB: sentinel lymph node biopsy.

[i]CMF: cyclophosphamide, methotrexate, fluorouracil.

Primary Outcome Measures

In Table 2 and in Table 3, we have presented only effects that are of interest (ie, adjusted mean difference between the groups for each of the 3 time periods with 95% confidence intervals and P values; for all mean differences, a positive difference indicates a higher value in the intervention group than the corresponding value in the control group). The adjusted mean differences between the intervention and the control group after the first week of chemotherapy were statistically significant for both the global quality of life score (10.1, 95% CI 1.8 to 18.5, P=.02) and the summary score (8.9, 95% CI 3.1 to 14.7, P=.003). The adjusted mean difference for summary score at the end of treatment (10.6, 95% CI 3.9 to 17.3, P=.002) was also significant. Complete model results are presented in Multimedia Appendix 5 and Multimedia Appendix 6. Clinically meaningful differences in the adjusted means between groups were found for global quality of life score after the first week and for summary score at the end of treatment (ie, an adjusted mean difference greater than 10).

Table 2. Global quality of life score adjusted mean differences between intervention and control groups.

Time period	Estimate	95% CI	*P* value
First week	10.1	1.8 to 18.5	.02
First treatment cycle	4.7	–3.8 to 13.2	.27
Entire treatment period	7.0	–2.7 to 16.7	.16

Table 3. Summary score adjusted mean differences between intervention and control groups.

Time period	Estimate	95% CI	*P* value
First week	8.9	3.1 to 14.7	.003
First treatment cycle	5.3	–0.6 to 11.2	.08
Entire treatment period	10.6	3.9 to 17.3	.002

Secondary Outcome Measures

In the first week, 44% (20/46) of patients in the intervention group visited the doctor, and 60% (27/45) of patients in the control group visited the doctor. In the first cycle, 37% (16/43) of patients in the intervention group visited the doctor at least twice, and 54% (21/39) of patients in the control group visited the doctor at least twice. The differences between groups were not statistically significant (first week: $\chi^2_1=2.49$, $P=.12$; first cycle: $\chi^2_1=2.29$, $P=.13$; Table 4). The number of hospitalizations was low—3% (3/91) in the first week and 7% (6/84) in the first cycle—without substantial or statistically significant differences between the groups (first week: Fisher exact test, $P=.62$; first cycle: Fisher exact test, $P>.999$).

Table 4. Health care use during the first week and the first cycle of systemic therapy.

	All, n (%)	Intervention group, n (%)	Control group, n (%)	Chi-square[a] (*df*)	*P* value
Doctor visits during first week	91 (100)	46 (100)	45 (100)	2.49[a] (1)	.12
no visits	44 (48)	26 (56)	18 (40)		
1 visit or more	47 (52)	20 (44)	27 (60)		
Doctor visits during first cycle	82 (100)	43 (100)	39 (100)	2.29[a] (1)	.13
1 visit or less	45 (55)	27 (63)	18 (46)		
2 or more visits	37 (45)	16 (37)	21 (54)		
Hospitalizations during first week	91 (100)	46 (100)	45 (100)	N/A[b]	.62
No	88 (97)	45 (98)	43 (96)		
Yes	3 (3)	1 (2)	2 (4)		
Hospitalizations during first cycle	84 (100)	43 (100)	41 (100)	N/A[b]	>.999
No	78 (93)	40 (93)	38 (93)		
Yes	6 (7)	3 (7)	3 (7)		

[a]Denotes a chi-square test value.

[b]N/A: not applicable; in this instance, denotes the use of Fisher exact test.

## Exploratory Analysis of Other EORTC QLQ C-30 and BR-23 Scales

Scales with clinically important differences between the groups (baseline adjusted), and with $P<.10$ (Mann-Whitney U tests) were social functioning (first week: $P=.001$, adjusted $P=.04$; first cycle: $P=.005$, adjusted $P=.29$; end of treatment: $P=.003$, adjusted $P=.14$), pain (first week: $P=.03$, adjusted $P>.999$; end of treatment: $P=.005$, adjusted $P=.29$), physical functioning (first week: $P=.051$, adjusted $P>.999$; end of treatment: $P=.005$, adjusted $P=.27$), role functioning (end of treatment: $P=.01$, adjusted $P=.49$), cognitive functioning (end of treatment: $P=.09$, adjusted $P>.999$), appetite loss (first week: $P=.04$, adjusted $P>.999$), and systemic therapy side effects (first cycle: $P=.08$, adjusted $P>.999$). These scales are presented in Figures 3 and 4. Complete exploratory analysis results are available in the Multimedia Appendix 7. The only difference between the groups that could be generalized to the population after adjusting for multiple testing was in social functioning after the first week (adjusted $P=.04$).

Figure 3. Boxplots showing change from baseline for the selected functional scales of the EORTC QLQ C-30 and BR-23. EORTC: European Organisation for Research and Treatment of Cancer; QLQ BR-23: Quality of Life Questionnaire Breast Cancer Module; QLQ C-30: Quality of Life Questionnaire Core 30.
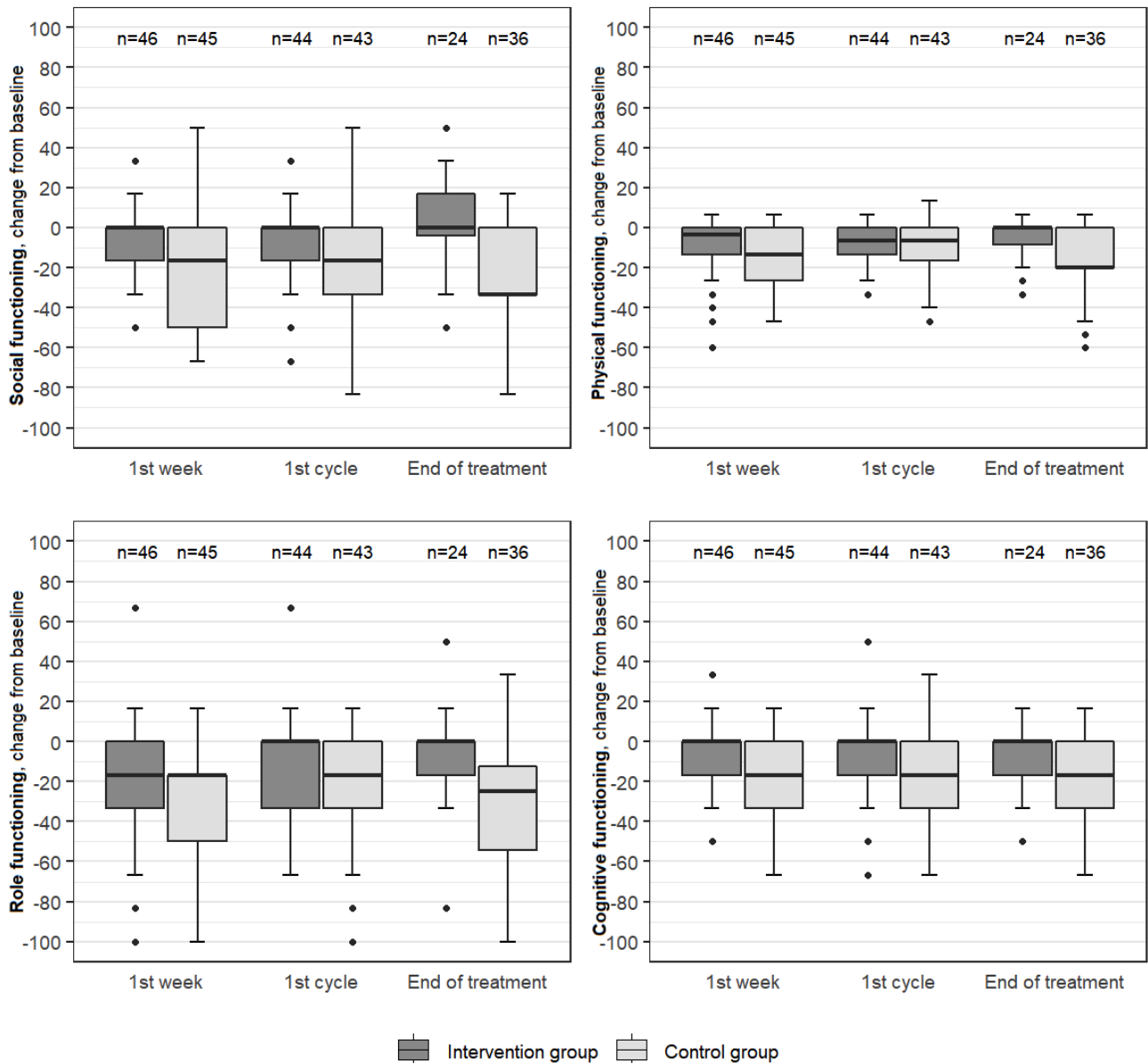
Figure 4. Boxplots showing change from baseline for the selected symptom scales of EORTC QLQ C-30 and BR-23. EORTC: European Organisation for Research and Treatment of Cancer; QLQ BR-23: Quality of Life Questionnaire Breast Cancer Module; QLQ C-30: Quality of Life Questionnaire Core 30.
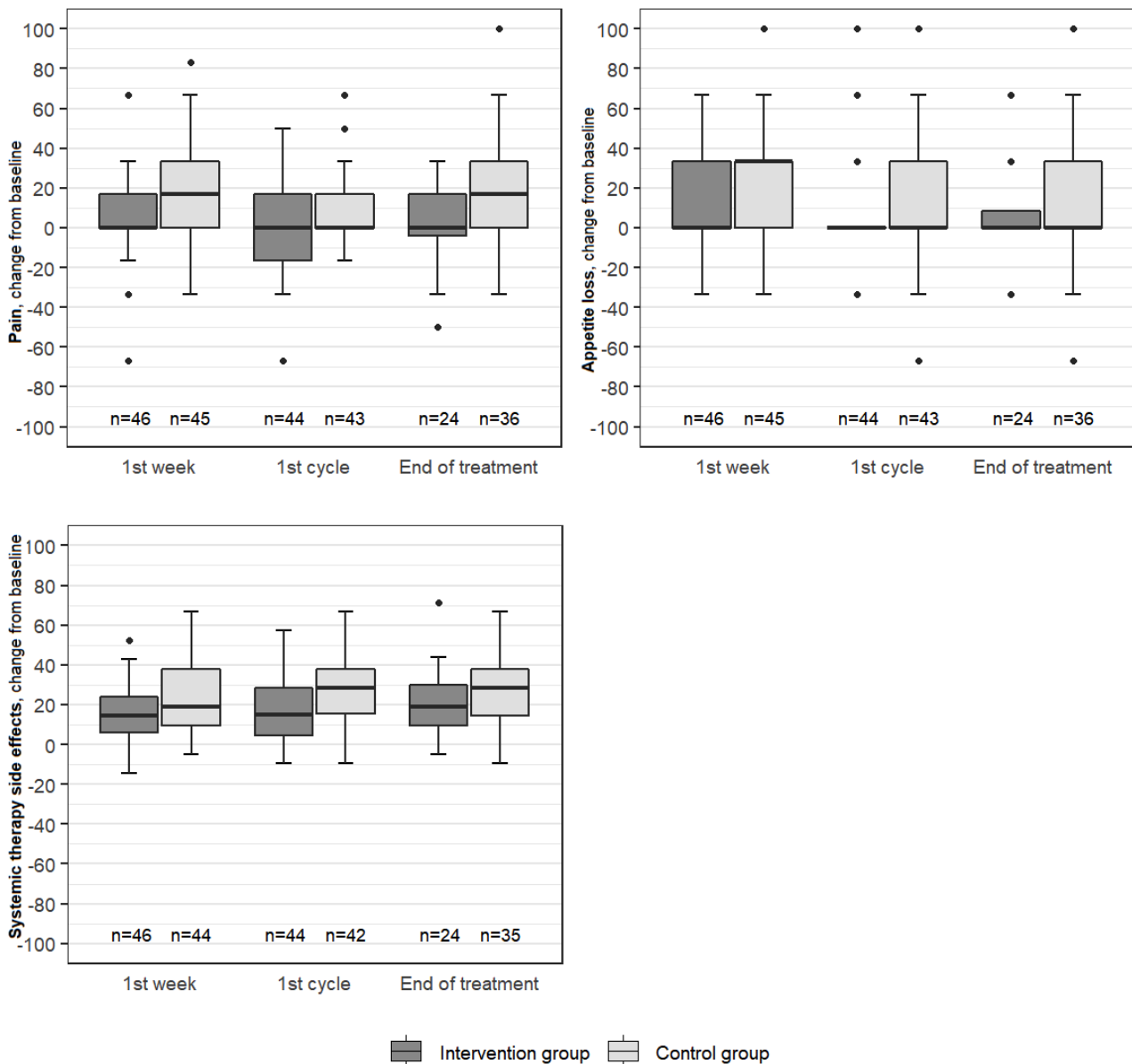
Discussion

Principal Findings

In this study, we investigated the impact of using of a mobile app for symptom reporting and self-management on patient-reported outcomes in early stage breast cancer patients who were receiving chemotherapy. Chemotherapy can cause numerous adverse effects which have a negative impact on health-related quality of life. Our hypothesis was that the intervention group (those using the mobile app) would cope better with adverse effects, which would result in the improved global quality of life and summary scores, especially at the beginning of chemotherapy, and would result in reduced use of health resources.

Our first hypothesis about the improved global quality of life and summary score in the intervention group using the mobile

app was confirmed. After the first week of treatment, global quality of life reported by the intervention group was better than that reported by the control group, with clinically important mean differences (Table 2). The global quality of life score represented patient perception of overall health and overall quality of life in the previous week, but was not sensitive enough to detect group differences over time [22]. This was confirmed in our analysis as both groups had similar global quality of life scores after the first cycle and at the end of treatment. Summary score, however, was a more robust alternative to the global quality of life score, as it took into account 13 scales that were calculated from 27 out of 30 EORTC QLQ C-30 items. We found statistically significant differences between groups in summary scores after the first week and at the end of treatment; the latter demonstrated a clinically important difference (Table 3). Our mobile app assisted patients as an interventional tool for the self-management of symptoms. As the patient reported

a symptom, the mobile app automatically provided the most suitable information for self-management according to the reported level of severity (Multimedia Appendix 1 and Multimedia Appendix 2). This improved patient operative capability and patient self-efficacy, which resulted in an improved quality of life in accordance with Bandura's theory of self-efficacy theory [29].

To investigate which EORTC QLQ C-30 scales most significantly contributed to the difference in summary score, we performed an exploratory analysis. QLQ BR-23 scales were also included in this analysis to shed light on symptoms and functioning that were specific to breast cancer. We found clinically important improvements in social, physical, role, and cognitive functioning in the intervention group versus the control group, as well as less severe pain, less appetite loss, and fewer systemic therapy side effects in the intervention group (Multimedia Appendix 7, Figure 3, and Figure 4). In a recently published paper on lung cancer [30], patients with access to a web-based health education program reported better global quality of life and emotional functioning, as well as a significant decrease in ten of the most important symptoms compared to those in the control group; however, no differences in physical, role, cognitive, or social functioning were found [30]. Patients with lung cancer, however, cannot be directly compared to patients with breast cancer. To our knowledge, comparable studies for breast cancer patients do not exist. We believe that the intervention group had better social, physical, role, and cognitive functioning due to the use of the mobile app (better self-management). For cancer patients, solving the problem (ie, self-managing of symptoms) correlates significantly with improvements in their levels of psychological distress, overall quality of life, role, and physical functioning [31,32].

We found the strongest effect was for social functioning after the first week, as the difference between groups remained statistically significant even after adjusting for multiple comparisons. Patients who receive chemotherapy treatments experience imposed limitations in social functioning, which is also influenced by family support, outlook on life and opportunities for social exchange [33]. The mobile app had a positive impact on social functioning, although, a bias as a result of nonrandom patient allocation may have been present.

Less severe systemic therapy side effects, pain, and appetite loss in the intervention group were probably due to self-management techniques employed by these patients. On the other hand, the surgery-related group differences may have contributed.

Our second hypothesis was not confirmed. We did not find any statistically significant difference between groups for use of health resources. Self-management strategies employed by patients themselves probably resolved mild and moderate side effects. For severe symptoms, we presumed that an option of immediate triggering alert to medical staff at its onset would reduce health resource use. Basch et al [5] reported significantly lower health resource usage in the group allocated to symptom reporting, probably because a severe symptom grade triggered an email alert to nurses.

Ginossar et al [34] performed a systematic review of literature and found over 100 breast cancer-related mobile apps; however, many of them did not lead to behavior change and many were not evidence-based. Recent mobile apps for women with breast cancer have a positive effect by promoting weight loss, decreasing stress, and improving the quality of life [35,36]. Apps for reporting symptoms and promoting self-care during cancer chemotherapy treatment are still rare [6-11,37,38]. Electronic capture and monitoring of patient-reported outcome for symptoms during systemic treatment of cancer has been found to be feasible, even though patients were not provided with feedback about their management [5]; however, in studies [6,7,9,10] where automated alerts for self-management of reported symptoms were available, a rapid benefit in the form of decreased symptom severity has been identified. It has been reported [6] that automated alerts were more effective than nurse-administered symptom management. In our study, we also rapidly found a benefit from the use of our mobile app, which was demonstrated by better global quality of life and summary scores in the first week after receiving chemotherapy. In our opinion, the first week after chemotherapy is a very important time to alleviate distress in patients. A recently published randomized trial from Japan [37] tested a mobile app similar to ours, to see whether the app affected anxiety and depression, but no improvement in anxiety, depression, or health literacy were found at the end of treatment [37].

Using a mobile app in collaboration with the treating physician improved patient well-being and their awareness of chemotherapy adverse effects [38]; however, the impact of the mobile app on quality of life outcomes was not assessed [37,38]. Zhu et al [9] reported findings similar to ours, namely that the e-support group had better quality of life, self-efficacy, and symptom control than that of the group receiving standard care. They found beneficial effects at 3 months, but which disappeared at 6 months [9]. In our study, even at the end of treatment (approximately 6 months), the beneficial effect in summary score (but not in global quality of life) remained; however, Zhu et al [9] assessed quality of life using a different questionnaire, ie, the Functional Assessment of Cancer Treatment-B. Recently, a mobile phone–based system for the remote monitoring and management of chemotherapy-related side effects was evaluated in Canadian patients with cancer; however, its impact on health-related quality of life has not yet been tested [11]. Several randomized controlled trials evaluating remote electronic monitoring and symptom management are still ongoing [14-16].

Strengths

The advantage of our study was the detailed patient-reported outcome analysis of the effect of mobile app usage, with data collected using validated questionnaires (EORTC QLQ C-30 and BR-23). In addition to using a standard tool (global quality of life score), we used summary score, a new tool suggested by EORTC. To our knowledge, our study is the first that uses both quality of life and summary score as primary outcomes when analyzing the impact of a mobile app on the care of patients with breast cancer who are receiving chemotherapy. Moreover, we appropriately planned the first assessment for after the first

week, indicated by the statistically significant difference in the global quality of life and summary score at this time.

Limitations

Our study has some limitations. First, it was a nonrandomized controlled cohort study. In order to avoid potential bias, it would have been better to conduct a randomized controlled trial. In fact, the groups in our study differed with respect to type of breast surgery, quality of life score, and summary score at baseline, so the results were adjusted for these differences using appropriate statistical methods. There may also have been seasonal differences affecting patient well-being since patients were enrolled at different times of the year. In addition, a possible bias may have existed as a result of different recruitment techniques used by the medical oncologists. At the beginning of the clinical study, we were still in the process of programming our mobile app which prevented us from randomly allocating patients.

Second, not all symptoms that can arise during the systemic therapy can be assessed with the QLQ C-30 and BR-23 questionnaires. Our app included approximately 50 of the 78 recommended PRO-CTCAE symptoms, so instructions for patient self-management for the missing symptoms were not included in the app; however, we aimed to include the symptoms that occur most commonly over the course of breast cancer treatments. Our collection was broader than that used by Zhang et al [39] who developed and evaluated a patient-reported outcome scale for breast cancer containing 38 items [39].

Our app could also be improved by utilizing wearable sensors (smart bracelets, watches, rings, etc) and incorporating additional questionnaires capable of measuring psychological distress, mental distress, and anxiety; it could be also equipped with alerts reminding patients to provide data at predetermined dates.

Future Work

Based on the positive effects on health-related quality of life measured in the systemic therapy part of breast cancer treatment, we plan to include other treatment modalities for breast cancer patients, ie, surgical interventions and radiotherapy. Another interesting upgrade would be to include a follow-up period. Our aim is to provide regular updates to ensure software compatibility as well as to avoid potential security issues. Since our app is likely most useful for mild and moderate symptoms where patients are empowered to manage symptoms themselves, we plan to also make the app useful for patients with severe symptoms. We could establish a system where alerts about patient symptoms are sent to a dedicated hospital server. If the alert contained any severe symptoms, it would be labelled as urgent and an appointed research nurse would be trained to react in accordance with the prespecified medical algorithms. Such algorithms are also incorporated in recently reported proposals of studies [14-16] to assess symptom burden, quality of life, supportive care needs, anxiety, self-care, self-efficacy, work limitations, and cost effectiveness.

Our findings regarding the impact of mobile app use on the quality of life and summary scores in combination with those regarding other scales that were only a part of our exploratory analysis, should be confirmed in a larger randomized controlled trial. Further research is also needed in order to extend these findings to a broader population of patients, including those suffering from other ailments, and to different organizations in the medical field.

Conclusions

Our mobile app has interventional value for the self-management of symptoms for patients with breast cancer who are receiving systemic therapy which was shown by a better global quality of life scores (in the first week of therapy) and a better summary scores (in the first week and at end of treatment) for the intervention group than for the control group. Based on the exploratory analysis, the app contributed to clinically important improvements in social, physical, role, and cognitive functioning while diminishing pain, appetite loss, and systemic therapy side effects.

Authors' Contributions

CGK obtained funding for the clinical trial. TK designed and programmed the mPRO Mamma mobile app, under the supervision of MK. CGK, TGC, and NRG conceived and designed the analysis. TGC and CGK were involved in collecting the data. NRG performed the statistical analysis and the interpretation of the results. CGK wrote the initial version of the manuscript and subsequent drafts were reviewed and edited by MK, NRG, and TGC.

Conflicts of Interest

TK, under the supervision of MK, was also the designer and developer of the mPRO Mamma app. Access to the app is free and open.

XSL•FO

RenderX

Multimedia Appendix 1
Screenshot of a page for the daily report of a symptom pain.
[PNG File , 76 KB - mhealth_v8i8e17408_app1.png]

Multimedia Appendix 2
Screenshot of a page for the daily report of symptoms of nausea and vomiting.
[PNG File , 99 KB - mhealth_v8i8e17408_app2.png]

Multimedia Appendix 3
Screenshot of a page for the daily report of symptoms with their severity.
[PNG File , 64 KB - mhealth_v8i8e17408_app3.png]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1) in pilot-testing.
[PDF File (Adobe PDF File), 1625 KB - mhealth_v8i8e17408_app4.pdf]

Multimedia Appendix 5
Linear mixed-effects model for global quality of life, whole model.
[DOCX File , 13 KB - mhealth_v8i8e17408_app5.docx]

Multimedia Appendix 6
Linear mixed-effects model for summary score, whole model.
[DOCX File , 13 KB - mhealth_v8i8e17408_app6.docx]

Multimedia Appendix 7
The explorative analysis of the other scales of the EORTC QLQ C-30 and BR-23.
[DOCX File , 16 KB - mhealth_v8i8e17408_app7.docx]

References

1. Fromme EK, Eilers KM, Mori M, Hsieh Y, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. J Clin Oncol 2004 Sep 1;22(17):3485-3490. [doi: 10.1200/JCO.2004.03.025] [Medline: 15337796]
2. Atkinson TM, Li Y, Coffey CW, Sit L, Shaw M, Lavene D, et al. Reliability of adverse symptom event reporting by clinicians. Qual Life Res 2012 Sep;21(7):1159-1164 [FREE Full text] [doi: 10.1007/s11136-011-0031-4] [Medline: 21984468]
3. Common terminology criteria for adverse events (CTCAE). Institute NIoHaNC. 2009. URL: https://www.eortc.be/services/doc/ctc/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf [accessed 2019-10-18]
4. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Natl Cancer Inst 2014 Sep;106(9) [FREE Full text] [doi: 10.1093/jnci/dju244] [Medline: 25265940]
5. Basch E, Iasonos A, Barz A, Culkin A, Kris MG, Artz D, et al. Long-term toxicity monitoring via electronic patient-reported outcomes in patients receiving chemotherapy. J Clin Oncol 2007 Dec 01;25(34):5374-5380. [doi: 10.1200/JCO.2007.11.2243] [Medline: 18048818]
6. Given CW, Sikorskii A, Tamkus D, Given B, You M, McCorkle R, et al. Managing symptoms among patients with breast cancer during chemotherapy: results of a two-arm behavioral trial. J Clin Oncol 2008 Dec 20;26(36):5855-5862 [FREE Full text] [doi: 10.1200/JCO.2008.16.8872] [Medline: 19029420]
7. Mooney KH, Beck SL, Wong B, Dunson W, Wujcik D, Whisenant M, et al. Automated home monitoring and management of patient-reported symptoms during chemotherapy: results of the symptom care at home RCT. Cancer Med 2017 Mar;6(3):537-546 [FREE Full text] [doi: 10.1002/cam4.1002] [Medline: 28135050]
8. Stefanovic S, Wallwiener M, Karic U, Domschke C, Katic L, Taran F, et al. Patient-reported outcomes (PRO) focused on adverse events (PRO-AEs) in adjuvant and metastatic breast cancer: clinical and translational implications. Support Care Cancer 2017 Feb;25(2):549-558. [doi: 10.1007/s00520-016-3437-2] [Medline: 27747478]
9. 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]

10. Sikorskii A, Given CW, Given BA, Banik A, Krauss JC. Patient Engagement With an Automated Telephone Symptom Management Intervention: Predictors and Outcomes. Ann Behav Med 2020 Jun 12;54(7):484-494. [doi: 10.1093/abm/kaz067] [Medline: 31925414]

11. Moradian S, Krzyzanowska MK, Maguire R, Morita PP, Kukreti V, Avery J, et al. Usability Evaluation of a Mobile Phone-Based System for Remote Monitoring and Management of Chemotherapy-Related Side Effects in Cancer Patients: Mixed-Methods Study. JMIR Cancer 2018 Dec 21;4(2):e10932 [FREE Full text] [doi: 10.2196/10932] [Medline: 30578238]

12. 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]

13. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. JAMA 2017 Jun 04. [doi: 10.1001/jama.2017.7156] [Medline: 28586821]

14. Maguire R, Fox PA, McCann L, Miaskowski C, Kotronoulas G, Miller M, et al. The eSMART study protocol: a randomised controlled trial to evaluate electronic symptom management using the advanced symptom management system (ASyMS) remote technology for patients with cancer. BMJ Open 2017 Jun 06;7(5):e015016 [FREE Full text] [doi: 10.1136/bmjopen-2016-015016] [Medline: 28592577]

15. Furlong E, Darley A, Fox P, Buick A, Kotronoulas G, Miller M, et al. Adaptation and Implementation of a Mobile Phone-Based Remote Symptom Monitoring System for People With Cancer in Europe. JMIR Cancer 2019 Mar 14;5(1):e10813 [FREE Full text] [doi: 10.2196/10813] [Medline: 30869641]

16. Beck SL, Eaton LH, Echeverria C, Mooney KH. SymptomCare@Home: Developing an Integrated Symptom Monitoring and Management System for Outpatients Receiving Chemotherapy. Comput Inform Nurs 2017 Oct;35(10):520-529 [FREE Full text] [doi: 10.1097/CIN.0000000000000364] [Medline: 28570285]

17. Kova? T. mPRO Mamma. Google Play Store: Google LLC; 2017. URL: https://play.google.com/store/apps/details?id=com.kovac.timotej.breastcare&hl=sl [accessed 2019-10-18]

18. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 03;85(5):365-376. [doi: 10.1093/jnci/85.5.365] [Medline: 8433390]

19. Michels FAS, Latorre MDRDDO, Maciel MDS. Validity, reliability and understanding of the EORTC-C30 and EORTC-BR23, quality of life questionnaires specific for breast cancer. Rev Bras Epidemiol 2013 Jun;16(2):352-363 [FREE Full text] [doi: 10.1590/S1415-790X2013000200011] [Medline: 24142007]

20. Material Design. Google. URL: https://material.io/ [accessed 2020-07-15]

21. Fayers P, Bjordal K, Groenvold M, Curran D, Bottomley A. EORTC QLQ-C30 Scoring Manual. 3rd ed. BrusselsORTC. Brussels: EORTC Brussels; 2001. URL: https://www.eortc.org/app/uploads/sites/2/2018/02/SCmanual.pdf [accessed 2020-06-22]

22. Giesinger JM, Kieffer JM, Fayers PM, Groenvold M, Petersen MA, Scott NW, EORTC Quality of Life Group. Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. J Clin Epidemiol 2016 Jan;69:79-88 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.08.007] [Medline: 26327487]

23. Scoring of the QLQ-C30 Summary Score. In: Cancer EOfRaTo. Brussels: EORTC Brussels; Jan 01, 2001.

24. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998 Jan;16(1):139-144. [doi: 10.1200/JCO.1998.16.1.139] [Medline: 9440735]

25. Cameron D, Morden JP, Canney P, Velikova G, Coleman R, Bartlett J, TACT2 Investigators. Accelerated versus standard epirubicin followed by cyclophosphamide, methotrexate, and fluorouracil or capecitabine as adjuvant therapy for breast cancer in the randomised UK TACT2 trial (CRUK/05/19): a multicentre, phase 3, open-label, randomised, controlled trial. Lancet Oncol 2017 Jul;18(7):929-945 [FREE Full text] [doi: 10.1016/S1470-2045(17)30404-7] [Medline: 28600210]

26. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982 Dec;38(4):963-974. [Medline: 7168798]

27. Core Team R. R: A Language and Environment for Statistical Computing. https://www.R-project.org/. Vienna: R Foundation for Statistical Computing; 2017. URL: https://www.R-project.org/ [accessed 2020-06-22]

28. 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]

29. Lev EL. Bandura's theory of self-efficacy: applications to oncology. Sch Inq Nurs Pract 1997;11(1):21-37; discussion 39. [Medline: 9188268]

30. Huang C, Kuo H, Lin Y, Chen S. Effects of a Web-based Health Education Program on Quality of Life and Symptom Distress of Initially Diagnosed Advanced Non-Small Cell Lung Cancer Patients: A Randomized Controlled Trial. J Cancer Educ 2019 Feb;34(1):41-49. [doi: 10.1007/s13187-017-1263-y] [Medline: 28780685]

31. Nezu AM, Nezu CM, Felgoise SH, McClure KS, Houts PS. Project Genesis: assessing the efficacy of problem-solving therapy for distressed adult cancer patients. J Consult Clin Psychol 2003 Dec;71(6):1036-1048. [doi: 10.1037/0022-006X.71.6.1036] [Medline: 14622079]

XSL•FO

RenderX

32. Kim YH, Choi KS, Han K, Kim HW. A psychological intervention programme for patients with breast cancer under chemotherapy and at a high risk of depression: A randomised clinical trial. J Clin Nurs 2018 Feb;27(3-4):572-581. [doi: 10.1111/jocn.13910] [Medline: 28557043]

33. Bloom JR, Spiegel D. The relationship of two dimensions of social support to the psychological well-being and social functioning of women with advanced breast cancer. Soc Sci Med 1984;19(8):831-837. [doi: 10.1016/0277-9536(84)90400-3] [Medline: 6505749]

34. Ginossar T, Shah SFA, West AJ, Bentley JM, Caburnay CA, Kreuter MW, et al. Content, Usability, and Utilization of Plain Language in Breast Cancer Mobile Phone Apps: A Systematic Analysis. JMIR Mhealth Uhealth 2017 Mar 13;5(3):e20 [FREE Full text] [doi: 10.2196/mhealth.7073] [Medline: 28288954]

35. 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]

36. 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 Aug 27;7(8):e13245 [FREE Full text] [doi: 10.2196/13245] [Medline: 31456578]

37. Handa S, Okuyama H, Yamamoto H, Nakamura S, Kato Y. Effectiveness of a Smartphone Application as a Support Tool for Patients Undergoing Breast Cancer Chemotherapy: A Randomized Controlled Trial. Clin Breast Cancer 2020 Jun;20(3):201-208. [doi: 10.1016/j.clbc.2020.01.004] [Medline: 32201165]

38. 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 06;18(9):e238 [FREE Full text] [doi: 10.2196/jmir.6414] [Medline: 27601354]

39. Zhang J, Yao Y, Zha X, Pan L, Bian W, Tang JH. Development and Evaluation of a Patient-Reported Outcome (PRO) Scale for Breast Cancer. Asian Pac J Cancer Prev 2015;16(18):8573-8578 [FREE Full text] [doi: 10.7314/apjcp.2015.16.18.8573] [Medline: 26745119]

Abbreviations

EORTC: European Organisation for Research and Treatment of Cancer
HER2: human epidermal growth factor 2
Ki-67: proliferation marker
PRO-CTCAE: patient-reported outcome version of the Common Toxicity Criteria of Adverse Effects
QLQ BR-23: Quality of Life Questionnaire Breast Cancer Module
QLQ C-30: Quality of Life Questionnaire Core 30

Original Paper

Developing a Heart Transplantation Self-Management Support Mobile Health App in Taiwan: Qualitative Study

Yi-Wen Chen[1][*], MS; Jeng Wei[1][*], MD, PhD; Hwei-Ling Chen[1], MS; Ching-Hui Cheng[1], MS; I-Ching Hou[2], PhD

[1]Heart Center, Cheng Hsin General Hospital, Taipei, Taiwan

[2]School of Nursing, National Yang-Ming University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
I-Ching Hou, PhD
School of Nursing
National Yang-Ming University
Nursing Building Room 407
Number 155, Section 2, Linong Street, Beitou District
Taipei
Taiwan
Phone: 886 28267000 ext 7315
Email: evita@ym.edu.tw

Abstract

Background: Heart transplantation (HTx) is the most effective treatment for end-stage heart failure patients. After transplantation, patients face physiological, psychological, social, and other health care problems. Mobile health (mHealth) apps can change the delivery of conventional health care to ubiquitous care and improve health care quality. However, a dearth of mHealth apps exists for patients with HTx worldwide, including in Taiwan.

Objective: The aim of this study was to investigate the information needed and to develop a preliminary framework for an mHealth app for post-HTx patients.

Methods: A qualitative approach with individual in-depth interviews was conducted at a heart center in the regional hospital of northern Taiwan from June to November 2017. Patients that had undergone HTx and their health professionals were recruited for purposeful sampling. A semistructured interview guideline was used for individual interviews and transcribed. Thematic analysis was used for data analysis.

Results: A total of 21 subjects, including 17 patients and 4 health professionals, were recruited for the study. The following five major themes were identified: reminding, querying, experience sharing, diet, and expert consulting. Minor themes included a desire to use the app with artificial intelligence and integration with professional management.

Conclusions: An intelligent mHealth app that addresses the five main themes and integrates the processes of using a mobile app could facilitate HTx self-management for Taiwanese patients.

KEYWORDS

heart transplantation; mobile health app; self-management

Introduction

Background

Heart transplantation (HTx) is a surgery for patients with heart failure who are not eligible for drug or other surgical treatments, involving mechanical assistance such as extracorporeal membrane oxygenation or a ventricular assist device, which is then replaced with a healthier heart [1,2]. Since Barnard performed the world's first HTx in 1967 [3], according to the International Society for Heart and Lung Transplantation, there have been approximately 5000 HTx (adult and pediatric) performed every year worldwide [4]. The first HTx in Taiwan was performed in 1987 [5], and according to reports of the Taiwan Organ Registry and Sharing Center, there have been approximately 668 HTx performed to date in the country, with 1-, 3-, 5-, and 8-year allograft survival rates of 79%, 71%, 65%, and 56%, respectively [6].

After HTx, the recipients have to take immunosuppressants to decrease the activity of their immune system and prevent it from attacking the donated heart. The current standard immunosuppressive regimens include calcineurin inhibitors (eg, cyclosporine, tacrolimus, Prograf), antiproliferative agents (eg, azathioprine and mycophenolate mofetil), and steroids (eg, methylprednisolone, prednisolone) [7,8]. The main side effects of immunosuppressants include nephrotoxicity, hypertension, trembling, body hair growth, gum hypertrophy, liver toxicity, high blood sugar, high cholesterol, myelosuppression, leukopenia, thrombocytopenia, hypocytopenia, Cushing signs, weight gain, high blood pressure, high blood lipids, and gastrointestinal bleeding [9,10].

Previous studies showed that approximately 20%-30% of patients with HTx exhibit drug noncompliance [11] or forget to take their drugs [12]. Patient drug noncompliance increases with a longer time after HTx [13]. The main health problems caused by taking immunosuppressants after HTx are physical deterioration, fatigue, foot cramps, hair hyperplasia, moon face, poor vision, and acne, sequentially [14-17]. Anxiety and depression are the most common psychological problems in HTx patients [18]. The reasons for psychological stress include unclear prognosis of the disease, fear of death, fear of rejection, fear of complications, fear of increasing family trouble, and pressure to use drugs [19,20].

Currently, wireless networks, smartphones, and mobile health (mHealth) apps are becoming increasingly more popular. According to a systematic review literature, patients with cardiovascular disease, acquired immunodeficiency syndrome, diabetes, and organ transplants have positive perceptions after using mHealth services to improve their medication compliance [21]. In 2016, a mobile app for heart failure patients was developed, which included functions for self-assessment of heart failure symptoms, exercise recommendations, vital signs records, and statistical data. The graphics helped physicians make decisions and the research findings showed good results for patients to self-manage the disease, improve compliance with medications, and implementation of diet and exercise [22]. Online and smartphone-based apps for cardiac rehabilitation programs can augment secondary prevention strategies compared with standard cardiac rehabilitation. In particular, the app was shown to improve risk factors (eg, weight, blood pressure, and diet) and to reduce the health care burden of repeat cardiovascular disease events (eg, rate of rehospitalizations/emergency department visits) [23]. Patients with chronic diseases who used mobile phones (eg, medication reminder apps, text messaging) showed better medication adherence compared with usual care [24,25]. In a randomized 3-month study, patients with coronary heart disease who were found to have an app showed increased drug compliance, and patients with positive app acceptance and participation with the app showed positive results [26]. These studies demonstrated that mHealth apps can support patients with specific diseases for health self-management.

To our knowledge, there is no mHealth app available for supporting patients with HTx in Taiwan. Most hospitals use traditional self-management education and clinic consultation to support their patients. There are some disadvantages of this approach, including the fact that patients may not fully understand the interventions of self-management within the limited clinic visiting time and they often have a physical burden after surgery (eg, pain, fatigue). Thus, a tailored HTx mHealth app could offer continuous support to patients anytime and anywhere to ultimately improve their quality of life in the long term. To support patients with HTx, Cheng Hsin General Hospital (an HTx-specific hospital in Taiwan) and National Yang-Ming University have been working together to design an HTx mHealth app since 2017.

Objectives

When developing an effective mHealth app to support health self-management, it is important to understand the expectations of end users in the early phase. The end users of the mHealth app designed in this study are patients with HTx. However, according to a previous study, HTx physicians and nurses were the major education providers for the self-management of HTx-related problems [17]. An mHealth app with management from health professionals may be beneficial to patients when developing a useful app. To our knowledge, there is a dearth of studies investigating the information needed for patients with HTx and their health professionals in the design of mHealth apps. To address these gaps, the objective of this study was to discover these needs to facilitate future development of an HTx mHealth app.

Methods

Study Design and Ethics

In this study, we used a qualitative approach involving individual interviews with HTx patients and their health professionals from June to November in 2017, with the goal of collecting and analyzing their information needs in an mHealth app for supporting HTx health self-management. The researchers (YW, IC, LF) are trained in qualitative academic research and have extensive experience in individual interviews. The clinical researchers from Cheng Hsin General Hospital included the cardiac surgeon and the superintendent (WJ), director of nursing (HL), and nurse supervisor (CH) who facilitated the process of this study.

This study complies with the Helsinki Declaration and was approved by the Institutional Review Board of the study site in Taipei before the study began [No. (596) 106-04].

Study Site and Service Processes

The study site is The Heart Center of Cheng Hsin General Hospital, which specializes in the diagnosis, management, and treatment of cardiovascular diseases in northern Taiwan. Up to February 2020, a total of 523 patients successfully received HTx operations, 46 of whom have survived postsurgery for more than 20 years. This survival rate is the highest in Taiwan. The HTx health professional team (of the Cardiovascular Surgery Division) consists of two surgeons who are qualified instructors of HTx and three coordinators who are responsible for communication between health professionals and patients through the service process. The study site also employs nurse practitioners, who are nurses that assist with frontline surgery/medical therapy and direct care.

The service process for patients with HTx includes three phases: before hospital admission, during hospitalization, and at discharge. The first phase includes cases with acceptable preevaluation information (eg, medical records, test results). Consultation, laboratory tests, preoperative evaluations, surgery/medical therapy, intensive care, recuperation, and cardiac rehabilitation are included in the second phase.

Follow-up services (eg, regular clinic appointments, telecommunication as needed) are provided in the third phase.

Samples

To fully understand the users' information needs for the mHealth app, a purposeful sampling approach was used to recruit patients with HTx and their health professionals for this study. The inclusion criteria are described below and a flowchart of the sampling process is shown in Figure 1.

Figure 1. Study flowchart.
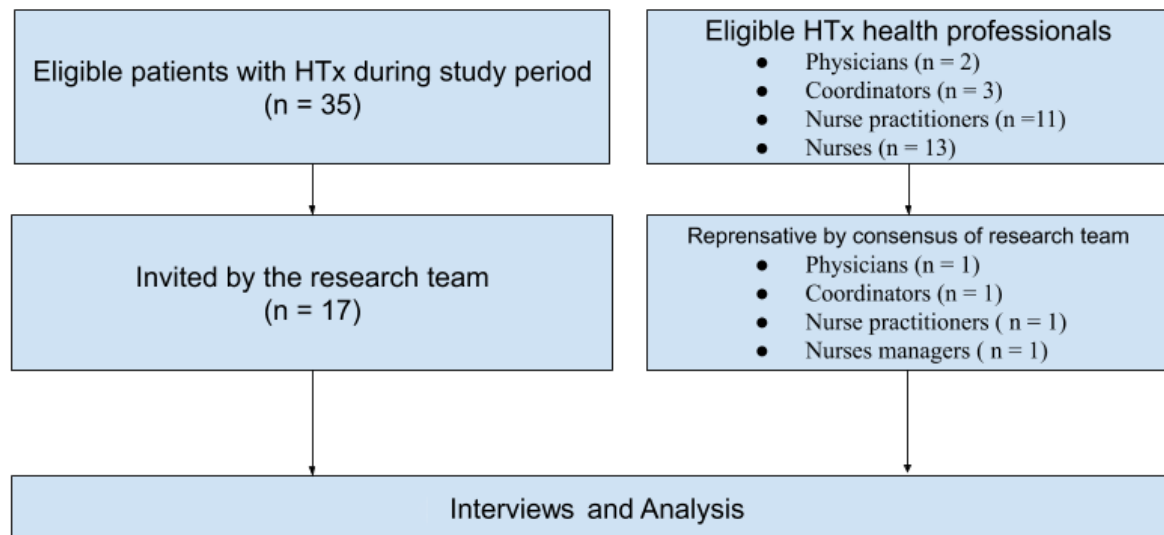
For patients with HTx, the inclusion criteria were: (1) HTx recipients who regularly visited the HTx clinic (Tuesday afternoons) during the study period (June to November 2017), (2) over 20 years old, (3) native Mandarin/Taiwanese speakers, and (4) mobile phone users. The exclusion criteria were: (1) mental illness and (2) slurred speech. The patients matched to the inclusion criteria were referred by the HTx coordinators to the researcher (YW) to confirm their willingness to participate and all agreed to have individual interviews. Before the formal interview, the consent form was filled out and the formal interview was held by the researcher for at least 30 minutes in the meeting room of the study site.

Four types of HTx health professionals (physicians, coordinators, nurse practitioners, and nurses) who were the primary health self-management education providers at the study site were recruited [17]. The inclusion criteria for each role were: (1) experience with HTx direct care, (2) experience with HTx administrative management (eg, director, head nurse, assistant nurse, leader), and (3) mobile phone users. The exclusion criteria were not current research team members (all

authors). After the consensus meeting of the research team, the representative health professionals were invited and all agreed to have individual interviews by the researcher (YW) at their office. Every subject was given US $15-30 for compensation after the interview according to their interview time.

Interview Guideline

One interview guideline was created to support the data collection. The first part was for patients, which included demographics (eg, education, marital status, occupation before and after HTx, time after HTx) and open questions for mHealth app self-management on HTx (eg, what difficulties have you faced after HTx and how do you think the mobile app could support you?). The second part was for HTx health professionals, with open questions about how mHealth apps could support patients with HTx. The interview guideline was reviewed by the nurse managers of the heart center at the study site (HL, CH), who confirmed the appropriateness before individual interviews. The details of the guideline are provided in Textbox 1.

Textbox 1. Details of the interview guideline for heart transplant (HTx) patients and professionals.

Interview guideline for patients with HTx

- Name:

- Age:

- Education:

- Marital status:

- Occupation before HTx:

- Occupation after HTx:

- Time since HTx:

- Mobile smartphone experience: Yes/No

- Open questions:

1. Would you please tell me what difficulties you have faced after HTx (eg, medication side effects, nutrition intake, etc)?

2. We would like to develop a mobile phone app to support patients with HTx on self-management in the future. How do you think the mobile health app could support you (eg, useful functions)?

Interview guideline for HTx professionals

- Name:

- Age:

- Education:

- HTx working years:

- Open questions:

1. Would you please tell me what difficulties you have faced in managing patients with HTx (eg, medication compliance, communication problems, etc)?

2. We would like to develop a mobile phone health app to support patients with HTx on health self-management in the future. How do you think that mobile health apps could support them (eg, useful functions)?

Data Collection

Before the interview, a brief introduction to the study and interviewer was provided by one of the researchers (YW). An electronic voice recorder was used with permission of the subjects. During individual interviews, the subjects were free to answer the questions and allowed to refuse when they felt uncomfortable. The researcher encouraged them to share their opinions without enforcement and interruption. Field notes were also taken by the interviewer to retrieve the key information from subjects. Before ending the interview, the subjects were asked whether they had additional comments to share and confirmed that there were no more data to be obtained from the subjects. After the interview, all voice records and nonverbal body language (eg, facial expressions, voice tone, and motions) were transcribed verbatim within 48 hours and returned (via email or post) to the subjects to confirm the semantic accuracy [27].

Information Needs Analysis Framework

We adopted the methodology of Vaismoradi et al [28] for trustworthy thematic analysis to summarize our subjects' information needs and to analyze the data from the individual interviews. The first step involved an overall reading of the text. All of the verbatim transcribed notes were read by researchers (YW, IC) with an open mind. Any meaningful text or researchers' reflections were written down as the fundamental content for theme retrieving after the initial reading. The second step involved generating initial codes. According to the purpose of the study, the researchers generated the initial codes with empathy to the mHealth app needs of the subjects. The third step involved generating confirming codes. To clarify the contradictions and find hidden messages from the original verbatim transcribed text, the researchers read the text again. The initial codes were reconfirmed, adjusted, added, and any redundancies were eliminated. The fourth step involved generating the initial themes. The codes with similar meaning were categorized into the same themes. The fifth step involved generating the confirmed themes. After rechecking the verbatim transcribed text and reflecting on the journal and codes, the final themes and their names were confirmed by the research team. Finally, the report was produced. The Consolidated Criteria for Reporting Qualitative Research guidelines were used to produce this report. Through the process of analysis, peer debriefing, reflective journaling, and consensus of research teams, theme saturation was achieved. With the names of codes, themes were created according to familiarity for HTx patients, and these names also helped with communication for mHealth app development. Table 1 provides an example of the thematic analysis process followed in this study.

XSL•FO

RenderX

Table 1. Example analysis of information needs framework for the Reminder theme.

Transcribed text from interview	Codes
"When I am busy, the (mobile) phone (app) could remind me to take four FK (tacrolimus; FK506) in the morning and three (FK) in the afternoon." -Patient B, 50-year-old man, college, business, CAD[a] with ICMP[b]	First reminder for medication
"I used the alarm clock with my (mobile) phone to remind me (to take the medication) regularly and so did my wife. When I forgot (to take the medication), my wife would remind me (to take the medication)." - Patient L, 54-year-old man, senior high school, retirement, DCM[c]	Second reminder for medication
"When we decided to adjust the medication (the dose of immunosuppressant), we would send out messages (to patients) through Line message. The patients would know their medication was adjusted. (The hospital side would know the patient read the message)." - P1, 48-year-old man, college, heart transplant physician	Reminder of medication dose adjustment
"When patients return home after the heart transplant, the app would remind them about influenza if they have poor immunity. In the fall and winter, there would be a reminder warning about the window of the peak of influenza. Patients should pay attention to it." - P3, 34-year-old woman, college, cardiac intensive care unit assistant head nurse	Reminder about influenza season

[a]CAD: coronary artery disease.

[b]ICMP: ischemic cardiomyopathy.

[c]DCM: dilated cardiomyopathy.

Results

Description of the Subjects

A total of 21 subjects, including 17 patients with HTx and 4 health professionals, were recruited from June to November 2017. Most of the patients were male (13/17, 77%), 51 to 60 years old (7/17, 41%), with a college education (8/17, 47%), married (11/17, 65%), had a job before HTx (15/17, 88%) and after HTx (8/17, 47%), were diagnosed with dilated cardiomyopathy (DCM) (14/17, 82%), and received HTx more than 2 years previously (10/17, 59%). The detailed demographic data for each patient are shown in Table 2.

Table 2. Demographic data of heart transplant (HTx) patients.

Patient ID	Sex	Age (years)	Education	Marital status	Pre/post HTx occupation	Pre HTx diagnosis	Time since HTx
A	M	53	Master	Married	Business	CAD[a] with AMI[b]	1-2 years
B	M	50	College	Married	Business	CAD with ICMP[c]	>2 years
C	M	39	College	Single	Service industry	DCM[d]	>2 years
D	F	43	College	Single	Service industry	DCM	>2 years
E	M	52	Junior high school	Single	Industry	DCM	>2 years
F	M	54	Junior high school	Married	Industry/Retired	DCM	<6 months
G	F	51	Junior college	Married	Service industry	DCM	>2 years
H	F	62	Elementary school	Married	Industry/Housewife	DCM	<6 months
I	M	55	College	Divorced	Service industry	DCM	>2 years
J	M	46	College	Divorced	Service industry/Retired	DCM	>2 years
K	F	50	Senior high school	Married	Housewife	DCM	>2 years
L	M	54	Senior high school	Married	Industry/Retired	DCM	>2 years
M	M	27	College	Single	Unemployed	HOCM[e]	>2 years
N	M	32	College	Married	Service industry	DCM	1-1.5 years
O	M	55	Senior high school	Married	Service industry	DCM	1-1.5 years
P	M	48	College	Married	Business	DCM	1-1.5 years
Q	M	42	Junior high school	Married	Industry/Unemployed	DCM	1-1.5 years

[a]CAD: coronary artery disease.

[b]AMI: acute myocardial infarction.

[c]ICMP: ischemic cardiomyopathy.

[d]DCM: dilated cardiomyopathy.

[e]HOCM: hypertrophic cardiomyopathy.

XSL•FO

RenderX

The health professionals included one physician and three nurses who all had college degrees. The physician had more than 10 years of experience with HTx and is the director of the Taiwan Heart Transplant Association. The first nurse is the coordinator (ie, case manager) who had 11 years of experience with HTx direct care and currently communicates with patients, families, and hospitals. She is also responsible for outpatient clinic services, patient medication education, and holds support group activities for patients with HTx. The second nurse is the assistant head nurse in the cardiac intensive care unit who had 9 years of experience in intensive care for post HTx patients. The third nurse is the cardiac surgery nurse practitioner who had 10 years of experience and is responsible for the HTx care plan and evaluation. Their demographic data are shown in Table 3.

Table 3. Demographic characteristics of cardiology professionals.

ID	Sex	Age (years)	Education	Job title	Years of heart transplant experience
P1	M	48	College	Heart transplant physician	10
P2	F	44	College	Heart transplant coordinator	11
P3	F	34	College	Cardiac intensive care unit assistant head nurse	9
P4	F	34	College	Cardiac surgery nurse practitioner	10

Main Themes

After the thematic analysis, a total of five major themes and 14 codes were identified (Textbox 2).

Theme 1: Reminder

Nearly every subject mentioned that the app would be helpful for reminding them to take their medication regularly. The alarm on a mobile phone is the most commonly used tool for preventing medication noncompliance. However, if they still forget to take the medication after turning off the alarm, the app with a medication reminder would be helpful for increasing their medicine compliance.

Textbox 2. Themes and codes as the framework for a heart transplant self-management app.

Theme 1: Reminder

- 1-1. First reminder for regular medication administration

- 1-2. Second reminder for regular medication administration

- 1-3. Reminder when medication dose is adjusted

- 1-4. Reminder to return to the clinic for examination

- 1-5. Reminder of influenza season

Theme 2: Query

- 2-1. Laboratory results

- 2-2. Plasma drug concentration

- 2-3. Record of heart rate and blood pressure

Theme 3: Experience Sharing

- 3-1. Asynchronous experience sharing

- 3-2. Synchronous experience sharing

Theme 4: Diet

- 4-1. Diet guideline

- 4-2. Recipes

Theme 5: Expert Consulting

- 5-1. Artificial intelligence consulting

- 5-2. Remote professional consulting

Code 1-1: First Reminder for Regular Medication Administration

It is good to have a reminder (in the app), the mobile phone can help patients to take medicine regularly and decrease the risk of transplant rejection. [P2, 44-year-old woman; heart transplant coordinator]

The (mobile) phone (app) could remind patients to take four FK (tacrolimus; FK506) in the morning and three (FK) in the afternoon. [Patient B, 50-year-old man, college, business, coronary artery disease (CAD) with ischemic cardiomyopathy]

For those elderly (patients), sometimes they forget (to take medication). If there is an app to remind you to take the dose of medicine in the mornings, afternoons, and evenings, it would be really helpful. [Patient I, 55-year-old man, college, service industry, DCM]

In addition to the reminder to take medication, the patients with HTx also mentioned that they need the app to provide drug information since there are so many different kinds of medications to take after HTx. They can better follow the instructions of taking the medication if they have more knowledge of them.

The app should not only remind the dose of medication for each person to take but also provide the image, action, and side effects of the drug. Then, I would feel more clarity. [Patient C, 39-year-old man, college, service industry, DCM]

There should be the label and image of the drug on the screen of the mobile phone to remind me to take the medicine. That would prevent me from taking the wrong medication. [Patient Q, 42-year-old man, junior high school, industry before HTx and unemployed after HTx, DCM]

Code 1-2: Second Reminder for Regular Medication Administration

Some patients with HTx also mentioned that they depend on their family to remind them to take medication. The secondary reminder would provide them with more confidence to prevent the undesired neglect to take medication.

I use the alarm clock with my (mobile) phone to remind me (to take the medication) regularly. My wife also does the same thing. When I forgot (to take the medication) at the time, my wife also reminds me (to take the medication). [Patient L, 54-year-old man, senior high school, retired, DCM]

I am afraid of forgetting to take so many medications and I need my husband to remind me. If the phone app rings (referring to having the reminder from the app), then I will take the medication. It is convenient for my husband's phone app to ring to make sure I take the medication. [Patient H, 62-year-old man, elementary school, industry before HTx and housewife after HTx, DCM]

Code 1-3: Reminder When Medication Dose is Adjusted

In addition to the reminder for needing to take medication regularly, having information on medication dose adjustment according to their plasma immunosuppressants concentration is also important for patients with HTx. The medication dose adjustment reminder is currently provided by the HTx coordinator via telephone when their laboratory report is received. After HTx patients return to the outpatient clinic and have their blood drawn, the physician then adjusts their immunosuppressants dose according to their plasma drug concentration. However, there can be problems such as the patients being out of reach (eg, too busy to answer the phone) or forgetting the verbal order, and thus taking the wrong dose. In these situations, the subjects would like the app to remind them when medication dose adjustment is needed and to improve communication between patients and health professionals.

I think it (the reminder of medication dose adjustment) is necessary, especially at the beginning of receiving HTx. It would be convenient to receive the drug adjustment message from the app because sometimes (I) would forget. [Patient G, 51-year-old man, junior college, service industry, DCM]

When we decide to adjust medication (the dose of immunosuppressant), we could send out a message (to patients) through the Line message. The patient would know how their medication should be adjusted. (The hospital side would know the patient read the message.) [P1, 48-year-old man, college, HTx physician]

Code 1-4: Reminder to Return to the Clinic for Examination

Some subjects also mentioned that the app should provide the reminder of the outpatient clinic and examination appointments to support the patients and health professionals.

I do not know if it's because of retirement, but my memory seems to be getting worse. So, if the phone (app) can remind me to return (to the clinic), my wife and I would have less trouble. [Patient J, 46-year-old man, college, service industry before HTx and retired after HTx, DCM]

For the reminders on clinic appointments, if the app could remind them (HTx patients) to return (clinic appointment) after making the reminder setting, the following reminders are automatic (without setting manually), which could save lots of time for the (HTx) coordinator. They only (need to) find out which patients had no response (those who do not return to the clinic for follow up) and focus on reminding them. [P1, 48-year-old man, college, HTx physician]

Code 1-5: Reminder of Influenza Season

The subjects also mentioned that the app could send a warning or voice reminder during influenza season to support HTx self-care taking into account their lower immune system and provide self-protection procedures.

I am worried about getting an infection. So, (the app) could remind me to wear a mask during influenza

season. [Patient O, 55-year-old man, senior high school, service industry before HTx and unemployed after HTx, DCM]

It is easy to catch a cold when the seasons alternate. The app can remind me there is an influenza outbreak. [Patient P, 48-year-old man, college, business, DCM]

I think when patients return home, there should be a reminder about the influenza season when their immune systems are weak. In the fall and winter seasons, the app should have a pop-up reminder that it's the influenza season and to be careful. [P3, 34-year-old woman, cardiac intensive care unit assistant head nurse]

Theme 2: Query

Over half of the subjects indicated that they would like the app to provide query functions for laboratory results (eg, liver function index, renal function index, hemoglobin, hematocrit, blood sugar), plasma drug concentrations, heart rate records, and blood pressure. Such functions are convenient for patients with HTx to take care of themselves and they also could show their laboratory data to their physician when they return for a clinic visit. Some patients with HTx also felt that such functions would be more helpful than the oral report from health professionals.

Code 2-1: Laboratory Results

I hope I could query about my (laboratory) report of blood drawing in my phone (app) such as hemoglobin and liver function index. (Referring to his mobile phone with his finger and nodding.) [Patient M, 27-year-old man, college, unemployed, hypertrophic cardiomyopathy (HOCM)]

If queries are provided in the app, that would be better than the (HTx coordinator) and ribbit (nagging) you. You can remind yourself what the query of the last examination report was. (The subject refers to the coordinator with their finger with a smile on their face.) [Patient B, 50-year-old man, college, business, CAD]

Code 2-2: Plasma Drug Concentration

Many subjects queried about the report of their plasma drug concentrations and indicated that showing the trend in the app was very important to them. Such functions could show them the importance of taking immunosuppressants regularly and prevent rejection.

I could see data of my drug plasma concentration (from the app). It is very important for me to know the importance of taking medication. [Patient M, 27-year-old man, college, unemployed, HOCM]

To see the figure curves of drug plasma concentration would be very convenient for the patient to return to the clinic and show their physicians. [P4, 34-year-old man, college, cardiac surgery nurse practitioner]

Code 2-3: Record of Heart Rate and Blood Pressure

Some of the subjects mentioned that the app should provide records of heart rate and blood pressure as an important index of heart function, which would help them to better control their body.

If the app could integrate with other devices then I can check my blood pressure and heart rate, which would be better. (The subject refers to the data check in the monitor in the hospital which could be integrated in the app) [Patient C, 39 year-old-man, college, service industry, DCM]

I think there should be basic blood pressure and heart rate monitoring and then I could see the results. It would be very convenient. [Patient D, 43-year-old man, college, service industry, DCM]

Theme 3: Experience Sharing

About half of the subjects mentioned that would like to share their experience and information with other HTx patients asynchronously (eg, via a blog) or synchronously (eg, in an online chatting room). For new patients with HTx, asynchronous experience sharing would help them learn more about how to take care of themselves from a more senior patient with HTx.

Code 3-1: Asynchronous Experience Sharing

I told my psychologist about the panic attack (the subject felt chest tightness and nervous) at midnight. He encouraged me to write down the severity of anxiety before the heart biopsy. Now, I am creating a blog and hope to record the process through texts. The blog could share important messages with patients and families. Such peer groups are very important. [Patient A, 53-year-old man, master degree, business, CAD]

It is good (referring to the experience sharing post HTx). We can learn from other people how to protect their heart. My husband and I can pay attention to these experiences. [Patient H, 62-year-old woman, elementary school, housewife, DCM]

Code 3-2: Synchronous Experience Sharing

Synchronous experience sharing could support patients mentally (eg, to decrease the anxiety on the uncertainty of a rejection response) and prevent social isolation.

We understand that new people (referring to new patients with HTx) on the medication and other problems would be more anxious and worry about their condition. To have such (experience sharing), would be helpful. [Patient G, 51-year-old man, junior college, service industry, DCM]

You may have less contact with your friends. I participated in many activities before, but after HTx, there may be foods I cannot eat. Maybe friends go to climb a (mountain) or have high-intensity activities and we may not be able to participate. Then, we become far away from friends. If the app can provide a chatting room, it allows the patients to have a place

where they can decompress. [Patient K, 50-year-old woman, senior high school, housewife, DCM]

Theme 4: Diet

Some of subjects mentioned they would like to know about the diet guidelines after HTx. They were told that they cannot eat certain foods without cooking them first such as raw fish and salads, which they liked to eat before HTx. They felt the limitations of food choices in their daily life. They would like to know what foods they can eat and to tailor the food recommendations according to their physiques.

Code 4-1: Diet Guidelines

I think there should be customized functions in the app which provides recommendations on the nutrition according to your physique. [Patient A, 53-year-old man, master degree, business, CAD]

There should be a notice (in the app) on which raw foods to avoid and if grapefruits would affect the metabolizing of the medication. [Patient E, 52-year-old man, junior high school, industry, DCM]

We, the earlier HTx patients, were told by the health professionals that we could not take Chinese herbs if we needed to inactivate our immune system (referring to preventing an auto rejection response). Some Chinese herbs would activate the immune system of the body like wheat grass juice. Such knowledge should be provided earlier for our specific body condition and prevent us from eating those foods. [Patient G, 51-year-old man, junior college, service industry, DCM]

I hope there would be a list of foods that we can eat and I can see the detailed information when I click on it. [Patient N, 32-year-old man, college, service industry before HTx and unemployed after HTx, DCM]

Code 4-2: Recipes

Some of the subjects had the problem of becoming overweight after HTx from taking steroids. They need the app to provide recipes, calculate food calories, and determine the daily intake of calories.

I think you can provide food recipes. Recently, I read the recipes for patients after HTx. The app could provide such recipes like a low-fat diet or low carbohydrate diet of your choice. You could provide recipes for 1 week and they could try them out. You can retrieve some recipes or food photos from the internet which I think would be good. [Patient K, 50-year-old woman, senior high school, housewife, DCM]

The app could provide recipes or videos of cooking. Then, the patients would know what to eat when they come home. [P3, 34-year-old woman, cardiac intensive care unit assistant head nurse]

Theme 5: Expert Consulting

The subjects indicated that the app could provide artificial intelligence consulting. The patients could upload a text or photo to the app for consulting. The app could then automatically answer the patients with efficiency.

Code 5-1: Artificial Intelligence Consulting

Like a bank app, when typing in simple questions about specific items, the app could automatically answer the questions. [Patient Q, 42-year-old man, junior high school, industry before HTx and unemployed after HTx, DCM]

The app could let the patients ask questions with a short text message. For those simple questions, the app could automatically respond to patients. [P4, 34-year-old woman, college, cardiac surgery nurse practitioner]

Code 5-2: Remote Professional Consulting

For urgent situations, the subjects mentioned that they could remotely consult the health professionals through the telephone using photos, which can then be sent to the physician or coordinators for assessment and problem solving. Such functions would help patients who live in more remote areas to save on travel time.

If you have a problem, you can ask questions (with the app). For example, at midnight, when there is an urgent situation, some health professionals (in the app) can help. [Patient D, 43-year-old man, college, service industry, DCM]

I think the phone number of HTx health professionals could be shown in the app for contact. [Patient N, 32-year-old man, college, service industry before HTx and unemployed after HTx, DCM]

When patients come home, there could be a health problem like limb edema from their heart function getting worse. Patients who live in the middle or south of Taiwan could send photos to ask the physician or coordinator. That would save on their travel time and solve the problem immediately. [P3, 34-year-old woman, cardiac intensive care unit assistant head nurse]

Discussion

Principal Findings and Comparison to Previous Studies

According to the results, the information needed for patients with HTx on mHealth apps included five main themes: reminder, query, experience sharing, and diet and expert consulting. The results were mainly consistent with a previous study showing that patients with HTx need self-management education such as about regular medication taken, regular exercise, diet control, infection/rejection signs observation, and regular clinic visiting, which could support patients with HTx on health self-management [29].

Most subjects indicated needing the app to support them on reminders for medication taken (code 1-1, code 1-2, code 1-3), clinic/examination appointments (code 1-4), and influenza

season (code 1-5). With regard to the reminder for medication taken, this is focused on immunosuppressants management to prevent rejection of the grafted heart [7,8], supporting the same importance and difficulty in medication compliance as highlighted in previous studies [21]. To our knowledge, there are existing mHealth apps for supporting taking medication regularly [30] but these were not adopted by our subjects. Instead, they used the original alarm in the mobile phone but sometimes also forgot to take the medication. In such situations, some of their family members would remind them. The design of an mHealth app that could support secondary medication reminders from family or health professionals might be useful for such unexpected noncompliant medication situations.

To resolve the ineffective communication process between discharged patients and their health professionals (eg, informing medication adjustment, clinic appointments), our health professionals suggested that the app could support them in knowing that their messages had been received by the patients. Currently, such functions are provided by instant message software in Taiwan. Therefore, the newly developed app should consider involving the same function that may facilitate its adoption.

In addition, a reminder for influenza season (code 1-5) was indicated as a requirement of the app by both the patients and health professionals. The growing availability of big data in health care and public health opens up possibilities for infectious disease control in local settings. The detection and prediction (nowcasting) of influenza epidemics are now becoming possible [31].

With regard to the query of their laboratory report (code 2-1), plasma drug concentration (code 2-2), and their own health records (code 2-3), these needs are consistent with those of patients with other diseases [22,32]. Currently, most patients only know about their report when they come to the clinic and are told by the health professionals through an electronic health record query. Query from the app and the ability to read their report anytime may support patients in achieving self-control of their health behaviors rather than being passively monitored by health professionals.

With regard to the theme of experience sharing, our subjects had difficulties in facing activities with people owing to suffering from physical problems after HTx (eg, fatigue, Cushing syndrome) [14-17]. Therefore, the app may support patients in providing social connections with other patients to mentally support each other and decrease feelings of social distance anytime and anywhere (code 3-1, code 3-2). To our knowledge, social networks are popular (eg, Facebook, Instagram, Line) and most patients use them to exchange information. However, the individual privacy and accuracy of the content is a concern, as mentioned by some subjects: "There should be someone to manage the accuracy of the data" (Patient B, 50-year-old man, college, business, CAD); "If there is an online community, you need to manage the personal data, but who has that responsibility?" (Patient I, 55-year-old man, college, service industry, DCM).

With regard to the theme of diet, this is a common need expressed by patients who experience severe illnesses and recovery. To date, there have been many new technologies (eg, web-based tools, smartphone apps) developed to support diet self-management [33]. However, the diet guideline (code 4-1) specific to the needs of HTx patients (eg, avoid eating raw foods that may weaken their poor immune system) should be considered. Specifically, female subjects indicated a need for more advanced diet and caloric calculation functions in the app given concern for body weight control problems from taking steroid medications. Providing specific and customized recipes (code 4-2) in the app for patients with HTx may support them in consuming appropriate foods and obtaining good nutrition.

With regard to the theme of expert consulting, continuous support from health professionals was highlighted as very important to patients with HTx (code 5-1, code 5-2). A previous study showed that telemedicine apps can support communication about oral conditions among clinicians and patients [34]. Considering the busy work and limited human power of health professionals, the subjects referred to their experience of using artificial intelligence (eg, chat robot) and recommended the research team to design such functions in the app.

When patients with HTx have an emergency physical situation, an automatic quick response from the app when they ask questions can first help to screen the severity of their medical needs and save travel time on returning to the hospital. An app with artificial intelligence could increase the convenience. However, the validity and reliability of artificial intelligence should be evaluated carefully when adopted in the app to prevent risks. In addition, the app could facilitate communication between health professionals and patients. With the reports from the hospital's side (eg, blood report, echo image, heart biopsy) and patient's side (eg, body temperature, heart rate, blood pressure), both sides could easily communicate with each other through the app (eg, medical advice when there is an abnormal report). Such communication depends on integration into the hospital management in the future. These two minor findings would facilitate end user intent to use the app.

In summary, five main themes were identified by the research team. The information based on these main themes and their codes were then used to form the framework (Textbox 2) for development of an HTx mHealth self-management support app. Most themes were consistent with previous studies, but some new advanced codes specific to HTx management were uncovered (code 1-2, second reminder for regular medication administration; code 1-5, reminder of influenza season; code 5-1, artificial intelligence consulting). We believe that individual interviews with patients and health professionals is a strong method for identifying mHealth technology information needs for Taiwanese patients with HTx.

Limitations

The first limitation of this study is that it was conducted in urban areas with more medical resources than are typically available in the more rural areas of the country, which might have resulted in geographical bias. We did not provide the simulation app, which might have resulted in different feedback. The second limitation of this study is the purposeful sampling approach, which prevented recruitment of more subjects and we did not consider the sex ratio among HTx patients, which might result

in different information needed in the mHealth app. The third limitation is that we did not include the perspective from managers and information communication technology specialists at the study site, who may have insight into the policy (eg, human power to support an mHealth app service) and technology feasibility (eg, artificial intelligence consulting) on developing a patient-centered app to support HTx self-management.

Summary and Conclusions

Our team used individual in-depth interviews to retrieve the information needed for use in an mHealth app from patients with HTx in Taiwan. A total of five main app needs were retrieved efficiently to facilitate developing an HTx self-management app. The next steps include building a real app and validating the self-management outcomes (eg, technology acceptance, medication compliance, rejection response, emergency visiting) from Taiwanese patients with HTx.

Acknowledgments

Authors' Contributions

YC and IH contributed to the study design; implementation, analysis, and interpretation of the findings; and preparation of the manuscript. JW, LL, HC, and CC contributed to data interpretation.

Conflicts of Interest

None declared.

References

1. Bhagra SK, Pettit S, Parameshwar J. Cardiac transplantation: indications, eligibility and current outcomes. Heart 2019 Feb;105(3):252-260. [doi: 10.1136/heartjnl-2018-313103] [Medline: 30209127]
2. Hunt SA. Taking heart--cardiac transplantation past, present, and future. N Engl J Med 2006 Jul 20;355(3):231-235. [doi: 10.1056/NEJMp068048] [Medline: 16855261]
3. Söderlund C, Rådegran G. [50 years of heart transplantations]. Lakartidningen 2018 Apr 12;115 [FREE Full text] [Medline: 29664541]
4. Stehlik J, Kobashigawa J, Hunt SA, Reichenspurner H, Kirklin JK. Honoring 50 Years of Clinical Heart Transplantation in : In-Depth State-of-the-Art Review. Circulation 2018 Jan 02;137(1):71-87. [doi: 10.1161/CIRCULATIONAHA.117.029753] [Medline: 29279339]
5. Wang S, Wang C, Chou N, Chi N, Huang S, Yu H, et al. Current status of heart transplantation in Taiwan. Transplant Proc 2014 Apr;46(3):911-913. [doi: 10.1016/j.transproceed.2013.11.029] [Medline: 24767379]
6. TORSC. URL: http://www.TORSC.org.tw [accessed 2019-10-08]
7. Chou C, Chou C, Huang Y, Wu M, Hsu C, Chou Y. Prescription trends of immunosuppressive drugs in post-heart transplant recipients in Taiwan, 2000-2009. Pharmacoepidemiol Drug Saf 2014 Dec;23(12):1312-1319. [doi: 10.1002/pds.3722] [Medline: 25335855]
8. Karam S, Wali RK. Current State of Immunosuppression: Past, Present, and Future. Crit Rev Eukaryot Gene Expr 2015;25(2):113-134. [doi: 10.1615/critreveukaryotgeneexpr.2015011421] [Medline: 26080606]
9. Vega E, Schroder J, Nicoara A. Postoperative management of heart transplantation patients. Best Pract Res Clin Anaesthesiol 2017 Jun;31(2):201-213. [doi: 10.1016/j.bpa.2017.06.002] [Medline: 29110793]
10. Song J, Ou SH, Lin PH, Liu CT, Sung SY, Tsai CS. [New life from the heart: assessment and management of heart transplant patients]. Hu Li Za Zhi 2014 Aug;61(4):10-14. [doi: 10.6224/JN.61.4.10] [Medline: 25116309]
11. De Geest S, Denhaerynk K, Dobbels F. Clinical and economic consequences of non-adherence to immunosuppressive drugs in adult solid organ transplantation. Compliance in solid organ transplantation. In: Oppenheimer F, editor. International Transplantation Updates. Barcelona: Permanyer Publications; 2011:63-81.
12. Blanca Martínez Pérez A, López Suárez A, Rodríguez Rodríguez J, Sobrino Márquez JM, Lage Gallé E. Medication adherence in patients who undergo cardiac transplantation. Transplant Proc 2013;45(10):3662-3664. [doi: 10.1016/j.transproceed.2013.11.006] [Medline: 24314989]
13. Korb-Savoldelli V, Sabatier B, Gillaizeau F, Guillemain R, Prognon P, Bégué D, et al. Non-adherence with drug treatment after heart or lung transplantation in adults: a systematic review. Patient Educ Couns 2010 Nov;81(2):148-154. [doi: 10.1016/j.pec.2010.04.013] [Medline: 20627643]
14. Wang SM, Ku NP, Lin HT, Wei J. The Relationships of Symptom Distress, Social Support and Self-Care Behaviors in Heart Transplant Recipients. J Nurs Res 1998;6(1):4-18. [doi: 10.7081/NR.199802.0004]

XSL•FO
RenderX

15. Stiefel P, Malehsa D, Bara C, Strueber M, Haverich A, Kugler C. Symptom experiences in patients after heart transplantation. J Health Psychol 2013 May;18(5):680-692. [doi: 10.1177/1359105312454909] [Medline: 22933576]

16. Chou Y, Lai Y, Wang S, Shun S. Impact of Fatigue Characteristics on Quality of Life in Patients After Heart Transplantation. J Cardiovasc Nurs 2017;32(6):551-559. [doi: 10.1097/JCN.0000000000000400] [Medline: 28306705]

17. Huang SH, Wang SS, Tai JJ, Lou MF. Effects of Self-Management on Symptom Distress of Heart Transplant Recipients. Formosan J Med 2009;13(2):122-132 [FREE Full text]

18. Conway A, Schadewaldt V, Clark R, Ski C, Thompson DR, Kynoch K, et al. The effectiveness of non-pharmacological interventions in improving psychological outcomes for heart transplant recipients: a systematic review. Eur J Cardiovasc Nurs 2014 Apr;13(2):108-115. [doi: 10.1177/1474515113519519] [Medline: 24399842]

19. Dew MA, DiMartini AF. Psychological disorders and distress after adult cardiothoracic transplantation. J Cardiovasc Nurs 2005;20(5 Suppl):S51-S66. [doi: 10.1097/00005082-200509001-00007] [Medline: 16160585]

20. Palmar-Santos AM, Pedraz-Marcos A, Zarco-Colón J, Ramasco-Gutiérrez M, García-Perea E, Pulido-Fuentes M. The life and death construct in heart transplant patients. Eur J Cardiovasc Nurs 2019 Jan;18(1):48-56. [doi: 10.1177/1474515118785088] [Medline: 29923416]

21. 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]

22. Athilingam P, Labrador MA, Remo EFJ, Mack L, San Juan AB, Elliott AF. Features and usability assessment of a patient-centered mobile application (HeartMapp) for self-management of heart failure. Appl Nurs Res 2016 Nov;32:156-163. [doi: 10.1016/j.apnr.2016.07.001] [Medline: 27969021]

23. Widmer RJ, Allison T, Lerman L, Lerman A. The augmentation of usual cardiac rehabilitation with an online and smartphone-based program improves cardiovascular risk factors and reduces hospitalizations. J Am Coll Cardiol 2014 Apr;63(12):A1296. [doi: 10.1016/s0735-1097(14)61296-1]

24. Santo K, Singleton A, Rogers K, Thiagalingam A, Chalmers J, Chow CK, et al. Medication reminder applications to improve adherence in coronary heart disease: a randomised clinical trial. Heart 2019 Feb;105(4):323-329. [doi: 10.1136/heartjnl-2018-313479] [Medline: 30150326]

25. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile Telephone Text Messaging for Medication Adherence in Chronic Disease: A Meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]

26. Santo K, Chow CK, Thiagalingam A, Rogers K, Chalmers J, Redfern J. MEDication reminder APPs to improve medication adherence in Coronary Heart Disease (MedApp-CHD) Study: a randomised controlled trial protocol. BMJ Open 2017 Oct 08;7(10):e017540 [FREE Full text] [doi: 10.1136/bmjopen-2017-017540] [Medline: 28993388]

27. Boyce C, Neale P. Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input. Pathfinder International Tool Series: Monitoring and Evaluation 2. 2006 May. URL: http://www2.pathfinder.org/site/DocServer/m_e_tool_series_indepth_interviews.pdf?docID=6301 [accessed 2020-08-10]

28. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]

29. Wang SM, Ku NP. Distress and Intervention of Heart Transplant Patients after Discharge. J Nurs Taipei 1998;45(5):85-90 [FREE Full text]

30. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile Phone Apps to Improve Medication Adherence: A Systematic Stepwise Process to Identify High-Quality Apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]

31. Spreco A, Eriksson O, Dahlström Ö, Cowling BJ, Timpka T. Evaluation of Nowcasting for Detecting and Predicting Local Influenza Epidemics, Sweden, 2009-2014. Emerg Infect Dis 2018 Oct;24(10):1868-1873. [doi: 10.3201/eid2410.171940] [Medline: 30226160]

32. Kessel KA, Vogel MM, Kessel C, Bier H, Biedermann T, Friess H, et al. Mobile Health in Oncology: A Patient Survey About App-Assisted Cancer Care. JMIR Mhealth Uhealth 2017 Jun 14;5(6):e81 [FREE Full text] [doi: 10.2196/mhealth.7689] [Medline: 28615159]

33. Cade JE. Measuring diet in the 21st century: use of new technologies. Proc Nutr Soc 2017 Aug;76(3):276-282. [doi: 10.1017/S0029665116002883] [Medline: 27976605]

34. Petruzzi M, De Benedittis M. WhatsApp: a telemedicine platform for facilitating remote oral medicine consultation and improving clinical examinations. Oral Surg Oral Med Oral Pathol Oral Radiol 2016 Mar;121(3):248-254. [doi: 10.1016/j.oooo.2015.11.005] [Medline: 26868466]

Abbreviations

CAD: coronary artery disease
DCM: dilated cardiomyopathy
HOCM: hypertrophic cardiomyopathy

XSL•FO
RenderX

HTx: heart transplantation
mHealth: mobile health

XSL•FO

RenderX

Original Paper

Beneficial Features of a mHealth Asthma App for Children and Caregivers: Qualitative Study

Misa Iio[1], RN, PHN, PhD; Yumiko Miyaji[2], MD, PhD; Kiwako Yamamoto-Hanada[2], MD, PhD; Masami Narita[3], MD, PhD; Mayumi Nagata[1], RN, PHN, PhD; Yukihiro Ohya[2], MD, PhD

[1]College of Nursing, Kanto Gakuin University, Yokohama, Japan

[2]Allergy Center, National Center for Child Health and Development, Setagaya, Japan

[3]Division of Allergy, Tokyo Metropolitan Children's Medical Center, Fuchu, Japan

Corresponding Author:
Misa Iio, RN, PHN, PhD
College of Nursing
Kanto Gakuin University
1-50-1 Mutsuurahigashi
Kanazawa-ku
Yokohama, 236-8503
Japan
Phone: 81 45 786 5641
Email: misaiio@kanto-gakuin.ac.jp

Abstract

Background: mHealth and uHealth apps are available for children with asthma and their caregivers. However, previous studies on mHealth apps for children older than 7 years old with asthma are limited, and most studies on asthma apps do not consider interactions involving communication between children and caregivers. Therefore, a prototype mHealth child asthma app was developed for children and their caregivers, with features of tailored feedback messages in continuing self-management and interactions between children and caregivers.

Objective: The aim of this study was to identify the beneficial features of a prototype mHealth app developed for children with asthma and their caregivers.

Methods: Children diagnosed with persistent asthma by allergy specialists at the National Center for Child Health and Development were recruited. The features of a prototype mHealth app for children with asthma and their caregivers were investigated using semistructured interviews after they tried the app. Data were analyzed using thematic analysis. Content-characteristic words were named and grouped together as categories to explore themes.

Results: We recruited 27 children with asthma aged 2 to 12 years and 26 their caregivers. Findings on the good aspects of the app for children older than 7 years old and caregivers suggested 4 themes (confirmation of asthma knowledge, child-caregiver interaction, design of the app, and child's interest), and 6 categories were identified. Findings on the good aspects of app for children 7 to 12 years old and caregivers suggested 5 themes (new knowledge, manga as a Japanese-style comic, child's interest, trigger of self-management, and design and operability), and 11 categories were identified. Findings on the beneficial features of app suggested 6 themes (asthma knowledge, elements for continuous, universal design, notification, monitoring, and functions), and 12 categories were identified.

Conclusions: Children with asthma and their caregivers perceived that the good aspects of the app were learning asthma knowledge with fun, including manga; interaction between child and caregiver; and easy-to-read design, such as colors. They wanted not only the asthma knowledge but also the universal design and enhanced elements, monitoring, and notification functions of the app.

(JMIR Mhealth Uhealth 2020;8(8):e18506) doi:10.2196/18506

KEYWORDS

children; caregivers; asthma; mobile app; proposed beneficial features

XSL•FO
RenderX

Introduction

As mobile technology and smartphones become widely used in the health care sector, several mobile medical and mobile health (mHealth) apps are being released—mHealth apps are most commonly developed to monitor a specific health disorder, to make medication management easier, or to inform users about health information and a specific health disorder. mHealth asthma apps are valuable assets for patients and caregivers alike because they offer immediate communication between patients and those responsible for providing care for their needs [1]. The asthma app's features include asthma education material, symptom forecast, asthma action plan, telemedicine, local specialist connection, what to do in an emergency, symptom monitoring, airborne trigger identification, and clinic notifications [1]. Contents of mHealth apps for asthma since 2011 have included comprehensive information and targeting specific skills, such as the use of action plans, recommended self-care procedures, inhaler technique, and inhaler instructions [2]. The mHealth app features were classified into 7 categories (inform, instruct, record, display, guide, reminders or alert, and communicate) [3]. The most commonly used behavior change techniques of the asthma management apps were instruction, behavior health association, self-monitoring, feedback, teaching to use prompts or cues, consequences, and others' approval [4].

A review [5] of mHealth apps for asthma in 2014 showed that 206 apps were available for patients with asthma, and 16 apps were available for children with asthma and their caregivers. There were 8 previous studies of smartphone apps to encourage asthma self-management in adolescents [6]. In pediatric asthma apps, children with asthma aged 6 to 16 years who utilized electronic adherence monitoring with daily reminder alarms together with the clinic's feedback regarding their inhaled corticosteroid use required significantly fewer courses of oral steroids and fewer hospital admissions than those who were in the usual care arm with adherence monitoring alone [7]. One self-management app contained reminder and notification features, such as tracking of medication, for young people with asthma (aged 15 to 24 years), which resulted in high satisfaction with usefulness and ease of use [8]. Another app determined the asthma phenotype in children through the asthma control test to monitor activity, sleep, peak expiratory flow, and indoor air quality [9].

Children with chronic conditions and their caregivers need to maintain self-management behavior for symptom control, and behavior change techniques have been incorporated to support their behavior continuity. A nonuniform and individualized support was tailored as per the patient's situation and behavior factors [10]. A meta-analysis [11] of tailoring studies that utilized print messages concluded that tailored interventions are more effective than nontailored ones. Tailoring is defined as "any combination of strategies and information intended to reach one specific person, related to the outcome of interest, and derived from an individual assessment [12]." Tailoring approaches to education have been referred to as "individual tailoring [13]" or "computer tailoring [13]" To date, tailoring has been mainly been provided through web-based programs. However, recently, tailored apps have been developed, and

studies have demonstrated the effects of tailoring on the medication adherence of children with asthma [14].

A systematic review of digital asthma self-management intervention interactions since the 2000s has been reported, and interventions were mostly implemented at age 7 to 17 years [6,15]. In addition, because children who have their own smartphones are often older than 12 years, participants in most studies on pediatric asthma apps are commonly adolescents older than 12 years [16,17]. Although digital self-management programs are available for children with asthma and their caregivers, few studies have investigated the effectiveness of these mHealth apps. Moreover, most studies on asthma apps did not consider interactions involving communication between children and caregivers. Furthermore, previous studies on mHealth apps associated with asthma self-management in 7- to 12-year-old children were limited. Therefore, we developed a prototype mHealth child asthma app for children and their caregivers, including features of tailored feedback messages in continuing self-management and interactions between children and caregivers.

This study aimed to identify the beneficial features of a prototype mHealth app under development for children with asthma and their caregivers. The findings of this study are important to help perfect this app and complete a mHealth child asthma app for children and their caregivers.

Methods

Prototype Development

Review of mHealth apps for adolescents with chronic conditions revealed a paucity of evidence-based apps, in contrast to the thousands of apps available on the app market that are not evidence-based or user or professional informed [18]. A prototype tailored mHealth asthma app was developed based on a past tailored program using touch-screen computer [19]. The social cognitive theory [20] and tailoring [10] were used to develop a new mHealth child asthma app. This is useful in children with insufficient self-management ability because their communication levels are developmentally immature. The interpersonal approach to education had a greater effect than the tailored approach [10]; however, establishing a one-on-one relationship is difficult. Therefore, this program used the tailored approach to communication as it was most similar to the interpersonal approach. Individual interviews with caregivers were used to collect information and develop tailored messages. The prototype app was developed based on tailored program issues from past research [19] that did not involve children with asthma and their caregivers in their app design.

The mHealth child asthma app comprised a combination of asthma knowledge, behavior change, target behaviors, and theory applications. There were 3 main goals for this app: to increase caregivers' feelings of self-efficacy in their asthma management, to increase asthma knowledge, and to continue self-management behavior of controlling asthma symptoms.

The app's protocols comprised 3 content areas: asthma knowledge, monitoring symptoms, and behavior change (Figure 1). This app was developed for infants and toddlers aged 0 to

6 years, school-going children aged 7 to 9 years and 10 to 12 years according to their developmental stages. Contents of the child asthma app for children aged 0 to 6 years and their caregivers were asthma knowledge (pictorial book about asthma clinical condition, causal factors, and complicating factors; quiz), self-monitoring of medications and symptoms, and tailored feedback according to the Japanese pediatric asthma

control test [21]. Contents of this app for children aged 7 to 12 years and their caregivers were asthma knowledge (manga as a Japanese-style comic regarding asthma, medication, exercise-induced asthma, and stress management; quiz), self-monitoring of medication and symptoms, and tailored feedback according to the Japanese pediatric asthma control test.

Figure 1. Examples of a mobile asthma app for children and their caregivers: screens commonly used for children <12 years old; setting screen (left), childhood asthma control test (middle), and self-monitoring (right).

Study Design

A qualitative study was conducted, and data were analyzed using thematic analysis [22]. As part of the study, we conducted semistructured interviews, and their narratives were analyzed.

Recruitment

Children participants aged 0 to 12 years diagnosed with persistent asthma by allergy specialists at the National Center for Child Health and Development were recruited. The types of persistent asthma, such as severity, treatment regimen, and treatment duration, were not considered. Participants were excluded if their involvement was deemed inappropriate by the pediatrician because of their mental and physical conditions. We used purposive sampling, which gathers information-rich cases that manifest the phenomenon under investigation. Participants who matched the inclusion criteria were selected by their electronic medical record. Researchers approached caregivers of children with asthma who regularly visit the hospital through telephone. A total of 31 caregivers who met the inclusion criteria by telephone were contacted. After obtaining the temporary consent to participate in the study over the telephone, caregivers and children with asthma provided consents during the next outpatient visit.

Data Collection

We contacted pairs of children with asthma and caregivers. Participants received a mobile phone to demonstrate the prototype of the child asthma app and tried the app for 10 to 15 minutes in the interview room during their outpatient visit. After the prototype app trial, each participant was interviewed once for approximately 7 to 22 minutes (average 14 minutes). Interviews were conducted in an outpatient private room by one researcher, a nurse specializing in pediatric allergies from March 2019 to August 2019.

Data were gathered through semistructured interviews. An interview guide that matched the study aim was prepared. The researcher presented the interview guide and items to participants. The first author (MI), a researcher trained to carry out qualitative interviews, conducted all interviews. During each interview, the interviewer encouraged participants to talk freely about the proposed beneficial features of the app based on their experiences and provided opportunities to express features they felt had not been covered. The interviewer had no prior relationship with the participants.

Demographic information of participants, such as age, sex, age at onset of asthma, and relationship with caregivers, was collected. Interview contents were assessed using 4 viewpoints

for each caregiver and children aged 7 to 12 years: (1) good aspects of the app, (2) proposed beneficial features of child asthma app, (3) notification (alert function) frequency and usage frequency, and (4) specific improvements of the app. Additional demographic information, such as age of asthma onset and treatment duration, was also collected. The app trial was conducted once for approximately 10 to 15 minutes before the interview; thereafter, these viewpoints points were mainly discussed to identify the portion of the app that needed modification and grasp the proposed beneficial features of the app. The interview guide was confirmed among 5 researchers before the study. The interview was recorded using a digital voice recorder after obtaining the assent of participants.

Data Analysis

All recorded data were transcribed verbatim in Japanese. A thematic analysis was used to identify codes and themes (subcategories and categories) from the qualitative data [20]. Thematic analysis can be used to develop a novel theoretical framework and identify the common meanings and themes of an existing model. The qualitative data were analyzed in 4 phases: (1) becoming familiar with the collected data, (2) generating the codes and collating similar data for each code, (3) naming content-characteristic words and grouping them as categories, and (4) exploring themes agreed upon by 4 research members who engaged in consensus decision making. Each theme in this study was derived from interview data. Research members included 2 pediatric nurses (MI and MN) and 3 pediatricians specializing in allergies (YM, KY, and MN). After initial coding of each transcript, the researchers discussed and identified a set of main themes, categories, and subcategories.

Informed Consent/Assent and Ethical Considerations

The Japanese National Center for Child Health and Development committee for ethics in research of social medicine (approval number: 2028) and the university committee for ethics in research involving human subjects approved this study. Participants who met the inclusion criteria together with their caregivers were informed verbally and in writing about the aim, significance, and methods of the study. They were informed of their rights as voluntary participants, including the right to withdraw from the study, data anonymity, protection of confidential information, handling and disposal of data, and the possibility of results being published. After receiving this information, participants 7 to 12 years old provided informed written assent, and caregivers of all participants provided written consent. COREQ guidelines [23] were followed.

Results

Participant Demographics

Out of the total 31 caregivers who were contacted by telephone, 4 caregivers of 1- to 12-year-old children (2 girls and 2 boys) did not participate because they were unable to contact us by telephone several times, and it was difficult to secure the time of the outpatient visit. A total of 27 pairs of caregivers and children were interviewed. Participant characteristics are shown in Table 1. Participants were divided into the developmental stage groups: children younger than 7 years old (n=10) and children 7 to 12 years old (n=17). The majority of caregivers were between the ages of 40 and 49 years. Two viewpoints in the prototype of the child asthma app were identified: good aspects of the app and proposed beneficial features for the app.

XSL•FO

RenderX

Table 1. Participants' characteristics.

Participant characteristics	Participants, n (%)
Patient	**27 (100)**
Developmental stage (years)	
Preschool children	
2-3	3 (11)
4-6	7 (26)
School-aged children	
7-9	9 (33)
10-12	8 (30)
Sex	
Boy	18 (67)
Girl	9 (33)
Age of onset (years)	
0-2	14 (52)
3-5	9 (33)
6-8	4 (15)
Caregiver[a]	**26 (100)**
Relationship	
Mother	25 (96)
Father	1 (4)
Age (years)	
30-39	8 (31)
40-49	17 (65)
50-59	1 (4)

[a]Patients included 2 sets of siblings.

Good Aspects for Caregivers of Children younger than 7 Years Old With Asthma

General

Coding and classifying phases revealed 6 categories and 4 themes from 25 codes in *good aspects of the app for the caregivers of children younger than 7 years old with asthma* (Table 2). By grouping the categories, 4 themes (ie, confirmation of asthma knowledge, child–caregiver interaction, app design, and child's interest) were identified.

XSL•FO
RenderX

Table 2. Good aspects for children younger than 7 years old with asthma and their caregivers.

Themes and categories	Representative caregivers' verbatim comment
Confirmation of asthma knowledge	
Reconfirmation of asthma knowledge	"I was able to re-recognize my knowledge with the quiz." [40s, mother of a 5-year-old girl]
Acquisition of asthma knowledge	"I learned quiz knowledge." [40s, mother of a 2-year-old girl]
	"I heard from my pediatrician that I realized that I should be careful about daily life and playing outside." [40s, mother of a 5-year-old girl]
	"I don't seem to have knowledge, or I know it, but there are things I don't know, so if it's a quiz format, I think it's easy to understand and increase knowledge." [40s, mother of a 5-year-old boy]
Child–caregiver interaction	
Initiatives for parents and children	"Since the child still does not know about asthma, the app was good for the parent and child to do it together." [30s, mother of a 4-year-old girl]
	"Although it is a difficult content such as the asthma mechanism, there is an illustration and the child became aware of asthma." [30s, mother of a 4-year-old girl]
Design	
Ease of viewing the screen	"I thought app was easy to understand because of illustrations." [40s, mother of a 5-year-old boy]
	"The size of quizzes was good." [30s, mother of a 6-year-old boy]
Presence of childhood asthma app	"The good thing was that there was an asthma app." [40s, mother of a 2-year-old boy]
Child's interest	
Child's interest	"The child just got interested in the body and bought a picture book, so I thought the app would be fun in order to ask questions about the structure that rubs the coughing part." [40s, mother of a 4-year-old girl]

Theme 1: Confirmation of Asthma Knowledge

The coding and classifying phases revealed 2 categories in this theme. Caregivers of children younger than 7 years old with asthma reported good aspects including *reconfirmation of asthma knowled*ge and *acquisition of asthma knowledge*. They reported that they learned and recognized asthma knowledge through the app.

Theme 2: Child–Caregiver Interaction

The coding and classifying phases revealed 1 category in this theme. Caregivers of children younger than 7 years old with asthma reported good aspects including *initiatives for parents and children*. They reported that the picture book helped ascertain asthma knowledge with the child younger than 7 years old, creating a positive child–caregiver interaction (Figure 2).

Figure 2. Examples of features for children <7 years old with asthma and caregivers: main home screen (left) and asthma pictorial book with scratch and search for asthma deterioration factors (middle and right).
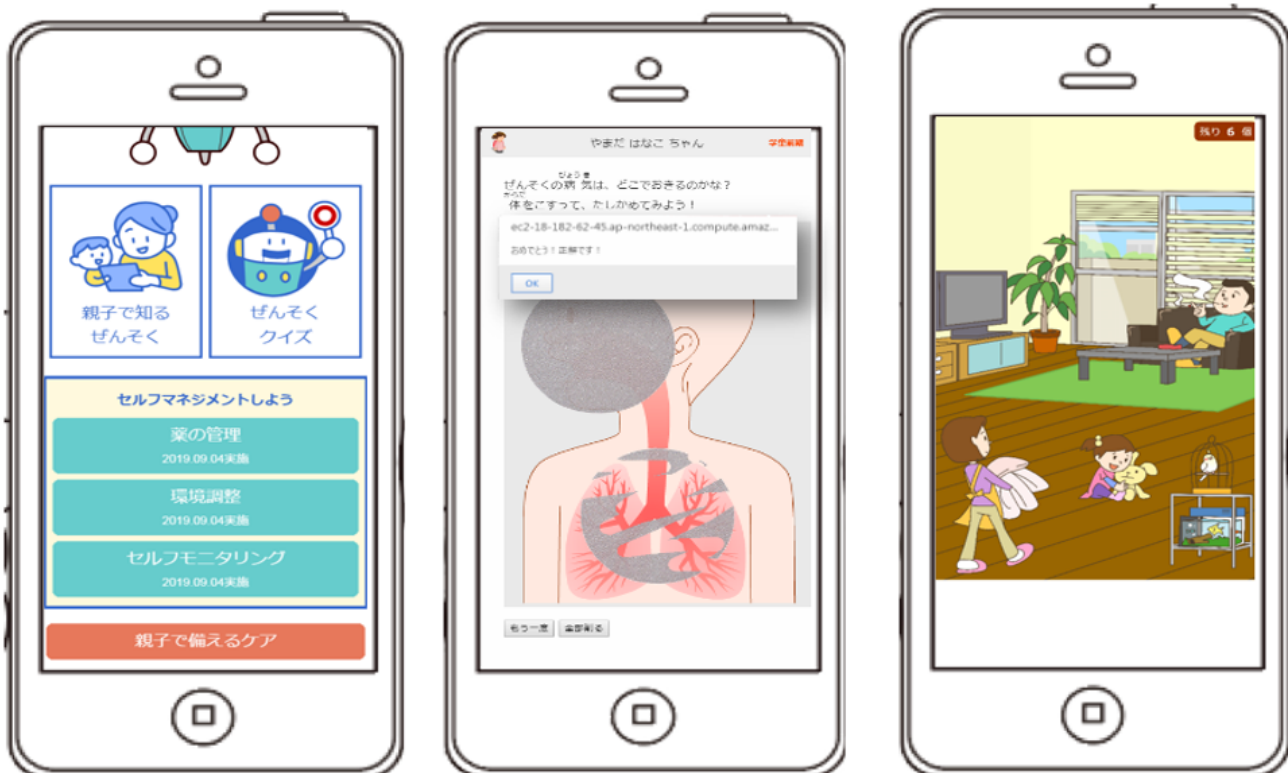
Theme 3: Design

The coding and classifying phases revealed 2 categories in this theme. Caregivers of children younger than 7 years old with asthma reported good aspects including *ease of viewing the screen* and *presence of childhood asthma app*. They reported that viewing the screen was easy and the font size was good. In addition, as the number of mobile child asthma apps is very limited in Japan, having this newly developed app was a good way to manage children's asthma.

Theme 4: Child's Interest

The coding and classifying phases revealed 1 category in this theme. Caregivers of children younger than 7 years old with asthma reported good aspects including *child's interest*. They reported that it was good that children 7 years old could learn with the caregiver as they began to be interested in how the body functions.

Good Aspects for 7- to 12-Year-Old Children With Asthma and Their Caregivers

General

The coding and classifying phases revealed 11 categories from 50 codes for *good aspects of the app for 7- to 12-year-old children with asthma and their caregivers* (Table 3). By grouping the categories, 5 themes were identified: new knowledge, manga, child's interest, app design, self-management promotion, and app operability.

Table 3. Good aspects for 7- to 12-year-old children and caregivers.

Themes and categories	Representative verbatim comment
New knowledge	
Acquisition of asthma knowledge	"The quiz gave me the knowledge I expected." [child aged 7, boy]
Specific explanation	"The app was easy to understand that I used to explain each difficult word." [child aged 12, boy]
New discovery	"Even though I knew the app, there was a new discovery (I thought I would die only through suffocation, but I knew that I usually suffer from a wound in the bronchi)." [child aged 9, boy]
Manga	
Manga	"I'm tired of writing alone, but I thought that manga was easy to understand." [child aged 12, girl]
	"Because the app for asthma knowledge was a manga, my child read it with interest." [40s, mother of a 7-year-old boy]
	"Manga was interesting." [child aged 10, boy]
	"It was easy to understand because there was manga that understood asthma well." [child aged 9, boy]
	Colored manga was easy for children to be interested in, and my child was also interested. [40s, mother of a 10-year-old boy]
Child's interest	
Child's interest	"My child was having fun." [40s, mother of a 7-year-old boy]
	"I thought I could play with the app." [child aged 9, boy]
	""My child likes playing with mobile phones, so I thought it would be interesting to do something with mobile phones." [40s, mother of a 7-year-old boy]
Trigger of self-management	
Opportunities for child self-management	The child still does not want to manage it with a paper diary; therefore, the app is an opportunity to do it himself. [40s, mother of a 7-year-old boy]
Opportunities for parent-child communication	"While checking with karuta [Japanese card game], I thought that the app would allow a proper communication between parents and children." [30s, mother of a 9-year-old boy]
Design and operability	
The color of the illustration	"The color is good and easy-to-read." [child aged 10, boy]
	"I thought the color of manga was really beautiful." [40s, mother of a 7-year-old boy]
Character size	"The size of the characters and kanji [Japanese-characters] was fine." [child aged 9, boy]
	"I could read the size of the character at karuta." [40s, mother of a 12-year-old girl]
	"The size of the letters could be large even if small." [child aged 10, boy]
Ordinariness	"The app was ordinariness." [child aged 7, boy]
Easy input operation for monitoring	"Asthma diary had to be issued individually, but the app was easy to use." [30s, mother of an 8-year-old boy]
	"The monitoring operation was easy." [30s, mother of an 8-year-old boy]

Theme 1: New Knowledge

The coding and classifying phases revealed 3 categories in this theme. Children aged 7 to12 years with asthma and their caregivers reported good aspects including *acquisition of asthma knowledge*, *specific explanation*, and *new discovery*. They reported that the app had provided asthma knowledge and they gained new information on childhood asthma.

Theme 2: Manga (Japanese-Style Comic)

The coding and classifying phases revealed 1 category in this theme. Children aged 7 to 12 years with asthma and caregivers reported good aspects including *manga about child asthma*. They reported that they enjoyed learning about asthma using manga (Figure 3).

Figure 3. Examples of the features of asthma knowledge for 7- to 12-year-old children: asthma quiz (left) and manga about stress management of asthma (middle and right).
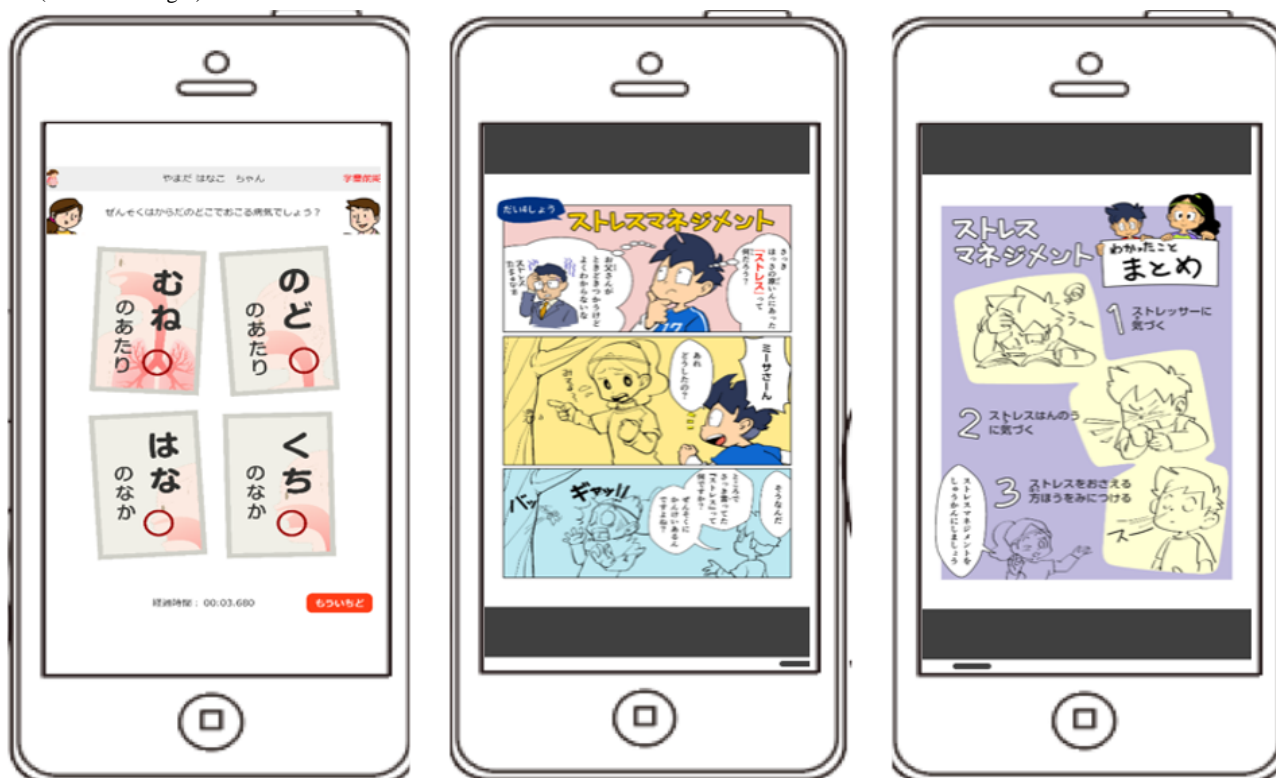
Theme 3: Child's Interest

The coding and classifying phases revealed 1 category in this theme. Children aged 7 to 12 years with asthma and their caregivers reported good aspects including *child's interest*. Caregivers reported that children had fun with the app, and 7- to 12-year-old children with asthma stated that they could play with the app.

Theme 4: Self-Management Promotion

The coding and classifying phases revealed 2 categories in this theme. Children aged 7 to 12 years with asthma and their caregivers reported good aspects including *opportunities for child self-management* and *opportunities for parent–child communication*. Caregivers reported that the app provided opportunities for the child's self-management and child–caregiver communication.

Theme 5: Design and Operability

The coding and classifying phases revealed 4 categories in this theme. Children aged 7 to 12 years with asthma and caregivers reported good aspects including *color of the illustration*, *character size*, *ordinariness*, and *easy input operation for monitoring*. They reported that the color of the illustration about manga and character size were good. Children aged 7 to 12 years with asthma stated that the app was not ordinary. In addition, caregivers reported that the operation of asthma diary and monitoring were easy.

Proposed App Features From the Perspective of Child or Caregiver

General

The coding and classifying phases revealed 44 subcategories and 12 categories from 214 codes in *proposed beneficial features for the app for children with asthma and their caregivers* (Table 4). By grouping the categories, we explored 6 themes: asthma knowledge, elements for continuity, universal design, monitoring, and functions.

XSL•FO
RenderX

Table 4. Proposed beneficial features for the app in children with asthma and caregivers.

Themes and categories	Subcategory	Developmental stage		
		2-6 years	7-9 years	10-12 years
Asthma knowledge				
Asthma knowledge	Asthma knowledge in general	✓	✓	✓
	Specific behavior of management		✓	
Regular provision of asthma knowledge	Regular provision of knowledge		✓	
	Newsletter		✓	
Elements for continuous				
Ideas that children can learn	Quiz	✓	✓	✓
	Manga		✓	✓
	Devices that can be used by younger children	✓		
	Sound/video	✓		✓
	Amusement	✓	✓	
	Character	✓		
Devices that can be used continuously	Game element		✓	✓
	Incentive for continuing		✓	
	Clear the stage/rank up			✓
	Increase self-efficacy	✓	✓	
	Update the information	✓		
	Not game element		✓	
	Consideration of developmental stage			✓
	Devices that do not end with a single use	✓		
	Clarification of what is ahead	✓	✓	
	Devices that can be used in the next step			✓
Character input for 7-12 years old children	Possible		✓	✓
	Impossible by developmental stage		✓	✓
Universal design				
Universal design	Illustration/color shade		✓	✓
	Visibility/readability			✓
Notification				
Notification function	Necessity of reminder notification	✓		✓
	Message notification function	✓		
Notification frequency	Every day at a fixed time	✓	✓	✓
	No notification required	✓	✓	
	Self-configuration	✓	✓	
	Notification once a week		✓	
Monitoring				
Monitoring function	Function of symptom input	✓	✓	✓
	Easy operation			✓
	Calendar function		✓	
	Remaining drug notification		✓	
	Remarks column settings		✓	

Themes and categories	Subcategory	Developmental stage		
		2-6 years	7-9 years	10-12 years
	Unnecessary	✓	✓	
	Self-configurable monitoring items		✓	
Monitoring frequency	Daily input	✓	✓	✓
	Entering all at once (eg, entering once a week)	✓		
Functions				
Used by 7-12 years old children alone	Use of apps by children alone		✓	✓
Operability and functions	Functions of the app	✓	✓	✓
	Family (sibling or parent) function	✓		
	Response	✓		
	What does not take time	✓		

Theme 1: Asthma Knowledge

The coding and classifying phases revealed 2 categories in this theme. Children with asthma and caregivers reported proposed beneficial features for the app including *asthma knowledge* and *regular provision of asthma knowledge*. They had needs for asthma knowledge in general knowledge and specific behavior management and the regular provision of asthma knowledge.

Theme 2: Elements for Continuity

The coding and classifying phases revealed 3 categories in this theme. Children with asthma and caregivers reported proposed beneficial features of the app including ideas that children can learn, *devices that can be used continuously* and character input for 7- to 12-year-old children. They expressed the need for fun elements for continuous use of the app, such as quiz, manga, game element, clear the stage, and so on.

Theme 3: Universal Design

The coding and classifying phases revealed 1 category in this theme. Children with asthma and caregivers reported proposed beneficial features of the app including *universal design*. They expressed the need for universal design, such as illustration or color shade and visibility or readability, that were easy to use for everyone, from children to adults.

Theme 4: Notification

The coding and classifying phases revealed 2 categories in this theme. Children with asthma and caregivers reported proposed beneficial features of the app including *notification function* and *notification frequency*. They expressed the need for message notification function and necessity of reminder notification every day at the fixed time or once a week or no notification.

Theme 5: Monitoring

The coding and classifying phases revealed 2 categories in this theme. Children with asthma and their caregivers reported proposed beneficial features of the app including *monitoring function* and *monitoring frequency*. They expressed the need for the easy operation, remaining drug notification, and daily input. In addition, a mother (in her 40s) of a 7-year-old boy reported that monitoring with the app can be shared with the pediatrician during their outpatient visit.

Theme 6: Functions

The coding and classifying phases revealed 2 categories in this theme. Children with asthma and caregivers reported proposed beneficial features of the app including *use by 7- to 12- year-old children alone* and *operability and functions*. Children aged 7 to 12 years generally do not own a mobile phone; therefore, they need to be able to use one. In addition, children with asthma and their caregivers needed operability and functions, such as family (sibling or parent) function among others.

Discussion

Principal Results

Four themes (confirmation of asthma knowledge, child-caregiver interaction, app design, and child's interest) comprising the good aspects of the app were identified for children older than 7 years old with asthma and their caregivers. Moreover, 5 themes (new knowledge, manga, child's interest, app design, self-management trigger, and app operability) comprising the good aspects of the app for 7 to 12 years old children with asthma and their caregivers were also determined. In addition, 6 themes (asthma knowledge, elements for continuous, universal design, monitoring, and functions) on the proposed beneficial features of the app were identified for children with asthma and their caregivers.

Proposed Beneficial Features for an mHealth Child Asthma App

Children with asthma and their caregivers captured the features of asthma knowledge as good aspects of the app. On the one hand, they wanted elements of continuity in the app, such as quiz, manga, and games, among others. Hospital outpatients identified requiring more knowledge about the diseases through the smartphone app [24]. In addition to providing asthma knowledge, ways should be devised for 7- to 12-year-old children with asthma and even their caregivers to learn asthma knowledge while continuously accessing the app. Manga about asthma in this study was highly acclaimed by 7- to 12-year-old

children with asthma. The concept of entertainment education [25], entertainment media that provide educational information intended to increase psychological readiness toward the desirable behavior change, such as television, digital games, and comics, has potentially become an effective communication strategy to implement the health message for the high-risk population [26]. Several educational programs have adapted manga as an effective learning tool, known as an educational manga [27,28]. Additionally, previous studies [29,30] challenged the application of manga or comic book for patient education targeting children. By incorporating Japanese culture manga into the app, 7- to 12-year-old children can easily accept new information that results in high educational effects. Serious games without entertainment, enjoyment, or fun as their primary purpose [31], have emerged as a new generation of videogames meant to provide education and training [32]. Although serious games designed for asthma education have evolved with advances in technology, results of their evaluation remained similar across studies, with clear improvements in knowledge but little to no change in behaviors and clinical outcomes [33]. Therefore, although game elements were useful in terms of gaining asthma knowledge, it was suggested that they were unsuitable for self-management and behavior change.

Two themes that were good aspects of this app were the child's interest and design and operability, which were commonly reported by 7- to 12-year-old children with asthma and their caregivers. Additionally, findings on the proposed beneficial features of the app suggested that children with asthma and their caregivers wanted a universal design that could easily be used by everyone. The design and development of a self-management app included validation of app features through user-centered design methods [34,35]. The patient-centered universal design of the app should also be improved, and its feasibility determined. Furthermore, caregivers of children with asthma should also enhance the basic functions and operability, not complex functions. Children aged 7 to 12 years with asthma and their caregivers required notifications, reminders, monitoring, and alerts. This supports the findings of previous studies [36] on asthma management mHealth app for adolescents. In addition, caregivers' needs for app monitoring were to be able to share the monitoring content with their pediatrician during their outpatient visits. The sharing of symptom monitoring in children's daily lives promotes communication between children with asthma and their caregivers and pediatricians and maintains good relationships. Moreover, sharing of monitoring content helps pediatricians make treatment decisions by understanding the appearance of detailed asthma symptoms of the children. The monitoring and alerting functions should be developed not only for the ease of use but also for the pace of the user. In future app improvements, ways to set monitoring items and notifications by the users themselves should be devised. In particular, 7- to 12-year-old children with asthma wanted to keep records in monitoring while having fun. As for the monitoring function of children with asthma, incentives obtained by the child's continuous medication intake and input symptoms are necessary.

Limitations

This study had several limitations. Findings derived from 27 children and 26 caregivers at one children's hospital were used to identify the proposed beneficial features of the app in children with asthma and their caregivers and are limited to that population. In addition, the majority of participants were boys and the majority of caregivers were mothers. Pediatric asthma has a high prevalence in boys in Japan, as is true worldwide. Furthermore, although the number of double-income families is increasing in Japan, mothers still attend the outpatient visit of children. However, in Japan, patients who go to children's hospital are those with severe asthma and are likely to use the app frequently for long term. Therefore, the results were useful in understanding the needs of the target population: children younger than 7 years old and 7 to 12 years old with asthma and their caregivers. In addition, this study has clarified caregiver-reported proposed beneficial features of the app that in children younger than 7 years old were not as they reported. Children less than 7 years old were having difficulty for answering the questions accurately given their limited language function and cognitive development; therefore, caregivers were asked to answer their needs.

Additionally, since school-aged children aged 7 to 12 years often do not have their own mobile phones, their caregivers' mobile phones were used. However, as shown in the results of this study, despite the child's inconvenience in using the app, using a caregiver's mobile phone can lead to communication and interaction between them and their caregivers. Furthermore, the study did not determine whether the app changed the participant's behavior. Moreover, there are very few apps for child asthma management in Japan; therefore, we did not include questions in the interview on the topics of "why they would not use an app" and "what they did not like." We also did not consider the 4 patients who declined to participate as being any different from the participants. However, considering that the app provides potentially more information is very important. Finally, it would be worth reiterating that the app was only tried for approximately 15 minutes at the time of the interview; the conditions would be very different if they were used at home. The next step should be to develop and complete the contents and features of the app including the results of this study. To continue the self-management behavior of asthma in children with asthma and their caregivers, communication and interactions between the child and caregiver should be enhanced through the app. In addition, it is necessary to consider behavior changes associated with the app usage.

Strengths and Future Research

One of the strengths of this study was that it included preschool children aged 0 to 6 years and school going children aged 7 to 12 years who were not previously targeted by app developers. Additionally, another strength was that it incorporated manga; therefore, children with asthma and their caregivers could have fun while learning about asthma. The app included monitoring symptoms and medication, asthma control, and tailored messages so that it could be used continuously rather than in a single sitting.

Based on the results of this study, the app was improved and completed. The completed app is now undergoing formal feasibility and usability testing in children with asthma and their caregivers. Data collection from 2 to 12 years old children, such as usage data of app features and their collection of behavior changes, will help evaluate its feasibility, usability, and operability. The process and content evaluation of the app will help identify the contents for effective patient education in preschool and school going children aged 0 to 12 years with asthma and their caregivers.

Conclusions

Children with asthma and their caregivers perceived the good aspects of the app about learning the asthma knowledge with fun, including manga, child–caregiver interaction, contents that interest children, and easy-to-read design such as colors. Children with asthma and their caregivers wanted not only the asthma knowledge but also the enhanced monitoring and notification functions of the app, elements that could be continuously used by children, and universal design that can easily be used for everyone.

Acknowledgments

Authors' Contributions

MI conducted the research idea, planning, and app development process; recruited patients, interviewed the participants; and wrote the manuscript. YM helped recruit the patients, developed the app content, discussed data analysis, and provided comments on the paper. KY, MN, and MN helped develop the app content, discussed data analysis, and provided comments on the paper. YO helped develop the app content and provided comments on the paper.

Conflicts of Interest

None declared.

References

1.   Kagen S, Garland A. Asthma and Allergy Mobile Apps in 2018. Curr Allergy Asthma Rep 2019 Feb 2;19(1). [doi: 10.1007/s11882-019-0840-z]

2.   Huckvale K, Morrison C, Ouyang J, Ghaghda A, Car J. The evolution of mobile apps for asthma: an updated systematic assessment of content and tools. BMC Med 2015 Mar;13:58 [FREE Full text] [doi: 10.1186/s12916-015-0303-x] [Medline: 25857569]

3.   Farzandipour M, Nabovati E, Sharif R, Arani MH, Anvari S. Patient Self-Management of Asthma Using Mobile Health Applications: A Systematic Review of the Functionalities and Effects. Appl Clin Inform 2017 Dec;8(4):1068-1081 [FREE Full text] [doi: 10.4338/ACI-2017-07-R-0116] [Medline: 29241254]

4.   Ramsey RR, Caromody JK, Voorhees SE, Warning A, Cushing CC, Guilbert TW, et al. A Systematic Evaluation of Asthma Management Apps Examining Behavior Change Techniques. The Journal of Allergy and Clinical Immunology: In Practice 2019 Nov;7(8):2583-2591. [doi: 10.1016/j.jaip.2019.03.041]

5.   Wu AC, Carpenter JF, Himes BE. Mobile health applications for asthma. J Allergy Clin Immunol Pract 2015;3(3):446-8.e1. [doi: 10.1016/j.jaip.2014.12.011] [Medline: 25725939]

6.   Alquran A, Lambert K, Farouque A, Holland A, Davies J, Lampugnani E, et al. Smartphone Applications for Encouraging Asthma Self-Management in Adolescents: A Systematic Review. IJERPH 2018 Oct 29;15(11):2403. [doi: 10.3390/ijerph15112403]

7.   Morton RW, Elphick HE, Rigby AS, Daw WJ, King DA, Smith LJ, et al. STAAR: a randomised controlled trial of electronic adherence monitoring with reminder alarms and feedback to improve clinical outcomes for children with asthma. Thorax 2016 Nov 04;72(4):347-354. [doi: 10.1136/thoraxjnl-2015-208171]

8.   Davis SR, Peters D, Calvo RA, Sawyer SM, Foster JM, Smith L. Kiss myAsthma: Using a participatory design approach to develop a self-management app with young people with asthma. J Asthma 2017 Oct 24:1-10. [doi: 10.1080/02770903.2017.1388391] [Medline: 29064746]

9.   Jaimini U, Thirunarayan K, Kalra M, Venkataraman R, Kadariya D, Sheth A. "How Is My Child's Asthma?" Digital Phenotype and Actionable Insights for Pediatric Asthma. JMIR Pediatr Parent 2018 Dec 14;1(2):e11988. [doi: 10.2196/11988]

10. Kreuter MW, Strecher VJ, Glassman B. One size does not fit all: The case for tailoring print materials. ann. behav. med 1999 Dec;21(4):276-283. [doi: 10.1007/bf02895958]

11. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]

12. Kreuter M, Farrell D, Olevitch L, Brennan L. Tailoring Health Messages: Customizing Communication with Computer Technology. New York: Routledge; 2000.

XSL•FO

RenderX

13. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]

14. Kenyon CC, Sundar KG, Gruschow SM, Quarshie WO, Feudtner C, Bryant-Stephens TC, et al. Tailored medication adherence incentives for high-risk children with asthma: a pilot study. J Asthma 2019 Aug 07:1-7. [doi: 10.1080/02770903.2019.1648503] [Medline: 31389724]

15. Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, Mackenzie AM, et al. Digital asthma self-management interventions: a systematic review. J Med Internet Res 2014 Feb;16(2):e51 [FREE Full text] [doi: 10.2196/jmir.2814] [Medline: 24550161]

16. Roberts C, Geryk L, Sage A, Sleath B, Tate D, Carpenter DM. Adolescent, caregiver, and friend preferences for integrating social support and communication features into an asthma self-management app. J Asthma 2016 Apr 26. [doi: 10.3109/02770903.2016.1171339] [Medline: 27116540]

17. Carpenter DM, Geryk LL, Sage A, Arrindell C, Sleath BL. Exploring the theoretical pathways through which asthma app features can promote adolescent self-management. Transl Behav Med 2016 Apr 26. [doi: 10.1007/s13142-016-0402-z] [Medline: 27118115]

18. 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;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]

19. Iio M, Hamaguchi M, Narita M, Takenaka K, Ohya Y. Tailored Education to Increase Self-Efficacy for Caregivers of Children With Asthma. CIN: Computers, Informatics, Nursing 2017;35(1):36-44. [doi: 10.1097/cin.0000000000000295]

20. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. New Jersey: Prentice-Hall; 1986.

21. Nishimuta Y, Watanabe H, Sato K, Nezu Y, Matsuura T, Suzuki S. A study on the usefulness if the Japanese pediatric asthma control program (JPAC) [Japanese]. Jpn J Pediatr Allergy Clin Immunolo 2008;22(1):135-145.

22. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

23. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

24. Micallef C, McLeod M, Castro-Sánchez E, Gharbi M, Charani E, Moore LS, et al. An Evidence-Based Antimicrobial Stewardship Smartphone App for Hospital Outpatients: Survey-based Needs Assessment Among Patients. JMIR Mhealth Uhealth 2016 Jul 06;4(3):e83. [doi: 10.2196/mhealth.5243]

25. Singhal A, Wang H, Roger E. The rising tide of entertainment-education in communication campaign. In: Rice RE, Atkin CK. editors. Public communication campaigns. 4th ed. London: SAGE; 2013:321-333.

26. Literat I, Chen NN. Communication Infrastructure Theory and Entertainment-Education: An Integrative Model for Health Communication. Commun Theor 2013 Jun 05;24(1):83-103. [doi: 10.1111/comt.12011]

27. Murakami S, Bryce M. Manga as an Educational Medium. The International Journal of the Humanities: Annual Review 2009;7(10):47-56. [doi: 10.18848/1447-9508/cgp/v07i10/42761]

28. Cantor A. Learning about the brain with comics. The Lancet Neurology 2014 Aug;13(8):764. [doi: 10.1016/s1474-4422(14)70025-0]

29. Maxwell E, Simmons M, Franklin L, Arnold J, Pall H. Impact of Educational Cartoon on Pediatric Bowel Preparation Quality at Time of Colonoscopy. Global Pediatric Health 2014 Aug 28;1:2333794X1454819. [doi: 10.1177/2333794x14548199]

30. Shigehatake Y, Yokota C, Amano T, Tomii Y, Inoue Y, Hagihara T, et al. Stroke Education Using an Animated Cartoon and a Manga for Junior High School Students. Journal of Stroke and Cerebrovascular Diseases 2014 Jul;23(6):1623-1627. [doi: 10.1016/j.jstrokecerebrovasdis.2014.01.001]

31. Michael D, Chen S. Serious games: games that educate, train, and inform. Association for Computing Machinery: ACM Digital Library; 2005.

32. Licari A, Ferrante G, Marseglia, MD GL, Corsello, MD G, La Grutta S. What Is the Impact of Innovative Electronic Health Interventions in Improving Treatment Adherence in Asthma? The Pediatric Perspective. The Journal of Allergy and Clinical Immunology: In Practice 2019 Nov;7(8):2574-2579. [doi: 10.1016/j.jaip.2019.08.008]

33. Drummond D, Monnier D, Tesnière A, Hadchouel A. A systematic review of serious games in asthma education. Pediatr Allergy Immunol 2017 Jan 27;28(3):257-265. [doi: 10.1111/pai.12690]

34. 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 Sep;Suppl:49-56 [FREE Full text] [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]

35. Goyal S, Morita P, Lewis GF, Yu C, Seto E, Cafazzo JA. The Systematic Design of a Behavioural Mobile Health Application for the Self-Management of Type 2 Diabetes. Can J Diabetes 2016 Feb;40(1):95-104. [doi: 10.1016/j.jcjd.2015.06.007] [Medline: 26455762]

XSL•FO
RenderX

36.   Sage A, Roberts C, Geryk L, Sleath B, Tate D, Carpenter D. A Self-Regulation Theory-Based Asthma Management Mobile
       App for Adolescents: A Usability Assessment. JMIR Hum Factors 2017 Feb 01;4(1):e5 [FREE Full text] [doi:
       10.2196/humanfactors.7133] [Medline: 28148471]

Original Paper

Predicting Early Warning Signs of Psychotic Relapse From Passive Sensing Data: An Approach Using Encoder-Decoder Neural Networks

Daniel A Adler[1], BSc; Dror Ben-Zeev[2], BA, MSc, PhD; Vincent W-S Tseng[1], BSc; John M Kane[3], MD, BA; Rachel Brian[2], MPH; Andrew T Campbell[4], BSc, MSc, PhD; Marta Hauser[5], PhD; Emily A Scherer[6], PhD; Tanzeem Choudhury[1], BSc, MSc, PhD

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

[2]BRiTE Center, Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

[3]Department of Psychiatry, The Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, United States

[4]Dartmouth College, Computer Science, Hanover, NH, United States

[5]Vanguard Research Group, Glen Oaks, NY, United States

[6]Biomedical Data Science Department, Dartmouth Geisel School of Medicine, Hanover, NH, United States

Corresponding Author:
Daniel A Adler, BSc
Cornell Tech
2 W Loop Rd
New York, NY, 10044
United States
Phone: 1 2155953769
Email: daa243@cornell.edu

Abstract

Background: Schizophrenia spectrum disorders (SSDs) are chronic conditions, but the severity of symptomatic experiences and functional impairments vacillate over the course of illness. Developing unobtrusive remote monitoring systems to detect early warning signs of impending symptomatic relapses would allow clinicians to intervene before the patient's condition worsens.

Objective: In this study, we aim to create the first models, exclusively using passive sensing data from a smartphone, to predict behavioral anomalies that could indicate early warning signs of a psychotic relapse.

Methods: Data used to train and test the models were collected during the CrossCheck study. Hourly features derived from smartphone passive sensing data were extracted from 60 patients with SSDs (42 nonrelapse and 18 relapse >1 time throughout the study) and used to train models and test performance. We trained 2 types of encoder-decoder neural network models and a clustering-based local outlier factor model to predict behavioral anomalies that occurred within the 30-day period before a participant's date of relapse (the near relapse period). Models were trained to recreate participant behavior on days of relative health (DRH, outside of the near relapse period), following which a threshold to the recreation error was applied to predict anomalies. The neural network model architecture and the percentage of relapse participant data used to train all models were varied.

Results: A total of 20,137 days of collected data were analyzed, with 726 days of data (0.037%) within any 30-day near relapse period. The best performing model used a fully connected neural network autoencoder architecture and achieved a median sensitivity of 0.25 (IQR 0.15-1.00) and specificity of 0.88 (IQR 0.14-0.96; a median 108% increase in behavioral anomalies near relapse). We conducted a post hoc analysis using the best performing model to identify behavioral features that had a medium-to-large effect (Cohen d>0.5) in distinguishing anomalies near relapse from DRH among 4 participants who relapsed multiple times throughout the study. Qualitative validation using clinical notes collected during the original CrossCheck study showed that the identified features from our analysis were presented to clinicians during relapse events.

Conclusions: Our proposed method predicted a higher rate of anomalies in patients with SSDs within the 30-day near relapse period and can be used to uncover individual-level behaviors that change before relapse. This approach will enable technologists and clinicians to build unobtrusive digital mental health tools that can predict incipient relapse in SSDs.

XSL·FO
RenderX

Introduction

Background

Schizophrenia spectrum disorders (SSDs) are complex chronic conditions characterized by a diverse set of symptoms that present themselves heterogeneously throughout the affected population. Symptoms are typically categorized into 2 groups: *positive* symptoms, which are an exaggeration of normal function (eg, hallucinations, disorganized speech or thought) and *negative* symptoms, described as a loss of normal function (eg, lack of expressiveness, apathy, and asociality) [1]. Symptom exacerbation in SSDs leads to a psychotic relapse. Relapse has serious potential consequences, jeopardizing many aspects of patients' lives, including personal relationships and employment, with an increased risk of patients causing harm to themselves or others [2]. Previous research has estimated that the annual direct medical cost of schizophrenia amounts to US $37.7 billion within the United States, with an even larger indirect cost (US $117.3 billion) [3]. Early detection of relapse could inform time-sensitive clinical efforts that may reduce the severity of relapses or prevent their occurrence altogether.

The heterogeneity of symptoms and the timing of symptom exacerbation make detecting early warning signs of relapse difficult. Relapse symptoms, unlike common first-episode psychosis symptoms, can appear abruptly [2]. In-depth interviews with patients with SSDs describing their prerelapse symptoms show that symptom manifestation is extremely idiosyncratic but often consistent within individuals. Each individual may have their own unique *relapse signature*, and identifying this signature could be the most effective manner of detecting incipient relapse [4]. Traditional measures of relapse come from clinician-administered rating scales that attempt to quantify a patient's current experience with an SSD [5,6]. However, it is often unlikely that patients present themselves to a clinician when their symptoms begin to re-emerge or worsen, particularly in an illness characterized by cognitive disorganization, loss of insight, and inconsistent treatment delivery systems where it can be difficult to access care [7]. To prevent symptom exacerbation, tools need to be developed that are able to detect early warning signs of relapse outside of the clinic.

Over the past decade, improvements in sensing technologies within smartphones, wearables, and other devices have created new opportunities for remote measurement of mental health symptoms [8,9]. Behavioral data collected with passive sensors from smartphones offer unobtrusive methods to measure trajectories of mental health and mental illness [10-13]. Smartphones can track a diverse set of behaviors and are owned and utilized by most individuals with SSDs [14,15].

The CrossCheck system was the first smartphone-based tool designed to collect passive sensing data as a method of tracking the symptoms of SSDs. The system combines passive sensing with triweekly self-reported survey measures [16]. Using CrossCheck, researchers were able to predict patient self-reported ecological momentary assessments (EMAs) from passive sensing data and combine both the passive sensing and self-reported data to predict clinician-administered Brief Psychiatric Rating Scale (BPRS) scores [17,18]. In addition, researchers were able to detect significant changes in patient smartphone social behavior during the 30 days preceding relapse [19]. Although these analyses provide a foundation for identifying symptom changes that contribute to relapse, it is an open question whether we can predict specific time points of symptom exacerbation that show a clear relapse signature.

Relapse is a rare event, and lack of available data near relapse can make prediction problematic [20]. Anomaly detection is a branch of data mining specifically for the prediction of peculiar, infrequent events [21,22]. Traditional approaches for anomaly detection within time series involve forecasting and use statistical measures based on cumulative sums, moving averages, and regression models that rely on predicting changes in the underlying distribution of the time series [22]. Forecasting human behavior is an extremely difficult problem, and behavioral data from patients with schizophrenia do not traditionally follow the circadian rhythms seen within a healthy population [23,24]. Algorithms designed to learn complex features within time series data are likely to have more success in finding anomalies and detecting behaviors associated with relapse.

More novel approaches to time series anomaly detection use encoder-decoder neural network models to identify anomalies in multivariate time series data. These algorithms have had success in learning complex features, specifically in highly irregular sensing data [25-27]. Unlike statistical approaches, neural networks do not require assumptions about the underlying distribution of the data and are often ideal compared with classical machine learning techniques because they can provide accurate predictions without the need for complex feature engineering. However, there is a tradeoff. It can be difficult to interpret the reasoning behind why neural networks make specific predictions, leading to the common description that neural networks are *black box* models. In medicine, specifically, interpretability is important because clinicians need to justify the risk of using new approaches; thus, it is challenging to introduce neural network–based decision making into the clinical workflow [28]. Machine learning researchers focused on model interpretability have offered approaches to analyze models *post hoc*, after model training, to uncover the relationships between the input features to the network and the network prediction [29]. To successfully implement a neural network–based anomaly detection system within behavioral health, one needs to not only show good results in detection but also provide a

process for uncovering the underlying behaviors that lead to an anomaly and provide a clinical translation for those behaviors.

Related Work

Researchers have begun to utilize anomaly detection to predict early warning sides of psychosis. A pilot study using a combination of mobile sensing features and self-reported survey responses was able to identify an increase in anomalies within 2 weeks of relapse in a small patient population [30]. Another recent study utilized retrospective features extracted from Facebook to create a classifier for detecting the 1-month period before relapse and then analyzed the behaviors that were significantly different within this period [31]. Developing an algorithm to predict specific days of symptom exacerbation before relapse using exclusively passive sensing data could provide clinicians an unobtrusive method to measure SSD symptoms without the need for patient self-reporting.

Contributions

This study makes the following contributions:

1. We created a variety of encoder-decoder neural network–based anomaly detection models to predict early warning signs of psychotic relapse using passive sensing data collected from a smartphone. To the best of our knowledge, these are the first models designed to predict early warning signs of relapse using exclusively passive sensing data.

2. We provided a post hoc analysis for clinical interpretation of the detected anomalies within the context of SSDs and demonstrated that our algorithm can detect participant-specific relapse signatures.

3. We analyzed how variations in participant data can change model performance to provide guidance for future researchers in digital mental health for model and study design.

Methods

CrossCheck System and Study

The CrossCheck system was an Android app combined with a cloud-based data collection and storage platform. The app continuously collected users' passive sensing data and prompted participants every 2 to 3 days to self-report EMAs to track both positive and negative symptoms of SSDs [17,32]. EMAs were not utilized in our anomaly detection system owing to low completion rates across relapse participants. Table 1 provides an overview of the raw passive sensing data collected using CrossCheck. Sensors also collected environmental data, including ambient sound and light. The ambient sound was utilized by the app to classify when conversations occurred near the participant, but the raw sound and light data were not used in this research. Refer to our previous work for more specific information about the data collected during this study [16,17].

Table 1. Summary of passive sensing behavioral data collected throughout the study.

Behavior	Description	Derived hourly features
Acceleration	3-axis acceleration data were collected from a smartphone, sampled from 50-100 Hz. Previous CrossCheck studies utilized the Android activity recognition API[a], which classifies activity data as follows: on bicycle, still, in vehicle, tilting, or unknown. In this study, we chose to use raw acceleration features to make our anomaly detection system independent of a specific activity recognition API platform	Mean acceleration over the hour
App use	CrossCheck recorded apps running on a user's smartphone every 15 min	Number of unique apps opened within an hour
Call	Phone calls can indicate social interaction and communication. We tracked when incoming, outgoing, missed, rejected, and blocked calls occurred	Number and duration of incoming, outgoing, missed, rejected, and blocked calls
Conversation	Previous studies have investigated the link between conversations, human voice, and mental health [12,33,34]. We detected human voices and conversational episodes using algorithms from our previous work [35]	Number and duration of conversations
Location	Previous research has shown that location can be associated with mental health [12,13,36]. We tracked location information from users through their smartphones	Time in primary, secondary, and all other locations as well as total distance travelled in the hour
Screen activity	The amount of time users spend on their phones can be tracked to learn normal daily behaviors. The time users' screens were on versus off was recorded	Number of times the phone was used as well as the duration of use
Sleep	On each day, the sleep duration, onset, and wake time were detected. These calculations occurred using a combination of information based upon users' screen time, physical activity, ambient sound, and light [12,37]	Sleep duration, onset. and wake time. As we estimated only the longest sleep episode per day, this is technically a daily feature. We replicated these features across all hours within a single day
Text	Text messages are another indicator of social interaction. We tracked when texts were received, sent, drafted, left in a user's outbox, failed to send, and were queued for sending	Number of received, sent, drafted, outbox, failed to send, and queued messages

[a]API: application programming interface.

The CrossCheck study was a randomized controlled trial (RCT) aimed at testing the efficacy of using passive sensing and self-reported data to identify digital indicators of relapse. The participants enrolled were randomized either into a smartphone arm for passive sensing data collection or into a control arm to receive treatment as usual. In this work, because our goal was to predict early warning signs of relapse from collected passive sensing data, we focused exclusively on the smartphone arm. Participants enrolled in the study were given an Android smartphone for 12 months and instructed to carry the device with them and complete the EMA. Trained clinical assessors met with participants to conduct a baseline assessment of symptoms and functioning. Clinical assessors also conducted follow-up assessments with participants during months 3, 6, 9, and 12 of the study to administer the 7-item BPRS, which measures psychiatric symptoms associated with SSDs [6,16]. Participants' electronic medical records (EMRs) were also made available to the clinical assessors. The following events, either reported during assessment or recorded within the EMR, were designated as relapse: psychiatric hospitalization, a significant increase in psychiatric care (including more intensive or frequent services, increased medication dosage, or additional medication prescribed) coupled with an increase of 25% from the baseline total BPRS score, suicidal or homicidal ideation with clinical relevance, self-injury, or violent behavior resulting in harm to another person or property [19]. The date of relapse, any notes surrounding the relapse event, and the reason for designating the event as a relapse were recorded. When corroborating evidence surrounding the relapse was not available within the EMR, clinicians worked with participants during the assessments to gain more information regarding the relapse event.

Relapse is an acute event, but when the early warning signs of relapse begin to surface is an open question. Consistent with previous research on early warning signs of relapse, we defined the 30-day period before relapse as the *30-day near relapse period* (NR30), and all data outside of this period were considered *days of relative health* (DRH) [19,31,38].

Study Protocol

The CrossCheck study was approved by the Committee for Protection of Human Subjects at the Dartmouth College and the Institutional Review Board of the Northwell Health System. The study was registered as a clinical trial (NCT01952041).

Participants

Participants were recruited into the RCT from several treatment programs at a psychiatric hospital in New York. Participants were recruited through flyers posted at the study site with the research coordinator's phone number. In addition, researchers reviewed the hospital's EMRs to identify potential participants. A potential participant's clinician was contacted by the investigative team, and after describing the study to the patient, clinicians referred patients interested in the study to the research team.

Eligible participants met the following inclusion criteria: (1) a chart diagnosis of schizophrenia, schizoaffective disorder, or psychosis not otherwise specified, (2) 18 years of age, and (3) an inpatient psychiatric hospitalization, daytime psychiatric

hospitalization, outpatient crisis management, or short-term psychiatric hospital emergency room visit within 12 months before beginning the study. Individuals were excluded if they had the following: (1) hearing, vision, or motor impairment that would impede smartphone usage (determined using a smartphone demonstration during screening), (2) a below sixth grade reading level (determined using the Wide Range Achievement Test–4th Edition), and (3) unable to provide informed consent (using a competency screener) [16,39].

A total of 1367 individuals were initially assessed for eligibility and 149 were enrolled in the study. Eligible individuals who did not enroll were no longer receiving care at the hospital (n=682), failed to meet the diagnostic criteria (n=131), did not want to participate (n=129), or did not meet the severity criteria (n=108). Of the 149 individuals enrolled, 62 were randomized into the smartphone arm of the study [19]. Participants included in this work (n=60) were required to have had at least 10 DRHs collected by the smartphone app.

Feature Extraction and Data Cleaning

An advantage of using neural networks for machine learning is that they have the ability to learn intricate features from raw data [40]. We sought to create features for our learning algorithm that were close to the raw data to exploit this fact. Hourly features were created from the raw sensor data. A summary of the hourly features used can be found in Table 1. In addition to the passive sensing features, we included the day of the week and the hour of the day as features in our model. The few features that require more complex calculations are described below.

Android phones track acceleration within a 3D *x, y, and z* coordinate system. This produces 3 values for every acceleration reading, namely $a=(a_x, a_y, a_z)$. We computed the mean hourly acceleration by taking the vector norm of each *a* within a specific hour and averaging over the values.

We also tracked the longitude and latitude locations over time for each participant. The locations for each participant were clustered using the density-based spatial clustering of applications with noise (DBSCAN) algorithm, implemented in the scikit-learn library [41,42]. DBSCAN clusters samples of high density together, requires a minimum number of samples per cluster, and requires a maximum distance, ε, between points to be specified as hyperparameters. We required a minimum of 10 samples per cluster and set ε=1 km. For each participant, the 2 majority clusters were tagged as the participant's *primary* and *secondary* locations, and all other data points were grouped together into a third cluster. Finally, we calculated the distance between each pair of longitude and latitude coordinates using the Haversine formula [43]. We then summed the distances over each hour.

Two types of missing data were identified. The first type of missing data (type 1) occurred when there was a sensor reading during an hour for one feature but there was no reading within the same hour for another feature. We imputed missing data for type 1 values with a "0," indicating our belief that the CrossCheck system was functioning during these hours, but an individual did not partake in specific behaviors that the system

records (eg, no texts were recorded within an hour). A second type of missing data (type 2) was identified during hours where all features were missing. We imputed features for the second type of missing data utilizing the mean value of a given feature for that hour. Location features (time spent in primary, secondary, and other locations) were filled differently. We assumed that the participant remained at their last recorded location and filled the features accordingly.

We assumed that by using mean filling for type 2 missing values, we would direct our anomaly detection models to focus on finding anomalies within the actual passive sensing data. That being said, missing values, specifically type 2 missing values, could have an implication for function. For example, if a participant stopped using their phone and the smartphone app, missing values could be an indication of asocial behavior, which may precede relapse. We added an additional feature to the model that indicated the percentage of features filled within a given hour. If this feature was <1, the hour was filled using the type 1 missing data procedure, but if the feature was equal to 1, the hour was filled using the type 2 missing data procedure.

Encoder-Decoder Models

We developed multiple algorithms to detect early warning signs of relapse using passive sensing data. Patients with SSDs are known to not experience normal circadian rhythms that are typically found within a healthy population [23,24]. Thus, we chose to apply a neural network approach to this problem that has been used for multivariate anomaly detection in irregular sensor data [25-27]. Specifically, we created a fully connected neural network autoencoder (FNN AD) model and a gated recurrent unit sequence-to-sequence (GRU Seq2Seq) model that learned to reconstruct an input time series [44]. A GRU network was used over a vanilla recurrent neural network (RNN) and other popular RNN architectures, such as a long short-term memory (LSTM) network, as the GRUs counter the vanishing gradient problem that occurs when training the vanilla RNNs, and they converge faster during training than LSTM networks [45]. After training the encoder-decoder models, our algorithm learned participant-specific anomaly thresholds based on the model reconstruction error. We discuss the architecture of the encoder-decoder models in this section and describe the thresholding procedure in the subsequent sections.

We considered each participant's data to be a time series of varying length L, $X=\{x^{(1)},...,x^{(L)}\}$, where each $x^{(i)}$ is a multivariate data point, $x^{(i)} \in R^m$. In our case, each $x^{(i)}$ represented a set of hourly features for a single participant. We created subsequences of data of length l starting at each i, $i=\{i,...,L-l+1\}$. Note that a given data point, $x^{(i)}$, could be potentially included within each of the $1,...,L$ subsequences. For the FNN AD model, we let $l=1$, and for the GRU Seq2Seq model, we let $l=24$.

The models were constructed as follows. This is also detailed in Figure 1. The FNN AD model comprised 2 fully connected encoder and decoder layers that compressed an input subsequence to a lower dimension and then recreated the initial subsequence. For the GRU Seq2Seq model, we first input a subsequence of data into a single *encoding* layer of a bidirectional GRU with a specified hidden unit size. A bidirectional layer was used for the encoder because previous research has shown that bidirectional layers improve the results over unidirectional layers [46]. The last cell in the encoding layer outputs a prediction for the next timestep, $x'^{(l+1)}$, and encodes hidden information from the entire sequence, $h^{(l+1)}$. We then passed this information as inputs into a unidirectional GRU *decoder* layer that reconstructed the subsequence in reverse order: $\{x'^{(l)},...,x'^{(1)}\}$.

Figure 1. Encoder-decoder neural network architectures. (a) the architecture for the fully connected neural network autoencoder (FNN AD) model. (b) the architecture for the GRU sequence-to-sequence (GRU Seq2Seq) model.
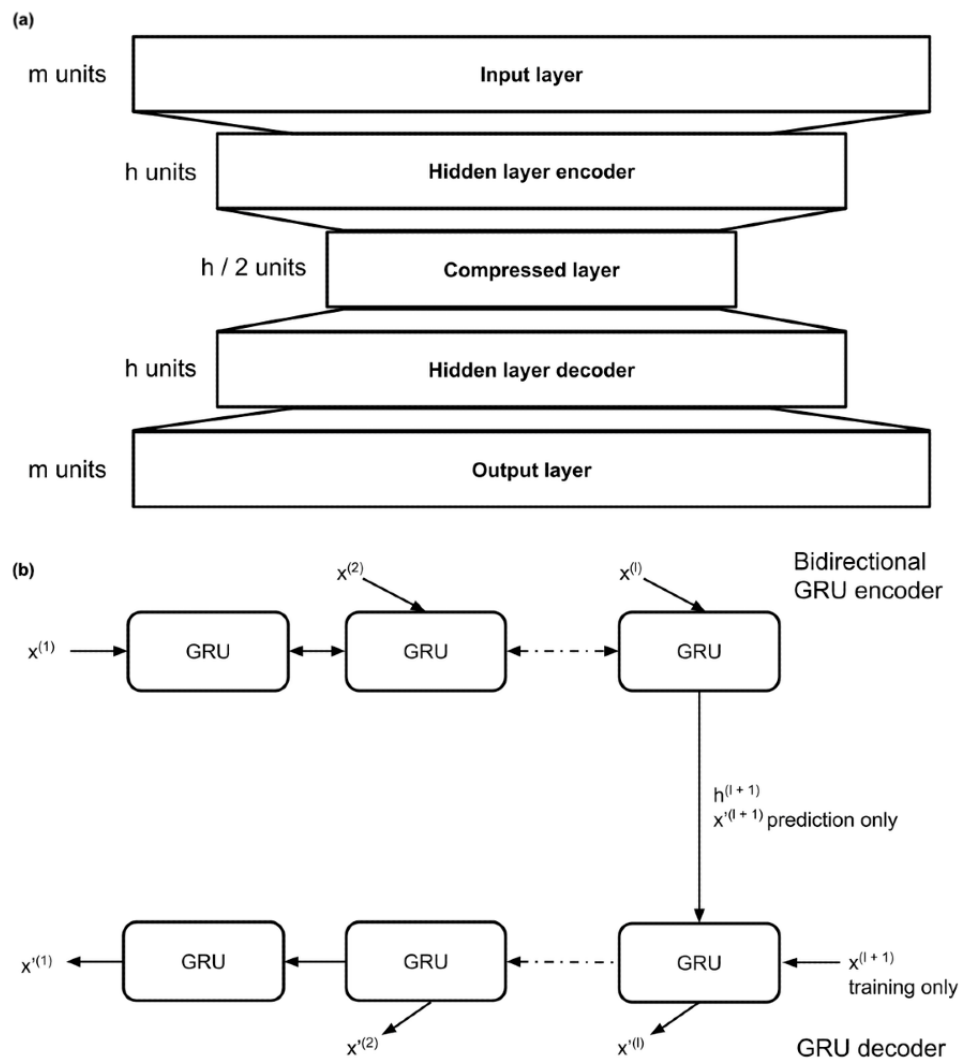
Model Training Procedure

We utilized a similar data-splitting and cross-validation procedure as described in previous work [25-27]. The data for each participant were first split into equal length nonoverlapping subsequences, and the subsequences were placed into 1 of 4 data sets. Defining NR30 as the 30-day near relapse period and DRH as days of relative health (ie, all days not in NR30), the data were split into the following:

1. A training data set, comprising *only* DRH, called H_R. These training data are utilized to train each model.
2. A cross-validation data set, comprising *only* DRH, called H_{CV}. These cross-validation data are utilized to validate the ability of the models to reconstruct sequences of new data.
3. A cross-validation data set, comprising DRH and NR30, called N_{CV}. These cross-validation data are used to tune the anomaly detection component of our algorithm as described in the following section.
4. A test data set, comprising *both* DRH and NR30, called N_T. The test data set are used to report the metrics of the anomaly detection algorithm described in the Results section.

We also experimented with the percentage of relapse participant data to include in each of these 4 data sets. Specifically, we experimented with placing different percentages of DRH within H_R and H_{CV}. We experimented with placing 0%, 20%, 40%, 60%, and 80% of relapse participants' DRH into H_R and H_{CV}. DRH for both relapse and nonrelapse participants were split such that 80% of DRH were placed into H_R and 20% into H_{CV}. Nonrelapse participant data were split entirely between H_R and H_{CV}.

Monte Carlo cross-validation was used to examine the robustness of the algorithm across different potential N_{CV} and N_T. We stratified each Monte Carlo sample to place equal amounts of NR30 data per participant within N_{CV} and N_T. The Monte Carlo procedure was repeated over 100 iterations and the median and IQR of the true-positive rate (TPR or sensitivity), and false-positive rate (FPR) of the current Monte Carlo test set N_T were recorded.

Anomaly Detection System

We used the trained encoder-decoder models to reconstruct H_{CV}, N_{CV}, and N_T, producing H'_{CV}, N'_{CV}, and N'_T. For a data

point $x^{(i)}$ in each of these data sets and its reconstructed counterpart $x'^{(i)}$, we calculated the absolute error of the data points: $e^{(i)}=|x^{(i)}-x'^{(i)}|$.

Within our algorithm, the full-time series was split into subsequences of length l, and any point $x^{(i)}$ could appear in at most l different subsequences. Thus, for a point $x^{(i)}$, there can exist l different predictions, $\{x_1'^{(i)},...,x_l'^{(i)}\}$. We filtered our data set to include only points that were predicted l times. The error vectors for these data were considered to be normally distributed, $e^{(i)} \sim N(\mu, \Sigma)$, and the error between H_{CV} and H'_{CV} was used to approximate (μ, Σ) parameterizing the *expected* error of our algorithm. We then calculated an anomaly score $s^{(i)} \in R$ for error vectors between N_{CV}, N'_{CV}, and N_T, N'_T using the Mahalanobis distance, which calculates the distance of a point to a distribution as follows: $s^{(i)}=((e^{(i)}-\mu)^T \Sigma^{-1} (e^{(i)}-\mu))^{1/2}$ [47].

The average anomaly score for a single day was calculated from the hourly scores, s^d. The Mahalanobis distance from data in N_{CV} was used to optimize an anomaly threshold, τ, for each participant over all s^d for that participant. A day was tagged as an anomaly if $s^d > \tau$ or normal if $s^d \le \tau$. τ was chosen to maximize the ratio between the TPR divided by the FPR, or TPR/FPR, defining a *true positive* as an anomaly detected within NR30 and a *false positive* as an anomaly detected on a DRH. Optimizing this ratio maximized the number of anomalies detected during the NR30 when minimizing the number of anomalies detected during DRH. This τ was applied to the Mahalanobis distances from the held-out test sample, N_T, and the final results using the best τ for each participant's N_T were recorded.

Evaluation Metrics

We used the TPR/FPR ratio as an evaluation metric to rank model performance. By maximizing this ratio, we subsequently maximized the sensitivity and specificity of our models. Sensitivity and specificity are metrics commonly used within medicine to assess the strength of a diagnostic test [48]. The sensitivity is equivalent to the TPR and the specificity is equivalent to the true negative rate (or 1–FPR). Thus, by maximizing the TPR/FPR, we found a model that maximized both sensitivity and specificity.

Anomalies are rare events; thus, it is unlikely that every day within NR30 would contain an anomaly. Clinically, we assumed that an anomaly detection system for early warning signs of relapse would be relevant as long as the anomalies were rare (low sensitivity and high specificity), but increased (TPR/FPR>1) within NR30. This increased signal could then be used to find passive sensing features that distinguished anomalies within NR30 from anomalies identified within DRH.

Baseline Model and Evaluation

We used a k-nearest neighbors local outlier factor (LOF) model as a baseline comparison against our neural network models [49]. The LOF model estimated the local density around each data point using a k-nearest neighbor algorithm and then compared the local density of a given data point with the local density of its neighbors. If the point was in a substantially less dense area, it had a higher calculated LOF. We initially fit an LOF model for each relapse participant utilizing H_R with the number of neighbors equal to 10, and we incremented the number of neighbors by 1 until the mean and SD of the LOF scores under H_{CV} converged. We could then utilize the approach described above to calculate anomalies by considering the distribution of LOF scores obtained under H_{CV} and learning an appropriate anomaly threshold for N_{CV}. The LOF model was trained and tested using scikit-learn [42].

Neural network models were created using TensorFlow and Keras libraries [50,51]. Models were trained until the validation loss from H_{CV} converged. We used cross-validation for all neural network models to determine the optimal hidden layer size (between 10 and 50 units), the percentage of DRH from relapse participant data to include within H_R and H_{CV} (between 0% and 80%), and the τ that maximized the TPR/FPR ratio on N_{CV} (between 0 and 20). For the LOF model, we also optimized the number of neighbors utilized for the local density within each relapse patient.

We applied 2 forms of regularization to train the neural networks. For both the GRU Seq2Seq and the FNN AD models, we used early stopping to terminate model training when the reconstruction error from H_{CV} increased. In addition, we applied dropout (rate=0.2) and recurrent dropout (rate=0.2) to the GRU Seq2Seq model. Dropout masks, or *drops*, inputs randomly within the network, whereas recurrent dropout adds this mask between the recurrent layers at each timestep [52]. This exposed the trained network to different permutations of the training data to prevent overfitting. Batch normalization was briefly used during model creation, but we found that batch normalization did not improve anomaly detection performance and was not used to train the final iteration of the models.

Results

Data Overview

We collected a total of 20,137 days of mobile sensing data from 60 patients with SSDs. Relapse events were recorded for 18 of 60 participants (30%) during the 1-year study, totaling 726 days of data collected within any NR30 data (0.037% of the total days of data collected). Table 2 provides a summary of the data collected from the relapse and nonrelapse groups.

Table 2. Summarized data characteristics for relapse and nonrelapse participants (continuous characteristics listed by median [IQR]).

Characteristics	Relapse	Nonrelapse
Patients, n	18	42
Age at beginning of study (years), median (IQR)	33 (23-47)	40 (26-50)
Female, n (%)	8 (44)	17 (40)
Number of days of data collected per participant, median (IQR)	335 (285-346)	295 (176-361)
Missing hours of data (type 2), median (IQR)		
Number of hours	2309 (1333-2551)	1785 (660-2871)
Percentage of total hours	25.73 (14.77-28.73)	27.17 (7.72-52.50)
Diagnosis, n (%)		
Schizophrenia	9 (50)	17 (40)
Schizoaffective disorder	7 (39)	18 (43)
Psychosis NOS[a]	2 (11)	7 (17)
Assessment at baseline, median (IQR)		
BPRS[b] (7-item) total	29 (23-33)	24 (21-29)
Lifetime hospitalizations, n (%)		
1-5	13 (72)	30 (71)
6-10	1 (6)	8 (19)
11-15	1 (6)	3 (7)
16-20	1 (6)	0 (0)
>20	1 (6)	1 (2)
Missing or declined	1 (6)	0 (0)
Distribution of relapse events, n (%)		
1 relapse event	14 (78)	N/A[c]
2 relapse events	1 (5)	N/A
3 relapse events	3 (17)	N/A

[a]NOS: not otherwise specified.

[b]BPRS: Brief Psychiatric Rating Scale.

[c]N/A: not applicable.

Anomalies Increased Near Relapse

The highest performing cross-validation results for each model, with hyperparameters, are shown in Table 3. All results are listed using median (IQR) sensitivity and specificity. Across all model architectures, the FNN AD model using 80% of the data from DRH with 40 hidden units had the highest rank across participants (9.28), achieving a median sensitivity of 0.25 (IQR 0.15-1.00) and specificity of 0.88 (IQR 0.14-0.96). LOF models did not show predictive power (sensitivity 1.0 and specificity 0.0) and were not included in our results. Figure 2 shows the resulting sensitivity and specificity achieved from models trained on different percentages of DRH. Adding a larger percentage of DRH to model training initially increased the sensitivity and decreased the model specificity, but then decreased the sensitivity and increased the specificity as more data were added. Figure 2 shows that the anomaly rate increased before the NR30 period but then remained fairly constant among participants.

Table 3. Cross-validation results per model type within relapse participants listed by median (IQR).

Model	Rank	Days of relative health in train, %	Hidden units	Sensitivity, median (IQR)	Specificity, median (IQR)
FNN AD[a]	9.28	80	40	0.25 (0.15-1.00)	0.88 (0.14-0.96)
GRU Seq2Seq[b]	12.72	80	50	0.29 (0.08-0.83)	0.86 (0.24-0.90)

[a]FNN AD: fully connected neural network autoencoder.

[b]GRU Seq2Seq: gated recurrent unit sequence-to-sequence.

XSL•FO

RenderX

Figure 2. Overall model results, the anomaly rate of the best performing model across the near relapse (NR30) period and in (a-c) split by the DRH used in model training. In (a-c), the bar heights describe the median value of the metric listed on the y-axis across study participants and the error bars show lower and upper quartile values (25% and 75% percentiles of the data). In (a) and (b), local outlier factor (LOF) models are not shown as they did not hold predictive power. (a) Sensitivity, or true positive rate, of the models and (b) specificity, or true negative rate. (c) Median number of DRH used to train each model from each study participant. (d) Average (blue line) and 95% CI (gray shading) anomaly rate across relapse participants beginning 35 days before relapse using the best performing model (fully connected neural network autoencoder, 80% of DRH in train, 40 hidden units). DRH: days of relative health.
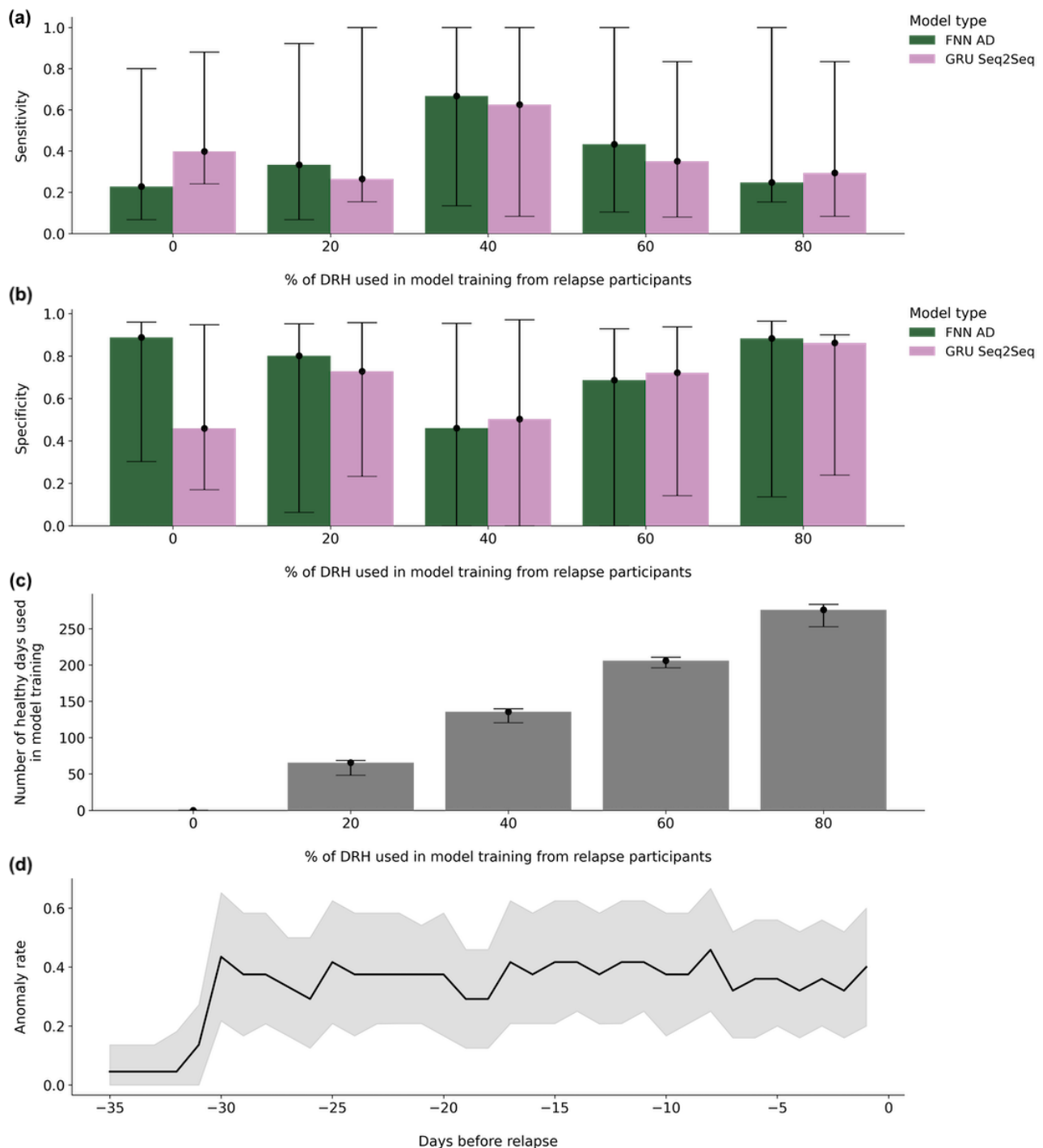
Anomaly Detection System Identified Specific Near Relapse Behaviors

Previous research has shown that individuals often report symptom exacerbation, which could be used to predict the onset of relapse [4]. Identifying participant-specific behaviors that are consistent during relapse would give clinicians a potential signature to identify when a patient needs clinical support. A total of 4 participants within our study relapsed multiple times. We performed a post hoc analysis using our best-performing algorithm across participants (FNN AD, 80% of DRH in train, hidden unit size=40) to compare features on NR30 anomalous days with DRH within multirelapse participants. We used Cohen d to calculate the effect of continuous features on discriminating an NR30 anomaly to any DRH and the OR for calculating whether type 2 missing data appeared more frequently in NR30

anomalies [53]. Figure 3 shows the distribution of the 5 features with the largest effect per participant on differentiating detected

anomalies within NR30 from DRH.

Figure 3. The hourly features that had the greatest effect on differentiating identified anomalous days near relapse (NR30) from all DRH within the 4 multirelapse participants. We used the Cohen d to identify the 5 features that were the most differentiated. Each subfigure, (a-d), displays boxplots comparing the distribution of these features on anomalous days within each NR30 period compared with all DRH. The center line in the boxplot is the median value, the box limits are the IQR, and the whiskers are 1.5 x the IQR. Points outside of the whiskers are greater than or less than 1.5 x the IQR. A lower IQR signifies that the median result is more generalizable. For example, in (a), we identified anomalies within 2 NR30 periods, described in the figure as Near relapse 1 and Near relapse 2. The 2 left boxes on each plot show the distribution of the feature for anomalies detected within each of these 2 NR30 periods and the right box shows the distribution of this feature on all DRH outside of the 2 NR30 periods. NR30: 30-day near relapse period. DRH: days of relative health.
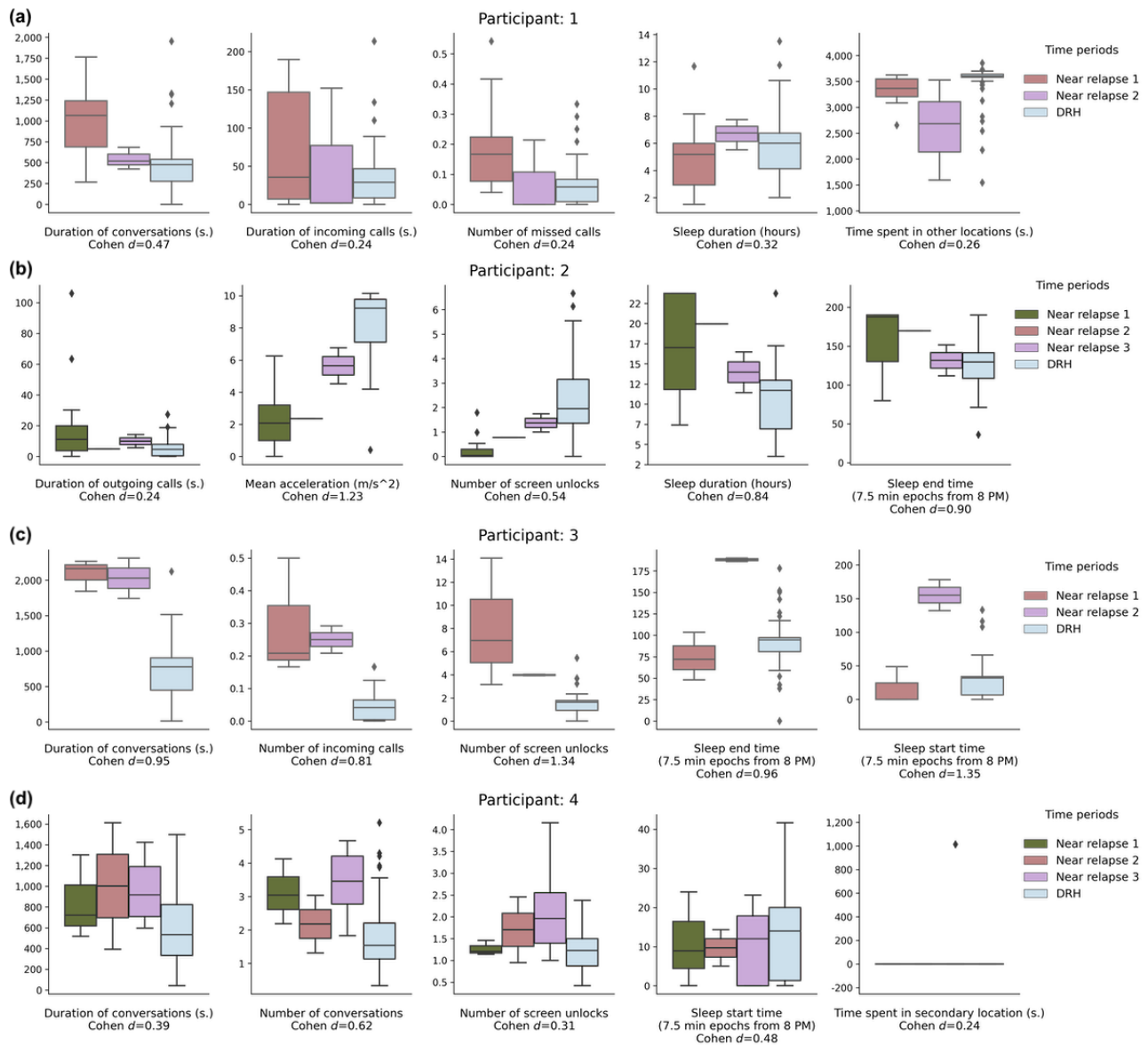
Notes surrounding each relapse, extracted from the participant's EMR or obtained during clinical visits, were compiled by a team of trained clinical assessors. We used these notes as a qualitative validation to understand whether the identified features from our analysis were presented to clinicians. We now briefly describe the results of each comparison for features that were identified to have a large effect (Cohen $d > 0.8$), medium effect ($0.5 <$ Cohen $d \leq 0.8$), or the feature with the largest effect if no features with a large or medium effect were identified [54]. The ORs indicated that type 2 missing data did not discriminate anomalies within NR30 from DRH for multirelapse participants

(OR<1 for all multirelapse participants). The features are described in detail in the Methods section.

Multirelapse Participant 1

We did not identify a feature with a large or medium effect for this participant. The conversation duration had the largest effect (Cohen $d = 0.47$), which increased before relapse, as shown in Figure 3. Clinical notes from the first relapse event indicate that the participant was hospitalized because she was tired of hearing voices, which suggested that her neighbors were constantly talking about her. Notes from the second relapse did not describe any participant behavior.

Multirelapse Participant 2

The participant's mean acceleration (Cohen *d*=1.23), sleep end time (Cohen *d*=0.90), and sleep duration (Cohen *d*=0.84) had a large effect. Figure 3 shows that the mean acceleration decreased in all 3 NR30 periods for this participant, whereas the sleep duration and end time increased. Clinical notes from the first relapse identified that the participant had been feeling ill, specifically that his "brain was shaking." On the second relapse, the participant stated that he felt like he "was going to die," and was feeling depressed. The clinician wrote that the patient "has had difficulty sleeping." Notes regarding the third relapse indicate that the participant had been disorganized, physically aggressive toward his mother, and was barely sleeping.

Multirelapse Participant 3

The participant's sleep start time (Cohen *d*=1.35), number of smartphone screen unlocks (Cohen *d*=1.34), sleep end time (Cohen *d*=0.96), duration of conversations (Cohen *d*=0.95), and number of incoming calls (Cohen *d*=0.81) had a large effect. Figure 3 shows abnormal behavior in the sleep start and end times for all relapse periods, but is inconsistent in the direction of how the behavior differs from the median value across each relapse. The number of screen unlocks, incoming calls, and duration of conversations increased in both relapse periods. Notes regarding the first relapse did not identify any specific behavioral changes. Clinical notes from the second relapse identified that the participant had been spending his days "making music and beats" and was sleeping less at night, but had increased sleep during the day. The notes also identified the participant as having auditory hallucinations.

Multirelapse Participant 4

One feature, the number of conversations, had a medium effect (Cohen *d*=0.62) for this participant. Figure 3 shows that the number of conversations increased during all 3 relapse periods. Notes from the first relapse did not describe any specific behavioral differences in the participant. Clinical notes from the second relapse indicated that the participant presented herself to outpatient psychiatry with "signs of catatonia" and that the participant had mostly stopped speaking, although she had occasional spontaneous speech. We were not able to obtain notes regarding the third relapse event.

Anomalies Contained Fewer Hours of Type 2 Missing Data

We found that type 2 missing data did not have an effect on distinguishing anomalies within NR30 for the 4 multirelapse participants. We wanted to examine this question more broadly to determine how missing data influenced all detected anomalies. We conducted a one-sided Mann-Whitney *U* test to test the following hypothesis: predicted anomalies contain a smaller number of type 2 hours filled compared with all other days. Individual participant factors were controlled for using participant-specific Mahalanobis distance thresholds for anomaly designation. Anomalies had a median of 0 (IQR 0-6) hours of data filled using the type 2 missing data procedure, and all other days had a median of 2 (IQR 0-16) type 2 hours of data filled. The one-sided test was significant (*U*=514,546; *P*<.001), indicating that anomalous days were significantly less likely to contain type 2 missing data.

Variations in Relapse Participant Data Affected the Quality of Anomaly Detection

We analyzed the participant-level anomaly detection results to determine how variations in data quality affect the generalizability of our model. Table 4 summarizes the results of using linear regression to assess the significance between the sensitivity and the specificity of the highest performing model (FNN AD, 80% of DRH in train, hidden unit size=40) and the data quality parameters. All data quality parameters were significant (*P*<.001). Increasing the number of days of raw data and the percentage of days within NR30 increased the sensitivity of the model (ß=.60, 95% CI 0.48 to 0.72; ß=.73, 95% CI 0.49 to 0.97) but decreased the specificity of the model (ß=−.69, 95% CI −0.81 to −0.57; ß=−.71, 95% CI −0.95 to −0.47). Increasing the number of days per NR30 period and the number of relapse events decreased the sensitivity of the model (ß=−.43, 95% CI −0.52 to −0.34; ß=−.82, 95% CI −1.02 to −0.62) but increased the specificity of the model (ß=.33, 95% CI 0.23 to 0.43; ß=.87, 95% CI 0.67 to 1.07).

Table 4. Linear regression results between sensitivity and specificity and different data parameters.

Parameters	Sensitivity		Specificity	
	Coefficient β	*P* value	Coefficient β	*P* value
Days of raw data	.60 (95% CI 0.48 to 0.72)	<.001	−.69 (95% CI −0.81 to −0.57)	<.001
Days per near relapse period	−.43 (95% CI −0.52 to −0.34)	<.001	.33 (95% CI 0.23 to 0.43)	<.001
Percentage of days near relapse	.73 (95% CI 0.49 to 0.97)	<.001	−.71 (95% CI −0.95 to −0.47)	<.001
Relapse events	−.82 (95% CI −1.02 to −0.62)	<.001	.87 (95% CI 0.67 to 1.07)	<.001
Intercept	.00 (95% CI −0.04 to 0.04)	>.99	.00 (95% CI −0.04 to 0.04)	>.99

Discussion

Principal Findings

In this study, we created the first model, exclusively using passive sensing data from a smartphone, to predict behavioral anomalies that could indicate early warning signs of psychotic relapse. Developing an anomaly detection system from exclusively passive sensing data requires minimal effort for data collection from the participant and could lead to more objective and unobtrusive ways of monitoring symptoms of SSDs. Our anomaly detection system achieved a median

sensitivity of 0.25 (IQR 0.15-1.00) and specificity of 0.88 (IQR 0.14-0.96; a 108% increase in anomalies near relapse), indicating that anomalies increased before relapse but were restricted to specific days within the defined NR30 period. Once we identified anomalous days within NR30, we demonstrated that our methodology can be used to identify participant-specific behavioral signatures that occur across multiple NR30. In the future, anomaly detection models could be used to identify days that contain these signatures and supervised learning approaches could then be deployed to detect these signals as early warning signs of relapse. Identifying patient-specific behaviors that change exclusively before relapse could provide clinicians an indicator to measure when patients are declining in health and create time for early intervention.

Anomalies increased within NR30, but with low sensitivity. We believe this low sensitivity could be owing to our choice of an NR30 period or our decision to use the TPR/FPR ratio as a validation metric for this work. We chose an NR30 period because past research has shown that early warning signs of relapse might begin to develop up to 1 month before the actual relapse event [19,31,38]. Our low sensitivity indicated that only specific days within this 30-day period were considered anomalies, and training algorithms in the future to target these specific days could increase sensitivity. Another approach to increase sensitivity would be to shorten the number of days included in the near relapse period. For example, previous work using a combination of smartphone social behavior and self-reported EMAs to detect anomalies before relapse identified a 14-day near relapse period [30]. We observed an increased anomaly rate 30 days before relapse (Figure 2), which remained fairly constant; therefore, we did not further investigate shortening the near relapse period. It is important to note that the algorithm we used may result in a constant anomaly rate during a near relapse period of any prespecified length as these algorithms are trained to look specifically for behavioral differences within these periods.

In addition, we used the TPR/FPR ratio for model selection rather than directly optimizing for sensitivity or specificity. Most machine learning algorithms use the area under the receiver operating curve to assess the predictive power of a model, which we did not feel was appropriate for our anomaly detection algorithm. Anomaly detection, by definition, searches for extremely rare events. To introduce this process into a clinical workflow, we would need to strike a balance between highlighting potential early warning signs of relapse without overburdening the healthcare system with a high anomaly rate. We felt this could be achieved using our modeling approach as we showed an increase in anomalies before relapse without sacrificing the specificity of our results.

To increase model sensitivity in this context, more clarity is needed on what should constitute behaviors that can be used to identify early warning signs of psychotic relapse. A process can then be created where we first use anomaly detection to identify candidate relapse signatures and then train supervised learning algorithms to identify these signatures. This would, in turn, limit the feature space to the behaviors per individual that were differentiated before relapse. In addition, identifying these specific signatures as a starting point for a positive signal would

allow us to clarify whether false positives were merely noise or hold clinical significance. In this study, a relapse event was indicated for most of our participants by either a psychiatric hospitalization or a significant increase in symptoms as reported by clinician-administered BPRS. It is possible that symptoms were exacerbated on days outside of NR30, and our system detected these days as anomalies. This symptom exacerbation was not given clinical oversight; thus, we had no way to validate whether these anomalies should be considered true positives.

Days that contained more hours of type 2 missing data, in which no passive sensing data for the entire hour existed, were significantly less likely to be tagged as anomalies. Our approach to using mean filling for type 2 data was based on 2 assumptions: (1) that we would like to prioritize behavioral features collected from passive sensors for anomaly detection and (2) that it is possible that a large quantity of missing data might be a sign of asocial behavior and we should also account for missing data with an additional feature that tracks the amount of data missing over an hour. Previous work on anomaly detection with missing data has analyzed the effects of various data filling methods on anomaly detection results. With the assumption that imputed data points should not be detected as anomalies, the work found that this assumption can hold true if the imputed values are located in high-density regions of the feature distribution [55]. In this work, anomalies were significantly less likely to contain missing values ($P<.001$), indicating that the Mahalanobis distance per individual was less for imputed hours and anomalies were more likely to include non–type 2 data points. With this first assumption in mind, it is possible that the potential effect of missing data on relapse was ignored. For example, the missing data feature did not distinguish anomalies within NR30 in the 4 multirelapse participants. More research needs to be conducted on how different missing data imputation procedures can affect mental health symptom prediction algorithms.

We observed that increasing the amount of relapse participant data used for model training did not always increase the resulting sensitivity and specificity. Figure 2 shows that our model performance increased in sensitivity and decreased in specificity when we increased the percentage of DRH from relapse patients used in model training from 0% to 40%. We then observed a reverse trend (decreased sensitivity, increased specificity) when we increased the amount of DRH from 40% to 80%. This demonstrates that as models learned participant-level behaviors, there was a threshold for the amount of data required for model training (approximately 135 days from data from Figure 2) before a model could begin to distinguish anomalous behaviors within NR30. We also found that our anomaly detection system is sensitive to the quality of the relapse participant data. Table 4 demonstrates that increasing the total percentage of NR30 days increased the sensitivity (ß=.73), but not if this increased the average number of days within NR30 (ß=−.43), and increased the total number of relapse events (ß=−.82). Subsequently, having a higher number of relapse events increased (ß=.87) the specificity of the model, but not if this increased the percentage of days within NR30 (ß=−.71), and increased the number of days of raw data (ß=−.69). Taken

together, our results depended on identifying homogeneous behavioral signals that occurred exclusively in NR30.

Given the importance of finding homogeneity in the signal, we examined whether we could uncover consistent signals in participants that might indicate SSD symptom exacerbation. It can be difficult to introduce neural network models within clinical practice owing to their *black box* nature, even though they can achieve higher performance than classical machine learning models [28]. In our work, interpretability is critical because clinicians need to develop an understanding and trust of an algorithm's decision-making process. We utilized a post hoc notion of interpretability to identify participant-specific features that differed during NR30 anomalies [29]. We chose the effect size, a metric traditionally used to measure the strength of a treatment in an RCT, to identify the most differentiated features within NR30 [54]. We identified features with a medium to large effect (Cohen $d>0.5$) in 3 of 4 multirelapse patients. The features we identified encompass different aspects of social behavior, sleep, and physical activity.

We searched the literature to interpret the behavior changes that we identified within NR30. Our previous work identified that smartphone social behavior decreased across participants before relapse [19]. We examined smartphone social behavior at the individual level. Multirelapse participant 3 increased their smartphone social behavior before relapse, but, from contextual notes, we discovered that this participant experienced auditory hallucinations, potentially explaining the increased conversation duration detected by the smartphone as well as other increased smartphone social behaviors found. Multirelapse participant 4's number of conversations increased with a medium effect (Cohen d=0.62), contrasting the physician's notes, which stated that the participant was barely speaking and potentially catatonic. Figure 3 shows that the elevated conversation signal, whether from the participant or the environment, was unique to anomalies within each of the 3 NR30 periods.

In addition, we detected that changes in sleeping behavior had a large effect on 2 participants and decreased acceleration had a large effect on 1 participant. Previous research has shown that patients with SSDs are at a significantly higher risk of developing a sleep disorder or worsened sleep near relapse [56,57]. Multirelapse participant 2's detected sleep duration increased before relapse. In addition, the participant's acceleration decreased. Social withdrawal and physical inactivity are common symptoms of SSDs. These symptoms interfere with functioning, potentially leading to relapse, and relapse can produce aggression [58,59]. The symptoms identified were consistent with the clinician's explanation of the second relapse event for this participant, which described changes in sleep and aggressive behavior. Thus, the features that were most differentiated are consistent with past research identifying early warning signs of relapse.

It is important to note that although we found differentiated features for each participant that were consistent with the notes surrounding relapse, the changes detected by the passive sensors were not always consistent with the changes described in the clinical notes. For example, we found that participant 2's sleep increased before relapse, with a large effect, whereas the

clinician's notes stated that sleep decreased. Similarly, for participant 4, the number of conversations increased before relapse with a medium effect, whereas the clinician's notes stated that the participant exhibited signs of catatonia. The smartphone algorithms we used to detect conversation relied on ambient sound to detect human voice and conversational exchanges, but do not necessarily detect the voice of the participant [34]. In addition, the sleep algorithm used detected sleep based on a combination of phone usage, ambient light, stationary behavior, and environmental silence, all features that might occur when someone is still but not necessarily sleeping [37]. When interpreting the result of a *black box* algorithm clinically, we need to interpret the algorithm's results in the context of the technical capabilities of the passive sensing system used before judging the outputs of the system literally. Thus, although smartphones can find meaningful relapse signatures, the interpretations of these signatures should be corroborated with the patient and other qualitative information to better understand the underlying behaviors that preceded relapse.

Designing a Relapse Prediction System

We hope that this work moves researchers one step closer to creating a clinical intervention system to predict early warning signs of relapse that can be deployed within the clinical workflow. We reviewed digital mental health and mobile health (mHealth) literature to understand how such a system could be deployed. The MONARCA system was created to help individuals with bipolar disorder track disease symptom trajectories using a combination of both passive sensor data and self-assessment [10]. A field trial of the MONARCA system demonstrated the difficulties in obtaining both patient and clinician buy-in when forecasting mental health symptoms, as patients were not convinced of the accuracy of the passive sensing data and clinicians were unsure of steps to take if the system forecasted symptom exacerbation [60]. Although the possibility of a clinician having patient data *at their fingertips* seems appealing, it is also a liability for clinicians if they have 24/7 monitoring capabilities and choose not to act when a patient is potentially in danger [61]. One possible solution to this issue is to introduce a *clinical technology specialist* into a patient's care team whose responsibility is to successfully introduce and maintain technology-based services within the clinic [62]. It is evident that there is a gap between technology intervention creation and implementation.

Overall, acceptability will continue to play a large role in implementing mHealth interventions. The PD_Manager mHealth platform, a platform created to track symptoms of Parkinson's disease using passive sensing, provides an example of an mHealth tool where researchers specifically tested the acceptability of the platform to patients and clinicians before testing the effectiveness of the system [63]. To help increase the acceptability of mHealth tools, digital mental health and human computer interaction researchers are focused on solving mHealth implementation hurdles using a user-centered design framework, where technology is created and refined in an iterative process that places the proposed interventions directly in the hands of relevant stakeholders [64]. Creating the technology behind a relapse prediction system is a small piece

of the puzzle compared with the larger implementation challenges that will be faced when deploying it.

Figure 4 presents an example framework for an SSD behavioral monitoring and intervention system utilizing anomaly detection. Many questions remain regarding implementing this system, including the level of patient interaction with the system, how and when detected anomalies are presented to the clinician or other relevant parties, and a defined procedure that an individual should take to intervene in care when anomalies are detected. The implications of this work can only be truly justified when a system for detecting early warning signs of psychotic relapse has been deployed within the clinical workflow.

Figure 4. Example of an anomaly visualization and clinical intervention system. The dashed black lines in (a) each represent an hourly feature trajectory from the anomaly detection system, as identified on the y-axis, during a 30-day near relapse period (NR30). The gray line on each plot is the Mahalanobis distance, which can be interpreted as an anomaly score that increases as we are more likely to detect an anomaly. The 2 vertical thick black lines on each plot are detected anomalies. (b) Example of how this information could be utilized by a clinician or other individuals designated by the patient to intervene during symptom exacerbation. The system would be tuned to send alerts only when a patient is in crisis and not overburden the clinician and the healthcare system.
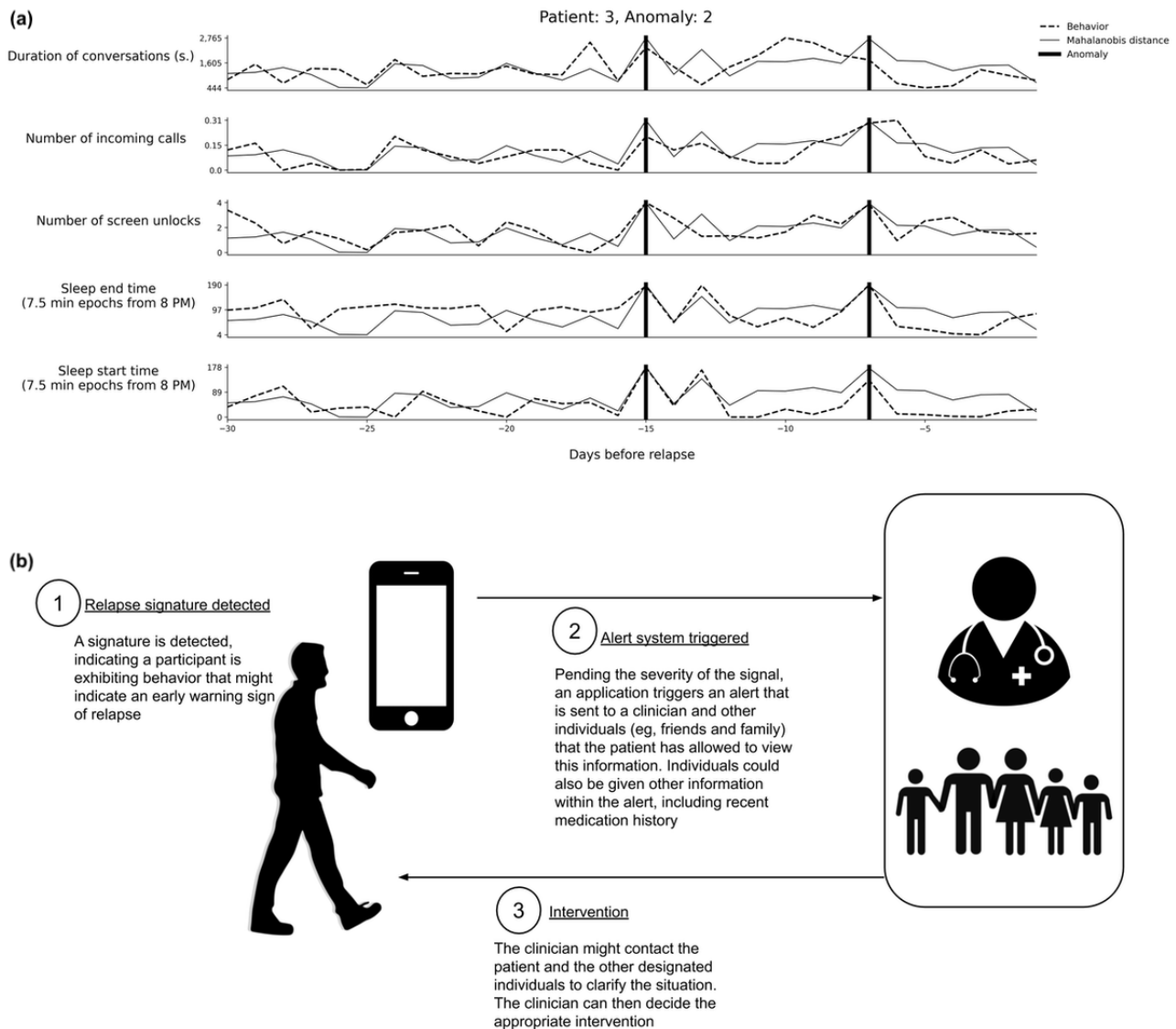
Limitations

The primary limitation of this study was the limited sample size. Our study consisted of 60 participants with SSDs, including 18 participants who relapsed. Most participants did not relapse at multiple points throughout the study and we could not assess whether the features underlying anomalies were consistent with relapse for those participants. To the best of our knowledge, this is still the largest study utilizing anomaly detection to predict early warning signs of relapse exclusively from smartphone behavior.

Comparison With Previous Work

Previous work using CrossCheck focused on identifying symptom exacerbation by predicting participant responses to the BPRS, a common tool for measuring symptoms of SSDs [6,18]. CrossCheck data have also been used to identify significant associations in smartphone social behavior between the 30-day period before relapse and all other days [19]. We showed that we can predict behavioral anomalies preceding relapse events recorded with and without BPRS. In addition, to the best of our knowledge, only one previous study has utilized

anomaly detection to identify early warning of psychotic relapse using a combination of smartphone data and self-reported behaviors [30]. This study used a statistical approach to identify near relapse anomalies within 3 participants. We expand this work by showing that anomaly detection can be used to predict an increase in anomalies before relapse on a larger data set of 18 participants. In addition, we derived features from passive sensing data exclusively, creating an anomaly detection method that does not rely on patient self-report.

Future Work

Future work should develop approaches to identify early warning signs of relapse across larger and more diverse patient populations with SSDs. These approaches could be tested across different smartphone passive sensing apps such that they become platform independent. In addition, researchers should train models to detect patient-specific relapse signatures, which could increase model sensitivity. Finally, a tool should be codesigned with clinicians and patients for remote monitoring of SSD symptoms.

Conclusions

In summary, we created an anomaly detection model using encoder-decoder neural networks to predict early warning signs of psychotic relapse. Our model predicted an increase in anomalies within the 30-day period preceding relapse. We developed a methodology to uncover behaviors that change before relapse, which could be used to identify patient-specific relapse signatures. Finally, we discussed the implications of this work and showed an example visualization of a remote monitoring system for SSDs. We hope that this work advances the field of digital mental health to create effective remote monitoring systems for serious mental illness.

Conflicts of Interest

DBZ has an intervention content licensing agreement with Pear Therapeutics and has a financial interest in FOCUS technology. DBZ has consulted for eQuility, Trusst Health, and Otsuka Pharmaceutical Ltd. JK has received honoraria for lectures and/or consulting from Alkermes, Dainippon, Sumitomo, Sunovion, Janssen, Intracellular Therapies, Lundbeck, Otsuka, Roche, Teva, Neurocrine, LB Pharma, and Merck. JK is a shareholder of Vanguard Research Group and LB Pharma. TC is a co-founder and equity holder of HealthRhythms, Inc. The other authors declare no competing interests.

References

1. Andreasen N. Symptoms, signs, and diagnosis of schizophrenia. Lancet 1995 Aug 19;346(8973):477-481. [doi: 10.1016/s0140-6736(95)91325-4] [Medline: 7637483]
2. Emsley R, Chiliza B, Asmal L, Harvey BH. The nature of relapse in schizophrenia. BMC Psychiatry 2013 Feb 8;13:50 [FREE Full text] [doi: 10.1186/1471-244X-13-50] [Medline: 23394123]
3. Cloutier M, Aigbogun MS, Guerin A, Nitulescu R, Ramanakumar AV, Kamat SA, et al. The economic burden of schizophrenia in the United States in 2013. J Clin Psychiatry 2016 Jun;77(6):764-771. [doi: 10.4088/JCP.15m10278] [Medline: 27135986]
4. Eisner E, Drake R, Lobban F, Bucci S, Emsley R, Barrowclough C. Comparing early signs and basic symptoms as methods for predicting psychotic relapse in clinical practice. Schizophr Res 2018 Feb;192:124-130 [FREE Full text] [doi: 10.1016/j.schres.2017.04.050] [Medline: 28499766]
5. Overall JE, Gorham DR. The brief psychiatric rating scale. Psychol Rep 1962 Jun 1;10(3):799-812. [doi: 10.2466/pr0.1962.10.3.799]
6. Kopelowicz A, Ventura J, Liberman RP, Mintz J. Consistency of brief psychiatric rating scale factor structure across a broad spectrum of schizophrenia patients. Psychopathology 2008;41(2):77-84. [doi: 10.1159/000111551] [Medline: 18033976]
7. Corrigan PW, Liberman RP, Engel JD. From noncompliance to collaboration in the treatment of schizophrenia. Hosp Community Psychiatry 1990 Nov;41(11):1203-1211. [doi: 10.1176/ps.41.11.1203] [Medline: 1979044]
8. Insel TR. Digital phenotyping: a global tool for psychiatry. World Psychiatry 2018 Oct;17(3):276-277 [FREE Full text] [doi: 10.1002/wps.20550] [Medline: 30192103]
9. Torous J, Onnela J, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. Transl Psychiatry 2017 Mar 7;7(3):e1053 [FREE Full text] [doi: 10.1038/tp.2017.25] [Medline: 28267146]

10. Frost M, Marcu G, Hansen R, Szaántó K, Bardram J. The MONARCA Self-Assessment System: Persuasive Personal Monitoring for Bipolar Patients. In: 5th International Conference on Pervasive Computing Technologies for Healthcare and Workshops. 2011 Presented at: PervasiveHealth'11; May 23-26, 2011; Dublin, Ireland. [doi: 10.4108/icst.pervasivehealth.2011.246050]

11. Osmani V. Smartphones in mental health: detecting depressive and manic episodes. IEEE Pervasive Comput 2015 Jul;14(3):10-13. [doi: 10.1109/MPRV.2015.54]

12. 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. [doi: 10.1145/2632048.2632054]

13. 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]

14. Gay K, Torous J, Joseph A, Pandya A, Duckworth K. Digital technology use among individuals with schizophrenia: results of an online survey. JMIR Ment Health 2016 May 4;3(2):e15 [FREE Full text] [doi: 10.2196/mental.5379] [Medline: 27146094]

15. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile phone ownership and endorsement of 'mhealth' among people with psychosis: a meta-analysis of cross-sectional studies. Schizophr Bull 2016 Mar;42(2):448-455 [FREE Full text] [doi: 10.1093/schbul/sbv132] [Medline: 26400871]

16. 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 Sep;40(3):266-275 [FREE Full text] [doi: 10.1037/prj0000243] [Medline: 28368138]

17. Wang R, Scherer EA, Tseng VW, Ben-Zeev D, Aung MS, Abdullah S, et al. CrossCheck: Toward Passive Sensing and 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, 2-16; Heidelberg, Germany. [doi: 10.1145/2971648.2971740]

18. Wang R, Scherer EA, Walsh M, Wang W, Aung MS, Ben-Zeev D, et al. Predicting Symptom Trajectories of Schizophrenia Using Mobile Sensing. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2017 Presented at: ACM'16; September 17, 2017; Heidelberg, DE. [doi: 10.1145/3130976]

19. Buck B, Scherer E, Brian R, Wang R, Wang W, Campbell AT, et al. Relationships between smartphone social behavior and relapse in schizophrenia: a preliminary report. Schizophr Res 2019 Jun;208:167-172 [FREE Full text] [doi: 10.1016/j.schres.2019.03.014] [Medline: 30940400]

20. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G. Learning from class-imbalanced data: review of methods and applications. Expert Syst Appl 2017 May;73:220-239. [doi: 10.1016/j.eswa.2016.12.035]

21. Chandola V, Banerjee A, Kumar V. Anomaly detection. ACM Comput Surv 2009 Jul 1;41(3):1-58. [doi: 10.1145/1541880.1541882]

22. Basseville M, Nikiforov I. Detection of Abrupt Changes. New York, USA: Prentice-Hall Inc; 1993.

23. Subrahmanian VS, Kumar S. Predicting human behavior: the next frontiers. Science 2017 Feb 3;355(6324):489. [doi: 10.1126/science.aam7032] [Medline: 28154052]

24. Wulff K, Dijk D, Middleton B, Foster RG, Joyce EM. Sleep and circadian rhythm disruption in schizophrenia. Br J Psychiatry 2012 Apr;200(4):308-316 [FREE Full text] [doi: 10.1192/bjp.bp.111.096321] [Medline: 22194182]

25. Malhotra P, Vig L, Shroff G, Agarwal P. Long Short Term Memory Networks for Anomaly Detection in Time Series. In: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 2015 Presented at: ESANN'15; April 22-24, 2015; Bruges, Belgium URL: https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2015-56.pdf

26. Malhotra P, Ramakrishnan A, Anand G, Vig L, Agarwal P, Shroff G. LSTM-based encoder-decoder for multi-sensor anomaly detection. ArXiv 2016:- epub ahead of print [FREE Full text]

27. Cowton J, Kyriazakis I, Plötz T, Bacardit J. A combined deep learning GRU-autoencoder for the early detection of respiratory disease in pigs using multiple environmental sensors. Sensors (Basel) 2018 Aug 2;18(8):2521 [FREE Full text] [doi: 10.3390/s18082521] [Medline: 30072607]

28. Holzinger A, Biemann C, Pattichis C, Kell D. What do we need to build explainable AI systems for the medical domain? arXiv 2017 Dec 28:- epub ahead of print [FREE Full text]

29. Lipton ZC. The mythos of model interpretability. ArXiv 2017 Mar 6:- epub ahead of print [FREE Full text] [doi: 10.1145/3233231]

30. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela J. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018 Jul;43(8):1660-1666 [FREE Full text] [doi: 10.1038/s41386-018-0030-z] [Medline: 29511333]

31.   Birnbaum ML, Ernala SK, Rizvi AF, Arenare E, Van Meter AR, de Choudhury M, et al. Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. NPJ Schizophr 2019 Oct 7;5(1):17 [FREE Full text] [doi: 10.1038/s41537-019-0085-9] [Medline: 31591400]

32.   Granholm E, Loh C, Swendsen J. Feasibility and validity of computerized ecological momentary assessment in schizophrenia. Schizophr Bull 2008 May;34(3):507-514 [FREE Full text] [doi: 10.1093/schbul/sbm113] [Medline: 17932087]

33.   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 Jun 9;19(3):345-359. [doi: 10.1007/s11036-013-0484-5]

34.   Rabbi M, Ali S, Choudhury T, Berke E. Passive and in-situ assessment of mental and physical well-being using mobile sensors. Proc ACM Int Conf Ubiquitous Comput 2011;2011:385-394 [FREE Full text] [doi: 10.1145/2030112.2030164] [Medline: 25285324]

35.   Wyatt D. Measuring and Modeling Networks of Human Social Behavior. University of Washington. 2010. URL: http://www.dannywyatt.com/pubs/Wyatt-Dissertation.pdf [accessed 2020-08-14]

36.   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. [doi: 10.1145/2750858.2805845]

37.   Chen Z, Lin M, Chen F, Lane N, Cardone G, Wang R, et al. Unobtrusive Sleep Monitoring Using Smartphones. In: Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare. 2013 Presented at: PervasiveHealth'13; May 5-8, 2013; Venice, Italy. [doi: 10.4108/icst.pervasivehealth.2013.252148]

38.   Birchwood M, Smith J, Macmillan F, Hogg B, Prasad R, Harvey C, et al. Predicting relapse in schizophrenia: the development and implementation of an early signs monitoring system using patients and families as observers, a preliminary investigation. Psychol Med 1989 Aug;19(3):649-656. [doi: 10.1017/s0033291700024247] [Medline: 2798634]

39.   Wilkinson GS, Robertson GJ. Wide range achievement test 4 (WRAT 4). Score 2007;29(1):4-5. [doi: 10.1037/e672132007-004]

40.   LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]

41.   Ester M, Kriegel H, Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: The Second International Conference on Knowledge Discovery and Data Mining. 1996 Presented at: KDD'96; August 2-4, 1996; Portland, Oregon.

42.   Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. ArXiv 2018 Jun 5:- epub ahead of print [FREE Full text] [doi: 10.1145/2786984.2786995]

43.   Smith MJ, Goodchild MF, Longley PA. Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools. New York, USA: The Winchelsea Press; 2018.

44.   Baldi P. Autoencoders, Unsupervised Learning, and Deep Architectures. In: Proceedings of Machine Learning Research. 2012 Presented at: PMLR'12; June 26-July 1, 2012; Edinburgh, Scotland, UK URL: http://proceedings.mlr.press/v27/baldi12a/baldi12a.pdf

45.   Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. ArXiv 2014 Dec 11:- epub ahead of print [FREE Full text]

46.   Schuster M, Paliwal K. Bidirectional recurrent neural networks. IEEE Trans Signal Process 1997;45(11):2673-2681. [doi: 10.1109/78.650093]

47.   Mahalanobis PC. On the generalised distance in statistics. Sankhya A 1936 Apr 15;80(1):1-7. [doi: 10.1007/s13171-019-00164-5]

48.   Altman DG, Bland JM. Diagnostic tests. 1: sensitivity and specificity. Br Med J 1994 Jun 11;308(6943):1552 [FREE Full text] [doi: 10.1136/bmj.308.6943.1552] [Medline: 8019315]

49.   Breunig M, Kriegel H, Ng R, Sander J. LOF: Identifying Density-Based Local Outliers. In: International Conference on Management of Data and Symposium on Principles of Database Systems. 2000 Presented at: ACM'00; May 16-18, 2000; Dalles, TX. [doi: 10.1145/342009.335388]

50.   Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: A System for Large-Scale Machine Learning. In: Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation. 2016 Presented at: OSDI'16; November 2–4, 2016; Savannah, GA URL: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

51.   Chollet F. Keras: the Python Deep Learning API. 2015. URL: https://keras.io [accessed 2020-01-31]

52.   Gal Y, Ghahramani Z. A theoretically grounded application of dropout in recurrent neural networks. ArXiv 2016 Oct 5:- epub ahead of print [FREE Full text]

53.   Sullivan GM, Feinn R. Using effect size-or why the p value is not enough. J Grad Med Educ 2012 Sep;4(3):279-282 [FREE Full text] [doi: 10.4300/JGME-D-12-00156.1] [Medline: 23997866]

54.   Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. Law Hum Behav 2005 Oct;29(5):615-620. [doi: 10.1007/s10979-005-6832-7] [Medline: 16254746]

55.   Dietterich TG, Zemicheal T. Anomaly Detection in the Presence of Missing Values. Carnegie Mellon University. 2018. URL: https://www.andrew.cmu.edu/user/lakoglu/odd/accepted_papers/ODD_v50_paper_13.pdf [accessed 2020-08-10]

56.   Kaskie RE, Graziano B, Ferrarelli F. Schizophrenia and sleep disorders: links, risks, and management challenges. Nat Sci
       Sleep 2017;9:227-239 [FREE Full text] [doi: 10.2147/NSS.S121076] [Medline: 29033618]
57.   Herz MI, Melville C. Relapse in schizophrenia. Am J Psychiatry 1980 Jul;137(7):801-805. [doi: 10.1176/ajp.137.7.801]
       [Medline: 6104444]
58.   Herz MI, Szymanski HV, Simon JC. Intermittent medication for stable schizophrenic outpatients: an alternative to maintenance
       medication. Am J Psychiatry 1982 Jul;139(7):918-922. [doi: 10.1176/ajp.139.7.918] [Medline: 6124133]
59.   Rastad C, Martin C, Asenlöf P. Barriers, benefits, and strategies for physical activity in patients with schizophrenia. Phys
       Ther 2014 Oct;94(10):1467-1479. [doi: 10.2522/ptj.20120443] [Medline: 24830718]
60.   Frost M, Doryab A, Faurholt-Jepsen M, Kessing L, Bardram J. 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 7-11, 2013; Zurich, Switzerland. [doi:
       10.1145/2493432.2493507]
61.   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]
62.   Ben-Zeev D, Drake R, Marsch L. Clinical technology specialists. Br Med J 2015 Feb 19;350:h945. [doi: 10.1136/bmj.h945]
       [Medline: 25697164]
63.   Antonini A, Gentile G, Giglio M, Marcante A, Gage H, Touray MM, PD_Manager Consortium. Acceptability to patients,
       carers and clinicians of an mHealth platform for the management of Parkinson's disease (PD_Manager): study protocol for
       a pilot randomised controlled trial. Trials 2018 Sep 14;19(1):492 [FREE Full text] [doi: 10.1186/s13063-018-2767-4]
       [Medline: 30217235]
64.   Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design
       and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153 [FREE Full text]
       [doi: 10.2196/jmir.7725] [Medline: 28490417]

Abbreviations

BPRS: Brief Psychiatric Rating Scale
DBSCAN: density-based spatial clustering of applications with noise
DRH: days of relative health
EMA: ecological momentary assessment
EMR: electronic medical record
FNN AD: fully connected neural network autoencoder
FPR: false-positive rate
GRU: gated recurrent unit
GRU Seq2Seq: gated recurrent unit sequence-to-sequence
LOF: local outlier factor
LSTM: long short-term memory
NR30: 30-day near relapse period
RCT: randomized controlled trial
RNN: recurrent neural network
SSD: schizophrenia-spectrum disorder
TPR: true-positive rate

XSL•FO

RenderX

Original Paper

Monitoring Occupational Sitting, Standing, and Stepping in Office Employees With the W@W-App and the MetaWearC Sensor: Validation Study

Judit Bort-Roig[1], PhD; Emilia Chirveches-Pérez[2], PhD; Francesc Garcia-Cuyàs[3], PhD; Kieran P Dowd[4], PhD; Anna Puig-Ribera[1], PhD

[1]Sport and Physical Activity Research Group, Centre for Health and Social Care Research, University of Vic-Central University of Catalonia, Vic, Spain

[2]Research Group on Methodology, Methods, Models and Outcomes of Health and Social Sciences, Centre for Health and Social Care Research, University of Vic-Central University of Catalonia, Vic, Spain

[3]Digital Care Research Group, Centre for Health and Social Care Research, University of Vic-Central University of Catalonia, Vic, Spain

[4]Department of Sport and Health Sciences, Athlone Institute of Technology, Athlone, Ireland

Corresponding Author:
Anna Puig-Ribera, PhD
Sport and Physical Activity Research Group
Centre for Health and Social Care Research
University of Vic-Central University of Catalonia
Miquel Martí i Pol, 1
Vic, 08500
Spain
Phone: 34 938863342
Email: annam.puig@uvic.cat

Abstract

Background: Replacing occupational sitting time with active tasks has several proposed health benefits for office employees. Mobile phones and motion sensors can provide objective information in real time on occupational sitting behavior. However, the validity and feasibility of using mobile health (mHealth) devices to quantify and modify occupational sedentary time is unclear.

Objective: The aim of this study is to validate the new Walk@Work-Application (W@W-App)—including an external motion sensor (MetaWearC) attached to the thigh—for measuring occupational sitting, standing, and stepping in free-living conditions against the activPAL3M, the current gold-standard, device-based measure for postural behaviors.

Methods: In total, 20 office workers (16 [80%] females; mean age 39.5, SD 8.1 years) downloaded the W@W-App to their mobile phones, wore a MetaWearC sensor attached to their thigh using a tailored band, and wore the activPAL3M for 3-8 consecutive working hours. Differences between both measures were examined using paired-samples t tests and Wilcoxon signed-rank tests. Agreement between measures was examined using concordance correlation coefficients (CCCs), 95% CIs, Bland-Altman plots (mean bias, 95% limits of agreement [LoA]), and equivalence testing techniques.

Results: The median recording time for the W@W-App+MetaWearC and the activPAL3M was 237.5 (SD 132.8) minutes and 240.0 (SD 127.5) minutes, respectively ($P<.001$). No significant differences between sitting ($P=.53$), standing ($P=.12$), and stepping times ($P=.61$) were identified. The CCC identified substantial agreement between both measures for sitting (CCC=0.98, 95% CI 0.96-0.99), moderate agreement for standing (CCC=0.93, 95% CI 0.81-0.97), and poor agreement for stepping (CCC=0.74, 95% CI 0.47-0.88). Bland-Altman plots indicated that sitting time (mean bias –1.66 minutes, 95% LoA –30.37 to 20.05) and standing time (mean bias –4.85 minutes, 95% LoA –31.31 to 21.62) were underreported. For stepping time, a positive mean bias of 1.15 minutes (95% LoA –15.11 to 17.41) was identified. Equivalence testing demonstrated that the estimates obtained from the W@W-App+MetaWearC and the activPAL3M were considered equivalent for all variables excluding stepping time.

Conclusions: The W@W-App+MetaWearC is a low-cost tool with acceptable levels of accuracy that can objectively quantify occupational sitting, standing, stationary, and upright times in real time. Due to the availability of real-time feedback for users, this tool can positively influence occupational sitting behaviors in future interventions.

Trial Registration: ClinicalTrials.gov NCT04092738; https://clinicaltrials.gov/ct2/show/NCT04092738

XSL·FO
RenderX

Introduction

Replacing sedentary time (ie, sitting, lying, or reclining postures that involve an energy expenditure of ≤1.5 metabolic equivalent units during waking hours) [1] with physical activity (PA) or movement of any kind has proposed health benefits for adults [2]. Positive associations have been reported with cardiometabolic biomarkers, mortality risk reduction, and body composition [3]. Many adults accumulate large amounts of daily sitting time at work, with white-collar workers being the most likely to engage in extensive occupational daily sitting [4]. Given that leveraging the time-inverse relationship between sedentary behaviors (SB) and PA could achieve important public health benefits [5], interventional efforts should target this high-risk subgroup [6] in the setting where daily sitting mostly occurs [5].

Self-monitoring is a key element to increase individuals' awareness of and empowerment toward behavior change [7]. For PA and SB, self-reported questionnaires have traditionally been the most commonly employed tool in largescale population studies due to their low cost, simplicity, and feasibility [8-11]. However, technological advances over the last 2 years have enabled the use of device-based measures, such as accelerometers, for self-monitoring PA and SB [8].

Evidence has identified mobile phones as a potential alternative to accurately self-monitor PA and SB via inbuilt inertial sensors [12-15]. However, battery life and mobile phone location have been major issues that have compromised usability and long-term monitoring. While external devices, such as wearables, may have overcome such weaknesses [16], the most popular devices are commercial motion sensors that use acceleration data to recognize activity behaviors (ie, distance, time, and intensity). Unfortunately, such measures struggle to distinguish postures (ie, sitting and standing), primarily due to wear position (ie, where the device is placed) and the use of proprietary algorithms that do not accurately quantify such behaviors [7].

Commercially available devices that examine SB through postural positioning rather than lack of movement (ie, acceleration) are scarcer [17]. However, devices that quantify time spent sitting, standing, and light intensity PA are critical when self-monitoring occupational behaviors, as moderate-to-vigorous physical activity is less prevalent during working hours or transport time to and from work [18].

Mobile phones alone struggle with postural identification due to the nonattachment of phones to the body and the ubiquitous nature of phone use [12]. However, the use of mobile phones with external monitoring devices may have the potential to become an accurate, cost-effective self-monitoring tool [12]. The range of novel and engaging mobile phone–based intervention strategies, as well as the user's perceptions on their usefulness and viability, highlights the potential of such technology on PA promotion [12].

In this context, the Walk@Work-Application (W@W-App) was developed to self-monitor occupational PA and SB with a high level of validity. The W@W-App communicates with a MetaWearC external sensor [19], attached via a band to the thigh, to quantify occupational sitting, standing, and stepping while offering real-time feedback on these behaviors, which is an essential component for changing behaviors at the time and place where they occur. This study examined the validity of the W@W-App+ MetaWearC tool to quantify time spent in occupational sitting, standing, and stepping against the current gold-standard, device-based measure for postural behaviors.

Methods

Measurement Tools

The new W@W-App was developed from a previous version [20], adding a commercially available sensor (MetaWearC; MbientLab Inc) to gather postural and movement information. The MetaWearC is a small sensor (24 mm × 6 mm; 5.6 g) covered with a waterproof round case. The sensor is a triaxial accelerometer with an amplitude range of ±16 g and a sampling rate of 6.25 Hz. Key features of the MetaWearC sensor are shown on the MbientLab web page [19]. Raw sensor data are synchronized with the W@W-App software via a low-energy Bluetooth system with a long battery life (>30 days) and a range of up to 10-15 meters. The data are directly processed and displayed in real time by the app on the phone and securely stored on the backend server. Figure 1 depicts the W@W-App (login page) and the MetaWearC sensor.

The algorithm for the W@W-App+MetaWearC (Figure 2) was designed to analyze accelerometer output from the MetaWearC sensor. The MetaWearC sensor is worn within a small bag inside an elastic and adjustable band (Figure 1) that is attached to the participant's right thigh. The algorithm is based on two primary requirements: (i) data can only be recorded during the defined recording period (ie, working hours) and (ii) data can only be collected when both the device and the software are connected via Bluetooth. When these criteria have been met and the sensor detects an acceleration, the step counter begins and the sitting and standing counters are reset to 0. Stepping time is initiated when the sensor identifies a balance between false positives (ie, counting a step when the step has not happened) and false negatives (ie, not counting a step when the step has occurred). There are three sensitivity modes for the step detector: normal, sensitive, and robust. These modes balance sensitivity (false negatives) and robustness (false positives). Normal mode is used in most applications as it provides a balance between false positives and false negatives. An example of a false positive would be the detection of a step while an individual is in a sitting position, possibly as a result of stretching one's leg.

Figure 1. (A) Initial W@W-App interface and (B) the MetaWearC sensor with the waterproof case and the thigh band.

Figure 2. Flow diagram illustrating the algorithm for decision-rules used by the W@WApp + MetaWearC sensor to determine sitting, standing and stepping time.
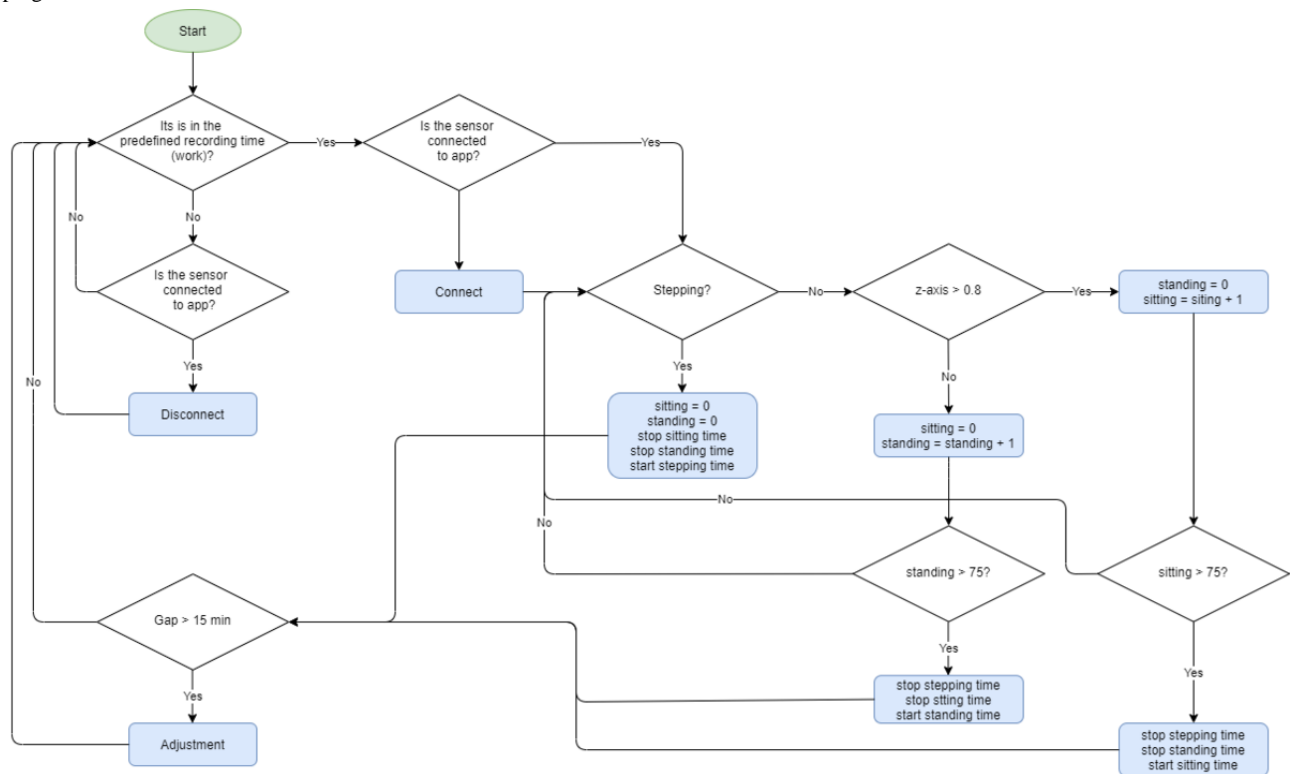
The recognition of postures (sitting and standing) is based on the angle of the z-axis, where 0 indicates a completely vertical posture (standing) and 1 indicates a completely horizontal posture (sitting). When the sensor detects a value higher than 0.8 in the z-axis, the sitting counter is initiated while the standing counter remains to 0. When the sensor detects a value equal to or lower than 0.8 in the z-axis, the standing counter begins while the sitting counter returns to 0. If either the sitting or standing counters reach 75 readings (approximately every 2 seconds), this indicates that the sensor has not detected stepping during those 75 readings, and therefore the step counter stops and assumes that the user is either sitting or standing depending on which of these counters reaches 75.

Finally, if there is a difference greater than 15 minutes between the time counters for the W@W-App-MetaWearC (stepping, sitting, and standing) and the elapsed time, a weighted adjustment is completed. Normally, this difference is due to temporal disconnections of the sensor if it is kept more than 20 meters away from the mobile phone. For example, if the W@W-App-MetaWearC has counted for 100 minutes (75 minutes stepping, 10 minutes standing, and 15 minutes sitting), but the real time elapsed is 115 minutes, the weighted adjustment

will correct the W@W-App-MetaWearC to 86 minutes stepping, 12 minutes standing, and 17 minutes sitting.

The activPAL3M (PAL Technologies Ltd) is referred to as the gold-standard, device-based measure for postural recognition in free-living conditions [21]. The activPAL3M was employed as the criterion measure for sitting, standing, and stepping times. The activPAL3M (25 mm × 45 mm × 5mm; 9g) was placed in a waterproof nitrile sleeve and attached on the midline of the anterior aspect of the participant's thigh using a transparent film (10 cm × 10cm of hypoallergenic Tegaderm Foam Adhesive Dressing).

Participants and Procedures

Office workers from the University of Vic-Central University of Catalonia (UVic-UCC) who owned a mobile phone (Android version 6.0.0/iOS version 10.0.0 or higher) were invited to participate in the study. A convenience sample was recruited (N=23). All volunteers provided written informed consent prior to participation. This study was conducted within a Spanish national project (W@WApp-Diab; PI17/01788) led by the UVic-UCC. Ethical approval was obtained by the research ethics committee of the Research Institute of Primary Care Jordi Gol (IDIAP).

Participants installed and configured the W@W-App, following guidance provided by the researchers: (i) registration on the Walk at Work web platform [22], (ii) user verification through email, (iii) W@W-App installation and initialization, (iv) recording day and time period configuration (ie, between 3 and 8 working hours), and (v) recognition of the MetaWearC sensor via Bluetooth. In adherence to the European Union General Data Protection Regulation, participants could read the private policy of the W@W-App, which is written using clear and straightforward language, on the Walk at Work website [22]. In addition, participants provided affirmative consent prior to using the W@W-App when they voluntarily registered on the web platform.

Researchers initialized the activPAL3M and the W@W-App+MetaWearC and placed both devices on the midpoint (ie, one over the other) of the anterior aspect of the thigh of the same leg to avoid measurement bias due to asymmetric leg positions and movements. To ensure that the timestamp of the W@W-App+MetaWearC and the activPAL3M aligned for data analysis, they were initialized from the same PC.

Participants wore the W@W-App+MetaWearC and the activPAL3M sensor in occupational free-living conditions for 3-8 hours. They were required to keep their mobile phone within a 5-meter radius throughout the measurement period (ie, participants were asked to keep their mobile phones with them at all times).

Variables and Statistical Analysis

The variables recorded and quantified by the W@W-App+MetaWearC were time spent in sitting and standing postures and time spent stepping. These variables were extracted from the W@W-App software. For the activPAL3M, files were processed via the activPAL Professional Software (version 7.2.32) upon completion of data collection. Data were then exported to a Microsoft Excel (Microsoft Excel 2016; Microsoft Corporation) file format, providing data on sitting, standing, and stepping in 15-second epochs. This enabled the quantification of the number of minutes spent sitting, standing, and stepping. In addition, the time spent sitting and standing were added together to quantify stationary time, while the amount of time spent standing and stepping were added together to compute upright time. Total recording time (ie, minutes) from both devices was calculated by summing the amount of time spent sitting, standing, and stepping.

Descriptive characteristics (mean [SD] and median [IQR]) were used to describe the data. Differences between the W@W-App+MetaWearC and the activPAL3M were examined using paired-samples t tests and Wilcoxon signed-rank tests. Pearson correlation coefficients were used to determine the strength and direction of association between variables quantified by the two measures when the data were normally distributed. Spearman rank-order correlation coefficients were employed when data were not normally distributed. The concordance correlation coefficient (CCC), using Lin's approach [23], was used to examine the level of agreement between the W@W-App+MetaWearC variables and the activPAL3M determined variables. The CCC values were interpreted using the categorization recommended by McBride [24]. Bland-Altman plots with mean bias and limits of agreement (LoA) were constructed to examine the agreement between the W@W-App+MetaWearC variables and the activPAL3M determined variables using similar approaches reported previously [25]. Equivalence was determined using two one-sided paired t tests (90% CI) for the mean difference between the W@W-App+MetaWearC variables and the activPAL3M determined variables [26]. Equivalence was supported if the CI for the mean difference was within 15% of the activPAL3M-determined time spent sitting, standing, and stepping. The equivalence region was arbitrarily defined, as limited biologically and analytically relevant criteria can be defined for the equivalence regions for sitting, standing, and stepping. Less conservative equivalence regions were also tested in case the equivalence was not supported for the 10% level. Additional tests to determine the region of equivalence were completed using increments of 5%. This approach was selected to provide a clear estimation of the accuracy of the W@W-App+MetaWearC [27]. Measures were expected to differ by no more than 30 minutes for sitting, 11 minutes for standing time, and 4 minutes for stepping. All statistical analysis was conducted using IBM SPSS Statistics 25 (IBM Corporation) and Microsoft Excel (Microsoft Corporation).

Results

In total, 23 office workers participated in the study, whereby activity behavior information was recorded by both measures during workplace free-living conditions between October and November 2018. After excluding 3 participants because of technical problems with the mobile phone, data from 20 participants were included in the analyses (age: mean 39.5 years, SD 8.1, range 27-60; 16 [80%] women). A total of 115 hours of data was recorded, with an average of 5 hours per participant.

Of all participants, 13 used an Android smartphone (Samsung, n=5; BQ Aquaris, n=4; Xiaomi, n=2; Xperia, n=1; and Huawei, n=1) with an operational system version ranging from 6.0.1 to 8.0.0. The other 7 participants employed an iPhone 6 or iPhone 7 with an operational system version higher than 10.3.3.

Descriptive characteristics for variables of interest from the W@W-App+MetaWearC and the activPAL3M, as well as the

statistical differences between the two measures for each variable, are described in Table 1. The median recording time for the W@W-App+MetaWearC was 237.50 (SD 132.75) minutes, while the activPAL3M median recording time was 240.00 (SD 127.50) minutes. No significant differences between the W@W-App+MetaWearC and the activPAL3M were observed for sitting time (*P*=.53), standing time (*P*=.12), and stepping time (*P*=.54).

Table 1. Descriptive characteristics and statistical significance (*P* value) of the differences between the W@W-App+MetaWearC and the activPAL3M for minutes spent in different activity behaviors (N=20).

Variable	W@W-App	activPAL3M	*P* value
Recording time (min), median (IQR)	237.5 (132.8)[a]	240.0 (127.5)[a]	<.001
Sitting time (min), median (IQR)	191.0 (132.0)[a]	180.5 (124.3)[a]	.53
Standing time (min), mean (SD)	70.3 (38.1)	75.4 (36.1)	.12
Stepping time (min), median (IQR) or mean (SD)	22.0 (24.0)[a]	24.0 (10.5)	.61
Stationary time (min), median (IQR)	223.5 (147.3)[a]	227.0 (138.0)[a]	.002
Upright time (min), mean (SD)	47.7 (23.6)	49.7 (21.8)	.25

[a]Data presented as median (IQR) due to nonnormality.

The W@W-App+MetaWearC showed strong to very strong correlations with activPAL3M-determined activity variables. CCCs identified substantial agreement between the two measures for sitting (CCC=0.98, 95% CI 0.96-0.99), moderate

agreement for standing (CCC=0.93, 95% CI 0.81-0.97), and poor agreement for stepping (CCC=0.74, 95% CI 0.47-0.88). The correlation coefficients, CCC values, and associated 95% CI are shown in Table 2.

Table 2. Correlation coefficients (*r*), concordance correlation coefficients (CCC), and 95% CIs between the W@W-App+MetaWearC and the activPAL3M for minutes spent in different activity behaviors (N=20). All *P*<.001.

Variable	*r* (95% CI)	CCC (95% CI)
Recording time (min)	0.89 (0.73-0.95)	0.99 (0.99-0.99)
Sitting time (min)	0.97 (0.92-0.99)	0.98 (0.96-0.99)
Standing time (min)	0.93 (0.83-0.97)	0.92 (0.82-0.97)
Stepping time (min)	0.74 (0.44-0.89)	0.74 (0.47-0.88)
Stationary time (min)	0.96 (0.90-0.98)	0.99 (0.99-1.00)
Upright time (min)	0.95 (0.88-0.98)	0.95 (0.87-0.98)

The mean bias and LoA from the Bland-Altman analysis are provided in Table 3. The Bland-Altman plots, which compare the mean sitting, standing, stepping, stationary, and upright times measured by the W@W-App+MetaWearC and the activPAL3M are presented in Figures 3 and 4. The Bland-Altman plots present a graphical description of the means for sitting, standing, stepping, stationary, and upright times as measured by the W@W-App+MetaWearC and the activPAL3M against the difference of the time spent in each of these behaviors between both measures. For sitting, a smaller mean bias was observed (–1.66 minutes) with a relatively wide LoA (–30.37 to 27.05). The equivalence procedure indicated that the 90% CI for the mean difference was 0.2 and 20.8 and was within the 15% equivalence region (–30.0 to +30.0 minutes). The

estimates obtained from the two measures were considered equivalent for sitting time. The largest observed mean bias for a specific behavior was observed for standing time (–4.85 minutes; LoA –31.31 to 21.62). The 90% CI for the mean difference was –10.5 and 0.3 and was within the 15% equivalence region (–11.0 to +11.0 minutes). The estimates obtained from the two measures were considered equivalent for standing time. For stepping time, a small mean bias was observed (1.15 minutes; LoA –15.11 to 17.41). However, the equivalence procedure indicated that the 90% CI for the mean difference was –4.5 and 2.1, which was not significantly within the 15% equivalence region (–4.0 to +4.0 minutes). The estimates obtained from the two measures were not considered equivalent for stepping time.

Table 3. Mean bias and limits of agreement (LoA) for sitting, standing, and stepping times.

Variable	Mean bias	Lower LoA	Upper LoA
Recording time (min)	–5.37	–13.56	2.81
Sitting time (min)	–1.66	–30.37	27.05
Standing time (min)	–4.85	–31.31	21.62
Stepping time (min)	1.15	–15.11	17.41
Stationary (min)	–6.52	–20.81	7.78
Upright time (min)	–1.85	–15.76	12.06

Figure 3. Bland-Altman plots of absolute agreement for (A) sitting, (B) standing, and (C) stepping, derived from the W@W-App+MetaWearC with the equivalent outcome derived from the activPAL3M.
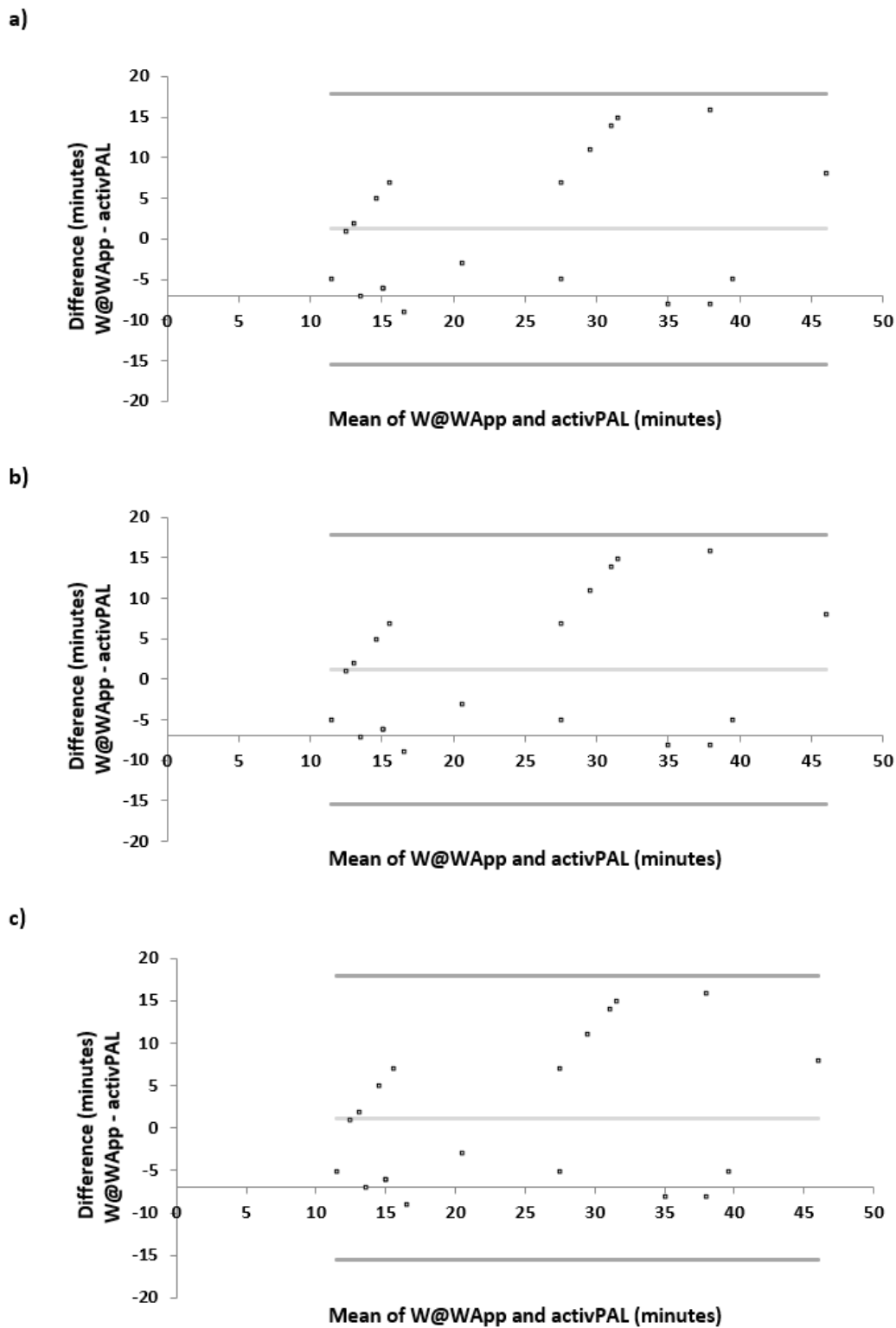
Figure 4. Bland-Altman plots of absolute agreement for (A) stationary time (sitting+standing) and (B) upright time (standing+stepping), derived from the W@W-App+MetaWearC with the equivalent outcome derived from the activPAL3M.

a)
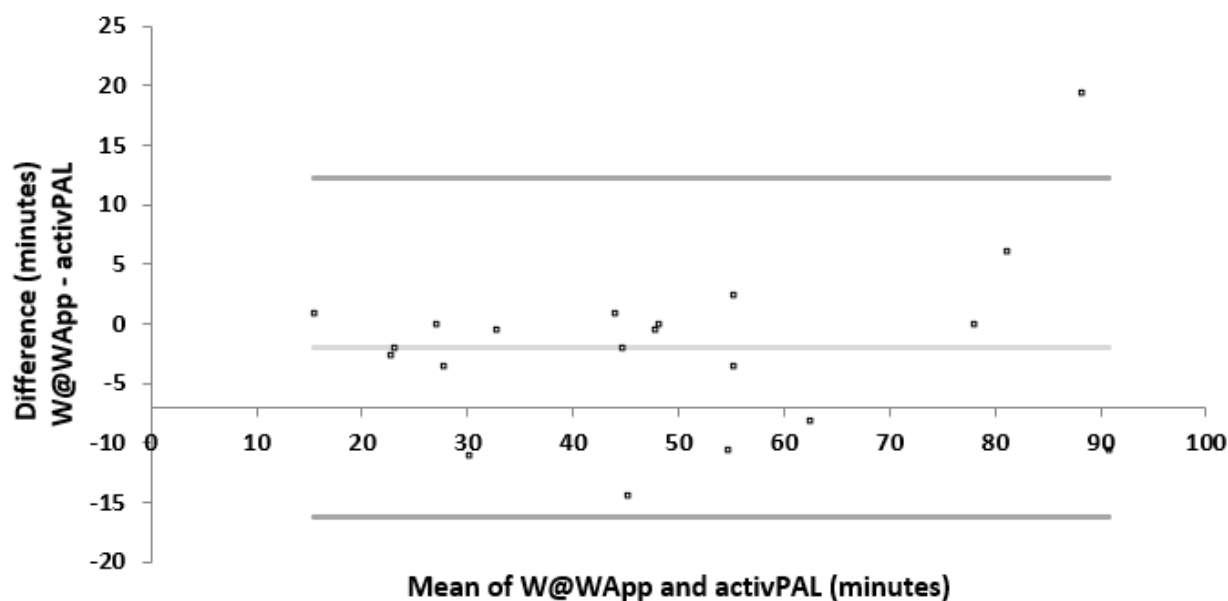
b)
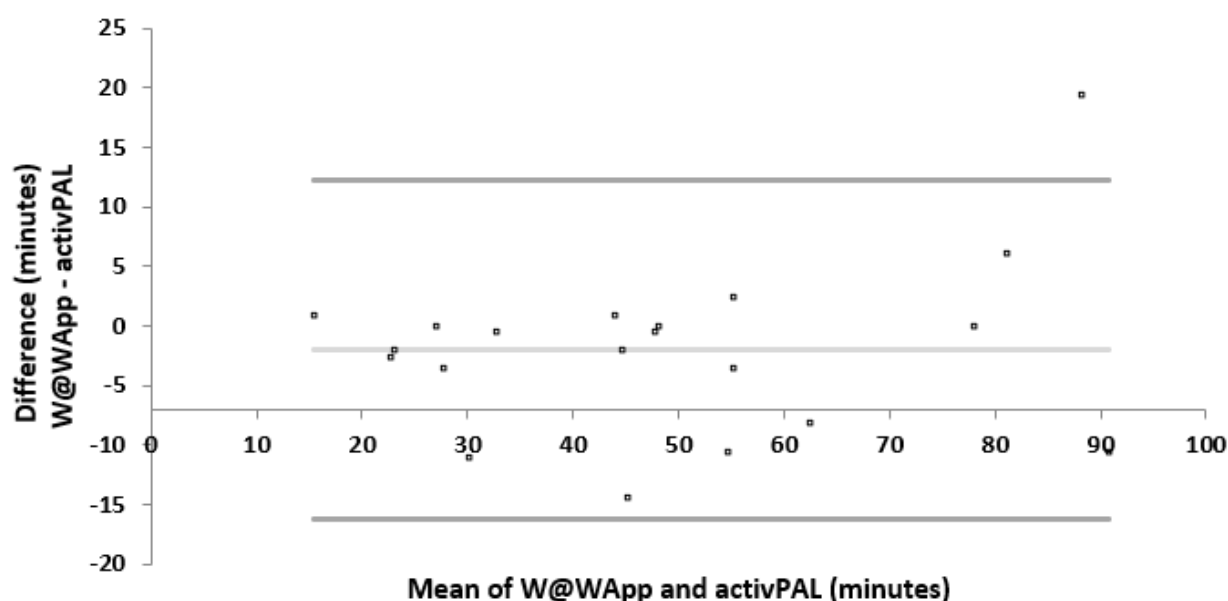
When combining variables, stationary time significantly differed between the two measures (*P*=.002), while no differences were observed for upright time (*P*=.25). However, stationary and upright times were strongly correlated with the criterion measure (activPAL3M) (*P*<.001). Time spent on stationary activities was underestimated with a mean bias of –6.52 minutes, with a relatively small LoA (–20.81 to 7.78 minutes). The 90% CI for the mean difference was 1.3 and 12.4 and was within the 15% equivalence region (–41 to +41 minutes). The estimates obtained from the two measures were considered equivalent for stationary time. A mean bias of –1.85 minutes was identified for time spent upright, with a relatively small LoA (–15.76 to 12.06 minutes) compared to the noncombined postural/activity variables. The equivalence procedure indicated that the 90% CI

for the mean difference was –0.9 and 4.8 and was within the 15% equivalence region (–7.0 to +7.0 minutes). The estimates obtained from the two measures were considered equivalent for upright time.

Discussion

Principal Findings

This study examined the validity of the W@W-App+MetaWearC to measure occupational sitting, standing, and stepping times in a free-living workplace environment. Our findings indicated that the W@W-App+MetaWearC is a valid tool for self-monitoring occupational sitting, standing, stationary, and upright times,

demonstrating moderate to very strong validity when compared to the criterion measure (activPAL3M). However, the analysis demonstrated that the findings for stepping from the W@W-App+MetaWearC are not equivalent to those from the activPAL3M.

Although a small mean bias of 1.15 minutes for stepping was observed between the W@W-App+MetaWearC and the activPAL3M, poor agreement, wide CIs and nonequivalence would suggest that the W@W-App+MetaWearC should not be recommended for use in detecting stepping time. However, it is quite plausible that these observations can be attributed not only to variance from the W@W-App+MetaWearC tool but also variance in the activPAL3M device. The activPAL3M is primarily used to examine postural position, namely sitting and standing, and has demonstrated high levels of accuracy in the detection of these behaviors in lab-based and free-living conditions [28], justifying its use as a device-based comparison for the measurement of sitting and standing times. However, lower levels of validity for the activPAL3M have been highlighted for stepping time and step count, particularly during activities of daily living. Therefore, future research should aim to utilize more accurate methods of movement when validating the W@W-App+MetaWearC tool. It should be acknowledged that W@W-App+MetaWearC performs relatively well in the detection of steps in free-living conditions when compared with findings from other commercially available activity monitors [29,30].

For stationary time (ie, sitting and standing), the W@W-App+MetaWearC demonstrated high levels of accuracy when compared with previous validation studies employing a range of activity monitors [31]. This is likely due to the ability of the W@W-App+MetaWearC to detect sitting and standing postures based on thigh acceleration. Recent studies have developed and validated self-monitoring devices that also provide real-time feedback on an integrated display, including the SitFit [32] or through a mobile phone app via Bluetooth synchronization such as the VitaBit [33] and Chair&App [34]. Similar to the findings presented here, the SitFit and Chair&App devices reported that sitting time was highly accurate when compared to the activPAL3M in free-living conditions. However, the W@W-App+MetaWearC reported a lower mean bias (W@W-App+MetaWearC) in comparison to other studies (SitFit). In contrast, the VitaBit device did not accurately distinguish between sitting and standing in free-living conditions but was accurate in the detection of movement. These findings are unsurprising, since the device used as the comparison measure (ActiGraph) struggles to accurately distinguish sitting and standing behaviors [35]. Both the SitFit and the VitaBit were designed to be worn in the pocket of a user's trousers, which may be a usability barrier when wearing clothes without pockets. Chair&App, as well as the W@W-App, focused on office-based jobs, but Chair&App used a regular office chair equipped with pressure sensors instead of a thigh-attached device. That may remove compliance issues related to recording time but standing, stepping, and sitting away from one's personal desk cannot be captured. The W@W-App has demonstrated

high levels of validity for sitting, standing, stationary, and upright times, while the wearer's location and attachment site may increase compliance with wearing a self-monitoring tool in the workplace.

The W@W-App+MetaWearC is a novel tool that simulates the activPAL3M activity monitor in accurately recognizing postural position at the workplace. The output from the W@W-App+MetaWearC tool for sitting, standing, stationary, and upright times were identified as equivalent to the current gold-standard, device-based postural measure, the activPAL3M. This suggests that this self-monitoring tool, which provides real-time feedback to users, is worthwhile for use in interventions that aim to reduce sitting behavior in the workplace. Self-monitoring is vital for increasing individuals' awareness and empowerment toward behavior change [7]. This may result in a more accurate, affordable, and accessible device than those currently available, enabling the more cost-effective inclusion of SB self-monitoring as a function of SB interventions in the future.

Strengths and Limitations

The strengths of this study include (i) the examination of the complete range of occupational sedentary and activity behavior types (sitting/lying, standing, and stepping), (ii) the examination of the validity of these measures in occupational free-living conditions, and (iii) the use of a gold-standard, objective measurement device to determine the validity of the W@W-App+MetaWearC.

The present study is not without limitations. Although the activPAL3M has been described as the gold standard for measuring sitting time [21] and is an acceptable field-based measure for activity behaviors in youth and adult populations [36,37], it is not the gold standard for comparison of stepping time. This should then be considered when interpreting the Bland-Altman plots, as these are designed to support the comparison of a new measure to a previous gold standard. The relatively small sample size with a large percentage of females (16/20) and the homogeneity of the workplace setting (ie, all sampled from a university context) might differ from the general office population. Furthermore, the data gathered included an average of 5 hours per subject, providing a limited timeframe. Additionally, the wide range of operating systems and hardware available added complexity to app development and subsequent validation.

Conclusions

The W@W-App+MetaWearC self-monitoring system demonstrates high levels of accuracy in determining postural position. This tool is a low-cost alternative tool for the examination of occupational sitting and standing times. It has demonstrated high levels of validity in detecting postural position and provides real-time feedback to users. Future research should examine the interventional effect of utilizing this system as a self-monitoring tool for modifying activity behaviors in office-based workers.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]

2. Del Pozo-Cruz J, García-Hermoso A, Alfonso-Rosa RM, Alvarez-Barbosa F, Owen N, Chastin S, et al. Replacing Sedentary Time: Meta-analysis of Objective-Assessment Studies. Am J Prev Med 2018 Sep;55(3):395-402. [doi: 10.1016/j.amepre.2018.04.042] [Medline: 30122216]

3. Grgic J, Dumuid D, Bengoechea EG, Shrestha N, Bauman A, Olds T, et al. Health outcomes associated with reallocations of time between sleep, sedentary behaviour, and physical activity: a systematic scoping review of isotemporal substitution studies. Int J Behav Nutr Phys Act 2018 Jul 13;15(1):69 [FREE Full text] [doi: 10.1186/s12966-018-0691-3] [Medline: 30001713]

4. Lakerveld J, Loyen A, Schotman N, Peeters CF, Cardon G, van der Ploeg HP, et al. Sitting too much: A hierarchy of socio-demographic correlates. Prev Med 2017 Aug;101:77-83 [FREE Full text] [doi: 10.1016/j.ypmed.2017.05.015] [Medline: 28551361]

5. Keadle S, Conroy D, Buman M, Dunstan D, Mattews C. Targeting Reductions in Sitting Time to Increase Physical Activity and Improve Health. Med Sci Sport Exerc 2017;49(8):82. [doi: 10.1249/mss.0000000000001257]

6. Smith L, McCourt O, Sawyer A, Ucci M, Marmot A, Wardle J, et al. A review of occupational physical activity and sedentary behaviour correlates. Occup Med (Lond) 2016 Apr 26;66(3):185-192. [doi: 10.1093/occmed/kqv164] [Medline: 27016747]

7. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The Effect of Physical Activity Interventions Comprising Wearables and Smartphone Applications on Physical Activity: a Systematic Review and Meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]

8. Dowd KP, Szeklicki R, Minetto MA, Murphy MH, Polito A, Ghigo E, et al. A systematic literature review of reviews on techniques for physical activity measurement in adults: a DEDIPAC study. Int J Behav Nutr Phys Act 2018 Feb 08;15(1):15 [FREE Full text] [doi: 10.1186/s12966-017-0636-2] [Medline: 29422051]

9. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exerc Sport 2000 Jun 13;71 Suppl 2(sup2):1-14. [doi: 10.1080/02701367.2000.11082780] [Medline: 25680007]

10. van Poppel MN, Chinapaw MJ, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. Sports Med 2010 Jul 01;40(7):565-600. [doi: 10.2165/11531930-000000000-00000] [Medline: 20545381]

11. Neilson H, Robson P, Friedenreich C, Csizmadi I. Estimating activity energy expenditure: how valid are physical activity questionnaires? Am J Clin Nutr 2008 Feb;87(2):279-291. [doi: 10.1093/ajcn/87.2.279] [Medline: 18258615]

12. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May 5;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]

13. Della Mea V, Quattrin O, Parpinel M. A feasibility study on smartphone accelerometer-based recognition of household activities and influence of smartphone position. Inform Health Soc Care 2017 Dec 22;42(4):321-334. [doi: 10.1080/17538157.2016.1255214] [Medline: 28005434]

14. Hekler EB, Buman MP, Grieco L, Rosenberger M, Winter SJ, Haskell W, et al. Validation of Physical Activity Tracking via Android Smartphones Compared to ActiGraph Accelerometer: Laboratory-Based and Free-Living Validation Studies. JMIR Mhealth Uhealth 2015 Apr 15;3(2):e36 [FREE Full text] [doi: 10.2196/mhealth.3505] [Medline: 25881662]

15. Lemaire ED, Tundo MD, Baddour N. Evaluation of a Smartphone-based Human Activity Recognition System in a Daily Living Environment. J Vis Exp 2015 Dec 11(106):e53004 [FREE Full text] [doi: 10.3791/53004] [Medline: 26710275]

16. Mortazavi B, Nemati E, VanderWall K, Flores-Rodriguez H, Cai J, Lucier J, et al. Can smartwatches replace smartphones for posture tracking? Sensors (Basel) 2015 Oct 22;15(10):26783-26800 [FREE Full text] [doi: 10.3390/s151026783] [Medline: 26506354]

17. Sanders JP, Loveday A, Pearson N, Edwardson C, Yates T, Biddle SJ, et al. Devices for Self-Monitoring Sedentary Time or Physical Activity: A Scoping Review. J Med Internet Res 2016 May 04;18(5):e90 [FREE Full text] [doi: 10.2196/jmir.5373] [Medline: 27145905]

XSL•FO
RenderX

18. Puig-Ribera AM, Bort-Roig J, Chirveches-Perez E, Puigdoriol-Juvanteny E, Navarro-Blasco L, Bausa-Peris R, et al. Sitting less at work as an input to the 24-hours movement behaviours model. Journal of Physical Activity and Health 2018;15(10):S237-S238.

19. MetwearC Product Specification v1. MbientLab Inc. URL: https://mbientlab.com/documents/MetaWearC-CPRO-PS.pdf [accessed 2019-11-15]

20. 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 Nov;32(6):563-566 [FREE Full text] [doi: 10.1016/j.gaceta.2017.05.004] [Medline: 28923337]

21. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of Wearable Monitors for Assessing Sedentary Behavior. Medicine & Science in Sports & Exercise 2011;43(8):1561-1567. [doi: 10.1249/mss.0b013e31820ce174]

22. Walk@Work: Sit less, move more. University of Vic-Central University of Catalonia. URL: http://walkatwork.uvic.cat/en/ [accessed 2019-11-15]

23. Lin LI. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 1989 Mar;45(1):255. [doi: 10.2307/2532051]

24. McBride G. A proposal for strength-of-agreement criteria for Lins concordance correlation coefficient. NIWA Client Report: HAM-062. 2005. URL: https://www.medcalc.org/download/pdf/McBride2005.pdf [accessed 2019-12-15]

25. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 1986 Feb;327(8476):307-310. [doi: 10.1016/s0140-6736(86)90837-8]

26. Lakens D. Equivalence Tests: A Practical Primer for Tests, Correlations, and Meta-Analyses. Soc Psychol Personal Sci 2017 May 05;8(4):355-362 [FREE Full text] [doi: 10.1177/1948550617697177] [Medline: 28736600]

27. Saint-Maurice PF, Welk GJ. Validity and Calibration of the Youth Activity Profile. PLoS One 2015 Dec 2;10(12):e0143949 [FREE Full text] [doi: 10.1371/journal.pone.0143949] [Medline: 26630346]

28. Sellers C, Dall P, Grant M, Stansfield B. Agreement of the activPAL3 and activPAL for characterising posture and stepping in adults and children. Gait Posture 2016 Jul;48:209-214. [doi: 10.1016/j.gaitpost.2016.05.012] [Medline: 27318305]

29. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015 Mar 27;12(1):42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]

30. Gomersall SR, Ng N, Burton NW, Pavey TG, Gilson ND, Brown WJ. Estimating Physical Activity and Sedentary Behavior in a Free-Living Context: A Pragmatic Comparison of Consumer-Based Activity Trackers and ActiGraph Accelerometry. J Med Internet Res 2016 Sep 07;18(9):e239 [FREE Full text] [doi: 10.2196/jmir.5531] [Medline: 27604226]

31. Rosenberger M, Buman M, Haskell W, Mcconnell M, Carstensen L. Twenty-four Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices. Medicine & Science in Sports & Exercise 2016;48(3):457-465. [doi: 10.1249/mss.0000000000000778]

32. Gill J, Hawari N, Maxwell D, Louden D, Mourselas N, Bunn C. Validation of a Novel Device to Measure and Provide Feedback on Sedentary Behavior. Med Sci Sports Exerc 2018;50(3):32. [doi: 10.1249/mss.0000000000001458]

33. Berninger N, ten Hoor GA, Plasqui G. Validation of the VitaBit Sit-Stand Tracker: Detecting Sitting, Standing, and Activity Patterns. Sensors (Basel) 2018 Mar 15;18(3):877 [FREE Full text] [doi: 10.3390/s18030877] [Medline: 29543766]

34. Arrogi A, Boen F, Seghers J. Validation of a smart chair and corresponding smartphone app as an objective measure of desk-based sitting. J Occup Health 2019 Jan 21;61(1):121-127 [FREE Full text] [doi: 10.1002/1348-9585.12033] [Medline: 30698337]

35. Clemes S, Edwardson C, Connelly J, Konstantinidis T, Koivula R, Yates T, et al. Validity of the ActiGraph GT3X+inclinometer and different counts per minute cut-points for the assessment of sedentary behaviour. Journal of Science and Medicine in Sport 2012 Dec;15:S68. [doi: 10.1016/j.jsams.2012.11.164]

36. Powell C, Carson BP, Dowd KP, Donnelly AE. Simultaneous validation of five activity monitors for use in adult populations. Scand J Med Sci Sports 2017 Dec 01;27(12):1881-1892. [doi: 10.1111/sms.12813] [Medline: 27905148]

37. Dowd KP, Harrington DM, Donnelly AE. Criterion and concurrent validity of the activPAL™ professional physical activity monitor in adolescent females. PLoS One 2012 Oct 19;7(10):e47633 [FREE Full text] [doi: 10.1371/journal.pone.0047633] [Medline: 23094069]

Abbreviations

CCC: concordance correlation coefficient
LoA: limits of agreement
mHealth: mobile health
PA: physical activity
UVic-UCC: University of Vic-Central University of Catalonia
W@W-App: Walk@Work-Application

Original Paper

Rams Have Heart, a Mobile App Tracking Activity and Fruit and Vegetable Consumption to Support the Cardiovascular Health of College Students: Development and Usability Study

Michelle C Krzyzanowski[1], PhD; Paul N Kizakevich[1], MSc, PE; Vanessa Duren-Winfield[2], MSc, PhD; Randall Eckhoff[1], BSc; Joel Hampton[1], MSc; Loneke T Blackman Carr[2,3], PhD, RD; Georgia McCauley[2], MT(AMT), MBA, PhD; Kristina B Roberson[2], FNP-C, ARNP, MPA, DNP; Elijah O Onsomu[2], MCHES, MSc, MPH, PhD; John Williams[2], PhD; Amanda Alise Price[2], PhD

[1]RTI International, Research Triangle Park, NC, United States
[2]Winston-Salem State University, Winston-Salem, NC, United States
[3]University of Connecticut, Storrs, CT, United States

Corresponding Author:
Michelle C Krzyzanowski, PhD
RTI International
3040 Conwallis Rd
Research Triangle Park, NC, 27709
United States
Phone: 1 919 485 5648
Email: mkrzyzanowski@rti.org

Abstract

Background: With the increasing use of mobile devices to access the internet and as the main computing system of apps, there is a growing market for mobile health apps to provide self-care advice. Their effectiveness with regard to diet and fitness tracking, for example, needs to be examined. The majority of American adults fail to meet daily recommendations for healthy behavior. Testing user engagement with an app in a controlled environment can provide insight into what is effective and not effective in an app focused on improving diet and exercise.

Objective: We developed Rams Have Heart, a mobile app, to support a cardiovascular disease (CVD) intervention course. The app tracks healthy behaviors, including fruit and vegetable consumption and physical activity, throughout the day. This paper aimed to present its functionality and evaluated adherence among the African American college student population.

Methods: We developed the app using the Personal Health Informatics and Intervention Toolkit, a software framework. Rams Have Heart integrates self-reported health screening with health education, diary tracking, and user feedback modules to acquire data and assess progress. The parent study, conducted at a historically black college and university-designated institution in southeastern United States, consisted of a semester-long intervention administered as an academic course in the fall, for 3 consecutive years. Changes were made after the cohort 1 pilot study, so results only include cohorts 2 and 3, comprising a total of 115 students (n=55 intervention participants and n=54 control participants) aged from 17 to 24 years. Data collected over the study period were transferred using the secure Hypertext Transfer Protocol Secure protocol and stored in a secure Structured Query Language server database accessible only to authorized persons. SAS software was used to analyze the overall app usage and the specific results collected.

Results: Of the 55 students in the intervention group, 27 (49%) students in cohort 2 and 25 (45%) in cohort 3 used the Rams Have Heart app at least once. Over the course of the fall semester, app participation dropped off gradually until exam week when most students no longer participated. The average fruit and vegetable intake increased slightly, and activity levels decreased over the study period.

Conclusions: Rams Have Heart was developed to allow daily tracking of fruit and vegetable intake and physical activity to support a CVD risk intervention for a student demographic susceptible to obesity, heart disease, and type 2 diabetes. We conducted an analysis of app usage, function, and user results. Although a mobile app provides privacy and flexibility for user participation in a research study, Rams Have Heart did not improve compliance or user outcomes. Health-oriented research studies relying on apps in support of user goals need further evaluation.

XSL·FO
RenderX

KEYWORDS

exercise; cardiovascular disease; diary; diet; mHealth; mobile phone

Introduction

Background

By 2025, nearly three-fourth of all internet users will access the web exclusively by mobile phone. Young people, in particular, show heavy usage: 53% own a smartphone by 11 years of age, and 84% of teenagers own a smartphone [1]. This increasing reliance on mobile phones has sparked the creation of self-help health apps, known as mobile health (mHealth) [2]. They aim to improve public health by supporting patient-led self-care [3,4]. They provide interactive, educational modules, and around 4 million versions can be readily downloaded through both the Google Play and Apple platforms [5]. They can be used for fitness tracking to address obesity, journal writing to address diabetes or mental health, and tracking other areas of concern, such as the menstrual cycle [6,7]. App developers see a growing youth market.

A popular mHealth focus is on fitness and diet tracking [6,7]. Fewer than 18% of American adults meet the daily recommendations for fruit and vegetable servings, and only one-third engage in the recommended amount of physical activity each week [8]. Although the transition from adolescence to young adulthood can be crucial in shaping weight-related behavior and lasting health patterns [9], only 10% of young adults entering college (18 to 30 years of age) are likely to adhere to dietary guidelines [10]. Studies also report a significant decrease in physical activity and an increase in sedentary behavior in this period [11]. Methods to address these health concerns include the promotion of healthy nutritional and dietary habits and physical activity [12-14].

The risk of obesity and high blood pressure (BP) is especially pronounced in minorities, leading to negative health outcomes when not detected early [15]. The incidence rate of diabetes among African Americans (AAs) is approximately twice that among non-Hispanic whites. Risk factors for cardiovascular disease (CVD) in college students include obesity and physical inactivity, which are disproportionately higher among AAs [16,17], and studies have shown that AA college students do not consistently engage in physical activity [18,19]. In a study conducted at a historically black college and university (HBCU), average BP was higher among AA students than their non-Hispanic white counterparts [20]. Whereas studies of the obesity epidemic have increased in recent years, few target AA college students.

To address this gap, we used the Personal Health Informatics and Intervention Toolkit (PHIT) framework to support a CVD risk factor intervention. mHealth apps have demonstrated success in improving medication adherence and supporting mental health interventions [21-23]. In particular, self-care methods have shown promising results in helping with depression and anxiety and encouraging a healthier lifestyle [24,25]. The Rams Have Heart mobile app targets healthy behaviors, tracking fruit and vegetable consumption, and physical activity throughout the day. It was not designed to actively support behavior change *per se* but to monitor the adaptation of concepts from the CVD-oriented college course. This study aimed to present the functionality of the Rams Have Heart mobile app and evaluated its usability and adherence in a vulnerable student population.

Methods

Ethics Approval

This study was approved by the Winston-Salem State University institutional review board (IRB). In addition, the Research Triangle International (RTI) Committee for the Protection of Human Subjects approved the RTI's role in the analysis of deidentified data.

Study Design and Population

This secondary analysis aimed to evaluate the usability and acceptability of the Rams Have Heart app, a tool developed as part of a larger pilot study to test an evidence-based CVD intervention in a susceptible demographic-AA college students.

This study was conducted at an HBCU in southeastern United States, where 73.6% of the student body is AA [20]. After the university's IRB reviewed and approved the study, the IRB at the supporting contracting institution approved its study personnel's access to the deidentified data for analysis. The CVD course was made available to students during the fall semester for 3 consecutive years; participants in each iteration are referred to as cohorts 1, 2, and 3. Cohort 1 was essentially in a pilot study, as the course and app were evolving, and a variety of revisions were made before starting cohort 2. Therefore, the study population, analyses, and results are limited to cohorts 2 and 3.

A total of 109 students participated across cohorts 2 and 3. Of these, 55 (50.4%) were assigned to the intervention group, who took the CVD risk-reduction course and were directed to use the Rams Have Heart app, and 54 (49.5%) were assigned to the control group, who took the traditional health course and did not use the app. All were full-time students at the institution, aged from 17 to 24 years. Cohort 2 was under study from fall 2017 to spring 2018, and cohort 3, from fall 2018 to spring 2019. The main course activity took place during the fall semester.

App Development

Personal Health Informatics and Intervention Toolkit

Rams Have Heart was developed using PHIT, a software-development framework geared toward research-oriented mobile apps [21,22]. PHIT apps comprise 5 primary elements: (1) user interactions involving forms, diaries, cognitive tasks, and games; (2) optional sensor data acquisition and processing; user-interaction scheduling, notification, and

XSL•FO

RenderX

management; and data visualization; (3) configuration and control scripts, virtual advisor processing, and activity and intervention logic; (4) database, privacy, and security support; and (5) study assets, such as educational materials and audiovisual media.

PHIT interactive instrument modules and configuration settings were implemented using extensible markup language and a custom language called PHITScript to construct program logic and activate such app functions as dynamically altering menus, controlling data collection, or scheduling notifications. The framework runtime loaded such modules and settings to render the mobile app dynamically tailored to the user and adaptable to the study protocol. The architecture has contributed reusable data collection and intervention modules to a growing library for mobile app studies.

Apps using the PHIT framework run on the user's device without an active internet connection. Raw and derived data are tagged with the protocol, participant, and other contextual information, encrypted, and stored locally in the app space. Whenever the internet is available, data are uploaded to a study-specific server via the secure https protocol. Individual and aggregate data can then be off-loaded via a password-protected portal for review and analysis. PHIT is based on Apache Flex and Adobe Integrated Runtime (AIR) technology, both open source and widely used for mobile game development. By using Apache Flex AIR, PHIT-based apps executed on both Android and iOS devices perform on smartphones and tablets and can be adapted for desktop platforms.

Rams Have Heart Mobile App

App Startup and BMI Calculation

Rams Have Heart integrated health screening with dietary and exercise education, food consumption and activity tracking, and graphic feedback showing progress. A simple, user-friendly home screen menu (Figure 1) listed app functions—tracking diaries, health information, personal feedback, and support modules—with an illustrative icon, a function label, and a brief descriptive statement to aid understanding.

When first running the app, the study participants entered their assigned identifier (ID) twice for verification. It was displayed on the blue home screen header above the function list for easy reference. All data records were tagged with this ID, and no personally identifiable information (PII) was saved. The research investigators maintained a study roster linking the ID to student contact information, and any other PII (offline and separate from the app), and any app-acquired data.

Next, participants were asked to set a time when notifications reminding them to keep up with dietary, activity, and evening diary entries would be posted (Figure 1). The *daily reminders* function was also listed on the home screen menu, allowing users to cancel or reschedule them to suit their needs.

Figure 1. Home screen menu for the Rams Have Heart app (a) and screens for selecting (b) and setting (b) daily diary reminders.
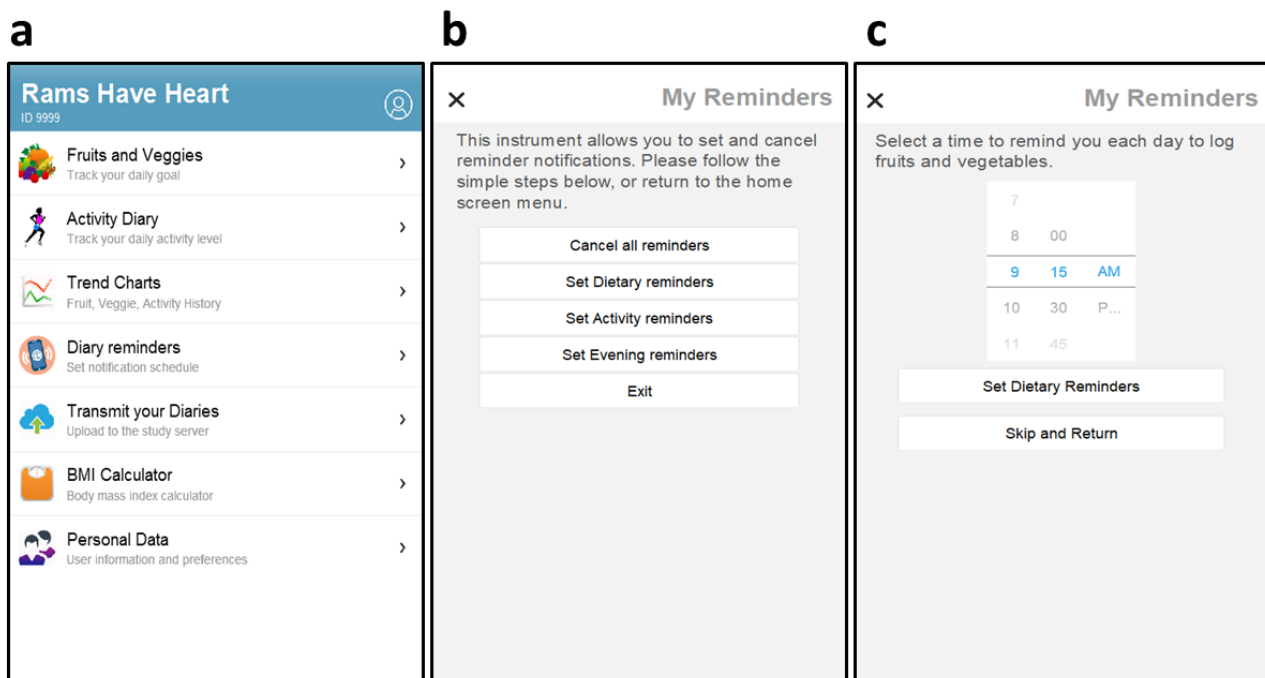
The final startup task was to enter self-reported personal data (ie, sex, age, height, weight) and save them for subsequent BMI calculations. Participants then needed only to enter their weight periodically to recalculate BMI and observe progress (Figure 2). Should any other personal data change (eg, age) during the study, it could be re-entered via the home screen menu. To support participant understanding, a series of graphic materials on BMI and its interpretation were available for reference (Figure 2).

XSL•FO
RenderX

Figure 2. BMI data entry and calculator (a), along with sample screens (b, c) from the five-slide set of BMI educational materials.
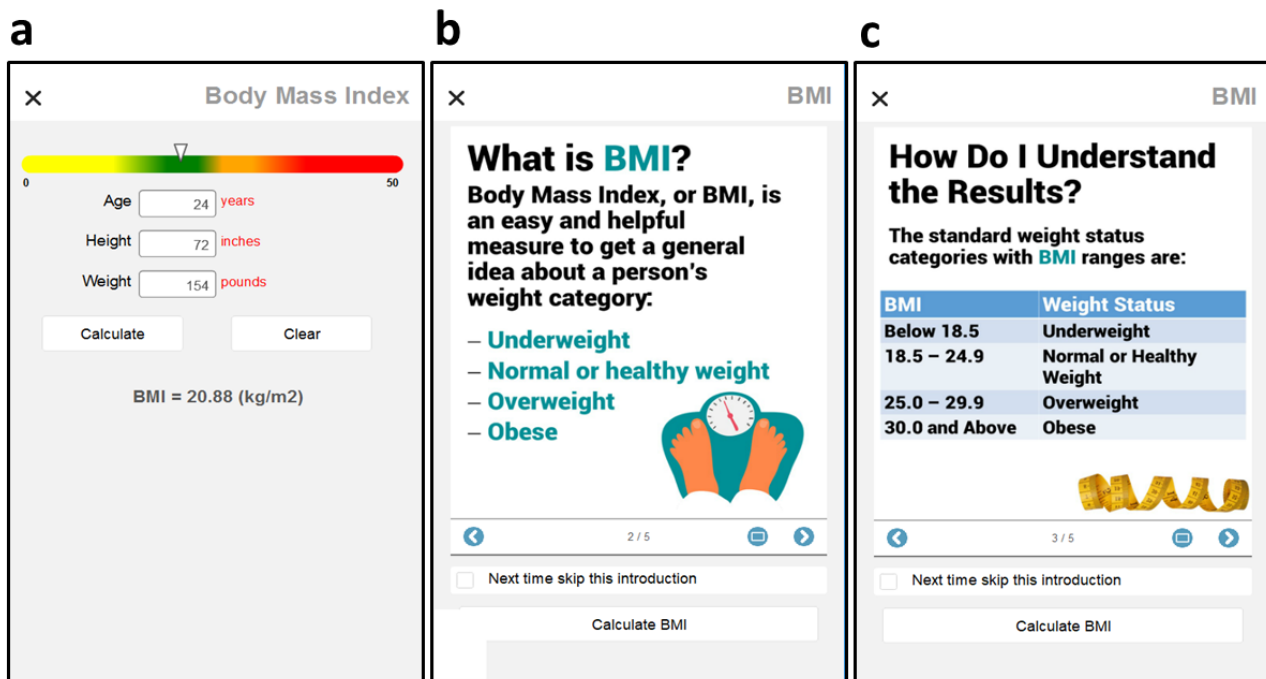
Dietary and Physical Activity Diaries

The primary function of the app was to support the daily recording of fruit and vegetable consumption and physical activity. The fruit and vegetable diary provided simple tap entry for up to four servings at each meal or when snacking across the day (Figure 3) for a range of 0 to 32 servings of fruits and vegetables per day. Servings data could be entered throughout the day or by recall at the end of the day and revised any time during the day. At the top of the screen, a color-coded progress bar advances toward a maximum of 10 servings. The green zone indicates whenever participants achieved at least six total servings, with a yellow transition to the 4 to 6 serving range. A graphic cluster of fruits and vegetables marks progress toward meeting the daily goal of 6 fruit or vegetable servings.

The physical activity diary used a similarly simple design, providing 5 levels of vigorous activity, moderate activity, and walking (Figure 3) across each day. The concept was modeled on the International Physical Activity Questionnaire (IPAQ). Participants could enter these data immediately after completing the activities or by recall later in the day and update or edit them at any time on the current day. At the top of the screen, a color-coded progress bar shows the cumulative weighted minutes of physical activity across the day. Weighted minutes were calculated using an IPAQ-adapted formula as follows:

Weighted minutes = (2.0 × vigorous minutes) + moderate minutes + (0.5 × walking minutes)

As new entries were made, the progress indicator advanced toward a maximum of 60 weighted min. The green zone indicated whenever participants achieved at least 45 weighted min, with a yellow transition for at least 30 weighted min. A graphic icon of a runner was used to indicate cumulative effort compliance.

Diary trend charts provided feedback to help students meet personal goals throughout the study intervention (Figure 3). A concise single screen presented a 3-week history of diary entries, allowing users to reflect on their eating and physical activity behaviors. This approach also aimed to support behavior change and encourage compliance with a daily diary entry.

During the first year of our study, participants frequently did not provide diary data every day. As they may have simply forgotten to make entries on a certain day, we revised the app for cohorts 2 and 3 to allow them to *catch up.* They could now fill in data for days they missed or forgot to enter for up to two days via recall. We added a form that appeared when participants entered either diary, asking whether the next entries will record (1) the current day or (2) 1 day or (3) 2 days earlier (Figure 4). Once the day was selected, either the fruit and vegetable dairy (Figure 4) or the physical activity diary (Figure 4) was presented. Labeling at the top of the diary screens was changed to remind the participant that recall data were to be entered for the specified day.

Figure 3. Example diary entry screens for daily tracking of fruit & vegetable consumption (a) and physical activity (b). Trend charts provide feedback on behavior goals over three weeks of recent entries (c).
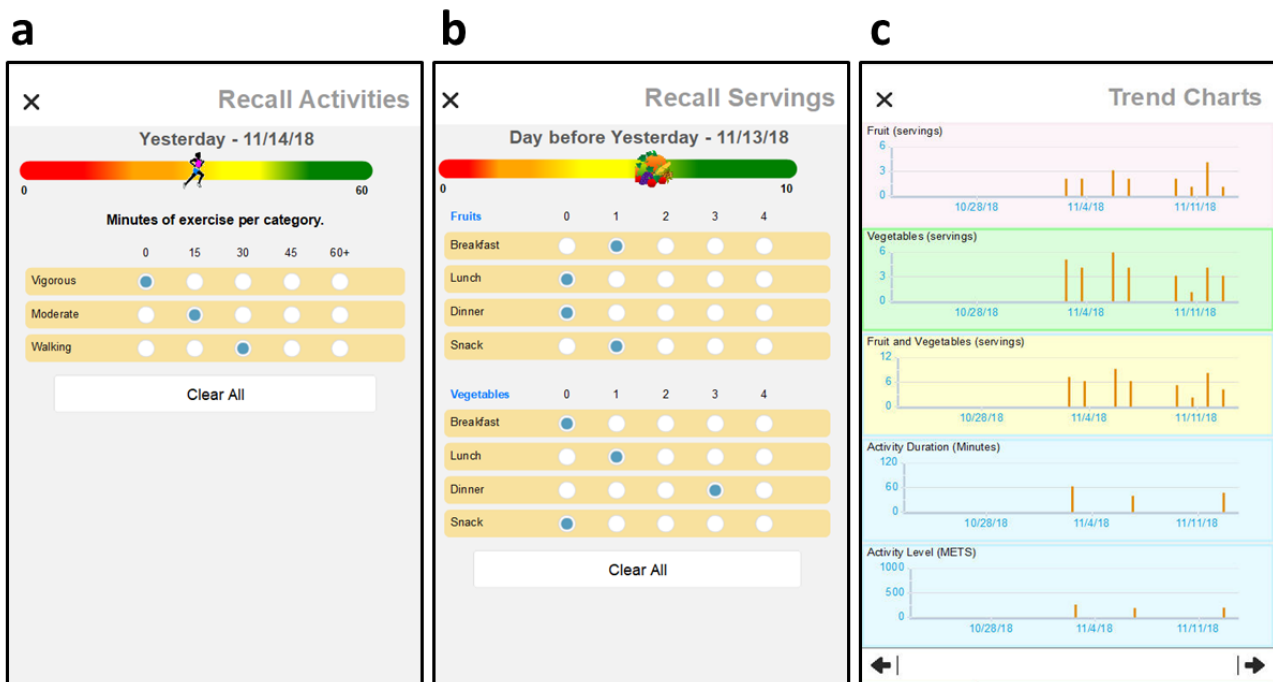
Figure 4. Users may catch up on missed diary entries for one or two days behind schedule (a). Labeling is changed to remind users that recall data are to be entered for specified day (b, c).
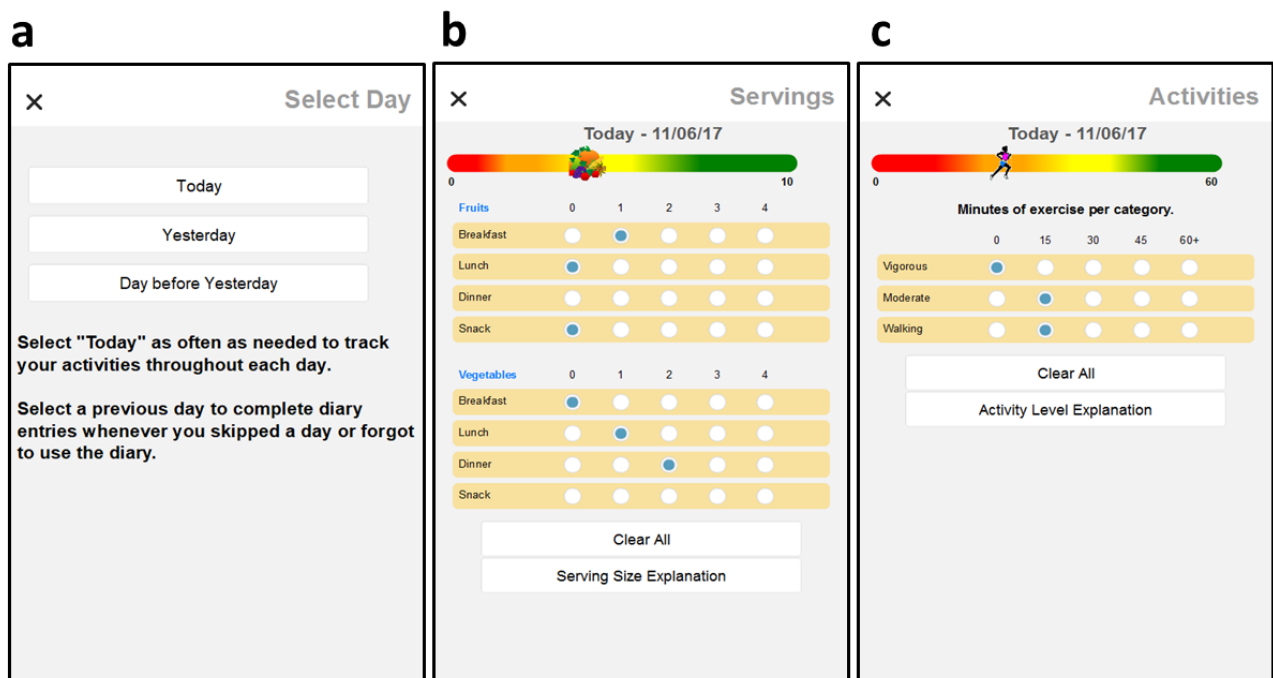
The first-time participants entered either the diary module, a series of slides would explain how to fill in the ensuing diary entries and reinforce the educational content and its relationship to the CVD risk–reduction course (Figures 5 and 6). Once reviewed, they could be disabled for the participants' future data entry, but they were always available as reference materials via a simple button tap.

Figure 5. Explanation of levels of activity effort to support quality diary entries.
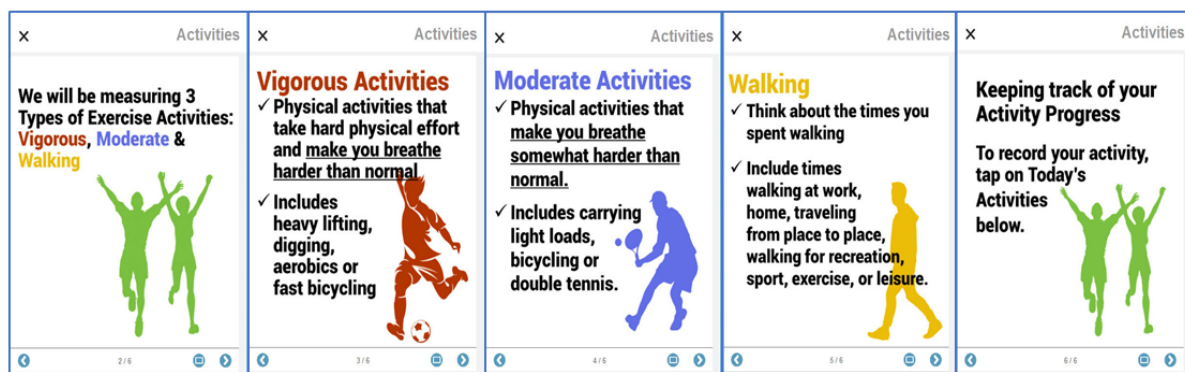
Figure 6. Explanation of fruit and vegetable serving size to support quality diary entries, along with educational materials supporting healthy eating choices and behaviors.
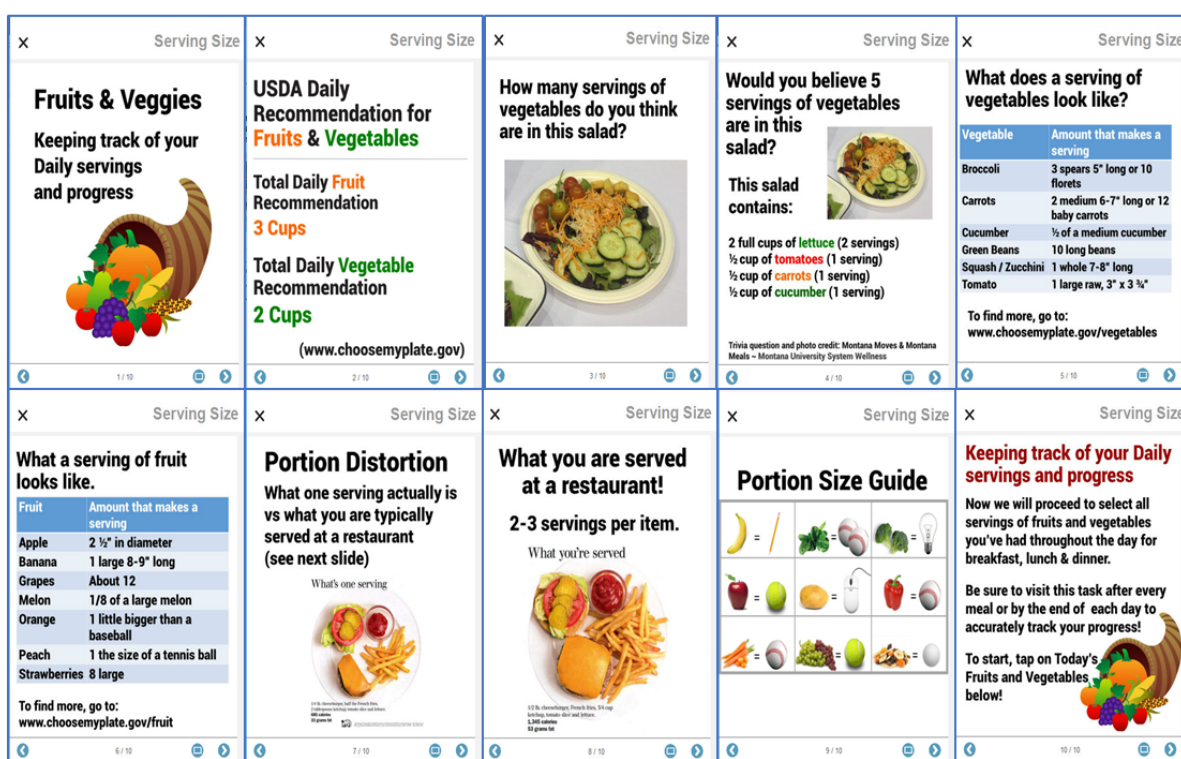
Data Storage, Privacy, and Security

Participants were provided with a unique ID to link all acquired data to that individual. They also entered a self-defined, secret, 4-digit Personal Identification Number (PIN) to prevent access by other individuals. When the app is in use, its screen will deactivate after a set period (eg, 2 min) of no interaction, and a screen cover displayed to hide current data or activity. Once deactivated, the secret 4-digit PIN must be entered to unlock the screen and allow the participant to continue.

All acquired and calculated data were stored locally in an encrypted Structured Query Language (SQL)ite database within the mobile app. Data reflecting app function, such as date-entry time stamps and the length of time spent entering diary information, were also saved to the secure in-app database. Each time the users exited a diary, they were offered the opportunity to upload the acquired data to the secure central study data server or to defer it. The app then returned the user to the home screen menu.

As data were stored locally on the device, the Rams Have Heart app operated offline, without requiring a continuous cellular or internet connection. Data were stored using a 128-bit advanced encryption standard algorithm with no PII. Whenever Wi-Fi internet access was available, they could be uploaded to a central, secure server to reduce the use of the participants' cellular data plans. They were transferred using the secure https protocol and stored in a secure SQL server database, which was only accessible to authorized persons via user ID and password authentication.

Results

Study Cohort

Upon study entry, participants completed a paper questionnaire to collect demographic data. Of the total 109, 104(95%)

completed the questionnaire; 83 (76%) were women, and 26 (24%) were men (Table 1). Most were 18 years old (76/109, 69.7%), the common age for first-year college students. Over half (56/109, 51.3%) had a normal BMI (18.5-24.9 kg/m^2; Table 1).

Table 1. Demographic characteristics of participants (N=109).

Demographics[a]	Values, n (%)
Age (years)	
17	8 (7.3)
18	76 (69.7)
19	13 (11.9)
20	8 (7.3)
≥21	4 (3.6)
Sex	
Male	26 (23.8)
Female	83 (76.1)
BMI (kg/m^2)	
Underweight (<18.5)	3 (2.7)
Normal (18.5-24.9)	56 (51.3)
Overweight (25.0-29.9)	28 (25.6)
Obese (≥30.0)	22 (20.1)

[a]Data represent cohort 2 and cohort 3 combined. Data were collected from separate questionnaires external to the mobile app.

The demographic composition of the entire participant population and those in the intervention group did not differ significantly (Table 2). Of the 55 participants in the intervention group, 27 (49%) in cohort 2 and 25 (45%) in cohort 3 used the Rams Have Heart app at least once.

Table 2. Demographic characteristics of the intervention group (N=55).

Demographics[a]	Values, n (%)	*P* value[b]
Age (years)		
17	4 (7)	.51
18	37 (67)	.62
19	4 (7)	.84
20	7 (13)	.15
≥21	3 (5)	.31
Sex		
Male	19 (35)	.08
Female	36 (65)	.92
BMI (kg/m^2)		
Underweight (<18.5)	2 (4)	.38
Normal (18.5-24.9)	29 (53)	.43
Overweight (25.0-29.9)	16 (29)	.32
Obese (≥30.0)	8 (15)	.82

[a]Data represent cohort 2 intervention and cohort 3 intervention combined. Data were collected from separate questionnaires external to the mobile app and are representative of the original pool of participants who provided diary data.

[b]*P* value refers to the significance of the difference between the entire population of cohorts and the intervention population (Table 1).

XSL•FO
RenderX

Use of Information Slides

For each module that requested manual user entry, participants could opt to view a brief introductory presentation, explaining, for example, what a serving of fruit looks like or the different levels of physical activity (ie, moderate vs vigorous). Of the 52 participants who used the app more than once, about 26 (50%)

watched it the first time they visited the fruit and vegetable intake module, and 22 (42%) watched the physical activity module (Table 3). They could also watch it again whenever they visited the module. Throughout the entire study period, 37 participants (71%) watched the fruit and vegetable intake introductory presentation, and 34 (65%) watched the physical activity introductory presentation at least once (Table 3).

Table 3. Viewership of information slides (N=52).

Module[a]	Values, n (%)
Fruit and vegetable intake	
Watched before the first entry	26 (50)
Watched at least once	37 (71)
Physical activity	
Watched before the first entry	22 (42)
Watched at least once	34 (65)

[a]Data were calculated based on participant ID and a true or false record of viewing the module in the dataset.

App Use Over the Study Period

The study ran from late September 2017 to spring 2018 for cohort 2 and September 2018 to spring 2019 for cohort 3, with the course conducted in the fall semester and follow-up data collection in the spring. Both cohorts received frequent in-class reminders to adhere to the data collection; they were encouraged to upload data during class and offered an incentive after the

fall 2017 semester ended. Cohort 2 participants gradually tapered their use of the app until around day 68 when 78% stopped, corresponding to the final exam week. Cohort 3 also displayed a sharp drop in the days just before the final exam week; participants recording fruit and vegetable diary entries fell from 16 to 5, and those recording their physical activity from 10 to 2 (Figure 7). App use did not increase afterward.

Figure 7. The number of active participants recording diary entries over the course of the study (a) active participants providing activity diary entries and (b) active participants providing diary entries of fruit and vegetables. The study day of drop-off was determined by the latest date of data entry for each module.
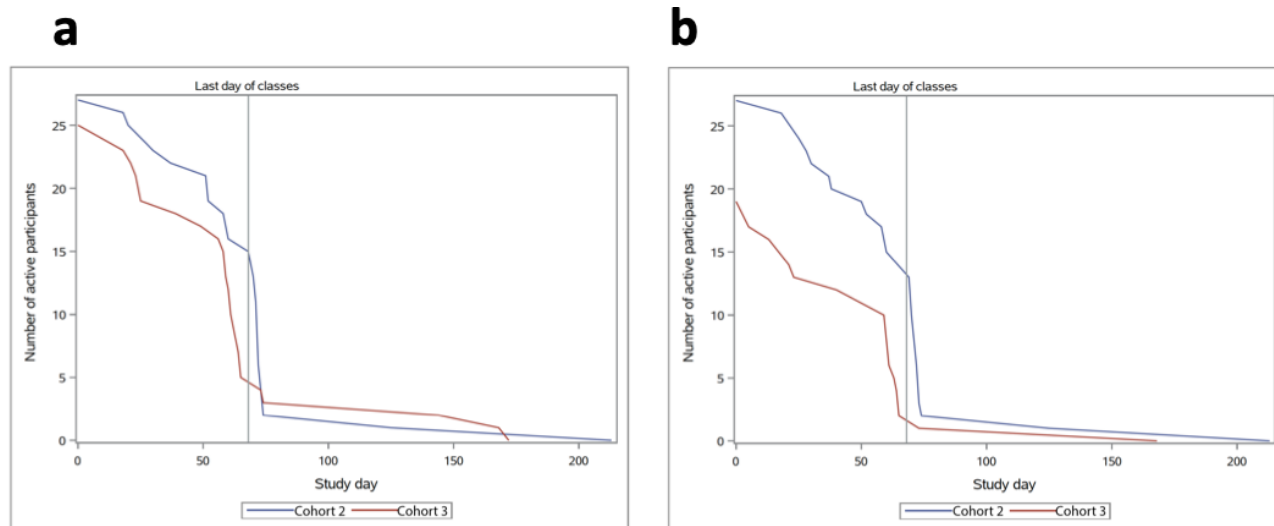
On average, the time participants took to record their fruit and vegetable intake was 16.3 seconds, and 10.1 seconds for activity (Figure 8). On average, they entered their recall data a full day

late (Figure 9). Most entered their data for both diaries in the afternoon (1-3 PM). This trend was similar for recall data (Tables 4 and 5).

Figure 8. Average time participants took to enter a diary entry. Error bars represent Upper Confidence Level of the Mean (UCLM) and Lower Confidence Level of the Mean (LCLM). The average time was calculated by pooling all entries across all participants.
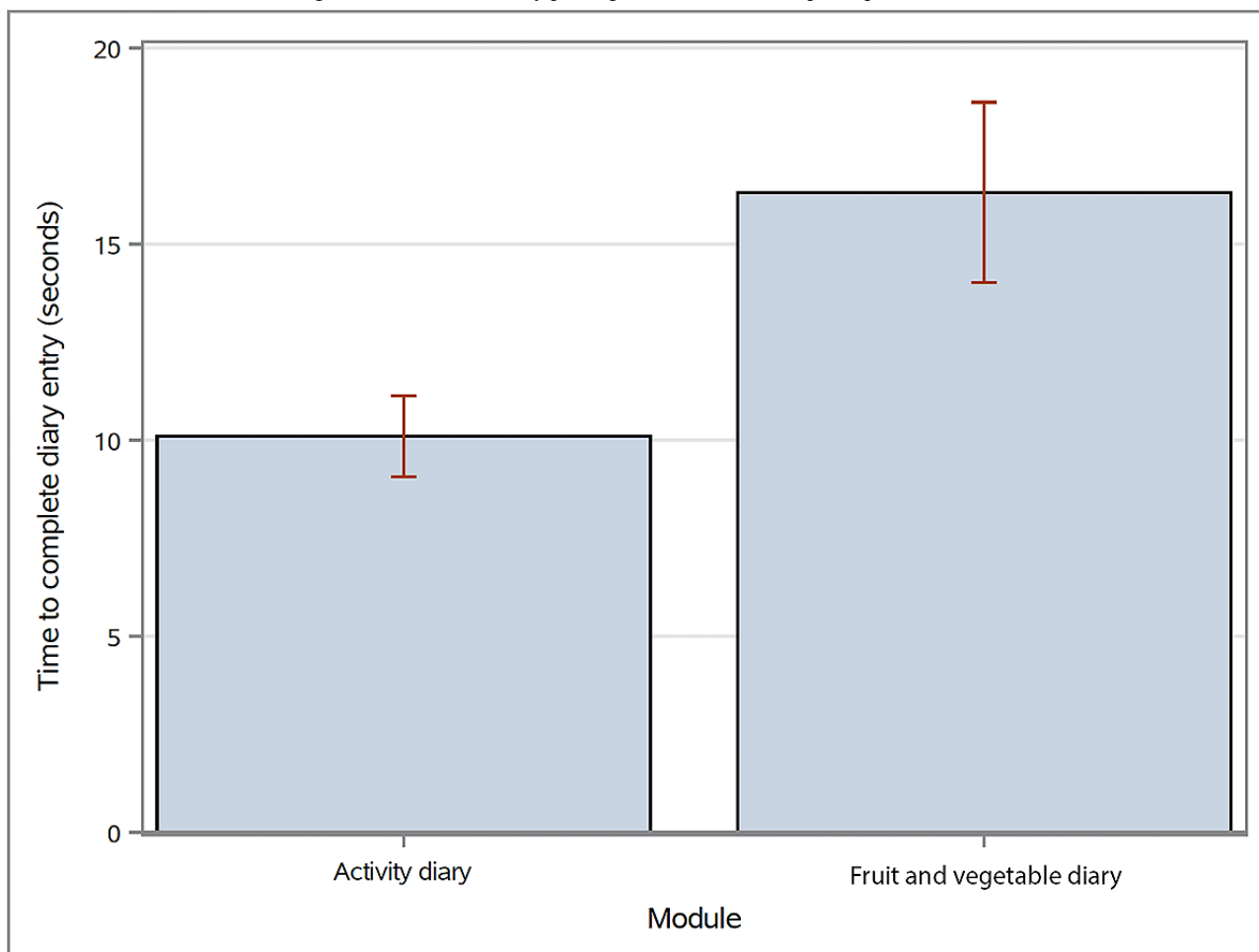
Figure 9. Average time participants took to enter a diary entry after the actual diary date. Error bars represent UCLM and LCLM. The average time was calculated by pooling all entries across all participants.
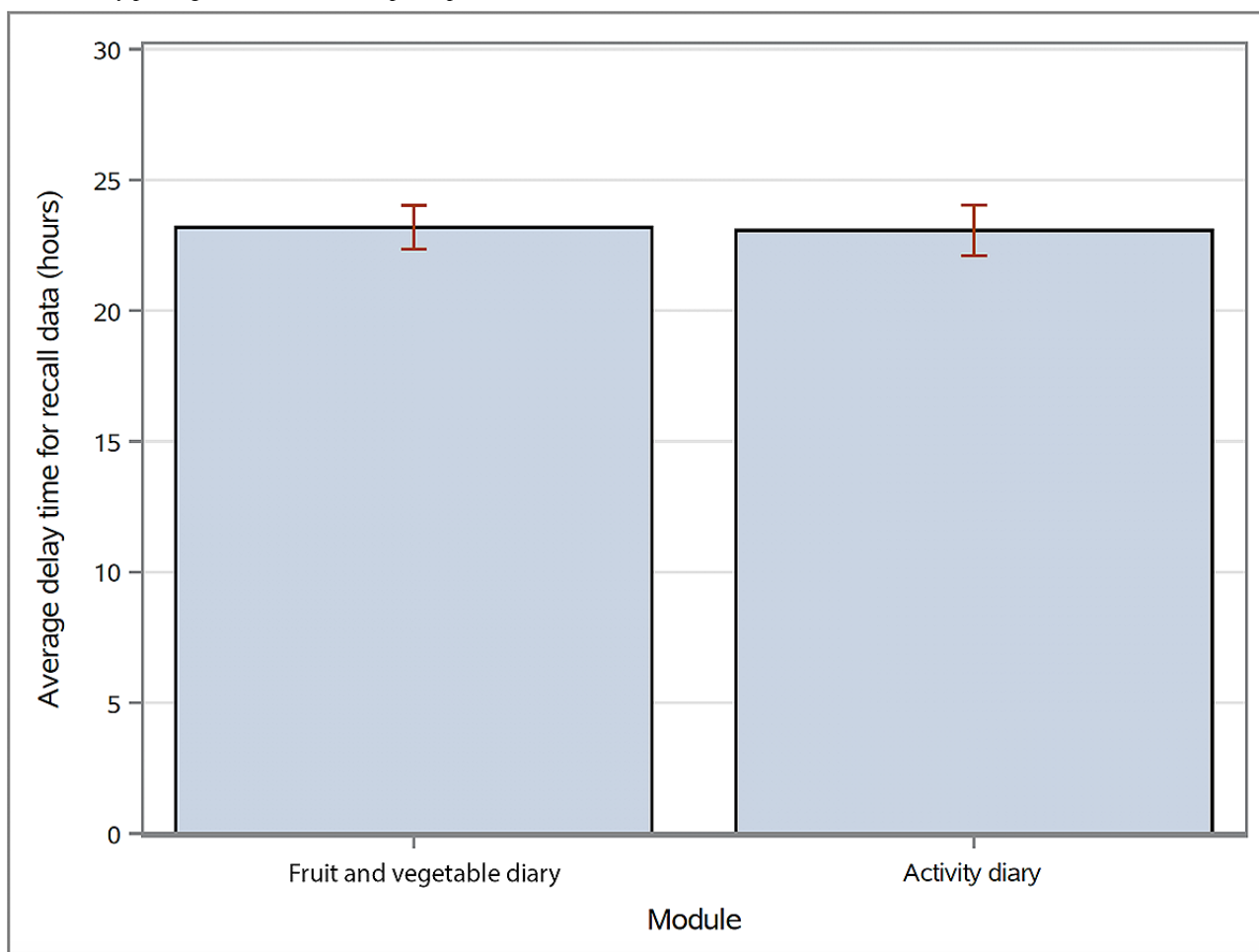
Table 4. Time of day data were recorded.

Time point	Fruit and vegetable intake n (%)[a], N=1968	Activity n (%)[a], N=1449
Midnight to 1 AM	114 (5.79)	120 (8.28)
1 AM to 2 AM	83 (4.22)	62 (4.28)
2 AM to 3 AM	120 (6.10)	95 (6.56)
3 AM to 4 AM	81 (4.12)	52 (3.59)
4 AM to 5 AM	66 (3.35)	56 (3.86)
5 AM to 6 AM	55 (2.79)	41 (2.83)
6 AM to 7 AM	19 (0.97)	17 (1.17)
7 AM to 8 AM	20 (1.02)	12 (0.83)
8 AM to 9 AM	5 (0.25)	3 (0.21)
9 AM to 10 AM	0 (0.00)	1 (0.07)
10 AM to 11 AM	1 (0.05)	1 (0.07)
11 AM to noon	11 (0.56)	10 (0.69)
Noon to 1 PM	42 (2.13)	30 (2.07)
1 PM to 2 PM	293 (14.89)	165 (11.39)
2 PM to 3 PM	269 (13.67)	208 (14.35)
3 PM to 4 PM	150 (7.62)	88 (6.07)
4 PM to 5 PM	67 (3.40)	58 (4.00)
5 PM to 6 PM	55 (2.79)	35 (2.42)
6 PM to 7 PM	86 (4.37)	60 (4.14)
7 PM to 8 PM	96 (4.88)	79 (5.45)
8 PM to 9 PM	69 (3.51)	56 (3.86)
9 PM to 10 PM	87 (4.42)	61 (4.21)
10 PM to 11 PM	108 (5.49)	81 (5.59)
11 PM to midnight	71 (3.61)	58 (4.00)

[a]Percentages indicate the proportion of entries recorded for the time period indicated. Data indicate all entries.

Table 5. Time of night data were recorded.

Time point	Fruit and vegetable intake (%)[a] (N=1024)	Activity (%)[a] (N=801)
Midnight to 1 AM	43 (4.20)	53 (6.62)
1 AM to 2 AM	39 (3.81)	30 (3.75)
2 AM to 3 AM	42 (4.10)	37 (4.62)
3 AM to 4 AM	41 (4.00)	24 (3.00)
4 AM to 5 AM	38 (3.71)	36 (4.49)
5 AM to 6 AM	35 (3.42)	29 (3.62)
6 AM to 7 AM	14 (1.37)	12 (1.50)
7 AM to 8 AM	13 (1.27)	9 (1.12)
8 AM to 9 AM	3 (0.29)	2 (0.25)
9 AM to 10 AM	0 (0.00)	1 (0.12)
10 AM to 11 AM	1 (0.10)	1 (0.12)
11 AM to noon	7 (0.68)	6 (0.75)
Noon to 1 PM	28 (2.73)	21 (2.62)
1 PM to 2 PM	178 (17.38)	111 (13.86)
2 PM to 3 PM	131 (12.79)	103 (12.86)
3 PM to 4 PM	81 (7.91)	58 (7.24)
4 PM to 5 PM	37 (2.61)	30 (3.75)
5 PM to 6 PM	33 (3.22)	22 (2.75)
6 PM to 7 PM	43 (4.20)	33 (4.12)
7 PM to 8 PM	48 (4.69)	44 (5.49)
8 PM to 9 PM	35 (3.42)	32 (4.00)
9 PM to 10 PM	47 (4.59)	35 (4.37)
10 PM to 11 PM	49 (4.79)	43 (5.37)
11 PM to midnight	38 (3.71)	29 (3.62)

[a]Percentages indicate the proportion of entries recorded for the time period indicated. Data indicate only entries entered as recall data.

Healthy Behavior Adaptation

Using the data entered on fruit and vegetable intake and activity level, we tracked individual and general trends over the study period. Note that the *n* of data (the total number of participants) was not constant and might skew the analysis of the results. Regression analysis showed that fruit and vegetable intake, as well as minutes and metabolic equivalents (METS) of activity (Figure 10) were all *P*<.001, indicating a positive correlation between days into the study and the measure studied. Activity levels in both minutes and METS decreased over time (Figure 10). Fruit and vegetable intake trended slightly upward as the study progressed (Figure 10).

Figure 10. Diary entry values recorded over the study period: (a) activity data (in minutes) over time; (b) activity data (in METS) over time; (c) fruit and vegetable servings over time. METS: metabolic estimates.
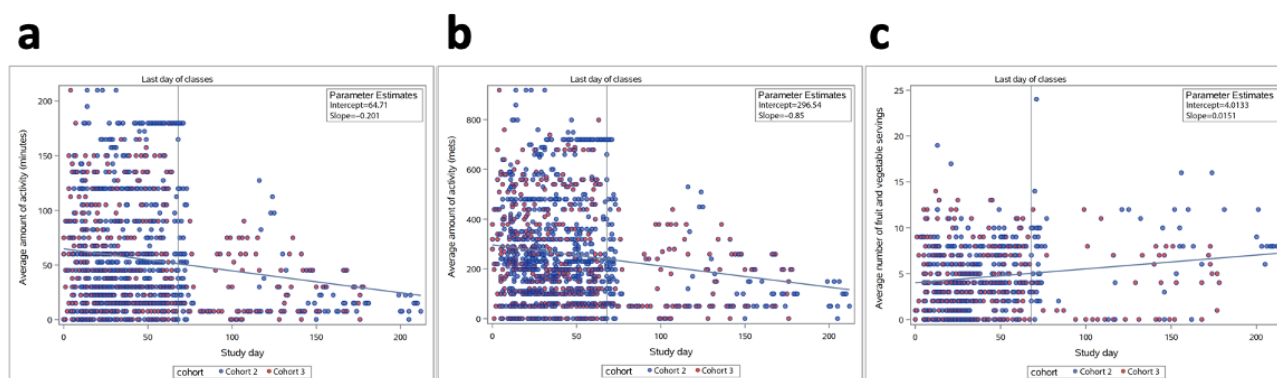
We ran an analysis of variance (ANOVA) using the Tukey two-tailed *t* test to determine whether the pooled data for cohorts 2 and 3 changed from week to week (Figure 11). The only significant finding (*P*=.03) occurred between weeks 1 and 2, with an average increase of 1.8 fruit and vegetable servings (average consumption for week 1=3.24, SD 3.50 and week 2=5.05, SD 4.15), but we found no significant change when comparing week 1 to week 3 (Table 6). Throughout the course of the study, average fruit and vegetable servings and minutes and METS of activity fluctuated (Figure 8).

Figure 11. Analysis of variance of diary entries grouped per study week: (a) activity data (in minutes) over time; (b) activity data (in METS) over time; (c) fruit and vegetable servings over time. METS: metabolic estimates.
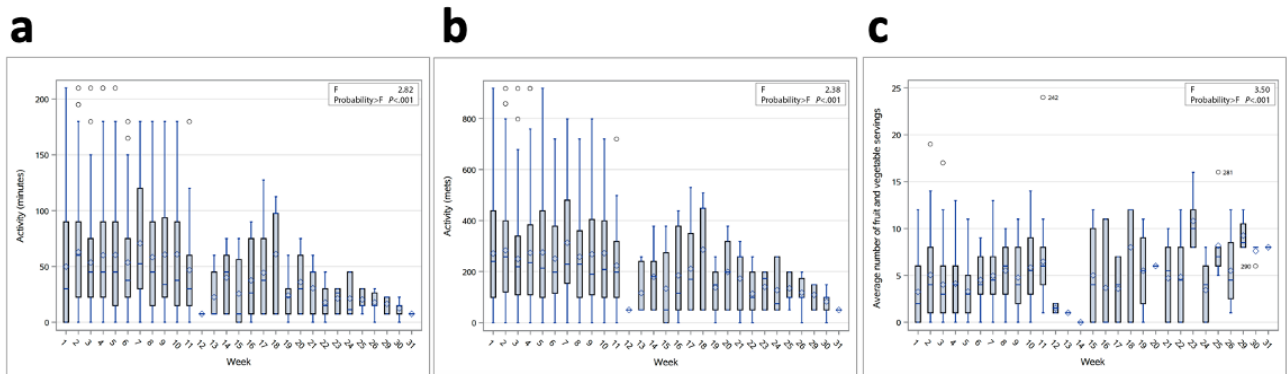
Table 6. Trend of fruit and vegetable and activity results over time.

Study week[a]	Fruit and vegetable	P value	Activity (min)	P value	Activity (metabolic equivalents)	P value
2 and 1	1.8089	.03	13.074	.94	12.06	>.99
3 and 2	−1.0330	.97	−9.476	>.99	−33.44	>.99
4 and 3	0.1032	>.99	6.675	>.99	24.01	>.99
5 and 4	−0.8430	>.99	0.188	>.99	1.38	>.99
6 and 5	1.2351	.95	−6.796	>.99	−24.25	>.99
7 and 6	0.4428	>.99	17.207	.63	62.44	.84
8 and 7	0.5558	>.99	−12.434	.98	−55.25	.95
9 and 8	−0.7438	>.99	2.348	>.99	10.20	>.99
10 and 9	1.0697	>.99	0.315	>.99	4.21	>.99
11 and 10	0.6354	>.99	−14.137	>.99	−48.68	>.99
12 and 11	−4.9688	.98	−39.324	>.99	−174.16	>.99
13 and 12	−0.5000	>.99	15.000	>.99	66.50	>.99
14 and 13	−1.0000	>.99	17.500	>.99	63.50	>.99
15 and 14	5.0000	>.99	−14.375	>.99	−46.13	>.99
16 and 15	−1.3333	>.99	11.875	>.99	51.73	>.99
17 and 16	−0.0952	>.99	6.818	>.99	25.85	>.99
18 and 17	4.4286	.99	16.753	>.99	75.12	>.99
19 and 18	−2.5000	>.99	−37.071	>.99	−148.47	>.99
20 and 19	0.5000	>.99	12.000	>.99	61.00	>.99
21 and 20	−1.3000	>.99	−5.500	>.99	−25.90	>.99
22 and 21	0.1333	>.99	−12.773	>.99	−59.25	>.99
23 and 22	5.9667	.50	3.523	>.99	26.80	>.99
24 and 23	−7.3714	.07	0.000	>.99	−13.00	>.99
25 and 24	4.7381	.80	−0.893	>.99	7.11	>.99
26 and 25	−2.6667	>.99	−2.545	>.99	−16.79	>.99
27 and 26	__[b]	—	—	—	—	—
28 and 27	—	—	—	—	—	—
29 and 28	—	—	—	—	—	—
30 and 29	−1.6500	>.99	−115.217	>.99	−26.40	>.99
31 and 30	0.4000	>.99	−5.000	>.99	−33.00	>.99
31 and 1	4.7624	>.99	−42.361	>.99	−221.49	>.99

[a]Numbers represent the difference between means calculated using ANOVA with the Tukey studentized range (honestly significant difference) test, pooling data from cohort 2 and cohort 3.

[b]Data not available.

Discussion

With the rise of mobile phone technology, several apps have been developed to tackle health problems. Many enable self-monitoring to achieve health-related goals and are promoted by medical practitioners to improve patient health [21,22,26]. Although Rams Have Heart featured modules similar to those used by other health-related mobile apps—for example, one counting daily fruit and vegetable servings—it was designed with a research question in mind. We did not assess whether participants were using any other apps for dietary or exercise improvement but based on the low adherence to our protocol, even as part of a college course, we deemed it unlikely.

Rams Have Heart aimed to monitor the adaptation of concepts from the CVD-oriented college course, not to actively support behavioral change. The app might have done more to support participant adherence, but the larger objective was to design a stand-alone course that would improve the cardiovascular health of students with or without this or any other app. Therefore, the app's functionality is centered on its ability to collect data for

further analysis of its usability and methods for improvement but, more importantly, achievement of course goals.

Studies of participant retention have suggested using mobile devices to collect data rather than burdening participants with phone calls or in-person appointments [27]. Flexibility is a key to promoting user engagement but can also lead to forgetfulness [28]. After the cohort 1 pilot study, increasing in-class reminders to use the app and adding incentives resulted in the longer adherence found in cohorts 2 and 3. Retention methods specific to the cohort are sometimes necessary in longitudinal studies [29]. In years 2 and 3, providing in-class reminders and time to upload data increased participant retention and the number of entries. Studies reviewing health research using mobile apps have shown that the average retention period is around 25.62 (SD 18.41) weeks, and participants who stick with the app over the long term typically have a strong incentive, such as a serious health problem [23]. Programmatic incentives include virtual rewards, such as accomplishment badges [30] and the ability to share on social media accounts [31-33], as well as more tangible rewards [34]. In spring 2018, our study introduced a monetary incentive: an Amazon gift card to the student who had the highest number of diary entries at the end of each month for the period after the course ended to the 6-month follow-up. However, only 1 student continued to provide entries. Introducing the incentive late in the process was not very effective.

As part of the diary modules, students could first watch informational slides. Although most skipped them when they first recorded a diary entry, 71% watched the fruit and vegetable servings presentation at least once, and 65% watched the activity presentation at least once. The physical activity module did not require much explanation, asking merely for activity time in minutes. College students are less likely to understand serving sizes and numbers of servings [35-37], which may account for the higher viewership of this presentation.

On average, fruit and vegetable serving entries took 5 seconds longer than the activity diary entries (Figure 8), possibly because they involved more rows—fruits and/or vegetables consumed at breakfast, lunch, dinner, and as snacks—or 8 rows as compared with only 3 rows for activity (Figure 3). However, this difference did not affect the average time of recall data entry (Figure 9) nor account for the difference in the number of participants inputting diary entries over the study period for either the fruit and vegetable serving diary or the activity diary (Figure 7).

All of the diary data were acquired using the app either in near real time or the app-supported recall within 1 or 2 days. No data were acquired via nondigital means, such as a paper diary. Although the student participants often failed to make daily diary entries, no attempt was made to estimate and *fill in* missing entries.

We saw no significant improvement in average fruit and vegetable servings or activity in minutes or METS when comparing the last week of the study to the first week (Table 6). Mobile apps with self-reported health assessments (eg, calorie counting and food diaries) rely on user input to provide useful feedback on improvements or when to seek professional help. When these measures are used for research data collection, accuracy and a large data pool are required [38]. In this study, most participants did not continue to record their food intake and activity levels because other life events, such as the end of the semester and the start of final exams, took precedence [39]. Other retention methods, such as text messages, might have improved the frequency of entries and reduced the number of dropouts.

Conclusions

Rams Have Heart was developed to enhance fruit and vegetable intake and physical activity for a student demographic susceptible to obesity, heart disease, and type 2 diabetes. As part of the study, we conducted an analysis of the functionality and usage of a health-related mobile app. Although Rams Have Heart provides privacy and flexibility for user participation in a research study, it did not improve participant retention or user outcomes. This finding calls for further evaluation to determine more effective retention methods.

Conflicts of Interest

None declared.

References

1. Common Sense Media. 2019. The Common Sense census: media use by tweens and teens URL: https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-tweens-and-teens-2019 [accessed 2020-03-07]
2. McDonald J. WARC. 2019. Almost three quarters of internet users will be mobile-only by 2025 URL: https://www.warc.com/content/paywall/article/warc-datapoints/almost_three_quarters_of_internet_users_will_be_mobileonly_by_2025/124845 [accessed 2020-03-08]

XSL•FO

RenderX

3.   Rehalia A, Prasad S. Global harnessing of advanced mHealth for community mobilization. Mhealth 2016;2:7 [FREE Full text] [doi: 10.21037/mhealth.2016.03.02] [Medline: 28293585]

4.   Finkelstein J, Knight A, Marinopoulos S, Gibbons MC, Berger Z, Aboumatar H, et al. Enabling patient-centered care through health information technology. Evid Rep Technol Assess (Full Rep) 2012 Jun(206):1-1531. [Medline: 24422882]

5.   Statista. 2017. App stores: number of apps in leading app stores URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2020-03-07]

6.   Årsand E, Frøisland DH, Skrøvseth SO, Chomutare T, Tatara N, Hartvigsen G, et al. Mobile health applications to assist patients with diabetes: lessons learned and design implications. J Diabetes Sci Technol 2012 Sep 01;6(5):1197-1206 [FREE Full text] [doi: 10.1177/193229681200600525] [Medline: 23063047]

7.   Mertz L. Ultrasound? Fetal monitoring? Spectrometer? There's an app for that!: biomedical smart phone apps are taking healthcare by storm. IEEE Pulse 2012 Mar;3(2):16-21. [doi: 10.1109/MPUL.2011.2181019] [Medline: 22481741]

8.   Centers for Disease Control and Prevention. 2015. Healthy People 2010 URL: http://www.cdc.gov/nchs/healthy_people/hp2010.htm [accessed 2020-05-05]

9.   Nelson MC, Story M, Larson NI, Neumark-Sztainer D, Lytle LA. Emerging adulthood and college-aged youth: an overlooked age for weight-related behavior change. Obesity (Silver Spring) 2008 Oct;16(10):2205-2211 [FREE Full text] [doi: 10.1038/oby.2008.365] [Medline: 18719665]

10. Krebs-Smith S, Guenther P, Subar A, Kirkpatrick S, Dodd K. Americans do not meet federal dietary recommendations. J Nutr 2010 Oct;140(10):1832-1838 [FREE Full text] [doi: 10.3945/jn.110.124826] [Medline: 20702750]

11. Small M, Bailey-Davis L, Morgan N, Maggs J. Changes in eating and physical activity behaviors across seven semesters of college: living on or off campus matters. Health Educ Behav 2013 Aug;40(4):435-441 [FREE Full text] [doi: 10.1177/1090198112467801] [Medline: 23232092]

12. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]

13. Onvani S, Haghighatdoost F, Surkan PJ, Larijani B, Azadbakht L. Adherence to the healthy eating index and alternative healthy eating index dietary patterns and mortality from all causes, cardiovascular disease and cancer: a meta-analysis of observational studies. J Hum Nutr Diet 2017 Apr;30(2):216-226. [doi: 10.1111/jhn.12415] [Medline: 27620213]

14. Li W, Wang D, Wu C, Shi O, Zhou Y, Lu Z. The effect of body mass index and physical activity on hypertension among Chinese middle-aged and older population. Sci Rep 2017 Aug 31;7(1):10256. [doi: 10.1038/s41598-017-11037-y] [Medline: 28860562]

15. Herbert R. ScholarWorks: Walden University Research. 2015. Race/Ethnicity, Diet, and Physical Activity Behavior Among College Students URL: https://scholarworks.waldenu.edu/dissertations/1647/ [accessed 2020-05-14]

16. Arts J, Fernandez M, Lofgren I. Coronary heart disease risk factors in college students. Adv Nutr 2014 Mar 1;5(2):177-187 [FREE Full text] [doi: 10.3945/an.113.005447] [Medline: 24618758]

17. der Ananian C, Winham D, Thompson S, Tisue M. Perceptions of heart-healthy behaviors among African American adults: a mixed methods study. Int J Environ Res Public Health 2018 Nov 1;15(11):2433 [FREE Full text] [doi: 10.3390/ijerph15112433] [Medline: 30388803]

18. Bancks MP, Kershaw K, Carson AP, Gordon-Larsen P, Schreiner PJ, Carnethon MR. Association of modifiable risk factors in young adulthood with racial disparity in incident type 2 diabetes during middle adulthood. J Am Med Assoc 2017 Dec 26;318(24):2457-2465 [FREE Full text] [doi: 10.1001/jama.2017.19546] [Medline: 29279935]

19. Durant NH, Joseph RP, Cherrington A, Cuffee Y, Knight B, Lewis D, et al. Recommendations for a culturally relevant internet-based tool to promote physical activity among overweight young African American women, Alabama, 2010-2011. Prev Chronic Dis 2014 Jan 16;11:130169 [FREE Full text] [doi: 10.5888/pcd11.130169] [Medline: 24433625]

20. Winston-Salem State University. About WSSU URL: https://www.wssu.edu/about/index.html [accessed 2020-05-05]

21. Adu MD, Malabu UH, Callander EJ, Malau-Aduli AE, Malau-Aduli BS. Considerations for the development of mobile phone apps to support diabetes self-management: systematic review. JMIR Mhealth Uhealth 2018 Jun 21;6(6):e10115 [FREE Full text] [doi: 10.2196/10115] [Medline: 29929949]

22. Ernsting C, Dombrowski SU, Oedekoven M, Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 5;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]

23. 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]

24. Clough BA, Casey LM. The smart therapist: a look to the future of smartphones and mhealth technologies in psychotherapy. Prof Psychol Res Pr 2015 Jun;46(3):147-153. [doi: 10.1037/pro0000011]

25. Kelli HM, Witbrodt B, Shah A. The future of mobile health applications and devices in cardiovascular health. Euro Med J Innov 2017 Jan;2017:92-97 [FREE Full text] [Medline: 28191545]

26. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]

27. van Gelder MM, Vlenterie R, IntHout J, Engelen LJ, Vrieling A, van de Belt TH. Most response-inducing strategies do not increase participation in observational studies: a systematic review and meta-analysis. J Clin Epidemiol 2018 Jul;99:1-13. [doi: 10.1016/j.jclinepi.2018.02.019] [Medline: 29518475]

28. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]

29. Abshire M, Dinglas VD, Cajita MI, Eakin MN, Needham DM, Himmelfarb CD. Participant retention practices in longitudinal clinical research studies with high retention rates. BMC Med Res Methodol 2017 Feb 20;17(1):30 [FREE Full text] [doi: 10.1186/s12874-017-0310-z] [Medline: 28219336]

30. Buchem I, Merceron A, Kreutel J, Haesner M, Steinert A. Gamification Designs in Wearable Enhanced Learning for Healthy Ageing. In: Proceedings of the International Conference on Interactive Mobile Communication Technologies and Learning. 2015 Presented at: IMCL'15; November 19-20, 2015; Thessaloniki, Greece p. 19-20. [doi: 10.1109/imctl.2015.7359545]

31. Naslund JA, Aschbrenner KA, Barre LK, Bartels SJ. Feasibility of popular m-health technologies for activity tracking among individuals with serious mental illness. Telemed J E Health 2015 Mar;21(3):213-216 [FREE Full text] [doi: 10.1089/tmj.2014.0105] [Medline: 25536190]

32. 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]

33. Ho K, Newton L, Boothe A, Novak-Lauscher H. Mobile digital access to a web-enhanced network (mDAWN): assessing the feasibility of mobile health tools for self-management of type-2 diabetes. AMIA Annu Symp Proc 2015;2015:621-629 [FREE Full text] [Medline: 26958197]

34. Cushing A, Manice M, Ting A, Parides M. Feasibility of a novel mhealth management system to capture and improve medication adherence among adolescents with asthma. Patient Prefer Adherence 2016;10:2271-2275 [FREE Full text] [doi: 10.2147/PPA.S115713] [Medline: 27853357]

35. Bryant R, Dundes L. Portion distortion: a study of college students. J Consum Aff 2005;39(2):399-408. [doi: 10.1111/j.1745-6606.2005.00021.x]

36. Ryan E. University of Wisconsin-Stout. 2008. College Students' Knowledge of Cereal Serving Sizes URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.591.7096&rep=rep1&type=pdf [accessed 2020-05-14]

37. Lambert M, Chivers P, Farringdon F. In their own words: a qualitative study exploring influences on the food choices of university students. Health Promot J Austr 2019 Jan;30(1):66-75. [doi: 10.1002/hpja.180] [Medline: 29892990]

38. Munos B, Baker PC, Bot BM, Crouthamel M, de Vries G, Ferguson I, et al. Mobile health: the power of wearables, sensors, and apps to transform clinical trials. Ann N Y Acad Sci 2016 Jul;1375(1):3-18. [doi: 10.1111/nyas.13117] [Medline: 27384501]

39. Deliens T, Deforche B, de Bourdeaudhuij I, Clarys P. Determinants of physical activity and sedentary behaviour in university students: a qualitative study using focus group discussions. BMC Public Health 2015 Feb 28;15:201 [FREE Full text] [doi: 10.1186/s12889-015-1553-4] [Medline: 25881120]

Abbreviations

AAs: African Americans
AIR: Adobe integrated runtime
ANOVA: analysis of variance
BP: blood pressure
CVD: cardiovascular disease
HBCU: historically black college and university
ID: identifier
IPAQ: International Physical Activity Questionnaire
IRB: institutional review board
METS: metabolic equivalents
mHealth: mobile health
PHIT: Personal Health Informatics and Intervention Toolkit
PII: personally identifiable information
PIN: Personal Identification Number
SQL: Structured Query Language

Original Paper

User Perception of a Smartphone App to Promote Physical Activity Through Active Transportation: Inductive Qualitative Content Analysis Within the Smart City Active Mobile Phone Intervention (SCAMPI) Study

Anna-Karin Lindqvist[1*], MSc, PhD; Stina Rutberg[1*], MSc, PhD; Emmie Söderström[2], MSc; Anna Ek[3], PhD; Christina Alexandrou[2,3], MSc; Ralph Maddison[4], PhD; Marie Löf[2,5], PhD

[1]Division of Health, Medicine and Rehabilitation, Department of Health Sciences, Luleå University of Technology, Luleå, Sweden

[2]Department of Health, Medical and Caring Sciences, Linköping University, Linköping, Sweden

[3]Department of Clinical Science Intervention and Technology, Karolinska Institutet, Huddinge, Sweden

[4]Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Melbourne, Australia

[5]Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden

[*]these authors contributed equally

Corresponding Author:
Anna-Karin Lindqvist, MSc, PhD
Division of Health, Medicine and Rehabilitation
Department of Health Sciences
Luleå University of Technology
Luleå, 971 87
Sweden
Phone: 46 0725390660
Email: anna-karin.lindqvist@ltu.se

Abstract

Background: Physical inactivity is globally recognized as a major risk factor for morbidity, particularly the incidence of noncommunicable diseases. Increasing physical activity (PA) is therefore a public health priority. Engaging in active transportation (AT) is a viable approach for promoting daily PA levels. Mobile health interventions enable the promotion of AT to a larger population. The Smart City Active Mobile Phone Intervention (SCAMPI) study was a randomized controlled trial designed to evaluate the ability of a behavior change program delivered via a smartphone app to motivate participants to increase their PA by engaging in AT.

Objective: This qualitative study aims to examine the acceptance and user experience of the app promoting AT that was used in the SCAMPI trial (the TRavelVU Plus app).

Methods: A total of 17 residents of Stockholm County (13 women; age range 25-61 years) who completed the 3-month app-based behavioral change program (delivered through the TRavelVU Plus app) in the SCAMPI randomized controlled trial during 2018 agreed to participate in a semistructured telephone-based interview. These participants were well representative of the whole intervention group (n=127) in terms of baseline characteristics such as age, sex, and area of residence. The interviews were audiorecorded, transcribed verbatim, and analyzed using an inductive qualitative content analysis.

Results: The content analysis revealed 2 themes and 4 subcategories. The first theme, "main motivators: monitoring and messages," highlighted that monitoring AT and being able to set weekly goals using the app were the primary motivators reported by study participants. The second theme, "acceptable but modifiable," reflects that the app was well accepted and effectively encouraged many participants to use more AT. Nevertheless, there were functions in the app that require modification. For example, while the semiautomated travel tracking feature was appreciated, participants found it time-consuming and unreliable at times.

Conclusions: This study contributes novel insight into adults' experiences of using a mobile app to promote the use of AT. The results showed that the app was well accepted and that self-monitoring and goal setting were the main motivators to engage in more AT. The semiautomated tracking of AT was appreciated; however, it was also reported to be energy- and time-consuming when it failed to work. Thus, this feature should be improved going forward.

XSL·FO
RenderX

Introduction

Over 30% of the global population is physically inactive [1]. This is alarming, as physical inactivity is one of the greatest risk factors for mortality worldwide [2] and contributes to the development of most noncommunicable diseases [3]. Therefore, promoting physical activity (PA) is a public health priority [4]. In Sweden, there has been a pronounced decline in mean cardiorespiratory fitness in adults between 1995 and 2017, and the proportion of people with low cardiorespiratory fitness has almost doubled [5]. One factor potentially associated with this fitness loss is preferred mode of transportation; over the past two decades, the use of active transportation (AT) has decreased while motorized transportation use has increased [6]. Municipalities over the world have tried different strategies, including banning cars from their city centers and redesigning urban streets in order to increase walking and cycling to ensure individuals can transport themselves actively and become healthier [7]. As health benefits can occur even with modest improvements in PA levels [8], engaging in AT is an effective, impactful way to increase the total amount of PA [3,9]. Moreover, AT has been inversely associated with noncommunicable diseases, including obesity, type 2 diabetes, cardiovascular disease, and falls [10-12]. However, changing people's behaviors toward AT is especially challenging because external factors influence travel mode, such as convenience and weather conditions [13]. This is particularly true in Sweden and other Nordic countries, where winters are long, cold, dark, and snowy [14]. Thus, new and innovative approaches to changing the attitudes and behaviors regarding AT are needed. In this project, AT was defined as walking or biking parts of or the entire way to a destination.

The extensive development of information and communication technologies characteristic of modern society enables interventions to reach large populations [15]. The benefits of a mobile health (mHealth)–based intervention are that it can be delivered at scale and at any time, can be tailored to meet peoples´ needs, and is more cost-effective than face-to-face counseling [16]. Specifically, mHealth programs have already proven to be useful in promoting PA and weight loss [17,18].

Despite their utility, it is important to understand how people perceive and use mHealth interventions in order to improving the effectiveness, engagement, and acceptance of existing and future mHealth apps. Increasingly, studies are now including an assessment of user perceptions. For example, Dennison et al explored user perceptions of mobile apps for behavior change among healthy young adults and showed that users expected apps to be accurate, legitimate, secure, and able to record and track behavior and goals [19]. Participants also expected health

apps to require minimal effort to operate [19]. Another study examined the design and content elements of health apps that facilitate or impede usage from the user perspective [20]. Their findings are largely similar to Dennison et al [19], with the addition of participants requesting individually tailored information to meet their needs. Additionally, two studies described user perceptions of mobile apps in the health care sector, one for medication management among older adults [21] and one regarding the acceptability of an app for diabetes self-management [22]. Through assessing users' perceptions of mHealth apps, it is possible to improve the design of future mHealth interventions that are effective and accepted by end users [20]. To date, no study that we are aware of has explored user perceptions of an app aiming to promote AT.

Recently, the main results from the Smart City Active Mobile Phone Intervention (SCAMPI) randomized controlled trial (ClinicalTrials.gov NCT03086837) [23] were published [24]. The SCAMPI trial aimed to evaluate the effectiveness of a 3-month behavior change program delivered through a mobile phone app to promote AT (TRavelVU Plus app) on moderate-to-vigorous PA. The results showed that there was a moderate effect of the intervention on moderate-to-vigorous PA at 6 months after baseline, which corresponded to around 30% of the weekly recommendation [24]. In this qualitative study, we examine the acceptance and user experience of the app used to promote AT in the SCAMPI trial (ie, the TRavelVU Plus app).

Methods

Study Design and Recruitment

The SCAMPI trial was conducted between September 2017 and September 2018 in Stockholm County, Sweden (N=254) [23]. Full details on the trial's design and methods, as well as the main intervention results, are published elsewhere [23,24]. Briefly, all participants downloaded the TRavelVU app and used it for baseline and follow-up measurements of their AT. The TRavelVU app automatically registered participants' use of transportation (time, duration, and mode) using GPS coordinates. In the evening, participants were asked to review and, if necessary, manually revise travel and locations in the app (ie, the recordings of travel were semiautomated). Travel behaviors were presented graphically for daily, weekly, and monthly AT use. After baseline measurements, participants were randomized into either the control (n=127) or intervention group (n=127). Participants allocated to the intervention group received the TRavelVU Plus app, which included a behavior change program grounded in Bandura's social cognitive theory and the principles of social ecology [25,26]. The behavior

change program comprised a weekly goal-setting function and feedback on the weekly goal to inspire additional AT. Messages also provided encouragement to engage in AT and offered strategies to do so. Detailed descriptions of the TRavelVU and the TRavelVU Plus apps are available in the study protocol [23].

After completing the intervention, participants in the intervention group were contacted by email and asked to participate in a semistructured telephone-based interview regarding their experiences using the app. Participants received information about the purpose of the interview and were informed that participation was voluntary. In total, 17 participants agreed to participate, and a telephone interview was scheduled with each. The 17 participants interviewed were representative of the intervention group, with the majority of them being women (13/17, 76%), having a mean age of 51 (SD 11) years, having a university degree (10/17, 59%), and spending on average 59 (SD 21) minutes per day in moderate-to-vigorous PA. None of these results were statistically different (*P* values ranging from .16 to .89) from the corresponding results for the entire intervention group (n=127), of which the majority of participants were women (78/127, 61.4%), with a mean age of 47 (SD 11), a university degree (80/127, 63.0%), and a mean moderate-to-vigorous PA of 60 (SD 28) minutes per day. Furthermore, the entire intervention group and the participants in the interviews lived in similar residential areas (urban) of Stockholm County. At baseline, there was a wide range in their use of active transportation, as reported previously [24], and no difference between the entire group and the participants in this qualitative study (intervention group: n=127; mean AT 58 min/day, SD 29; qualitative study: n=17; mean AT 55 min/day, SD 22; *P*=.68). Informed consent was obtained at the start of the interview. The study and the consent procedures were approved by the regional ethical board in Stockholm (January 11, 2017: dnr 2016/2403-31 and March 22, 2018: dnr 2018/615-32).

Data Collection

To explore the study participants´ perceptions of the TRavelVU Plus app, a semistructured interview guide was developed. This guide included questions regarding the app's layout and function, as well as the feasibility and acceptability of using the app. We explicitly asked whether participants had experienced problems with the app and if they had suggestions for improvements. To maximize the amount and depth of the data collected in the interview, follow-up questions were posed, such as "Could you tell me more?" The interviews lasted between 16 and 30 minutes, were audiorecorded, and were transcribed verbatim. ES, a research assistant, conducted all interviews.

Data Analysis

The interviews were analyzed using an inductive qualitative content analysis inspired by Graneheim and Lundman [27]. Initially, the transcribed interviews were carefully read several times by A-KL and SR for greater understanding of the material. The text was then divided into meaning units guided by the aim of the study. The meaning units were coded close to the original text. The codes were compared, contrasted, and sorted into preliminary subcategories and themes. After rereading the interviews, to verify that no important information had been omitted, 2 themes containing 2 subcategories were created. A-KL and SR have extensive experience in qualitative research, and they were primarily responsible for the analysis; however, all authors contributed to the final result. The authors have different backgrounds and areas of expertise, such as physiotherapy (A-KL, SR), physical activity and public health nutrition (ES, CA, ML), and clinical nutrition (AE). This variety of experiences and perspectives increases the likelihood of providing a more nuanced interpretation of the results [28]. In accordance with Graneheim and Lundman [27], quotations were included to strengthen the credibility of the study.

Results

The analysis resulted in 2 themes with 2 subcategories each to describe participants' perceptions of using the TRavelVU Plus app.

Main Motivators: Monitoring and Messages

Monitoring AT and setting weekly goals in the app were perceived to be the main motivators to use AT, for example, by choosing transportation options other than a car. A majority of participants also reported being motivated by the daily messages, which served as reminders to engage in AT. However, at the time of the interview, they could not always remember the actual content of the messages.

Going for Goals

Two of the most appreciated features of the app were the self-monitoring function and the ability to set a weekly AT goal. Participants highlighted how goal setting served as a motivator for behavioral change or for maintaining their current level of AT. Some participants set goals that they could easily reach and, therefore, did not become more active than before. In contrast, others challenged themselves and strived to constantly improve their behavior. When the set goal became within reach, it especially triggered some participants to use more AT to ensure that the goal was achieved. For participants who had not set challenging enough goals or fell too far behind with no chance of recovering, the goal-setting function was not perceived to be motivating. A feedback message was sent to participants who had not reached their goal by the end of the week. Some participants reported that this feedback on the set goal was provided too late and suggested that, instead, messages should be sent earlier in the week; then the messages would motivate participants to reach their goal, as it was seen by some participants as a competition against themselves. In addition, some participants suggested including a platform for connecting with other app users, sharing their progress, and maybe even competing against each other.

> *The weekly goal setting affected me a lot since it made me more physically active and gave me a sense of pride. I made sure to reach my goal, especially if I was very close to achieving it.* [Woman, 45 years old, interview 6]

Receiving feedback and getting graphical statistics on their transportation mode increased participants' awareness of their behavior and inspired several participants to make better

decisions and use AT more often. For example, participants reported that they checked the map of the route they had walked during the day and reflected on possible longer or shorter routes. Additionally, the graphics were used to see whether they could have saved time or somehow been more physically active. Even though some reported that the app made them more active, especially through promoting AT despite poor weather or other discouraging circumstances, many said that the app had little or no effect on their PA level. These participants were usually highly active prior to using the app or affected by external factors, such as large quantities of snow or slippery roads during the intervention, long transportation distances that obviate the use of AT, or practical things, such as dropping children off at preschool.

> *I think it is good to become more aware of my modes of transportation … If I walk about 10,000 steps a day, then I am satisfied, but if I also see that I take the car for 5 minutes, 10 times a day, then I realize that I could have walked even more. It provides more insights about how you travel during a day.* [Woman, 25 years old, interview 3]

Messages: Encouraging or Just a Frustration?

The app posted encouraging messages, to which participants had varying responses. Some expressed that messages affected their transportation choices and, therefore, found them useful. Participants with intentions to change their behavior also viewed the messages as helpful and effective reminders to engage in more AT, for example, by getting off the bus earlier and walking the remaining part of the way to their destination. They found that the messages prevented them from getting stuck in habits and forgetting about other alternatives.

In contrast, other participants reported that the messages made them feel bad for not using AT or not being physically active. Some participants found the messages to be somewhat annoying, especially when messages were similar or repeated. Others appreciated that the messages had a positive tone. Practical messages, for example, with information on how to take care of your bicycle, were brought up as helpful. Participants also said that it was motivating to receive facts on the health-related aspects of PA. One participant was especially inspired by a message about sedentary behavior that included information on the importance of breaking up sedentary portions of the day and how one could spread PA throughout the day rather than doing it all at once (eg, in the evening).

> *One of the most useful things was the messages that contained inspiration and suggestions on how to change behavior. Perhaps not the most fun, but the most useful, because then I realized that I have not done any active travel today. It made me think, perhaps I should get off the bus earlier and walk the rest of the way.* [Woman, 61 years old, interview 7]

Acceptable but Modifiable

Taken together, the interviews indicated that the app was generally well accepted by the participants and that it had encouraged many to use more AT and public transportation. The participants graded the app overall an average of 3.5 on a scale from 0 to 5 (range 2.0-4.5). Nevertheless, the participants also identified several issues and areas for improvement.

Semiautomated Registration: Heaven and Hell

Most participants appreciated the app's semiautomated registration of their transportation activity. However, registration was also the feature with the most reported problems. The app sometimes erroneously changed the mode of transportation, for example, when passing by a subway station. Additionally, transportation would sometimes incorrectly change modes, for example, from driving to biking, if the velocity suddenly changed from a relatively high speed to a lower one (eg, due to traffic). Some participants said it sometimes felt as if the app was just guessing their mode of transportation. They expressed that occasionally it would register a route as a straight line between A and B, which was probably due to the lack of GPS points in some areas. Resolving these issues by correcting the route and mode of transportation required a lot of work, according to the participants. One of the participants did not appreciate that the app could track your whereabouts and therefore deleted it as soon as the intervention was over. Participants expressed different experiences with respect to the workload required to manually change the registered data in the app. Some participants expressed it was difficult and time-consuming, while others found it annoying but manageable. Some participants reported that having to correct the registration became boring and that the corrections were, at times, less accurate. The app's ability to automatically choose an activity connected to a certain place was considered both a strength and a weakness. For example, if the same activity was always performed at one place, the feature was helpful, but when visiting a place where several activities could be performed, this default function required participants to spend more time on corrections than it would have taken them to register the activity themselves.

> *The app gave me some trouble and sometimes it was difficult to use. It has been random, sometimes it made correct assumptions and sometimes not. Sometimes it took 15 minutes to correct registrations and sometimes it just worked. Sometimes it says that I do not have a network connection even though I do, and without it the app does not work.* [Man, 63 years old, interview 2]

> *I think the app learned how to register my travels well. You did not need to do everything all over again every time, except maybe change the travel mode to bus instead of bike, but that was really easy to do.* [Man, 56 years old, interview 8]

Tailored Transportation

To increase the effectiveness of the app, many participants suggested that it should be more tailored to the individual to better address their specific needs and circumstances. Some participants wanted the app to not register travel modes that they never used or were not able to use (eg, not register a train ride if that was never an option in the area). One participant suggested that displaying the most used travel modes and the most frequently visited places at the top of the list in the app would make it more user-friendly. Some also suggested hiding

XSL•FO
RenderX

options that were never used (eg, remove motorbike as an option if the participant did not own one). Another example was more individually tailored messages. Some participants received messages about taking the subway even though it was not a travel option in their area or about riding a bike when they had no access to one. Some messages included season-specific information not suitable all year, which was highlighted specifically by many as relevant to improve. Some expressed that they wanted to turn off the messages and instead use the app for an overview of their travels (ie, time use statistics). One participant requested messages about traffic conditions, for example, when the subway was not working or a traffic accident had occurred, or real-time route optimization information, such as "If you get off the bus now you will reach your destination sooner." Other participants requested more startling information to catch their attention, like how health is affected by a mere 20 minutes of AT.

> *I would have graded the app higher if it was possible to adjust it to my preferences. I do not bike, and I have received a lot of information about cycling, and I do not take the bus, but the app often suggested that I travelled by bus when it registered my travels.* [Man, 47 years old, interview 1]

Some participants stated that they might have used the app more frequently if the statistics displayed were more appealing to them and if the app were more interactive. For example, many expressed that the app would have been more informative if it included measurement of total PA, since much of their activity never got included in the statistics, and some participants wanted to see their calorie expenditure as well. Participants also suggested connecting the TRavelVU Plus app to other apps, like Runkeeper (Asics Corp) or Moves (Facebook Inc), to summarize the amount of total PA. Some indicated that the app would benefit from a different start screen (eg, one that displays how much of a participant's goal remains). Some participants had trouble finding the feature used to set weekly goals if they did not do it shortly after receiving the descriptive message with a link to the goal-setting function. Along the same lines, some participants found only being able to set a goal for the upcoming week on Sundays and no later to be limiting. Moreover, several participants reported wishing that they had set more goals than they actually did and suggested incorporating a function that forced you to choose whether you wanted to set a goal or not for that week, which thereby served as a reminder to set a goal, ideally in a timely fashion.

> *I would have liked an app that includes a wide variety of health measurements. Now there are apps for movement and apps for eating, but if you got them all in one app I would use it a lot more. If the app included other health components, I could have set goals that were more attractive to me.* [Woman, 25 years old, interview 3]

Discussion

This study explored healthy adults' perceptions of using a smartphone app developed to promote the use of AT as a means to increase daily PA.

Principal Findings

Overall, results showed that the app was perceived as acceptable by study participants and was able to motivate them to choose active rather than motorized forms of transportation. Similar to the results published by Peng et al [20], the participants appreciated the semiautomated collection of AT data. However, as this function sometimes failed, participants had to make corrections in the app, and this was perceived as unnecessarily time-consuming and caused undue frustration. In this context, it is relevant to note that Dennison et al [19] found that self-monitoring, goal setting, and receiving feedback were important features, as long as the input effort required was not perceived as too burdensome. The participants in that study reported that automated, accurate, and detailed data registration programs were still highly needed. Indeed, a high-detail feature for the automatic recognition of AT was included in this study. However, our results indicated that some refinement is required to improve the accuracy level, which is an important topic for future research.

Understanding the mechanisms underlying health behavior changes inspired by a mobile app is relevant to connecting to the theories that the intervention is grounded in and the behavior change techniques included. Hoj et al [29] describe an association between self-reported app engagement and impact on the theory-based mechanism of behavior change. Our results indicate that self-monitoring their transportation mode increased participants' awareness of their behavior. This is consistent with the transtheoretical model, which describes health behavior change as occurring through multiple stages, of which awareness is the first stage [30]. Self-monitoring is also one of 3 steps in the behavior change process, according to Bandura's social cognitive theory [31]. This is important, as the quantitative outcome of this project showed an increase in moderate-to-vigorous PA at the 6-month follow-up [24]. By creating awareness, the app might have initiated a process of behavior change that was first seen after 6 months.

According to the participants, the most motivating features of the app were goal setting and self-monitoring, which correspond to the second step, judgment, in Bandura's [31] theory. Goal setting and self-monitoring are valuable in supporting and sustaining health behavior changes using mobile app interventions [19,20]. Our study adds to the evidence of the importance of putting extra effort into reaching a goal; however, some participants also reported losing motivation if they perceived the goal to be unreachable or easily achievable without changing their behavior. This is comparable to the results from a recent meta-analysis that concluded that goal setting robustly affects behavioral changes when the goal is specific and challenging enough [32]. The authors [32] also emphasized that goal setting is favored at the group level, where it is set and monitored publicly (preferably by another person). Along this line of thought, the participants in our study suggested that the app should allow and encourage connections between its users, a feature that could promote behavior change, as someone else is also monitoring the progress. Thus, in summary, the goal-setting and self-monitoring features should be kept in future versions of the TRavelVU Plus app. In addition, a feature that

allows participants the option to share their goals and achievements with other participants could be added.

The SCAMPI study aimed to promote AT among healthy adults to increase their PA. However, participants reported that they might have used the app more frequently if it had included statistics concerning their general health or all their types of PA (including physical exercise). This could be important to consider when designing behavior change interventions with a focus on AT. Goal setting is important when changing behavior, and the goals must correspond to participants' expectations [32]. Otherwise, it might affect both the use and the perception of the app. Thus, for future versions of the TRavelVU app, as well as for other similar tools, the ability to monitor other types of PA (eg, gym classes, running) should be considered as an additional feature.

In the present study, participants who intended to increase their PA perceived messages as motivating, as they reminded them to use more AT. This is in line with Prestwich et al [33], who described increased PA in participants who intended to change their behavior and received motivating text messages compared with a control group that did not receive reminders or did not have any intention of changing their behavior. Also, McDermott et al [34] found that combining self-monitoring with other behavior change techniques, such as information, more effectively promotes PA and healthy eating than interventions based on self-monitoring alone. Self-monitoring and reminders are valued features and possibly explain the results in Karppinen et al's [35] study on healthy habit formation. The participants in our study called for more tailored information in the messages promoting AT, as unnecessary information only irritated them. This is similar to the results published by Peng et al [20], which concluded that participants requested more personalized information. In addition, Peng et al [20] noted that the inclusion of tailored information and increased personalization can encroach on privacy protection issues and feel invasive to some users. This was also observed in our study, as one participant expressed concerns about the app collecting information on his travel habits and uninstalled the app as soon as the intervention was over, while others explicitly recommended including more personalized information. Maintaining the delicate balance between personalization and anonymity is a challenge facing the entire field of mHealth. With increasing numbers of commercially available apps for PA entering the market, users are required to share more personal and contextual data (eg, geolocation), which makes the balance between personalization and anonymity problematic. More research is required to understand how best to manage this balance.

Strengths and Limitations

Qualitative evaluations of novel intervention methods are critical, as user experience and acceptance are key in ensuring widespread adoption and implementation and in guiding the design of future interventions. Here we present a qualitative evaluation of the SCAMPI trial that further explains and complements the quantitative main findings of the study [24]. Given that participation in this follow-up study was voluntary, it is possible that the participants willing to be interviewed had a more positive attitude toward the project and the app than those who were unwilling to participate further by being interviewed. Therefore, we asked them to highlight problems and identify areas for improvement in the app, which allowed us to paint a more nuanced picture of their perceptions. As reported in our main outcome paper [24], a major strength of the SCAMPI trial was that participants were invited from a random sample drawn by Statistics Sweden. Nevertheless, as is common in research, participants in the entire intervention group (n=127) and participants in this qualitative evaluation (n=17) had higher educational attainment (10/17, 59% had a university degree) than the nonresponders. Moreover, participants were already active prior to joining the study, spending on average 59 (SD 21) minutes per day in moderate-to-vigorous PA; this level of interest in PA may have affected the result of this study. For instance, it is possible that individuals already engaging in PA may also be users of activity apps. Thus, future studies should also include populations that are more sedentary. Finally, this study was limited by having only 17 out of the 127 original participants participate in the interviews. However, we ensured that this subpopulation was representative of the original group and found that a sample of this size was sufficient to generate interviews that contained a rich, broad variety of experiences.

Conclusions

This article contributes novel information about healthy adults' experiences using an app that promotes AT as a means toward increased PA. The results showed that the app was acceptable and that participants who were ready to make a behavior change were motivated by self-monitoring, goal setting, and receiving reminder messages to use AT. In addition, all participants reported that the app increased their awareness of their travel habits, which is the initial step required for behavior change to occur. Taken together, our results showed that the app's features were appreciated by its users, who also identified modifications that would improve the usability of future versions of the app.

Authors' Contributions

ML and AE developed and designed the methodology with contribution from RM. ML is the principal investigator for the SCAMPI trial and acquired the funding for the trial, including this qualitative study. CA and ES recruited participants. ES performed the interviews. A-KL and SR conducted the qualitative content analysis and wrote the original draft of the manuscript. ML, ES, AE, CA, and RM gave feedback on the analysis and reviewed and edited the manuscript. All authors approved the final version of the manuscript as submitted.

Conflicts of Interest

None declared.

References

1. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, et al. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]

2. World Health Organization. Global recommendations on physical activity for health. Geneva, Switzerland: World Health Organization; 2010. URL: http://apps.who.int/iris/bitstream/10665/44399/1/9789241599979_eng.pdf [accessed 2020-07-07]

3. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. Geneva, Switzerland: World Health Organization; 2020. URL: https://apps.who.int/iris/handle/10665/94384 [accessed 2020-07-07]

4. Heath GW, Parra DC, Sarmiento OL, Andersen LB, Owen N, Goenka S, et al. Evidence-based intervention in physical activity: lessons from around the world. Lancet 2012 Jul 21;380(9838):272-281. [doi: 10.1016/S0140-6736(12)60816-2] [Medline: 22818939]

5. Ekblom-Bak E, Ekblom Ö, Andersson G, Wallin P, Söderling J, Hemmingsson E, et al. Decline in cardiorespiratory fitness in the Swedish working force between 1995 and 2017. Scand J Med Sci Sports 2019 Feb;29(2):232-239. [doi: 10.1111/sms.13328] [Medline: 30351472]

6. Götschi T, Garrard J, Giles-Corti B. Cycling as a Part of Daily Life: A Review of Health Perspectives. Transport Reviews 2015 Jun 30;36(1):45-71 [FREE Full text] [doi: 10.1080/01441647.2015.1057877]

7. Glazener A, Khreis H. Transforming Our Cities: Best Practices Towards Clean Air and Active Transportation. Curr Environ Health Rep 2019 Mar;6(1):22-37. [doi: 10.1007/s40572-019-0228-1] [Medline: 30701411]

8. Wen CP, Wai JPM, Tsai MK, Yang YC, Cheng TYD, Lee M, et al. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. Lancet 2011 Oct 01;378(9798):1244-1253. [doi: 10.1016/S0140-6736(11)60749-6] [Medline: 21846575]

9. Costa S, Ogilvie D, Dalton A, Westgate K, Brage S, Panter J. Quantifying the physical activity energy expenditure of commuters using a combination of global positioning system and combined heart rate and movement sensors. Prev Med 2015 Dec;81:339-344 [FREE Full text] [doi: 10.1016/j.ypmed.2015.09.022] [Medline: 26441297]

10. Mueller N, Rojas-Rueda D, Cole-Hunter T, de Nazelle A, Dons E, Gerike R, et al. Health impact assessment of active transportation: A systematic review. Prev Med 2015 Jul;76:103-114. [doi: 10.1016/j.ypmed.2015.04.010] [Medline: 25900805]

11. Dinu M, Pagliai G, Macchi C, Sofi F. Active Commuting and Multiple Health Outcomes: A Systematic Review and Meta-Analysis. Sports Med 2019 Mar;49(3):437-452 [FREE Full text] [doi: 10.1007/s40279-018-1023-0] [Medline: 30446905]

12. Celis-Morales CA, Lyall DM, Welsh P, Anderson J, Steell L, Guo Y, et al. Association between active commuting and incident cardiovascular disease, cancer, and mortality: prospective cohort study. BMJ 2017 Apr 19;357:j1456 [FREE Full text] [doi: 10.1136/bmj.j1456] [Medline: 28424154]

13. Buliung R, Faulkner G, Beesley T, Kennedy J. School travel planning: mobilizing school and community resources to encourage active school transportation. J Sch Health 2011 Nov;81(11):704-712 [FREE Full text] [doi: 10.1111/j.1746-1561.2011.00647.x] [Medline: 21972991]

14. Johansson C, Lövenheim B, Schantz P, Wahlgren L, Almström P, Markstedt A, et al. Impacts on air pollution and health by changing commuting from car to bicycle. Sci Total Environ 2017 Apr 15;584-585:55-63 [FREE Full text] [doi: 10.1016/j.scitotenv.2017.01.145] [Medline: 28135613]

15. Pratt M, Sarmiento OL, Montes F, Ogilvie D, Marcus BH, Perez LG, et al. The implications of megatrends in information and communication technology and transportation for changes in global physical activity. Lancet 2012 Jul 21;380(9838):282-293. [doi: 10.1016/S0140-6736(12)60736-3] [Medline: 22818940]

16. Tate EB, Spruijt-Metz D, O'Reilly G, Jordan-Marsh M, Gotsis M, Pentz MA, et al. mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. Transl Behav Med 2013 Dec;3(4):406-415 [FREE Full text] [doi: 10.1007/s13142-013-0222-3] [Medline: 24294329]

XSL•FO

RenderX

17. Flores MG, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]

18. Schippers M, Adam PCG, Smolenski DJ, Wong HTH, de Wit JBF. A meta-analysis of overall effects of weight loss interventions delivered via mobile phones and effect size differences according to delivery mode, personal contact, and intervention intensity and duration. Obes Rev 2017 Apr;18(4):450-459. [doi: 10.1111/obr.12492] [Medline: 28187246]

19. 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]

20. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]

21. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: a usability study. JMIR Mhealth Uhealth 2014 Mar 14;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]

22. Torbjørnsen A, Ribu L, Rønnevig M, Grøttland A, Helseth S. Users' acceptability of a mobile application for persons with type 2 diabetes: a qualitative study. BMC Health Serv Res 2019 Sep 06;19(1):641 [FREE Full text] [doi: 10.1186/s12913-019-4486-2] [Medline: 31492176]

23. Ek A, Alexandrou C, Delisle Nyström C, Direito A, Eriksson U, Hammar U, et al. The Smart City Active Mobile Phone Intervention (SCAMPI) study to promote physical activity through active transportation in healthy adults: a study protocol for a randomised controlled trial. BMC Public Health 2018 Jul 16;18(1):880 [FREE Full text] [doi: 10.1186/s12889-018-5658-4] [Medline: 30012116]

24. Ek A, Alexandrou C, Söderström E, Bergman P, Delisle Nyström C, Direito A, et al. Effectiveness of a 3-Month Mobile Phone-Based Behavior Change Program on Active Transportation and Physical Activity in Adults: Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Jun 08;8(6):e18531 [FREE Full text] [doi: 10.2196/18531] [Medline: 32510462]

25. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [doi: 10.1037/0003-066x.44.9.1175] [Medline: 2782727]

26. Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. Annu Rev Public Health 2006;27:297-322. [doi: 10.1146/annurev.publhealth.27.021405.102100] [Medline: 16533119]

27. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]

28. Graneheim UH, Lindgren B, Lundman B. Methodological challenges in qualitative content analysis: A discussion paper. Nurse Educ Today 2017 Sep;56:29-34. [doi: 10.1016/j.nedt.2017.06.002] [Medline: 28651100]

29. Hoj TH, Covey EL, Jones AC, Haines AC, Hall PC, Crookston BT, et al. How Do Apps Work? An Analysis of Physical Activity App Users' Perceptions of Behavior Change Mechanisms. JMIR Mhealth Uhealth 2017 Aug 03;5(8):e114 [FREE Full text] [doi: 10.2196/mhealth.7206] [Medline: 28778846]

30. Prochaska J, Velicer W. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]

31. Bandura A. Social cognitive theory of self-regulation. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):248-287 [FREE Full text] [doi: 10.1016/0749-5978(91)90022-L]

32. Epton T, Currie S, Armitage CJ. Unique effects of setting goals on behavior change: Systematic review and meta-analysis. J Consult Clin Psychol 2017 Dec;85(12):1182-1198. [doi: 10.1037/ccp0000260] [Medline: 29189034]

33. Prestwich A, Perugini M, Hurling R. Can the effects of implementation intentions on exercise be enhanced using text messages? Psychol Health 2009 Jul;24(6):677-687. [doi: 10.1080/08870440802040715] [Medline: 20205020]

34. McDermott MS, Oliver M, Iverson D, Sharma R. Effective techniques for changing physical activity and healthy eating intentions and behaviour: A systematic review and meta-analysis. Br J Health Psychol 2016 Nov;21(4):827-841. [doi: 10.1111/bjhp.12199] [Medline: 27193530]

35. Karppinen P, Oinas-Kukkonen H, Alahäivälä T, Jokelainen T, Teeriniemi A, Salonurmi T, et al. Opportunities and challenges of behavior change support systems for enhancing habit formation: A qualitative study. J Biomed Inform 2018 Aug;84:82-92 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.012] [Medline: 29936091]

Abbreviations

AT: active transportation
mHealth: mobile health
PA: physical activity
SCAMPI: Smart City Active Mobile Phone Intervention

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

Effects of a Novel Contextual Just-In-Time Mobile App Intervention (LowSalt4Life) on Sodium Intake in Adults With Hypertension: Pilot Randomized Controlled Trial

Michael P Dorsch[1], PharmD, MS; Maria L Cornellier[2], RD; Armella D Poggi[2], BS; Feriha Bilgen[2], MPH; Peiyu Chen[2], MSc; Cindy Wu[2], UG; Lawrence C An[2,3], MD; Scott L Hummel[3,4], MD, MS

[1]University of Michigan College of Pharmacy, Ann Arbor, MI, United States

[2]University of Michigan Health System, Ann Arbor, MI, United States

[3]University of Michigan Medical School, Ann Arbor, MI, United States

[4]Ann Arbor Veterans Affairs Health System, Ann Arbor, MI, United States

Corresponding Author:
Michael P Dorsch, PharmD, MS
University of Michigan College of Pharmacy
428 Church Street
Ann Arbor, MI, 48109
United States
Phone: 1 734 647 1452
Email: mdorsch@med.umich.edu

Abstract

Background: High dietary sodium intake is a significant public health problem in the United States. High sodium consumption is associated with high blood pressure and high risk of cardiovascular disease.

Objective: The aim of this study was to evaluate the effect of a just-in-time adaptive mobile app intervention, namely, LowSalt4Life, on reducing sodium intake in adults with hypertension.

Methods: In this study, 50 participants aged ≥18 years who were under treatment for hypertension were randomized (1:1, stratified by gender) into 2 groups, namely, the App group (LowSalt4Life intervention) and the No App group (usual dietary advice) in a single-center, prospective, open-label randomized controlled trial for 8 weeks. The primary endpoint was the change in the 24-hour urinary sodium excretion estimated from spot urine by using the Kawasaki equation, which was analyzed using unpaired two-sided *t* tests. Secondary outcomes included the change in the sodium intake measured by the food frequency questionnaire (FFQ), the 24-hour urinary sodium excretion, blood pressure levels, and the self-reported confidence in following a low-sodium diet.

Results: From baseline to week 8, there was a significant reduction in the Kawasaki-estimated 24-hour urinary sodium excretion calculated from spot urine in the App group compared to that in the No App group (–462 [SD 1220] mg vs 381 [SD 1460] mg, respectively; *P*=.03). The change in the 24-hour urinary sodium excretion was –637 (SD 1524) mg in the App group and –322 (SD 1485) mg in the No App group (*P*=.47). The changes in the estimated sodium intake as measured by 24-hour dietary recall and by FFQ in the App group were –1537 (SD 2693) mg and –1553 (SD 1764) mg while those in the No App group were –233 (SD 2150) mg and –515 (SD 1081) mg, respectively (*P*=.07 and *P*=.01, respectively). The systolic blood pressure change from baseline to week 8 in the App group was –7.5 mmHg while that in the No App group was –0.7 mmHg (*P*=.12), but the self-confidence in following a low-sodium diet was not significantly different between the 2 groups.

Conclusions: This study shows that a contextual just-in-time mobile app intervention resulted in a greater reduction in the dietary sodium intake in adults with hypertension than that in the control group over a 8-week period, as measured by the estimated 24-hour urinary sodium excretion from spot urine and FFQ. The intervention group did not show a significant difference from the control group in the self-confidence in following a low sodium diet and in the 24-hour urinary sodium excretion or dietary intake of sodium as measured by the 24-hour dietary recall. A larger clinical trial is warranted to further elucidate the effects of the LowSalt4Life intervention on sodium intake and blood pressure levels in adults with hypertension.

Trial Registration: ClinicalTrials.gov NCT03099343; https://clinicaltrials.gov/ct2/show/NCT03099343

International Registered Report Identifier (IRRID):   RR2-10.2196/11282

XSL•FO
RenderX

KEYWORDS

hypertension; sodium intake; geofencing; mHealth

Introduction

High sodium intake is a significant public health problem in the United States [1]. The current federal guidelines advocate a daily sodium intake of less than 2300 mg/day [2]. However, the average sodium intake for Americans is approximately 3460 mg/day [3]. In a meta-analysis of 177,025 patients, higher sodium intake was associated with higher risk of stroke (relative risk 1.23, 95% CI 1.06-1.43) and a trend toward higher risk of cardiovascular diseases (relative risk 1.14, 95% CI 0.99-1.32) [4]. Interventions that lower sodium intake can decrease blood pressure levels and cardiovascular outcomes [5-8]. Implementing these interventions can be complex as it involves food manufacturers, restaurant chains, and community organizations.

Smartphones and mobile apps offer a scalable and pervasive opportunity to provide an intervention to individuals who frequently eat at restaurants or shop for high-sodium foods at grocery stores. Over half of the Americans consume 1-3 restaurant meals per week and 23% of the Americans consume 4 restaurant meals per week [9,10]. People who eat at restaurants frequently and buy high-sodium foods at grocery stores are the prime targets for this intervention, as about 77% of the sodium intake in the average American diet originates from processed and restaurant foods. We have developed a mobile app named as LowSalt4Life, which is an intervention aimed at reducing the sodium intake. Using smartphone sensors, the mobile app can detect when the patient arrives home or enters a restaurant or a grocery store and the restaurant or the grocery store that the patient has entered. This app provides a highly adaptive intervention wherein tailored messages are provided exactly when the patient needs help, also known as "just-in-time," and with the precise information needed for that location, also known as "contextual." Just-in-time adaptive interventions have the potential to dramatically improve patient behaviors in a variety of clinical settings [11].

The purpose of this pilot study was to evaluate the effectiveness of this mobile app intervention in providing contextual just-in-time adaptive push messages on dietary sodium intake and in improving the patient's confidence in following a low-sodium diet.

Methods

Trial Design

The method of this trial has been published previously [12]. This clinical trial was a single-center, prospective, open-label randomized controlled trial that was conducted from June 2017 to March 2019. Participant recruitment was performed at Michigan Medicine, formerly the University of Michigan Health System, through the university recruitment platform and by sending letters to over 7000 patients who met our study criteria. Participants were randomized into the mobile app group (App group) or usual care group (No App group) in a 1:1 manner and

stratified by gender using the University of Michigan Consulting for Statistics, Computing, and Analytics Research randomization instrument. Participants randomized to the App group received a 30-minute in-person training session and were educated to use the mobile app for 8 weeks by the study coordinator. The training session included information on how to use the mobile app. Participants randomized to the No App group received the standard of care at the University of Michigan. The University of Michigan guideline for hypertension management recommends modification of dietary sodium intake to less than 2400 mg/day. This study was approved by the University of Michigan Medical School Institutional Review Board (approval received on April 4, 2017), and all the participants provided written informed consent. This trial was listed at ClinicalTrials.gov (NCT03099343, received March 28, 2017) and was sponsored by the Agency for Healthcare Research and Quality (R21 HS024567).

Participants in This Study

Patients older than 18 years diagnosed with hypertension, on antihypertensive therapy for at least 3 months, and using an iPhone were included. Patients were excluded if they had chronic kidney disease (CKD), heart failure, systolic blood pressure >180 mmHg, diastolic blood pressure >110 mmHg, insulin-requiring diabetes mellitus, or were taking loop diuretics, corticosteroids, or nonsteroidal anti-inflammatory medications. CKD was defined as known kidney damage (structural or functional abnormalities) or estimated glomerular filtration rate less than 60 ml/min/1.73 m^2 (CKD stage 3, 4, or 5). The participants who provided the initial consent completed the Block Food Frequency Questionnaire (FFQ, NutritionQuest Inc) [13], and those with estimated baseline dietary sodium intake of less than 2300 mg/day were excluded before randomization. On October 1, 2018, the estimated dietary sodium intake for exclusion was changed to less than 2000 mg/day, with patients consuming 2000-2300 mg/day allowed to participate if the estimated sodium to kilocalorie ratio was greater than 1, as estimated by the FFQ. This change led to 4 patients being recruited that did not meet our initial study inclusion criteria. The 110-item Block FFQ was electronically self-administered and it recorded commonly consumed foods to estimate the nutrient and energy intake.

Intervention

The mobile app intervention method has been published previously [12]. Briefly, the intervention began with a baseline assessment of foods containing high sodium levels by using the Block Sodium Screener (NutritionQuest Inc). Participants selected alternatives to their 5 high-sodium foods and geotagged the places in which these foods were consumed or purchased (home, restaurant, or grocery store). In addition to these geolocations, a cloud-based web service was used to predict when the participant was entering a grocery store, restaurant, or home. Contextual just-in-time adaptive messages were

provided to promote behavior change when a participant entered a grocery store, restaurant, or home. These messages were push notifications, which were tailored to the user's confidence in following a low-sodium diet and linked to content in the mobile app. The mobile app showed the participant-selected alternatives to their 5 high-sodium foods, curated a list of low-sodium meal options at restaurants, provided users the capability to search restaurant menus that were prioritized by low sodium contents

and/or the ability to scan the universal product codes of the grocery store items to find similar food options containing lower sodium contents. A standard nutrition database (Nutritionix) was used to provide the nutrition information at the grocery stores and restaurants through an application programming interface (API). Figure 1 shows an example of a push notification and product search in the app.

Figure 1. Example of a LowSalt4Life push notification and product search in the app.
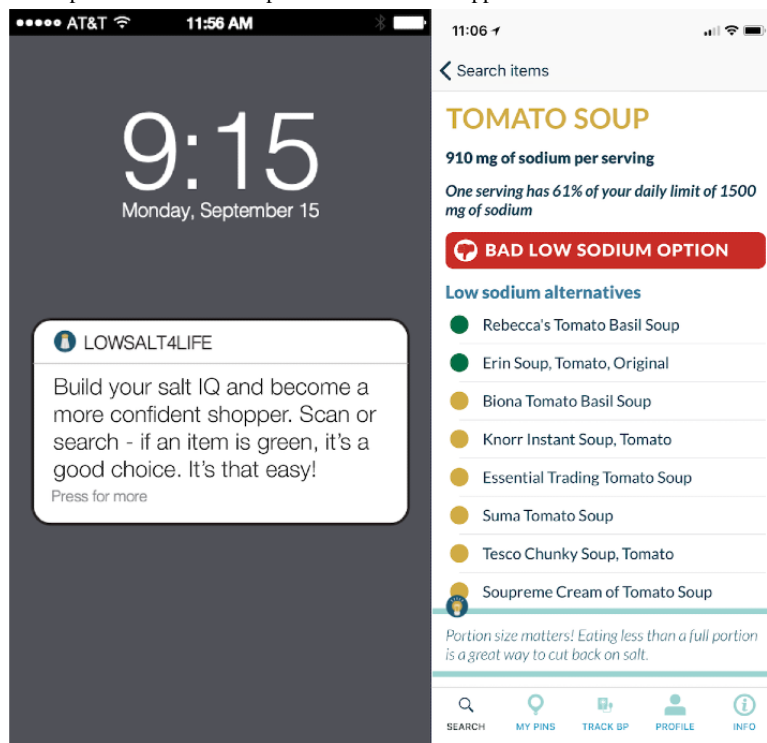
Outcomes

The change in the dietary sodium intake was measured by both subjective and objective measures. Twenty-four-hour dietary recall data were collected and analyzed using the Automated Self-Administered 24-hour (ASA24) dietary assessment tool, version 2016, developed by the National Cancer Institute, Bethesda [14]. The ASA24 is an electronic 24-hour recall

website that allows participants to self-administer the survey in a user-friendly manner. In addition to the Block FFQ 2014, a sodium screener survey was performed. This screener generates a score from the quantity and the frequency of consumption of high-sodium foods in the participant reports. The ASA24, FFQ, and sodium screener were administered at baseline and in week 8 of the study. Figure 2 shows the details of the timeline and all the outcomes.

Figure 2. Study timeline and outcomes. ASA24: Automated Self-Administered 24-hour dietary food recall; FFQ: food frequency questionnaire; SCFLDS: Self-care Confidence in Following a Low-sodium Diet Scale.

XSL•FO
RenderX
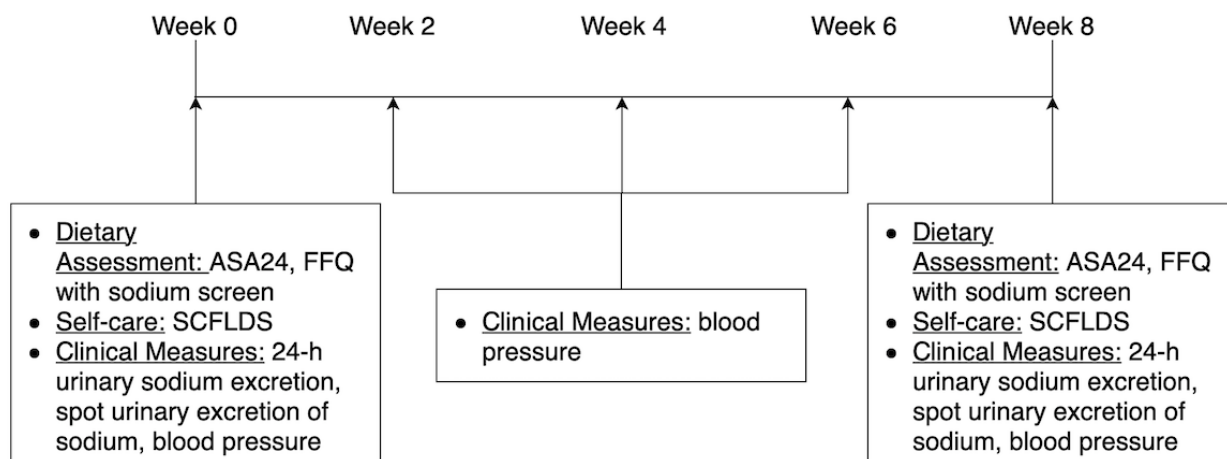
The urinary excretion of sodium was measured by the 24-hour urinary collection method for monitoring the sodium intake. Participants were instructed to collect all urine voids for 24 hours and return them for analysis for sodium excretion by using the standard University of Michigan laboratory processes. The method of collection from the participants was written in our study procedures and reviewed with the participants prior to the baseline in-person visit. Participants were mailed the supplies, a collection worksheet, and an instruction letter for collecting the 24-hour urine samples. The study coordinator reviewed the collection sheet and the 24-hour urine samples with the participants during the 2 in-person visits. The Knuimann strategy was used to assess the completeness of the 24-hour urine collection, which uses urinary creatinine values and body weight [15]. Twenty-four-hour urine excretion was collected at baseline and after 8 weeks. In addition, the estimated 24-hour urinary sodium excretion was accomplished with a morning spot urine excretion of sodium at baseline and after 8 weeks. The Kawasaki formula was used to estimate the 24-hour sodium urinary excretion from a fasting morning urine sample [16]. The estimated 24-hour urine excretion of sodium calculated from spot urine was the primary endpoint because 24-hour urine collection can be difficult and inconvenient for participants, thereby leading to a lack of data and decreased statistical power. All urine sodium measurements were done on the same day.

The 7-item Self-care Confidence in Following a Low-sodium Diet Scale (SCFLDS) was used to quantify the participant's confidence in following a low-sodium diet. It evaluates the patient's confidence in the ability to select and prepare foods with low sodium contents [17]. The SCFLDS was given at baseline and week 8 of the study.

Blood pressure was measured on a biweekly basis by the participant. Participants were trained to take 3 blood pressure readings at each time point by using the standard American Heart Association guideline–based recommendations on how to measure blood pressure at home [18]. An average of 3 measurements was used as the participants' blood pressure for that time point. Participants used their own automated blood pressure monitor at home for the study.

Mobile app data were collected to determine the extent to which the mobile app was used. The number of push notifications received by each participant and the use of the search functions for nutrition information was collected through the Nutritionix API. The Nutritionix API data is in aggregate and is not patient-specific. "Autocomplete" represents a search when typing in the app. "Universal Product Code Lookup" is the scanning function in the app. "Hits" represents the overall number of patient interactions with the nutrition information. Participants in the App group were also asked to complete a survey about their experience using the mobile app.

Statistical Analysis

Baseline demographics were compared using a two-sided t test or chi-square with Fisher's exact test, where appropriate, with the resulting P values for the tests. The primary endpoint was the estimated 24-hour urinary excretion of sodium from spot urine, which was estimated using the Kawasaki equation. Based on the data in the patients with hypertension, we expected the 24-hour urinary excretion of sodium to decline from 3400 (SD 1200) mg/day to 2400 (SD 1200) mg/day in the App group and a decline from 3400 (SD 1200) mg/day to 3300 (SD 1200) mg/day in the No App group [6]. A sample size of 24 patients (12 in each group) and 32 patients (16 in each group) would offer 80% and 90% power, respectively, for a 35% reduction in sodium intake. To account for patient dropout or incomplete follow-up, a total sample of 50 patients was identified as the study target enrollment. A two-sided unpaired t test was used to compare the change in each measure of sodium intake over time in the App versus No App groups.

Although not powered to determine the impact, other measures important for reducing sodium intake were evaluated. The Wilcoxon rank-sum test was used for continuous measures that were not distributed normally. The change in the self-confidence in following a low-sodium diet was analyzed using the two-sided t test. Repeated measures analysis of variance was used to determine the change in the blood pressure over time in each group. P values less than .05 were considered statistically significant. Data were represented as mean (SD) or number (%), unless otherwise noted. All analyses were performed using SAS software (version 9.4, SAS Institute, IBM Corp).

Results

Fifty patients were enrolled and randomized in the clinical trial (24 in the App group and 26 in the No App group). Figure 3 shows the CONSORT diagram for this clinical trial. There were no significant differences between the 2 groups, except for the SCFLDS. The mean age was 56.6 (SD 10) years in the App group and 58.2 (SD 11) years in the No App group (P=.58), and the baseline systolic blood pressure was 129.1 (SD 20) mmHg in the App group and 128.3 (SD 14) mmHg in the No App group. The measures of the sodium intake at baseline were similar between the 2 groups. All baseline measurements are represented in Table 1.

Figure 3. CONSORT flow diagram. ASA24: Automated Self-Administered 24-hour dietary food recall; FFQ: food frequency questionnaire; SCFLDS: Self-care Confidence in Following a Low-sodium Diet Scale.
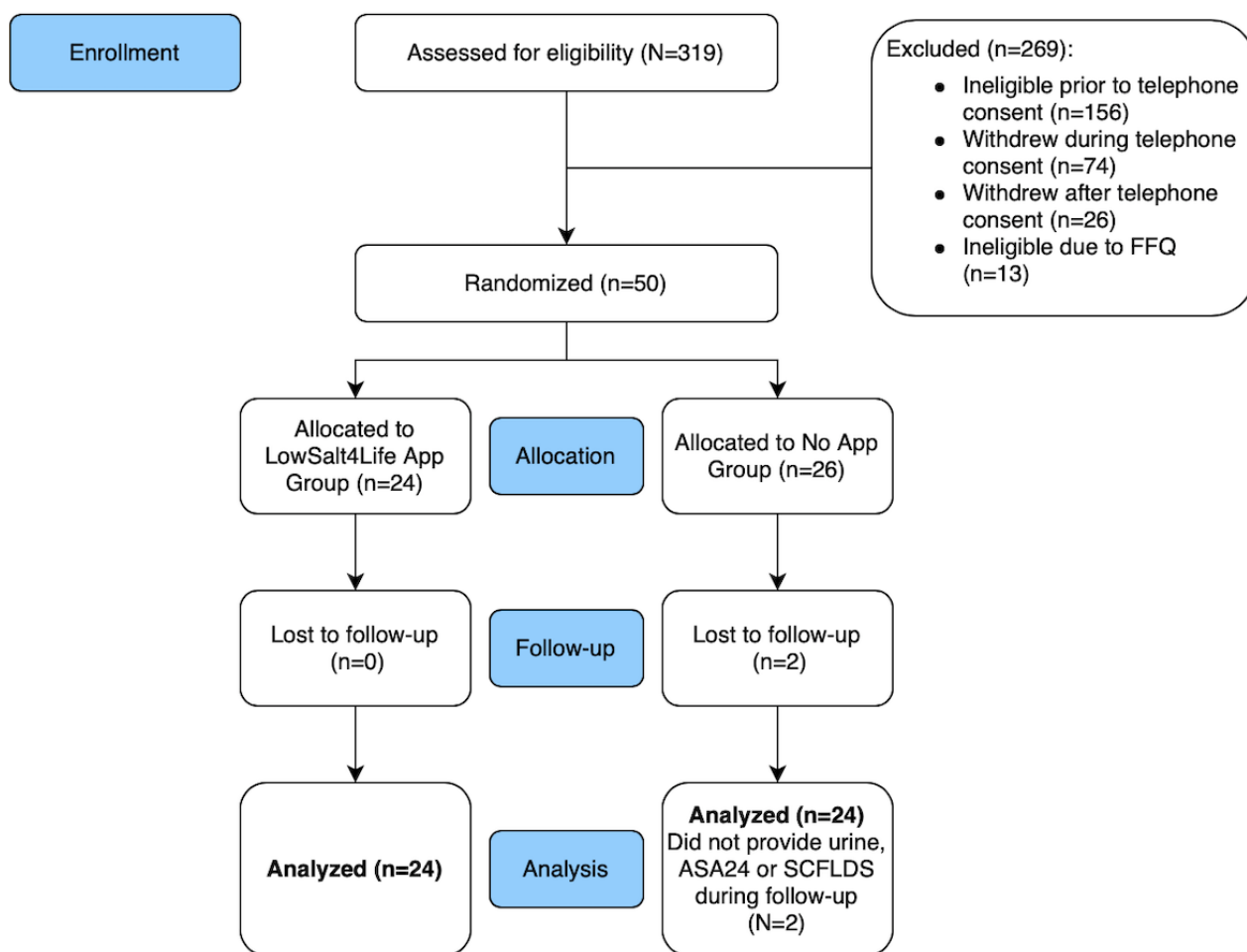
Table 1. Baseline demographic data.

Variable	App group (n=24)	No App group (n=26)	P value
Age (years), mean (SD)	56.6 (10)	58.2 (11)	.58
Females, n (%)	14 (58)	16 (61)	.82
Race, n (%)			.86
Caucasian	19 (79)	21 (81)	
African American	2 (8)	3 (12)	
Asian	2 (8)	1 (4)	
Other	1 (4)	1 (4)	
Ethnicity, n (%)			>.99
Latino	0 (0)	1 (4)	
Non-Latino	24 (100)	25 (96)	
Previous MI[a], n (%)	2 (8)	0 (0)	.22
DM[b], n (%)	0 (0)	1 (4)	>.99
Stroke or TIA[c], n (%)	0 (0)	1 (4)	>.99
Systolic BP[d] (mmHg), mean (SD)	129.1 (20)	128.3 (14)	.87
Diastolic BP (mmHg), mean (SD)	84.4 (12)	81 (8)	.23
SCr[e] (mg/dL), mean (SD)	0.88 (0.1)	0.89 (0.2)	.90
Kawasaki-estimated 24-h urine sodium (mg), mean (SD)	4026 (1514)	3798 (1463)	.59
24-h urine sodium (mg), mean (SD)	3607 (1755)	3561 (1924)	.93
FFQ[f] sodium (mg/day), mean (SD)	3995 (2119)	3660 (1314)	.51
ASA24[g] sodium (mg/day), mean (SD)	5127 (3118)	3877 (1773)	.09
Sodium screener (points), mean (SD)	30.3 (11)	31.4 (8)	.70
SCFLDS[h] (points), mean (SD)	20.9 (5)	18.4 (3)	.04

[a]MI: myocardial infarction.

[b]DM: diabetes mellitus type 2.

[c]TIA: transient ischemic attack.

[d]BP: blood pressure.

[e]SCr: serum creatinine.

[f]FFQ: food frequency questionnaire.

[g]ASA24: Automated Self-Administered 24-hour dietary food recall.

[h]SCFLDS: Self-care Confidence in Following a Low-sodium Diet Scale.

The change in the estimated 24-hour urinary sodium excretion calculated from spot urine from baseline to 8 weeks was –462 (SD 1220) mg in the App group and 381 (SD 1460) mg in the No App group (P=.03). The change in the sodium intake measured by the 24-hour urine was –637 (SD 1524) mg in the App group compared to –322 (SD 1485) mg in the No App group (P=.47). The change in the estimated sodium intake measured by the ASA24 was –1537 (SD 2693) mg in the App group compared to –233 (SD 2150) mg in the No App group (P=.07). The change in the estimated sodium intake by FFQ was –1553 (SD 1764) mg in the App group compared to –515 (SD 1081) mg in the No App group (P=.01). Figure 4 shows the change in the estimated sodium intake from baseline to 8 weeks. The sodium intake measurement values at baseline and at 8 weeks are presented in Multimedia Appendix 1. The change

in the sodium screen score in the App group was –9.5 (SD 9) points compared to –4.7 (SD 9) points in the No App group over the 8 weeks (P=.07). Over time, there was no difference in the change in the self-confidence between the 2 groups (App group 0.08 [SD 5] vs No App group –1.1 [SD 4]; P=.38). Analysis of the individual questions did not demonstrate a change in the confidence in reading food labels, shopping at grocery stores, or choosing low-sodium meal options at restaurants.

Over the 8 weeks of the study, participants in the App group had 7.5-mmHg reduction in systolic blood pressure (129 mmHg to 121.5 mmHg) compared to a 0.7-mmHg reduction in the No App group (128.4 mmHg to 127.7 mmHg; P=.12) based on the least squares means. Figure 5 demonstrates the change in the systolic and diastolic blood pressure levels over time.

XSL·FO

RenderX

Figure 4. Change in the estimated sodium intake from baseline to 8 weeks. Kawasaki is the 24-h urinary excretion of sodium estimated from spot urine, which was estimated by the Kawasaki equation. This was the primary outcome of the study.
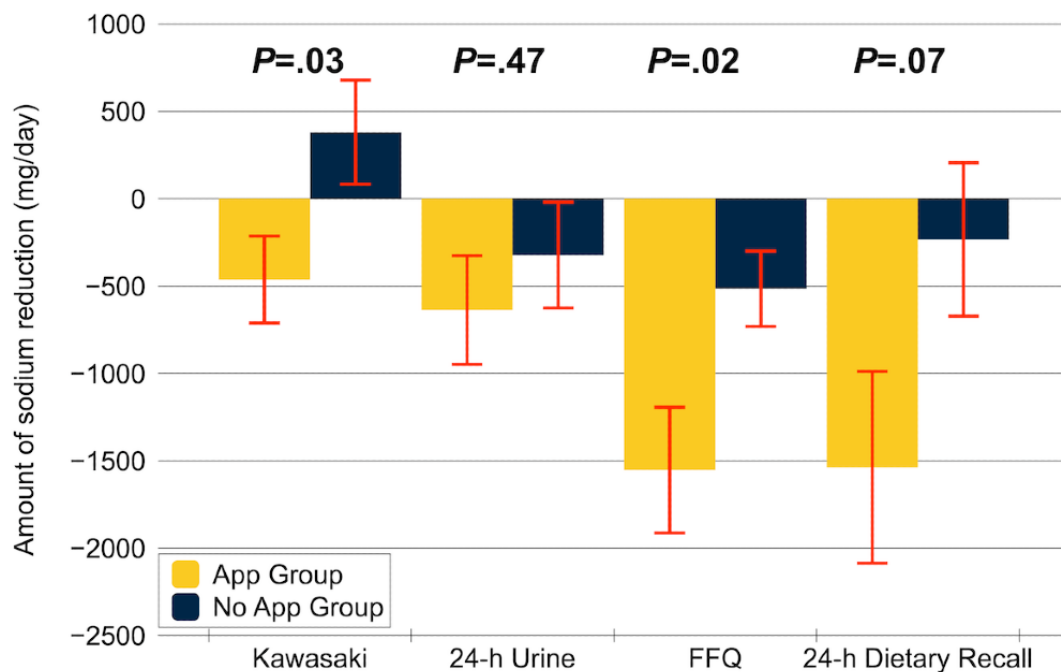

Figure 5. Blood pressure over time. The graph shows mean (SE) generated by least-squared means. DBP: diastolic blood pressure; SBP: systolic blood pressure.
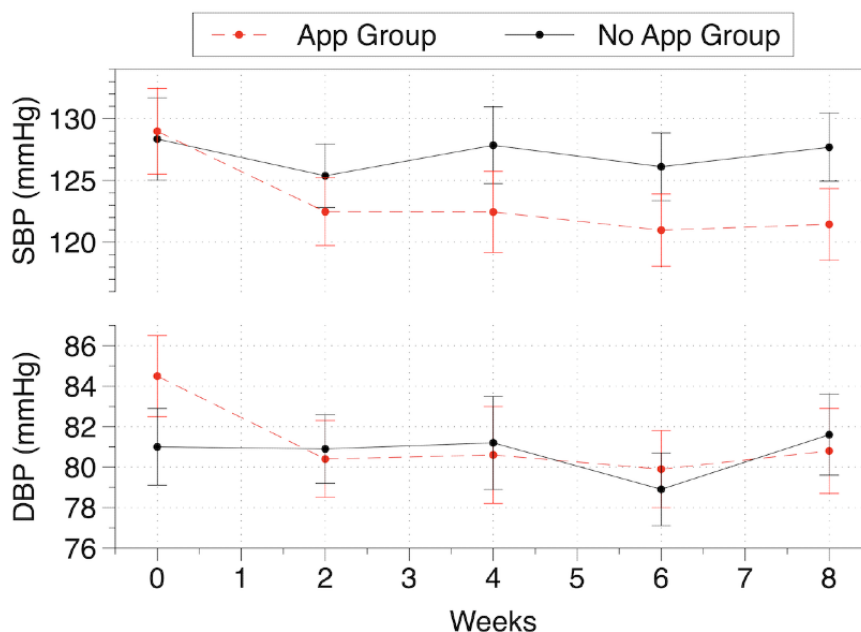
Over the 8 weeks of the study, participants received a median of 126 (IQR 75-186) push notifications: 6 (IQR 3-12) while arriving at the grocery store, 41 (IQR 9-60) while arriving at a restaurant, and 73 (IQR 37-96) while arriving home. Females received a median of 7.5 (IQR 4-15) notifications when arriving at a grocery store compared to 3 (IQR 1-9) notifications received by the males (Wilcoxon $P=.07$). Females received a median of 41 (IQR 9-60) notifications when arriving at a restaurant compared to 34 (IQR 4-61) notifications received by the males (Wilcoxon $P=.70$). There was substantial use of the nutrition information within the mobile app throughout the study. However, the individual participant-level data was not available.

The number of overall hits reached 1535 in June 2018. Autocomplete reached 836 searches in June 2018. Universal product code lookup was used 20-40 times per month during the clinical trial. Figure 6 shows the nutrition information searches in the mobile app during the clinical trial. In the survey of the 24 participants who used the app, 19 (79%) agreed that they found the app useful, 19 (79%) agreed that they used the information in the app in their daily life, 17 (71%) agreed that the information they received in the app was important to them, and 20 (83%) agreed the app was easy to use. Only 3 (13%) participants found the app confusing and 1 (4%) found the app difficult to understand.

Figure 6. Mobile app usage for nutritional information. UPC: universal product code.
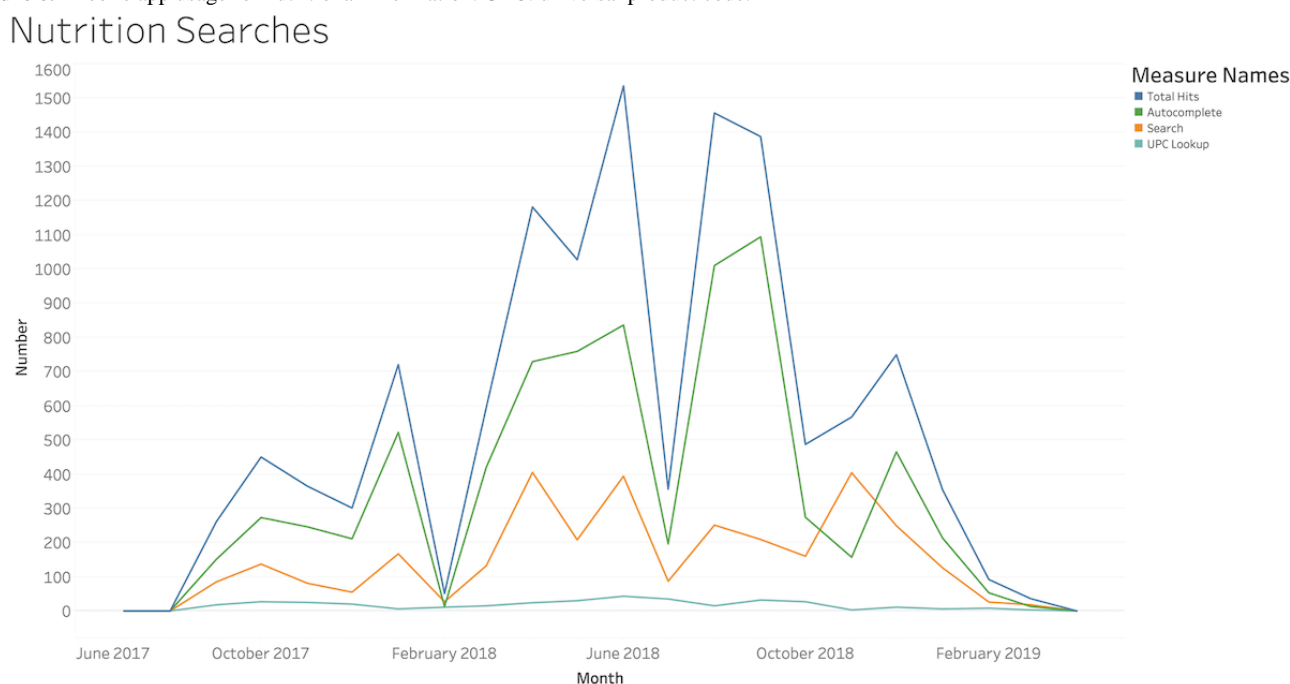
Discussion

In this pilot study, we found that LowSalt4Life, a just-in-time adaptive mobile app intervention that recommends lower dietary sodium food alternatives at home, restaurants, and grocery stores, holds promise in reducing the dietary sodium intake and systolic blood pressure. We demonstrated that compared to the control, the intervention group had a greater reduction in the dietary sodium intake over 8 weeks, as measured by the estimated 24-hour urinary sodium excretion from spot urine and FFQ. However, we did not observe a statistically significant difference between the 2 groups with regard to the self-confidence of the individuals in following a low-sodium diet, 24-hour urinary sodium excretion, dietary intake of sodium measured by 24-hour dietary recall or systolic blood pressure over 8 weeks. Although some of these measurements were not statistically significant, the measures of dietary sodium intake and systolic blood pressure showed clinically significant improvements in the intervention group compared to those in the control group. We believe the mechanism of the intervention is identifying high-sodium foods and providing lower sodium alternatives to those foods at the time of eating or purchasing the food. In a recent study on consumer understanding of sodium intake and food labeling, only half of the grocery store customers were able to correctly use the sodium label information to choose low-sodium foods [19]. The American Heart Association guideline for the dietary approach to prevent and treat hypertension states that "any meaningful strategy to reduce salt intake must involve the efforts of food manufacturers and restaurants" [20].

We are currently witnessing broad social changes in how individuals expect to find and use information about their health. According to the Pew Research Center's Internet and American Life Project, 73% of the households have broadband service and 81% of the Americans (53% of them >65 years) have a

smartphone [21], which supports the need for mobile app–based interventions for health care. In a recent systematic review of mobile health interventions to lower sodium intake, only 6 were randomized controlled clinical trials and only 2 of those trials were mobile app interventions published in English [22]. SaltSwitch (New Zealand) is a mobile app focused on supporting users to find food options with low sodium levels at grocery stores [23]. SaltSwitch provides low-sodium alternatives for items scanned at grocery stores but is not personalized for the user's requirements of high-sodium foods; moreover, it does not provide a push notification on entry to remind the user to scan the items and does not provide contextual food information at restaurants. The 4-week randomized controlled trial in 66 patients with cardiovascular disease demonstrated that those randomized to the SaltSwitch intervention group purchased foods with low sodium levels more often when compared to the control, but the intervention did not reduce urinary sodium excretion or blood pressure over time. Another study randomized 30 adults to the MyFitnessPal app for food logging to receive sodium content feedback or a paper journal of food logging for sodium content feedback for 4 weeks [24]. Although the MyFitnessPal app provided personalized dietary information about the sodium content of the foods, specific recommendations on what food substitutions could be done at different locations were lacking. The change in the predicted urinary sodium excretion from a spot urine test over 4 weeks was significantly greater in the MyFitnessPal group (–838 [SD 1093] mg) compared to that in the paper journal group (236 [SD 1333] mg). These mobile apps, in addition to our study app, show that mobile app–based interventions can improve patient health behaviors over a short time. The major difference between LowSalt4Life and these interventions is the just-in-time adaptive intervention that we deployed. The just-in-time adaptive intervention is a novel approach that makes mobile interventions "smarter" by incorporating real-time data streams to generate tailored notifications and then deliver these notifications at key

moments when there is a high likelihood of success. Just-in-time adaptive interventions have shown efficacy in multiple treatment domains, including smoking cessation, alcohol abuse, and mental health treatment [25,26].

Interestingly, the participants' confidence in following a low-sodium diet was different at baseline and was not affected by the intervention. There was no difference between the groups over time for specific questions from the survey about the grocery stores and restaurants. This supports the theory that the intervention does not affect patient self-confidence; in fact, the participants in our study were somewhat confident to very confident in following a low-sodium diet at baseline. Improvement in confidence may have been difficult, given such high confidence at the baseline. However, this confidence did not appear to be warranted, illustrating the unmet need for additional assistance with an intervention. It could also be that the participants require a longer duration of the intervention to feel more self-confident. Our intervention period was relatively short—only 8 weeks.

Our study had the following limitations. First, there were baseline differences in the confidence in following a low-sodium diet. It is unclear if this difference in confidence is clinically significant, but nevertheless, future research studies should investigate this finding further. Second, several of the sodium intake measures did not demonstrate statistically significant changes. However, all the measures of sodium intake present methodological challenges [27], and the trend toward improvement in the App group was consistent. Subjective measurements such as FFQ and 24-hour dietary recall have recall bias. Urinary measurements of sodium intake have high levels of random errors due to day-to-day variations. Third, all participants were required to have an iPhone, and enrollment was performed at 1 institution. This could have led to skewed socioeconomic backgrounds in our study compared to the general population diagnosed with hypertension. Fourth, the clinical trial was set up as an App versus No App study wherein the control group did not receive the app. This could have led to a lack of an attention control group. Participants could have felt left out of the intervention arm and not engaged enough in the study. However, the No App group participants did consistently participate by providing blood pressure measurements, dietary survey responses, and urinary collections over the 8-week study. To minimize potential bias, future studies could be designed with an attention control, sequential multiple assignment randomization, or microrandomization in the control group.

In conclusion, our randomized controlled 8-week pilot study in adults with hypertension showed that a contextual just-in-time mobile app intervention resulted in a greater reduction in dietary sodium intake as measured by the estimated 24-hour urinary sodium excretion from spot urine and FFQ compared to that in the control. The intervention group did not show statistically significant differences from the control in the self-confidence in following a low-sodium diet, 24-hour urinary sodium excretion, or dietary intake of sodium as measured by the 24-hour dietary recall over 8 weeks. A larger clinical trial is warranted to further elucidate the effects of the LowSalt4Life intervention on sodium intake and blood pressure.

Authors' Contributions

MPD, MLC, ADP, FB, PC, CW, LCA, and SLH contributed to the design/implementation of the research, analysis/interpretation of the results, and writing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Estimates of dietary sodium intake at baseline and week 8.
[DOCX File , 14 KB - mhealth_v8i8e16696_app1.docx]

References

1. Mugavero KL, Gunn JP, Dunet DO, Bowman BA. Sodium reduction: an important public health strategy for heart health. J Public Health Manag Pract 2014;20(1 Suppl 1):S1-S5 [FREE Full text] [doi: 10.1097/PHH.0b013e3182aa659c] [Medline: 24322810]
2. U.S. DOHAHSAUDOA. 2015-2020 Dietary Guidelines for Americans. 8th Edition. URL: http://health.gov/dietaryguidelines/2015/guidelines/ [accessed 2019-06-01]
3. Sebastian RS, Wilkinson Enns C, Steinfeldt LC, Goldman JD, Moshfegh AJ. Monitoring sodium intake of the US population: impact and implications of a change in what we eat in America, National Health and Nutrition Examination Survey dietary data processing. J Acad Nutr Diet 2013 Jul;113(7):942-949. [doi: 10.1016/j.jand.2013.02.009] [Medline: 23567248]

XSL•FO
RenderX

4. Strazzullo P, D'Elia L, Kandala N, Cappuccio FP. Salt intake, stroke, and cardiovascular disease: meta-analysis of prospective studies. BMJ 2009 Nov 24;339:b4567 [FREE Full text] [Medline: 19934192]

5. Sacks FM, Svetkey LP, Vollmer WM, Appel LJ, Bray GA, Harsha D, et al. Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. DASH-Sodium Collaborative Research Group. N Engl J Med 2001 Jan 4;344(1):3-10. [doi: 10.1056/NEJM200101043440101] [Medline: 11136953]

6. Taylor RS, Ashton KE, Moxham T, Hooper L, Ebrahim S. Reduced dietary salt for the prevention of cardiovascular disease: a meta-analysis of randomized controlled trials (Cochrane review). Am J Hypertens 2011 Aug;24(8):843-853. [doi: 10.1038/ajh.2011.115] [Medline: 21731062]

7. Cook NR, Cutler JA, Obarzanek E, Buring JE, Rexrode KM, Kumanyika SK, et al. Long term effects of dietary sodium reduction on cardiovascular disease outcomes: observational follow-up of the trials of hypertension prevention (TOHP). BMJ 2007 Apr 28;334(7599):885-888 [FREE Full text] [doi: 10.1136/bmj.39147.604896.55] [Medline: 17449506]

8. He FJ, Pombo-Rodrigues S, Macgregor GA. Salt reduction in England from 2003 to 2011: its relationship to blood pressure, stroke and ischaemic heart disease mortality. BMJ Open 2014 Apr 14;4(4):e004549 [FREE Full text] [doi: 10.1136/bmjopen-2013-004549] [Medline: 24732242]

9. Mattes RD, Donnelly D. Relative contributions of dietary sodium sources. J Am Coll Nutr 1991 Aug;10(4):383-393. [doi: 10.1080/07315724.1991.10718167] [Medline: 1910064]

10. Smith LP, Ng SW, Popkin BM. Trends in US home food preparation and consumption: analysis of national nutrition surveys and time use studies from 1965-1966 to 2007-2008. Nutr J 2013 Apr 11;12:45 [FREE Full text] [doi: 10.1186/1475-2891-12-45] [Medline: 23577692]

11. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2016 Sep 23. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]

12. Dorsch MP, An LC, Hummel SL. A Novel Just-in-Time Contextual Mobile App Intervention to Reduce Sodium Intake in Hypertension: Protocol and Rationale for a Randomized Controlled Trial (LowSalt4Life Trial). JMIR Res Protoc 2018 Dec 07;7(12):e11282 [FREE Full text] [doi: 10.2196/11282] [Medline: 30530462]

13. Block G, Woods M, Potosky A, Clifford C. Validation of a self-administered diet history questionnaire using multiple diet records. J Clin Epidemiol 1990;43(12):1327-1335. [doi: 10.1016/0895-4356(90)90099-b] [Medline: 2254769]

14. Subar AF, Kirkpatrick SI, Mittl B, Zimmerman TP, Thompson FE, Bingley C, et al. The Automated Self-Administered 24-hour dietary recall (ASA24): a resource for researchers, clinicians, and educators from the National Cancer Institute. J Acad Nutr Diet 2012 Aug;112(8):1134-1137 [FREE Full text] [doi: 10.1016/j.jand.2012.04.016] [Medline: 22704899]

15. Murakami K, Sasaki S, Takahashi Y, Uenishi K, Watanabe T, Kohri T, et al. Sensitivity and specificity of published strategies using urinary creatinine to identify incomplete 24-h urine collection. Nutrition 2008 Jan;24(1):16-22. [doi: 10.1016/j.nut.2007.09.001] [Medline: 17996421]

16. Kawasaki T, Itoh K, Uezono K, Sasaki H. A simple method for estimating 24 h urinary sodium and potassium excretion from second morning voiding urine specimen in adults. Clin Exp Pharmacol Physiol 1993 Jan;20(1):7-14. [doi: 10.1111/j.1440-1681.1993.tb01496.x] [Medline: 8432042]

17. Heo S, Moser DK, Lennie TA, Payne-Emerson H, Welch JL, Weaver M. Development and testing of the feasibility and acceptability of a tailored dietary intervention in patients with heart failure. J Cardiovasc Nurs 2015;30(3):213-221. [doi: 10.1097/JCN.0000000000000148] [Medline: 24695076]

18. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison HC, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2017 Nov 13. [doi: 10.1161/HYP.0000000000000065] [Medline: 29133356]

19. Grimes CA, Riddell LJ, Nowson CA. Consumer knowledge and attitudes to salt intake and labelled salt information. Appetite 2009 Oct;53(2):189-194. [doi: 10.1016/j.appet.2009.06.007] [Medline: 19540891]

20. Appel LJ, Brands MW, Daniels SR, Karanja N, Elmer PJ, Sacks FM, et al. Dietary approaches to prevent and treat hypertension: a scientific statement from the American Heart Association. Hypertension 2006 Feb;47(2):296-308 [FREE Full text] [doi: 10.1161/01.HYP.0000202568.01167.B6] [Medline: 16434724]

21. Pew RC. Mobile Technology and Home Broadband 2019. Internet & Technology. 2013. URL: https://www.pewinternet.org/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2019-08-01]

22. Ali SH, Luo R, Li Y, Liu X, Tang C, Zhang P. Application of Mobile Health Technologies Aimed at Salt Reduction: Systematic Review. JMIR Mhealth Uhealth 2019 Apr 17;7(4):e13250 [FREE Full text] [doi: 10.2196/13250] [Medline: 30994467]

23. Eyles H, McLean R, Neal B, Jiang Y, Doughty RN, McLean R, et al. A salt-reduction smartphone app supports lower-salt food purchases for people with cardiovascular disease: Findings from the SaltSwitch randomised controlled trial. Eur J Prev Cardiol 2017 Dec;24(13):1435-1444. [doi: 10.1177/2047487317715713] [Medline: 28631933]

24. Ipjian ML, Johnston CS. Smartphone technology facilitates dietary change in healthy adults. Nutrition 2017 Jan;33:343-347. [doi: 10.1016/j.nut.2016.08.003] [Medline: 27742102]

XSL•FO
RenderX

25.   Naughton F. Delivering "Just-In-Time" Smoking Cessation Support Via Mobile Phones: Current Knowledge and Future Directions. Nicotine Tob Res 2017 Mar;19(3):379-383. [doi: 10.1093/ntr/ntw143] [Medline: 27235703]

26.   Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]

27.   Cobb LK, Anderson CAM, Elliott P, Hu FB, Liu K, Neaton JD, American Heart Association Council on LifestyleMetabolic Health. Methodological issues in cohort studies that relate sodium intake to cardiovascular disease outcomes: a science advisory from the American Heart Association. Circulation 2014 Mar 11;129(10):1173-1186. [doi: 10.1161/CIR.0000000000000015] [Medline: 24515991]

Abbreviations

API: application programming interface
ASA24: Automated Self-Administered 24-hour dietary recall
CKD: chronic kidney disease
FFQ: food frequency questionnaire
SCFLDS: Self-care Confidence in Following a Low-sodium Diet Scale

Original Paper

Mobile App Use for Insomnia Self-Management in Urban Community-Dwelling Older Korean Adults: Retrospective Intervention Study

Kyungmi Chung[1,2,3], PhD; Seoyoung Kim[4], MA; Eun Lee[3,5], MD, PhD; Jin Young Park[1,2,3], MD, PhD

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

[2]Center for Digital Health, Yongin Severance Hospital, Yonsei University Health System, Yongin, Republic of Korea

[3]Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Yonsei University Health System, Seoul, Republic of Korea

[4]Nowon Eulji Medical Center, Eulji University, Seoul, Republic of Korea

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

Corresponding Author:
Jin Young Park, MD, PhD
Department of Psychiatry, Yonsei University College of Medicine
Yongin Severance Hospital
Yonsei University Health System
Department of Psychiatry, Yongin Severance Hospital
363, Dongbaekjukjeon-daero, Giheung-gu
Yongin, 16995
Republic of Korea
Phone: 82 2 2228 0972
Fax: 82 2 2227 6558
Email: empathy@yuhs.ac

Abstract

Background: As an evidence-based psychotherapy for treating insomnia, cognitive behavioral therapy for insomnia (CBT-I), which helps people with sleep problems to change their unhelpful sleep-related beliefs and habits, has been well-established in older adults. Recently, the utilization of mobile CBT-I apps has been getting attention from mental health professionals and researchers; however, whether mobile CBT-I apps are usable among older users has yet to be determined.

Objective: The aims of this study were to explore the relationships between subjective sleep quality and subjective memory complaints and depressive symptoms; to explore the relationship between perceived difficulty in mobile app use and usability of the mobile phone–based self-help CBT-I app, named MIND MORE, in urban community-dwelling Korean older adults; to compare changes in subjective sleep quality from pre-intervention to post-intervention, during which they used the mobile app over a 1-week intervention period; and evaluate adherence to the app.

Methods: During the 2-hour training program delivered on 1 day titled "Overcoming insomnia without medication: How to use the 'MIND MORE' mobile app for systematic self-management of insomnia" (pre-intervention), 41 attendants were asked to gain hands-on experience with the app facilitated by therapists and volunteer workers. They were then asked to complete questionnaires on sociodemographic characteristics, subjective evaluation of mental health status (ie, depression, memory loss and impairment, and sleep problems), and app usability. For the 1-week home-based self-help CBT-I using the app (post-intervention), 9 of the 41 program attendants, who had already signed up for the pre-intervention, were guided to complete the given questionnaires on subjective evaluation of sleep quality after the 1-week intervention, specifically 8 days after the training program ended.

Results: Due to missing data, 40 of 41 attendants were included in the data analysis. The main findings of this study were as follows. First, poor subjective sleep quality was associated with higher ratings of depressive symptoms (40/40; ρ=.60, P<.001) and memory complaints (40/40; ρ=.46, P=.003) at baseline. Second, significant improvements in subjective sleep quality from pre-intervention to post-intervention were observed in the older adults who used the MIND MORE app only for the 1-week intervention period (9/9; t_8=3.74, P=.006). Third, apart from the program attendants who did not have a smartphone (2/40) or

XSL•FO
RenderX

withdrew from their MIND MORE membership (3/40), those who attended the 1-day sleep education program adhered to the app from at least 2 weeks (13/35, 37%) to 8 weeks (2/35, 6%) without any further contact.

Conclusions: This study provides empirical evidence that the newly developed MIND MORE app not only is usable among older users but also could improve subjective sleep quality after a 1-week self-help intervention period.

Introduction

Background

Insomnia and sleep disturbance symptoms negatively influence mental health-related quality of life and functional abilities in older adults, with the perception of nonrestorative and poor-quality sleep [1]. According to the finding of a systematic review and meta-analysis of population-based prospective cohort studies [2], insomnia increases the risk of incident all-cause dementia in elderly individuals. Poor sleep quality may result in increased mortality and psychiatric comorbidities in both nondemented and demented elderly, causing stress for their caregivers [3]. Greater sleep disturbance has been correlated with more severe impairments in cognitive function in two cohort groups, particularly patients with Alzheimer disease and even healthy, nondemented elderly; furthermore, both groups with sleep problems experienced more depressive symptoms [4]. Persistent insomnia also appears to be a risk factor for onset of depression in the elderly [5]. With the rapid growth of the aging population in South Korea, sleep problems have increasingly been recognized as an important public health concern among community-dwelling older Korean adults. In particular, the prevalence of insomnia is estimated to be at least 32% in the elderly Korean population aged ≥60 years, with a significantly higher prevalence in women than in men [6,7]. The prevalence rates of insomnia in patients with subjective memory impairment, mild cognitive impairment, and dementia were 23.2%, 19.6%, and 31.0%, respectively, in a community sample of elderly Korean individuals aged ≥65 years [8]. Additionally, more than 50% of patients with insomnia report depressive symptoms [7]. Given the high prevalence of late-life insomnia worldwide, there is a need to better understand the associations between sleep quality and individual psychiatric comorbidities such as cognitive dysfunction and depression.

As approximately one-third of Korean older adults experience insomnia [7], its treatment is of great importance for both patients with insomnia and those who are concerned about their sleep-related health behaviors — to effectively manage sleep problems and improve sleep quality as well as prevent or reduce late-life disability [9]. It is important to consider long-term outcomes of any treatments for late-life insomnia even after treatment discontinuation as well as short-term outcomes during treatment administration because insomnia is often recurrent or persistent [10]. A number of previous studies have revealed that cognitive behavioral therapy for insomnia (CBT-I) alone was more effective than pharmacotherapy alone or combined treatment after treatment administration and discontinuation [10-15]. Although pharmacotherapy is well-known as the most frequently recommended intervention for insomnia, moderate efficacy has been shown while sleep medications were being used, and changes in sleep quality returned to baseline after discontinuing the medications [12,16]. In addition, the long-term use of sedative-hypnotics is contraindicated due to the diverse and unwanted adverse effects such as falls, nausea, confusion, dizziness, headache, daytime drowsiness, abuse and dependence, memory impairment, or rebound insomnia [12,13,17]. To deal with the potential adverse effects of sleep medicines, short-term, intermittent use or an alternative intervention based on CBT-I is recommended in older adults. Considering that psychological approaches produce sustained outcomes without the risk for tolerance or side effects related to pharmacologic approaches [13,15], CBT-I can be regarded as a more appropriate, evidence-based, first-line treatment option than pharmacotherapy, with clinically meaningful effect sizes. Defined as a multimodal therapy delivered in person, a combined CBT-I intervention, which incorporates at least 2 of the 5 widely accepted cognitive (cognitive strategies), behavioral (stimulus control, sleep restriction, relaxation), and educational (sleep hygiene education) components, is more preferred than a standalone component [15,18-20].

CBT-I results in significant improvements in sleep onset latency, wake after sleep onset, total sleep time, and sleep efficiency [15]; however, its efficacy might be reduced by such adherence issues as drop-out, premature termination, irregular attendance to sessions, and noncompliance to homework [21]. As mobile health (mHealth) technology has recently become more accessible, elderly adults are more familiar and comfortable with mobile devices in their everyday lives and natural settings [22]. Since adherence to a CBT-I protocol leads to better treatment outcomes [23], a promising, readily accessible approach to address these adherence issues can be the adoption of a mobile app–delivered CBT-I for the self-management of insomnia among older adults. Based on the stepped care model of health care delivery with two fundamental features [24,25], the pure self-help or guided self-help CBT-I app and its dashboard could serve as the "least restrictive" therapy with evidence-based, entry-level treatment to achieve significant health gains using a minimal intervention principle at the lowest cost and with the lowest required treatment intensity (ie, the least specialist therapist time) and as a "self-correcting" mechanism to systemically monitor and measure treatment progress and outcome, respectively. Despite the growing interest

XSL•FO

RenderX

in using mobile phones and services among the elderly, individual attributes, such as gender, education, technology self-efficacy and anxiety, and age-related health and ability characteristics, have direct effects on technology acceptance behavior [26]. In addition, 3 distinct factors, namely self-efficacy, conversion readiness, and peer support, affect older adults' technology acceptance behaviors, particularly the "intention to learn" phase [27]. Senior users tend to use their mobile phones for very limited purposes and discourage themselves from learning new technology through trial and error, thereby encountering more learning difficulties [28]. To enhance adherence to a newly developed CBT-I app installed on their own mobile phones, it is important to help senior users overcome the fear of learning a new mHealth technology and to provide them with more appealing features and useful, easy-to-use functionality.

Previous studies have focused on assessing the efficacy, feasibility, and usability of CBT-I apps to complement standard treatment for insomnia, such as sleep medication use or traditional CTB-I programs [29-32]; they have not focused on how to lower the barriers to the acceptance of a mobile app–based CBT-I by older adults, who often have learning difficulties in the initial phase. In terms of age-related barriers to usability evaluations with older adults, half of the identified high-level usability issues with mHealth app use were related to motivational barriers, such as low computer literacy and low trust in their own ability to use the apps, more than cognitive and perceptional barriers, such as working memory and visual acuity [33]. Particularly, when it comes to using digital technology to support mental health, the following may negatively affect the readiness of older adults to engage with technology: low mood, fear of consequences, perceived superiority of human contact, and a lack of prior knowledge, skills, or experience applicable to mobile devices and mHealth app usage [34]. Based on findings from relevant studies, motivational barriers should receive more attention in usability research in the field of mobile public health for older people. As there has been a high uptake of educational programs in community centers (eg, Community Mental Health Welfare Center and Seoul Metropolitan Center for Dementia) to improve quality of life in older community dwellers and to promote geriatric mental health for the prevention or early detection of a transition from nonclinical to clinical psychotic states [35], utilizing this public health infrastructure can allow older adults to easily access and efficiently learn and adhere to new CBI-I apps without active, constant intervention from health professionals.

Objectives

To examine the treatment effects of a mobile insomnia self-management intervention, this retrospective study analyzed data sourced from an urban community center for dementia prevention in South Korea. In the center, a mobile phone–based CBT-I app (hereafter "MIND MORE"), which was designed and developed by our research team, was used during a 1-day sleep education program. The primary purpose of this study was to determine whether self-management of insomnia is facilitated with the MIND MORE mobile app over a 1-week intervention period by comparing changes in subjective sleep

quality from pre-intervention to post-intervention, particularly in community-dwelling Korean older adults. We hypothesized that there would be improvements in subjective sleep quality from pre-intervention to post-intervention. Secondary aims of this study were to evaluate adherence to the MIND MORE app, explore the relationship between perceived difficulty in mobile app use and usability of the MIND MORE app, and explore the relationships between subjective sleep quality and subjective memory complaints and depressive symptoms.

Methods

Participant Recruitment

A total of 41 older adults, all of whom registered at the Seoul Metropolitan Center for Dementia (SMCD) located in Seodaemun-gu, Seoul, South Korea, signed up for a 2-hour 1-day training program titled "Overcoming insomnia without medication: How to use the 'MIND MORE' mobile app for systematic self-management of insomnia" (pre-intervention). The training program was held on August 8, 2019 in an auditorium at the SMCD, which aims to provide preventative solutions for geriatric residents at normal or high risk of developing geriatric mental illness (not only dementia) as well as early detection for better treatment, rehabilitation support, and appropriate management in those at different stages of illness. For the 1-week home-based self-help CBT-I using the MIND MORE app (post-intervention), a therapist recruited a limited number of 9 volunteers from the entire 41 attendants of the 1-day training program. As the program was not provided for the purpose of academic research, program attendants gave their written consent for their participation, and those who lacked the capacity to provide their own consent to participate in the program were required to provide surrogate consent through the person who was a legally-authorized representative, particularly under the supervision of the SMCD.

To sign up for the center's program, the following requirements had to be satisfied: (1) male and female senior residents in Seodaemun-gu (aged ≥60 years); (2) either possess a smartphone or a feature phone (with more limited computing capabilities than smartphones) but were interested in smartphone-based self-help treatment for insomnia; and (3) with adequate literacy. For this retrospective study, the study protocol was approved after the program ended by the Institutional Review Board of Severance Hospital, Yonsei University College of Medicine in Seoul, South Korea.

For sensitive health-related information to be protected, the final version of the digitized datasheet (n=41) had personally identifiable information such as name, birthdate, and mobile phone or home number removed and then was forwarded to our research team by the chief therapist who was in charge of the education program. According to the exclusion criteria for the data analysis, 1 of 41 attendants was excluded due to missing data because we had no contact information to follow-up on the missing responses to baseline clinical and other demographic characteristics for that person.

MIND MORE App

Mobile Phone–Based CBT-I App

MIND MORE is a multimodal CBT-I app, with a greater focus on sleep education, with 4 components: (1) sleep hygiene education, (2) sleep restriction, (3) stimulus control, and (4) cognitive therapy. In the MIND MORE app, the main sleep hygiene education program is composed of 3 sessions (Figure 1), quizzing its users at the end of each session to evaluate what they learned (Figure 2). The sleep hygiene education helps to better understand the perpetuating mechanisms that sustain insomnia and to correct unhelpful, inflexible sleep-related beliefs and anxiety, conditioned arousal to the bed and bedroom, and sleep-disruptive habits such as daytime napping and spending excessive time in bed [20], which contributes to adherence to treatment recommendations provided by MIND MORE. All the MIND MORE contents referred to the CBT-I program that is currently implemented in the Yonsei University Health System, particularly based on the CBT-I protocol by Edinger and Carney [20].

The education was supplemented by a sleep diary and cognitive therapy, as well as sleep restriction and stimulus control contents. As a valuable tool for assessing insomnia, a sleep diary can prospectively monitor sleep habits and patterns over time (Figure 3), thus identifying good candidates for implementing cognitive and behavioral therapy strategies based on the data [20]. More importantly, previous research addressing the issue found that some older adults had difficulty in keeping pencil-and-paper sleep diaries, thereby leading to missing values [36]. It was expected that this issue might be handled by technical support from mobile technology. In addition to keeping a sleep diary, the cognitive therapy included recording a "thought record" for cognitive restructuring and a "constructive worry worksheet" for worry control (Figure 4), helping elderly users to easily complete their sleep information in the formatted blank spaces.

To increase the accessibility of information, the MIND MORE app provided its users with additional features such as "learning progress management" to monitor the progress rate and quiz scores from each of the 3 sessions and "clipping" to add specific pages containing important information to their list of "Scrap" and to directly access clipped pages (eg, sleep diary, thought record, or constructive worry worksheet), particularly in the hamburger menu (Figure 5). The MIND MORE app can be downloaded in both the Google Play store for Android phone users and the App Store for Apple phone users. Multimedia Appendices 1–5 provide all MIND MORE app figures translated to English.

Figure 1. Sleep hygiene education program with 3 main sessions.
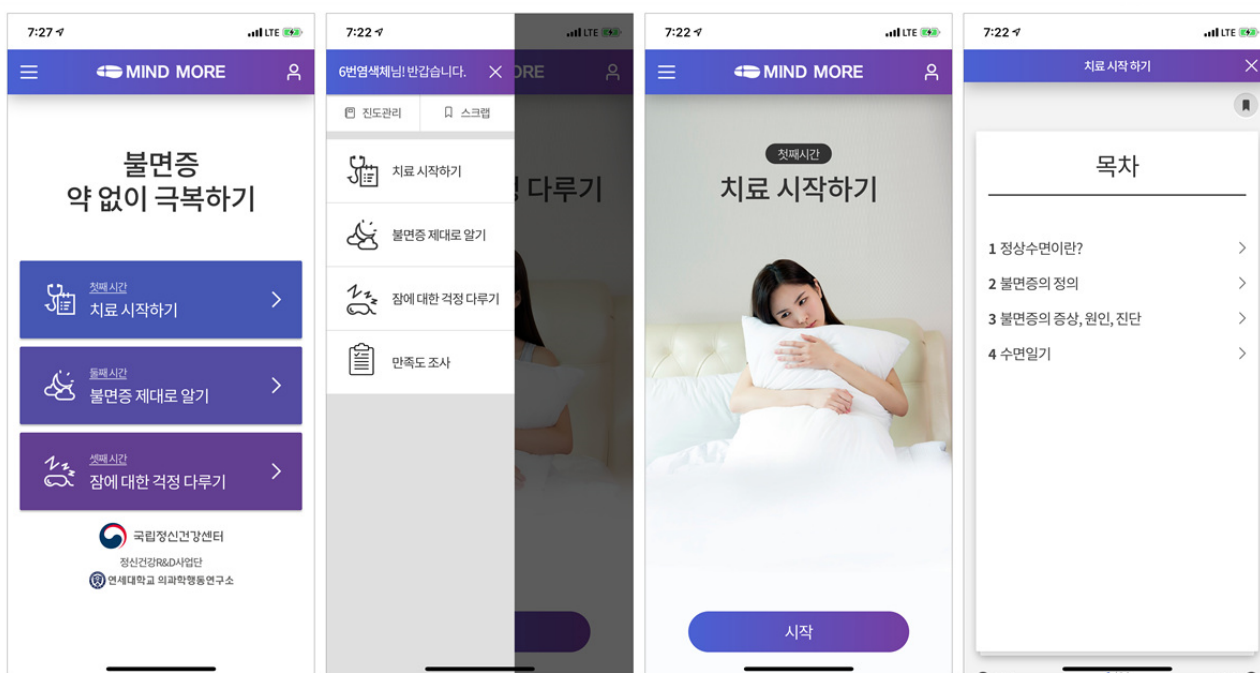
Figure 2. Quiz session after the end of each of the 3 main sleep education sessions.
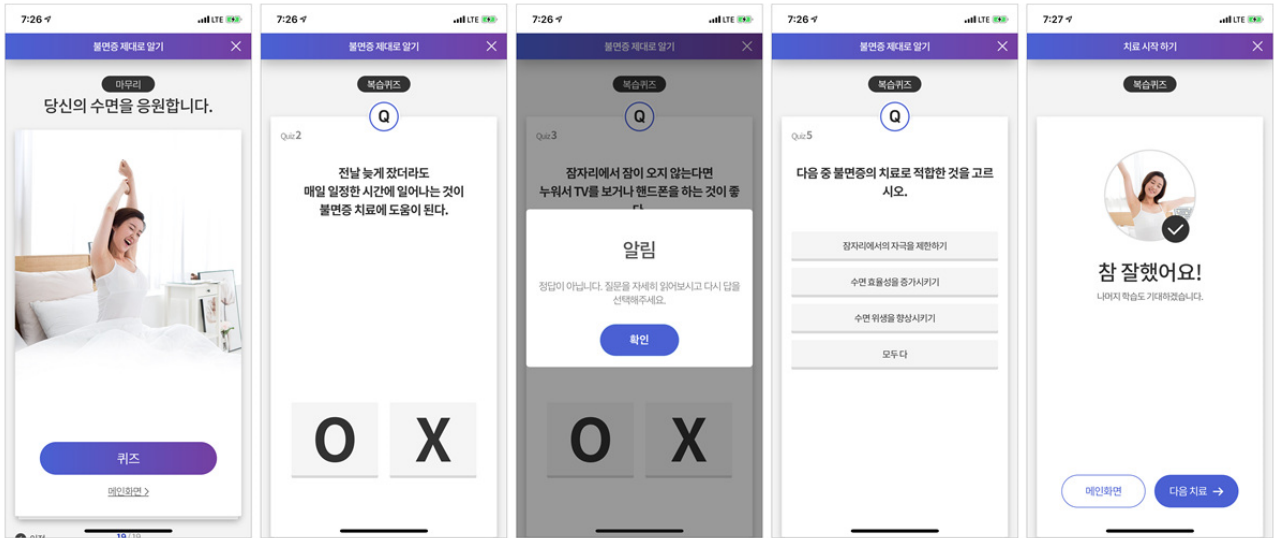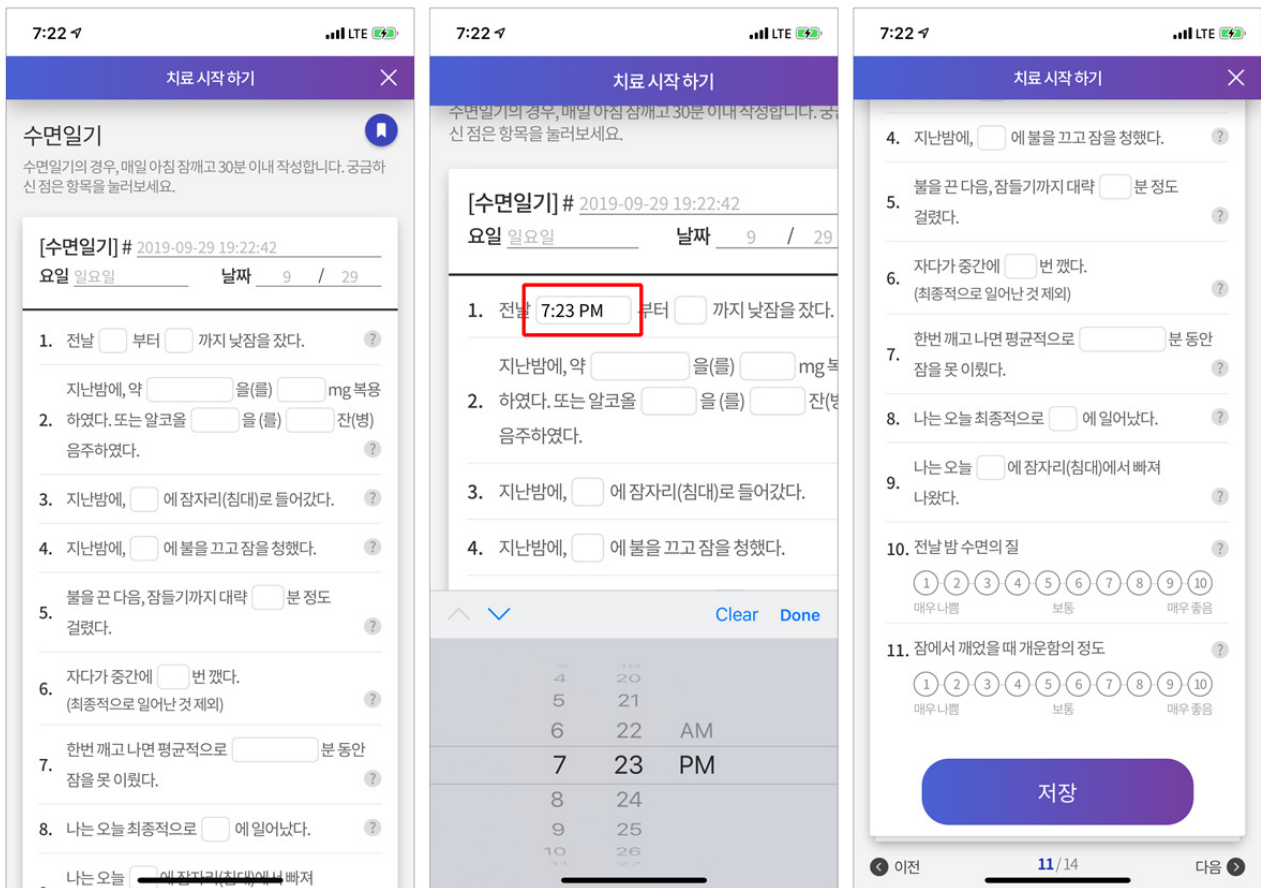

Figure 3. Sleep diary.

Figure 4. Cognitive therapy with (A) thought record and (B) constructive worry worksheet.
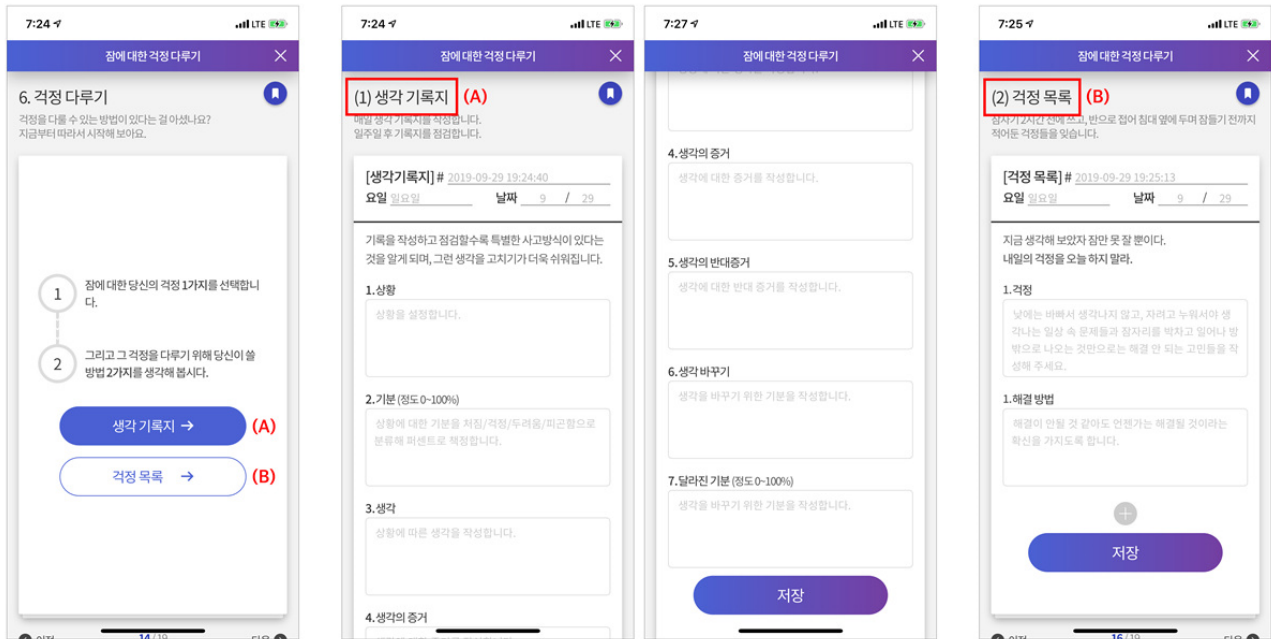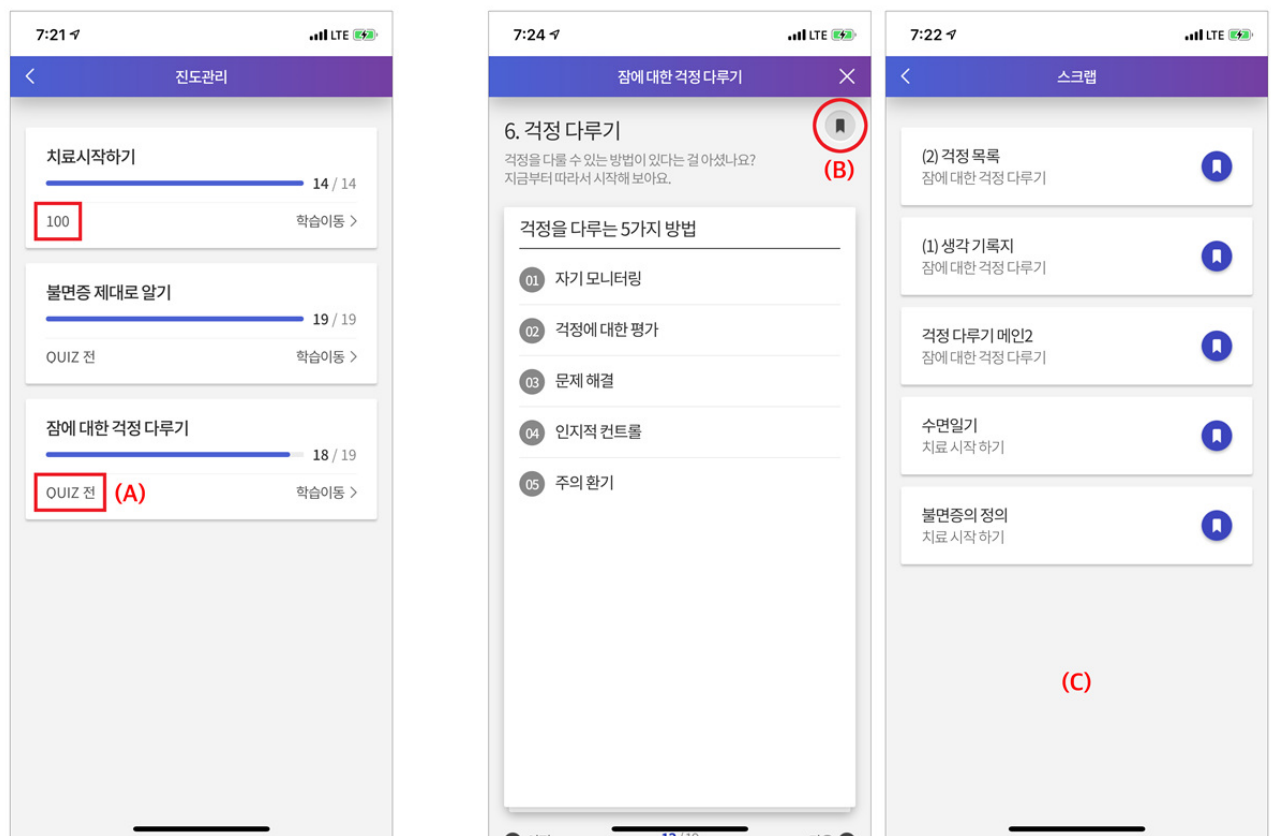

Figure 5. Additional features of the MIND MORE app: (A) learning progress management, (B) clipping button, and (C) list of clipped pages.

Dashboard

Based on the stepped care model [24,25], the MIND MORE app was initially designed to only allow users to reflect on the logs of their sleep diary, thought record, and constructive worry worksheet. If there were few changes in sleep quality, the users were expected to share all the logs with therapists or clinicians with whom they had established good rapport and then to receive more direct, systematic management, thus changing a pure self-help intervention into a therapist-guided self-help intervention. In this context, its dashboard was limited to display the following information: (1) registration and last login dates, (2) learning progress checks including last learning dates and quiz scores in each of the 3 sessions, and (3) behavioral and cognitive therapy records including the last dates and number of times users completed their sleep diary, thought record, and

constructive worry worksheet as well as the total number of written constructive worries. If the users withdrew from MIND MORE membership, their data were no longer displayed in the dashboard and were removed from the server.

Procedure

The entire procedure for this retrospective study was as follows. At the SMCD in Seodaemun-gu, a 2-hour training program for use of the MIND MORE CBT-I app was held over the course of 1 day; during the program, the attendants gained hands-on experience with the MIND MORE app with the help of therapists and volunteer workers, from installing and signing up for the app to learning its functions and features (pre-intervention). At the end of the program, all 41 attendants were required to complete the pencil-and-paper questionnaires on sociodemographic characteristics and subjective evaluation of mental health status (ie, depression, memory loss and impairment, and sleep problems) and usability of the MIND MORE. Of the 41 attendants, 9 were enrolled in the 1-week home-based self-help CBT-I where the 9 attendants used the MIND MORE app for a week based on what they learned during the 1-day training program. They were guided to fill out and submit the given pencil-and-paper questionnaires on subjective evaluation of sleep quality after the 1-week program (pre-intervention), specifically 8 days after the 1-day training program ended (post-intervention). For the 8-week follow-up period, adherence of all attendants to the app was monitored in the dashboard without any direct therapist contact. For this retrospective study, our research team was provided with a coded datasheet by the SMCD.

Subjective Measurements

Short-Form Geriatric Depression Scale

The 15-question brief version of the Geriatric Depression Scale, known as GDS-15, is a widely used screening measure for depression in the geriatric population [37,38]. In this study, the Korean version of the GDS-15 (SGDS-K) was employed in the community-dwelling elderly in Korea, with an optimal cut-off point of 8 for both normal and clinical populations [39,40]. All attendants were asked to answer each question of the 15-item SGDS-K with either "yes" (1) or "no" (0) in reference to how they felt over the past week.

Subjective Memory Complaints Questionnaire

The Subjective Memory Complaints Questionnaire (SMCQ) is a brief, reliable, and valid questionnaire for the global assessment of memory function (global memory) and the specific judgment of memories of particular events (everyday memory), with the ability to screen for dementia [41]. The SMCQ consists of a total of 14 items: 4 items for global memory and 10 items for everyday memory. Each item is answered with either "yes" (1) or "no" (0), and the optimal cut-off point for dementia is 5 for elderly without dementia and 6 for demented elderly.

Pittsburgh Sleep Quality Index

The Pittsburgh Sleep Quality Index (PSQI) is composed of 19 self-rated questions and 5 questions rated by a roommate or bed partner (if possible), and only self-rated questions are used for

the scoring [42]. According to the given scoring sheet, the 19 self-rated items are combined to calculate the following 7 components: (1) subjective sleep quality, (2) sleep latency, (3) sleep duration, (4) habitual sleep efficiency, (5) sleep disturbance, (6) use of sleeping medication, and (7) daytime dysfunction. Each component score is coded as 0, 1, 2, or 3. Then, the 7 component scores are summated to yield the global PSQI score. This study administered the Korean version of the PSQI (PSQI-K) using the best cutoff score of 8.5 [43]. All questions were related to usual sleep habits for the past month only, and attendants were guided to choose the appropriate reply for the majority of days and nights during the past month as accurately as possible.

Perceived Difficulty in Mobile App Use

To evaluate perceived difficulty in mobile app use, all elderly attendants both with and without their own smartphones responded to the question of "Do you have difficulty in installing, updating, and removing mobile apps on your own smartphone by yourself?" scored on a 5-point scale anchored by 0 (not at all), with scores ranging to 4 (always). In particular, feature phone users were guided to imagine that they would have their own smartphone and then answer the question. This question was included in the questionnaire for demographic information.

Usefulness, Satisfaction, and Ease of Use Questionnaire

To assess software usability, we used the Usefulness, Satisfaction, and Ease of Use (USE) questionnaire [44], consisting of 4 subfactors: usefulness (5 items), ease of use (4 items), ease of learning (3 items), and satisfaction (5 items). All attendants were asked to rate a series of attitude statements for agreement, using 5-point Likert rating scales ranging from 0 (strongly disagree) to 4 (strongly agree). The arithmetic mean was used to calculate each subfactor score and the total usability score.

Statistical Analysis

All statistical analyses were performed using PASW Statistics 18.0 software (SPSS Inc. Chicago, IL). To test the normal distribution of variables, Z scores for skewness (ie, skewness/SE of skewness) and kurtosis (ie, kurtosis/SE of kurtosis) were calculated. If the sample size is small (N<200), an absolute Z score value >1.96 is insufficient to assume normality of the data ($P<.05$), which means that a nonparametric statistical test should be used, based on the rule of thumb for normality testing by Ghasemi and Zahediasl [45]. According to the results of the normality tests for the 40 1-day program attendants, the normality assumptions of the SGDS-K, SMCQ, and PSQI-K scores were violated, while all the scores for the perceived difficulty in mobile app use and the USE questionnaire including the 4 subfactors and total usability scores met the assumption. For the 9 attendants in the 1-week home-based self-help CBT-I using the MIND MORE app, the PSQI-K scores for both pre-intervention and post-intervention were normally distributed. Hence, the correlation between variables was determined using statistical tests: the Spearman rank correlation as a nonparametric statistical test and Pearson correlation as a

parametric statistical test. The paired-samples t test, as a parametric method, was used for hypothesis testing.

Results

Participant Characteristics

Due to the occurrence of missing data, 40 (35 women) of 41 program attendants, aged 64 to 86 years (mean 75.75 years, SD 5.87 years), were included in the data analysis. Of the 40 attendants, the mean age of the 9 female attendants who were involved in the 1-week home-based self-help CBT-I using the MIND MORE app was 71.56 years (SD 4.36 years; range 65-78 years). This study sample was highly educated in that the majority of the attendants (25/40, 63%) reported that the highest level of education was high school. In terms of the prevalence of depressive symptoms, memory complaints, and sleep problems at baseline, 3 (3/40, 8%), 7 (7/40, 18%), and 15 (15/40, 38%) had scores greater than the cut-off points for the SGDS-K, SMCQ, and PSQI-K, respectively.

In terms of mobile phone possession, 38 had Android smartphones (38/40, 95%) including the brands Samsung (29/38, 76%) and LG (9/38, 64%), and the rest had feature phones (2/40, 5%) on which mobile apps were not available. Regardless of the type of phone the elderly attendants owned, the amount of time they spent using their own mobile phones was reported as follows: <1 hour (20/40, 50%; 2 feature phone users were included), 1-2 hours (12/40, 30%), 2-3 hours (2/40, 5%), 3-4 hours (3/40, 8%), and >4 hours (3/40, 8%). In addition to these basic demographic characteristics, more detailed information is presented in Table 1.

Table 1. Demographic characteristics of the study sample.

Characteristics	One-day training program for MIND MORE use (n=40)	One-week self-help intervention with MIND MORE (n=9)
Gender, n (%)		
Male	5 (13)	0 (0)
Female	35 (88)	9 (100)
Age, n (%)		
Young-old (60-69 years)	8 (20)	4 (44)
Old-old (70-79 years)	22 (55)	5 (56)
Oldest-old (80-89 years)	10 (25)	0 (0)
Marital status, n (%)		
Married	23 (58)	3 (33)
Widowed	17 (43)	6 (67)
Occupation, n (%)		
Employed (full-time)	2 (5)	2 (22)
Employed (part-time)	1 (3)	0 (0)
Housewife	25 (63)	6 (67)
Retired	3 (8)	0 (0)
Unemployed	9 (23)	1 (11)
Educational level, n (%)		
Elementary school	6 (15)	1 (11)
Middle school	9 (23)	3 (33)
High school	14 (35)	5 (56)
College (2-3 years)	3 (8)	0 (0)
University (4-5 years)	6 (15)	0 (0)
Postgraduate (Masters/Doctoral)	2 (5)	0 (0)
Psychiatric diagnosis (yes)[a], n (%)		
Alzheimer's disease	3 (8)	1 (11)
Major depressive disorder	5 (13)	0 (0)
Anxiety disorder	1 (3)	0 (0)
Sleep disorder	1 (4)	0 (0)
SGDS-K[b] (score), mean (SD)	2.00 (2.73)	1.00 (1.12)
SMCQ[c] (score), mean (SD)	2.33 (2.67)	1.67 (1.58)
PSQI-K[d] (score), mean (SD)	7.50 (4.12)	8.00 (2.50)
Perceived difficulty in mobile app use (score), mean (SD)	2.53 (1.28)	2.11 (1.17)
Social support for mobile phone use (yes)[a], n (%)		
Child	25 (63)	8 (2)
Grandchild	8 (20)	2 (5)
Daughter-in-law	3 (8)	1 (3)
Spouse	3 (8)	0 (0)
Acquaintance	7 (18)	2 (5)
None	2 (5)	0 (0)

[a]Multiple responses were allowed.

[b]SGDS-K: Korean version of the 15-question Geriatric Depression Scale (cut-off score of 8).

[c]SMCQ: Subjective Memory Complaints Questionnaire (cut-off score of 5 for elderly without dementia and 6 for demented elderly).

[d]PSQI-K: Korean version of the Pittsburgh Sleep Quality Index (cut-off score of 8.5).

Relationships Between Subjective Sleep Quality and Subjective Memory Complaints and Self-Reported Depressive Symptoms

As shown in Figure 6, the Spearman rank correlation coefficients between the PSQI-K and SGDS-K scores and between the PSQI-K and SMCQ scores were .60 ($P<.001$) and .46 ($P=.003$; 2-tailed), respectively, for all 40 elderly attendants. Both showed moderate positive correlations; however, subjective sleep quality of individuals was more strongly associated with change in affective state than in cognitive function.

Figure 6. Scatter plots showing the Spearman rank correlations (n=40) between (A) subjective sleep quality and subjective memory complaints and (B) subjective sleep quality and self-reported depressive symptom severity. PSQI-K: Korean version of the Pittsburgh Sleep Quality Index; SGDS-K: Korean version of the 15-question Geriatric Depression Scale; SMCQ: Subjective Memory Complaints Questionnaire.
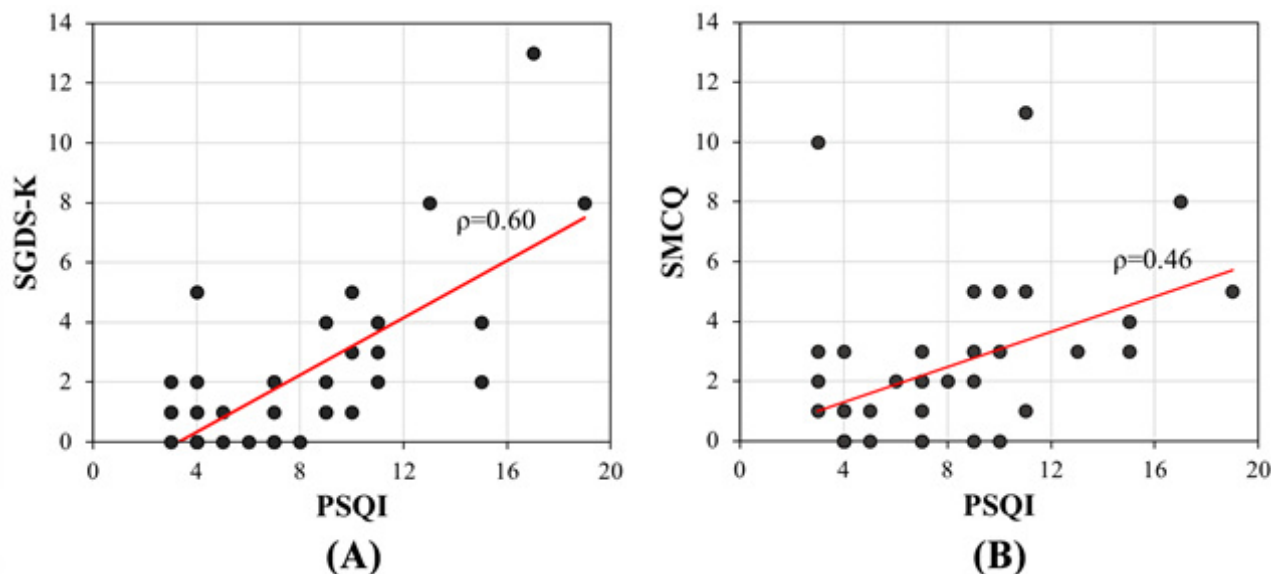
Relationship Between Perceived Difficulty in Mobile App Use and Usability of MIND MORE

Table 2 shows the scores for perceived difficulty in using the mobile app and usability (with the 4 subfactors: usefulness, ease of use, ease of learning, and satisfaction) of the MIND MORE app, and Table 3 shows the results of the Pearson correlation analysis (2-tailed) between the scores for perceived difficulty in using the mobile app and usability (with the 4 subfactors: usefulness, ease of use, ease of learning, and satisfaction). There was only the significant negative correlation between perceived difficulty and ease of learning; ease of learning was significantly, positively correlated with ease of use, satisfaction, and usability scores.

Table 2. Perceived difficulty in using the mobile app and Usefulness, Satisfaction, and Ease of Use (USE) questionnaire score and subfactor scores (n=40).

Factor	Scores, mean (SD)
PD[a]	2.53 (1.28)
Usefulness	2.90 (0.51)
EOU[b]	2.01 (0.86)
EOL[c]	1.85 (0.72)
SATIS[d]	2.56 (0.68)
Total USE score	2.33 (0.56)

[a]PD: perceived difficulty in using the mobile app.

[b]EOU: ease of use.

[c]EOL: ease of learning.

[d]SATIS: satisfaction.

XSL•FO
RenderX

Table 3. Pearson correlation coefficients of the relationships between perceived difficulty in using the mobile app and the Usefulness, Satisfaction, and Ease of Use (USE) questionnaire score and subfactor scores (n=40).

Factor	PD[a]	Usefulness	EOU[b]	EOL[c]	SATIS[d]	Total USE score
PD						
r	1	0.13	–0.17	–0.37	0.13	–0.12
P value	__e	.42	.28	.02	.44	.46
Usefulness						
r	0.13	1	0.32	0.29	0.50	0.60
P value	.42	__e	.047	.07	<.001	<.001
EOU						
r	–0.17	0.32	1	0.79	0.62	0.91
P value	.28	.047	__e	<.001	<.001	<.001
EOL						
r	–0.37	0.29	0.79	1	0.39	0.82
P value	.02	.07	<.001	__e	.013	<.001
SATIS						
r	0.13	0.50	0.62	0.39	1	0.79
P value	.44	<.001	<.001	.013	__e	<.001
Total USE score						
r	–0.12	0.60	0.91	0.82	0.79	1
P value	.46	<.001	<.001	<.001	<.001	__e

[a]PD: perceived difficulty in using the mobile app.

[b]EOU: ease of use.

[c]EOL: ease of learning.

[d]SATIS: satisfaction.

[e]Not applicable.

Comparison of Changes in Subjective Sleep Quality From Pre-Intervention to Post-Intervention

The result of a 2-tailed paired-samples *t* test revealed that there was a significant mean difference in the subjective evaluation of sleep quality between pre-intervention (mean 8.00, SD 2.50) and post-intervention (mean 5.11, SD 1.36), indicating that using the MIND MORE app for a week led to improved sleep quality in the elderly attendants (t_8=3.74, *P*=.006).

Adherence to the MIND MORE Mobile App

To quantify the adherence rate to the self-help MIND MORE CBT-I app during an 8-week follow-up, the registration and last login dates on the app for all 35 program attendants with their own smartphones were downloaded from the app dashboard. The total number of days from the registration date to the last login date within the 8-week period was counted. As listed in Table 4, total days of use were categorized into 6 groups: <1 day, 1-6 days (1 week), 7-13 days (1-2 weeks), 14-27 days (2-4 weeks), 28-55 days (4-8 weeks), and >56 days (>8 weeks). After that, the adherence rate was determined. In the same way, the adherence rate within the same period for the 9 attendants involved in the 1-week intervention was determined. In Table 4, the results of adherence to the app are presented more in detail.

Table 4. Elderly users' adherence to the MIND MORE app without any direct therapist contact.

Total days of use	One-day training program for MIND MORE use (n=35)[a], n (%)	One-week self-help intervention with MIND MORE (n=9), n (%)
<1 day (on-site experience)	18 (51)	3 (33)
1-6 days (<1 week)	2 (6)	2 (22)
7-13 days (1-2 weeks)	2 (6)	1 (11)
14-27 days (2-4 weeks)	5 (14)	2 (22)
28-55 days (4-8 weeks)	6 (17)	1 (11)
>56 days (>8 weeks)	2 (6)	0 (0)

[a]Of the 40 attendants, 2 were feature phone users, and 3 withdrew from MIND MORE app service membership; these 5 attendants were excluded.

Discussion

Principal Findings

The main findings of this study were as follows. First, the newly developed MIND MORE app was usable among community-dwelling older Korean adults. Second, as hypothesized, subjective sleep quality significantly improved from pre-intervention to post-intervention, particularly by using the mobile phone–based self-help CBT-I app we developed during only a 1-week intervention period. Third, except for the attendants who did not have their own smartphones (2/40) and withdrew from their MIND MORE membership (3/40), those in the 1-day sleep education program adhered to the MIND MORE app for at least 2 weeks (13/35, 37%) to 8 weeks (2/35, 6%) without any further contact.

In line with previous studies mentioned in the Introduction [7,8], this study showed that subjective sleep quality is more closely linked to depressive symptoms than to memory complaints. Given the high prevalence of depression in elderly women [6,7], it is important to help with timely prevention or management of sleep problems. Based on the characteristics of this study sample, most program attendants (35/40, 88%) were women, which is considered a positive outcome given that they are vulnerable to depression and sleep disturbance and are trying to self-manage their mental issues. According to the policy report of the 2017 National Survey of Older Koreans by the

Korea Institute for Health and Society Affairs [46], female older adults, who are more likely to have ≥3 comorbidities or be in a poorer health state, were the main users of senior citizen centers and senior welfare centers, compared with male counterparts. In contrast, male older adults were more satisfied with the status of their subjective health, as well as that of their social, leisure, and cultural activities, than female counterparts. In line with these survey findings, this study also revealed that elderly men had a relatively low participation rate in the community center programs, whereas elderly women tried to stay active and seek adequate treatment for their mental health concerns. In this regard, the issue of how to increase accessibility to the MIND MORE app for older male adults would be of great importance to be addressed in future studies.

To determine whether cognitive and behavioral changes were prompted through the mobile app–based CBT-I, the 9 attendants in the 1-week home-based self-help CBT-I intervention were asked to report their user experience with the MIND MORE app by using the given open-ended questionnaires, particularly on the effectiveness of mobile-based sleep hygiene education and the sleep diary. According to the following responses, sleep restriction and stimulus control were used to improve sleep quality (Textbox 1).

Moreover, the sleep diary played a key role in identifying appropriate candidates to implement CBT-I strategies based on the users' records of sleep habits and patterns over time [20] (Textbox 2).

Textbox 1. Responses regarding sleep restriction and stimulus control to improve sleep quality.

> *Since I haven't drunk coffee late in the evening, I could sleep well.* [A4]
>
> *I only drank a cup of coffee in the morning, and I didn't take a nap at all.* [A6]
>
> *When I was too tired, I used to take a nap. But…since I tried not to take too much naps, I got to fall asleep well at night.* [A3]
>
> *I was so hard to wake up too early in the morning, but sleep restriction improved my sleep a bit.* [A3]

Textbox 2. Responses regarding the usefulness of the sleep diary.

> *I tried to put what I've learned from the program into practice, so my insomnia almost went away these days.* [A6]
>
> *It doesn't seem to help very much. I don't take a nap and drink coffee at all, because I'm afraid that I won't be able to fall asleep.* [A7]

As reflected in the comments of our respondents, individual differences in sleep problems should be carefully considered when implementing a mobile app–based self-help CBT-I without therapist intervention. To enhance the treatment effects, it is

recommended to include a pretreatment assessment session to identify what causes or maintains sleep problems by conducting a clinical interview or completing a sleep history questionnaire or sleep diary before the start of the first session of sleep hygiene psychoeducation. After that, appropriate CBT-I components should be target outcomes, such as automatic thoughts or behaviors.

Most of all, because older adults perceive mobile app use as more difficult in everyday life, they were more likely to have difficulty learning how to use the MIND MORE app; therefore, it would be important to implement more education programs to familiarize elderly adults with mobile devices and apps in community centers, which in turn may positively influence perceived usability and mHealth technology self-efficacy. Consistent with the finding of the study by Lund [44], this study revealed that ease of learning was highly correlated with ease of use. Furthermore, once the elderly novice users perceived that they could easily and quickly learn to use the MIND MORE app (ease of learning), the MIND MORE app was more likely to be perceived as easy to use, as providing satisfaction, and as usable to achieve goals with effectiveness, efficiency, and satisfaction in a specified context to manage sleep problems (usability).

Although it was expected that the app's high learnability and mobility might contribute to easy access to this digital therapeutic from older users' own mobile phones, this study had a high attrition rate, similar to other recent studies [47-49] that also raised concerns about internet-based and mobile-based self-help CBT interventions. As one possible explanation for the high attrition rate, the MIND MORE app was designed as a multimodal, but more psychoeducation-focused, CBT-I app with less interactive, more informative educational content to provide app users with credible health information sourced from health professionals. For this reason, users might be less likely to continue using the main content of the sleep hygiene education after the completion of all 3 sessions, except when reviewing clipped educational content, keeping the sleep diary, or completing the thought record and constructive worry worksheet on a regular basis to track their progress over time. Another possible explanation is that the high attrition rate could result from the Google operating system or Play Store feature that might suggest the uninstallation of unused apps to optimize the operation of Android phones with a relatively low specification, even if this was slightly different depending on the mobile phone and iOS version. Another possible explanation is that this high attrition rate might be the outcome of learning how to withdraw from the MIND MORE app service, as well as how to install and uninstall the app from their devices during the 1-day training program. As this study failed to reveal the cause of the high attrition rate and what will contribute to improve retention, future research needs to collect the users' feedback as a result of their user experiences by conducting a focus group or in-depth interview after the intervention phase, thereby effectively managing expectations and enhancing treatment adherence.

Limitations and Future Direction

As a retrospective study, this study had some limitations based on the study protocol and sample. This sleep program protocol using the MIND MORE app was not designed for the purpose of academic research, but without particular focus on examining the efficacy of the mobile app–based CBT-I for a long-term intervention period. First, the assessment of subjective sleep quality was administered at baseline (1-day training program at the SMCD) and 1 week after the 1-week home-based self-help CBT-I. Considering that the PSQI-K is an explicit measure of usual sleep habits over the last month, it is possible that the intervention period was insufficient to examine improvements in subjective sleep quality related to use of the MIND MORE app. Moreover, PSQI-K scores should be compared between two intervention groups, namely a nonclinical or subclinical group and a group of patients with insomnia, before and after use of the MIND MORE app because the intervention group (n=9) who completed the 1-week home-based self-help CBT-I consisted of only normal-risk and high-risk older adults (Table 1).

Second, particularly when using this digital therapeutic for insomnia in clinical and academic research, a digital placebo effect and the need for an appropriate control group should be considered in future studies. Accordingly, an experimental group using the MIND MORE app should be compared with an active control group using a credible sham app, not with an inactive control group in a wait-list condition, in a double-blind, randomized controlled clinical trial. Regarding concerns that this study sample was biased to the elderly, who were more interested in digital technology use itself, mobile phone–based mental health intervention, or participation in community center activities, the effectiveness of the MIND MORE app will be investigated in a randomly selected community sample of the older population. Furthermore, the number of attendants involved in the 1-week intervention was too small to conduct statistical analysis, and the ratio of men to women was imbalanced.

Finally, the study sample was derived from a population of community-dwelling older adults in a single urban area. Differences might exist between interurban areas or between urban and rural areas. For these reasons, the findings of this study cannot be generalized to community-dwelling older adults in the general and clinical population without further research.

Conclusions

Despite this retrospective approach to data sourced from a mobile app–based CBT-I program, which was not designed for research purposes, in an urban community center for dementia, this study provides empirical evidence that the newly developed MIND MORE app was not only usable among older users but also could improve their subjective sleep quality after the 1-week self-help intervention period. Based on the findings of this study, more diverse research protocols should be designed and administered to examine the treatment effects of the mobile self-help CBT-I intervention as an alternative, first-line treatment for insomnia, thus leading to translation from research into practice.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1
English-translated version of screen shots for the MIND MORE app (Figure 1).
[PNG File , 130 KB - mhealth_v8i8e17755_app1.png]

Multimedia Appendix 2
English-translated version of screen shots for the MIND MORE app (Figure 2).
[PNG File , 195 KB - mhealth_v8i8e17755_app2.png]

Multimedia Appendix 3
English-translated version of screen shots for the MIND MORE app (Figure 3).
[PNG File , 236 KB - mhealth_v8i8e17755_app3.png]

Multimedia Appendix 4
English-translated version of screen shots for the MIND MORE app (Figure 4).
[PNG File , 249 KB - mhealth_v8i8e17755_app4.png]

Multimedia Appendix 5
English-translated version of screen shots for the MIND MORE app (Figure 5).
[PNG File , 123 KB - mhealth_v8i8e17755_app5.png]

References

1.   Schubert C, Cruickshanks K, Dalton D, Klein B, Klein R, Nondahl D. Prevalence of sleep problems and quality of life in an older population. Sleep 2002 Dec;25(8):889-893. [Medline: 12489896]

2.   de Almondes KM, Costa MV, Malloy-Diniz LF, Diniz BS. Insomnia and risk of dementia in older adults: Systematic review and meta-analysis. J Psychiatr Res 2016 Jun;77:109-115. [doi: 10.1016/j.jpsychires.2016.02.021] [Medline: 27017287]

3.   Bombois S, Derambure P, Pasquier F, Monaca C. Sleep disorders in aging and dementia. J Nutr Health Aging 2010 Mar;14(3):212-217. [doi: 10.1007/s12603-010-0052-7] [Medline: 20191256]

4.   Tractenberg RE, Singer CM, Kaye JA. Symptoms of sleep disturbance in persons with Alzheimer's disease and normal elderly. J Sleep Res 2005 Jun;14(2):177-185 [FREE Full text] [doi: 10.1111/j.1365-2869.2005.00445.x] [Medline: 15910515]

5.   Perlis ML, Smith LJ, Lyness JM, Matteson SR, Pigeon WR, Jungquist CR, et al. Insomnia as a risk factor for onset of depression in the elderly. Behav Sleep Med 2006;4(2):104-113. [doi: 10.1207/s15402010bsm0402_3] [Medline: 16579719]

6.   Kim KW, Kang S, Yoon I, Lee SD, Ju G, Han JW, et al. Prevalence and clinical characteristics of insomnia and its subtypes in the Korean elderly. Arch Gerontol Geriatr 2017;68:68-75. [doi: 10.1016/j.archger.2016.09.005] [Medline: 27665575]

7.   Kim WJ, Joo W, Baek J, Sohn SY, Namkoong K, Youm Y, et al. Factors Associated with Insomnia among the Elderly in a Korean Rural Community. Psychiatry Investig 2017 Jul;14(4):400-406 [FREE Full text] [doi: 10.4306/pi.2017.14.4.400] [Medline: 28845165]

8.   Kim W, Kim J, Kim B, Chang S, Lee D, Cho M, et al. The role of depression in the insomnia of people with subjective memory impairment, mild cognitive impairment, and dementia in a community sample of elderly individuals in South Korea. Int Psychogeriatr 2017 Apr;29(4):653-661. [doi: 10.1017/S1041610216002076] [Medline: 27921991]

9.   Spira AP, Kaufmann CN, Kasper JD, Ohayon MM, Rebok GW, Skidmore E, et al. Association between insomnia symptoms and functional status in U.S. older adults. J Gerontol B Psychol Sci Soc Sci 2014 Nov;69 Suppl 1:S35-S41 [FREE Full text] [doi: 10.1093/geronb/gbu116] [Medline: 25342821]

10. Beaulieu-Bonneau S, Ivers H, Guay B, Morin CM. Long-Term Maintenance of Therapeutic Gains Associated With Cognitive-Behavioral Therapy for Insomnia Delivered Alone or Combined With Zolpidem. Sleep 2017 Mar 01;40(3) [FREE Full text] [doi: 10.1093/sleep/zsx002] [Medline: 28364426]

11. Morin CM, Colecchi C, Stone J, Sood R, Brink D. Behavioral and pharmacological therapies for late-life insomnia: a randomized controlled trial. JAMA 1999 Mar 17;281(11):991-999. [doi: 10.1001/jama.281.11.991] [Medline: 10086433]

12. Jacobs GD, Pace-Schott EF, Stickgold R, Otto MW. Cognitive behavior therapy and pharmacotherapy for insomnia: a randomized controlled trial and direct comparison. Arch Intern Med 2004 Sep 27;164(17):1888-1896. [doi: 10.1001/archinte.164.17.1888] [Medline: 15451764]

13. Sivertsen B, Omvik S, Pallesen S, Bjorvatn B, Havik OE, Kvale G, et al. Cognitive behavioral therapy vs zopiclone for treatment of chronic primary insomnia in older adults: a randomized controlled trial. JAMA 2006 Jun 28;295(24):2851-2858. [doi: 10.1001/jama.295.24.2851] [Medline: 16804151]

14. Mitchell MD, Gehrman P, Perlis M, Umscheid CA. Comparative effectiveness of cognitive behavioral therapy for insomnia: a systematic review. BMC Fam Pract 2012;13:40 [FREE Full text] [doi: 10.1186/1471-2296-13-40] [Medline: 22631616]

15. Trauer JM, Qian MY, Doyle JS, Rajaratnam SMW, Cunnington D. Cognitive Behavioral Therapy for Chronic Insomnia: A Systematic Review and Meta-analysis. Ann Intern Med 2015 Aug 04;163(3):191-204. [doi: 10.7326/M14-2841] [Medline: 26054060]

16. Kales A, Manfredi RL, Vgontzas AN, Bixler EO, Vela-Bueno A, Fee EC. Rebound insomnia after only brief and intermittent use of rapidly eliminated benzodiazepines. Clin Pharmacol Ther 1991 Apr;49(4):468-476. [doi: 10.1038/clpt.1991.55] [Medline: 2015735]

17. Tariq SH, Pulisetty S. Pharmacotherapy for insomnia. Clin Geriatr Med 2008 Feb;24(1):93-105, vii [FREE Full text] [doi: 10.1016/j.cger.2007.08.009] [Medline: 18035234]

18. Morin CM, Espie CA. Insomnia: A clinical guide to assessment and treatment. New York, NY: Springer Science & Business Media; 2007.

19. Morin CM, Benca R. Chronic insomnia. Lancet 2012 Mar 24;379(9821):1129-1141. [doi: 10.1016/S0140-6736(11)60750-2] [Medline: 22265700]

20. Edinger J, Carney C. Overcoming insomnia: A cognitive-behavioral therapy approach, therapist guide. New York: Oxford University Press; 2014:A.

21. Kuhn E, Weiss BJ, Taylor KL, Hoffman JE, Ramsey KM, Manber R, et al. CBT-I Coach: A Description and Clinician Perceptions of a Mobile App for Cognitive Behavioral Therapy for Insomnia. J Clin Sleep Med 2016;12(4):597-606. [doi: 10.5664/jcsm.5700] [Medline: 26888586]

22. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]

23. Trockel M, Karlin BE, Taylor CB, Manber R. Cognitive Behavioral Therapy for insomnia with Veterans: evaluation of effectiveness and correlates of treatment outcomes. Behav Res Ther 2014 Feb;53:41-46. [doi: 10.1016/j.brat.2013.11.006] [Medline: 24412462]

24. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry 2005 Jan;186:11-17 [FREE Full text] [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]

25. Espie CA. "Stepped care": a health technology solution for delivering cognitive behavioral therapy as a first line insomnia treatment. Sleep 2009 Dec;32(12):1549-1558 [FREE Full text] [doi: 10.1093/sleep/32.12.1549] [Medline: 20041590]

26. Chen K, Chan AHS. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014;57(5):635-652. [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]

27. Kim S, Gajos K, Muller M, Grosz B. Acceptance of mobile technology by older adults: a preliminary study. 2016 Sep Presented at: 18th International Conference on Human-Computer Interaction with Mobile Devices and Services; September 6-9, 2016; Florence, Italy p. 147-157. [doi: 10.1145/2935334.2935380]

28. Tang C, Leung R, Haddad S, McGrenere J. What motivates older adults to learn to use mobile phones. 2012. URL: https://charlottetang.files.wordpress.com/2017/05/grand_mobile_phone_final.pdf [accessed 2020-08-02]

29. Babson KA, Ramo DE, Baldini L, Vandrey R, Bonn-Miller MO. Mobile App-Delivered Cognitive Behavioral Therapy for Insomnia: Feasibility and Initial Efficacy Among Veterans With Cannabis Use Disorders. JMIR Res Protoc 2015;4(3):e87 [FREE Full text] [doi: 10.2196/resprot.3852] [Medline: 26187404]

30. Chen Y, Hung Y, Chen H. Mobile Application-Assisted Cognitive Behavioral Therapy for Insomnia in an Older Adult. Telemed J E Health 2016 Apr;22(4):332-334. [doi: 10.1089/tmj.2015.0064] [Medline: 26393902]

31. Kang S, Kang JM, Cho S, Ko K, Lee YJ, Lee H, et al. Cognitive Behavioral Therapy Using a Mobile Application Synchronizable With Wearable Devices for Insomnia Treatment: A Pilot Study. J Clin Sleep Med 2017 Apr 15;13(4):633-640 [FREE Full text] [doi: 10.5664/jcsm.6564] [Medline: 28162145]

32. Reilly ED, Robinson SA, Petrakis BA, Kuhn E, Pigeon WR, Wiener RS, et al. Mobile App Use for Insomnia Self-Management: Pilot Findings on Sleep Outcomes in Veterans. Interact J Med Res 2019 Jul 24;8(3):e12408 [FREE Full text] [doi: 10.2196/12408] [Medline: 31342904]

33. Wildenbos GA, Jaspers MWM, Schijven MP, Dusseljee-Peute LW. Mobile health for older adult patients: Using an aging barriers framework to classify usability problems. Int J Med Inform 2019 Dec;124:68-77 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.01.006] [Medline: 30784429]

34. Andrews JA, Brown LJ, Hawley MS, Astell AJ. Older Adults' Perspectives on Using Digital Technology to Maintain Good Mental Health: Interactive Group Study. J Med Internet Res 2019 Feb 13;21(2):e11694 [FREE Full text] [doi: 10.2196/11694] [Medline: 30758292]

35. Lee D. Seoul dementia management project and Seoul metropolitan center for dementia. J Korean Geriatr Psychiatry 2007;11(1):8 [FREE Full text]

36. Oh E, Park K, An S, Namkoong K, Lee E. The Effects of a Brief Intervention for Insomnia on Community Dwelling Older Adults. Sleep Med Psychophysiol 2018;25(2):74. [doi: 10.14401/KASMED.2018.25.2.74]

37. Mitchell AJ, Bird V, Rizzo M, Meader N. Diagnostic validity and added value of the Geriatric Depression Scale for depression in primary care: a meta-analysis of GDS30 and GDS15. J Affect Disord 2010 Sep;125(1-3):10-17. [doi: 10.1016/j.jad.2009.08.019] [Medline: 19800132]

38. Pocklington C, Gilbody S, Manea L, McMillan D. The diagnostic accuracy of brief versions of the Geriatric Depression Scale: a systematic review and meta-analysis. Int J Geriatr Psychiatry 2016 Aug;31(8):837-857. [doi: 10.1002/gps.4407] [Medline: 26890937]

39. Bae JN, Cho MJ. Development of the Korean version of the Geriatric Depression Scale and its short form among elderly psychiatric patients. J Psychosom Res 2004 Sep;57(3):297-305. [doi: 10.1016/j.jpsychores.2004.01.004] [Medline: 15507257]

40. Lee S, Kim W, Chang S, Kim B, Lee D, Bae J. The use of the Korean version of Short Form Geriatric Depression Scale (SGDS-K) in the community dwelling elderly in Korea. J Korean Geriatr Psychiatry 2013;17(1):37 [FREE Full text]

41. Youn JC, Kim KW, Lee DY, Jhoo JH, Lee SB, Park JH, et al. Development of the Subjective Memory Complaints Questionnaire. Dement Geriatr Cogn Disord 2009;27(4):310-317. [doi: 10.1159/000205512] [Medline: 19252402]

42. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [Medline: 2748771]

43. Sohn SI, Kim DH, Lee MY, Cho YW. The reliability and validity of the Korean version of the Pittsburgh Sleep Quality Index. Sleep Breath 2012 Sep;16(3):803-812. [doi: 10.1007/s11325-011-0579-9] [Medline: 21901299]

44. Lund A. Measuring usability with the use questionnaire. Usability interface 2001 Jan;8(2):3-6 [FREE Full text]

45. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab 2012;10(2):486-489 [FREE Full text] [doi: 10.5812/ijem.3505] [Medline: 23843808]

46. Chung K. 2017 National Survey of Older Koreans: Findings and Implications. 2018 Policy Report of the Korea Institute for Health and Social Affairs (KIHASA). 2018. URL: https://www.kihasa.re.kr/web/publication/newbooks_pdsissue/view.do?menuId=46&tid=71&bid=200&ano=1487 [accessed 2020-08-02]

47. Antypas K, Wangberg SC. An Internet- and mobile-based tailored intervention to enhance maintenance of physical activity after cardiac rehabilitation: short-term results of a randomized controlled trial. J Med Internet Res 2014 Mar 11;16(3):e77 [FREE Full text] [doi: 10.2196/jmir.3132] [Medline: 24618349]

48. Mogoa e C, Cobeanu O, David O, Giosan C, Szentagotai A. Internet-Based Psychotherapy for Adult Depression: What About the Mechanisms of Change? J Clin Psychol 2017 Jan;73(1):5-64. [doi: 10.1002/jclp.22326] [Medline: 27684405]

49. Horsch CH, Lancee J, Griffioen-Both F, Spruit S, Fitrianie S, Neerincx MA, et al. Mobile Phone-Delivered Cognitive Behavioral Therapy for Insomnia: A Randomized Waitlist Controlled Trial. J Med Internet Res 2017 Apr 11;19(4):e70 [FREE Full text] [doi: 10.2196/jmir.6524] [Medline: 28400355]

Abbreviations

CBT-I: cognitive behavioral therapy for insomnia
EOL: ease of learning
EOU: ease of use
GDS-15: 15-question version of the Geriatric Depression Scale
mHealth: mobile health
PD: perceived difficulty in using the mobile app
PSQI: Pittsburgh Sleep Quality Index
PSQI-K: Korean version of PSQI
SATIS: satisfaction
SGDS-K: Korean version of the GDS-15
SMCD: Seoul Metropolitan Center for Dementia
SMCQ: Subjective Memory Complaints Questionnaire
USE: Usefulness, Satisfaction, and Ease of Use questionnaire

XSL•FO
RenderX

Viewpoint

Mobile Fotonovelas Within a Text Message Outreach: An Innovative Tool to Build Health Literacy and Influence Behaviors in Response to the COVID-19 Pandemic

Rena Brar Prayaga[1], MA, JD; Ram S Prayaga[1], MS

mPulse Mobile, Inc, Encino, CA, United States

Corresponding Author:
Rena Brar Prayaga, MA, JD
mPulse Mobile, Inc
16530 Ventura Blvd
Suite 500
Encino, CA, 91436
United States
Phone: 1 888 678 5735
Email: rena@mpulsemobile.com

Abstract

With all 50 US states reporting cases of coronavirus disease (COVID-19), people around the country are adapting and stepping up to the challenges of the pandemic; however, they are also frightened, anxious, and confused about what they can do to avoid exposure to the disease. Usual habits have been interrupted as a result of the crisis, and consumers are open to suggestions and strategies to help them change long-standing attitudes and behaviors. In response, a novel and innovative mobile communication capability was developed to present health messages in English and Spanish with links to fotonovelas (visual stories) that are accessible, easy to understand across literacy levels, and compelling to a diverse audience. While SMS text message outreach has been used to build health literacy and provide social support, few studies have explored the benefits of SMS text messaging combined with visual stories to influence health behaviors and build knowledge and self-efficacy. In particular, this approach can be used to provide vital information, resources, empathy, and support to the most vulnerable populations. This also allows providers and health plans to quickly reach out to their patients and members without any additional resource demands at a time when the health care system is severely overburdened.

KEYWORDS

Background

Coronavirus disease (COVID-19) has profoundly changed our experience of everyday life and our interpersonal connections. Both the US Centers for Disease Control and Prevention (CDC) and World Health Organization have recommended precautions to prevent and slow transmission of the disease; these include washing hands frequently with soap and water for 20 seconds, covering coughs and sneezes, and avoiding handshakes [1,2]. At the same time, social distancing and stay-at-home orders have required people to stay home and leave only for essential activities; this can lead to increasing social isolation and loneliness, especially for the 35.7 million Americans (28% of households) who live alone [3].

Health Literacy and Fotonovelas

Health literacy is the degree to which individuals can obtain, process, and understand basic health information and the services needed to make appropriate health decisions [4]. Approximately 80 million Americans have low health literacy; this has been associated with disparities in health care access, lower use of health care services, and poorer health outcomes [5]. Limited or insufficient health literacy has also been associated with lower adoption of protective behaviors, such as vaccinations, hand hygiene, and other self-care measures [6]. The urgent need to avoid exposure to COVID-19 would require a change in habits and behaviors, such as washing hands more frequently and maintaining a safe physical distance from others outside the home. It has been suggested that SMS text message

XSL•FO
RenderX

campaigns for public health are the "most effective medium for mass dissemination due to their reach, immediacy, opportunity for data collection and personalization, ability to tailor and adapt information, and opportunity to link to other sources [7]." Our goal was not only to use SMS text messages as a rapid deployment tool but also to build health literacy by including a visual story within the text messages that shows people modeling healthy behaviors to protect against the virus. While several innovative technologies have been developed to improve health literacy, a review of the literature revealed a lack of visual tools or mobile fotonovelas within text messages as a strategy to build health literacy and influence health behaviors.

There is substantial evidence that text messaging can be used for health outreach and to influence health behaviors [8]. At the same time, fotonovelas, a traditional print medium originally developed for Latin American audiences, have been used with great success with participants from other (eg, Dutch) cultures, particularly among low-literacy or underserved populations [9-11]. Similar to a comic book in format, fotonovelas typically tell a story using a dramatic or soap opera–style plot with illustrations or photographs and dialogue bubbles to capture the user's attention and share an important lesson. They have been used to help patients understand the value of preventive care and screenings and to improve self-management of chronic conditions. They are also more likely to be passed on to family or friends through social networks and can increase their reach beyond the targeted individual.

In this case, we wanted to adapt the traditional print format by creating a series of mobile (digital) fotonovelas or illustrated comic strips that could be used to improve health literacy around COVID-19, to fill knowledge gaps around perceived severity and susceptibility, and to help health plan members build new healthy habits to avoid exposure to the virus. We started with two mobile fotonovelas (in both English and Spanish) to (1)

build health literacy and awareness about simple changes in daily habits that could make a great difference in keeping safe, and (2) provide support to address the challenges of staying at home and social isolation as communities attempt to lower the risk of infection and "flatten the curve." These fotonovelas were delivered to health plan members and patients as a link within SMS text messages. SMS is an obvious and well-accepted channel for rapid deployment that has a high rate of adoption; it seemed particularly appropriate given the urgent need to communicate health risks about the virus. The text messages with links to fotonovelas were designed to be educational, with lighthearted content, and to reduce cultural and linguistic barriers. The readability of all content was at or below a sixth grade reading level.

Fotonovela 1: Promoting Healthy Behaviors to Prevent the Spread of COVID-19

The messages in Fotonovela 1 were created as an early response to the COVID-19 threat and were informed in part by the Health Belief Model (HBM) [12-14], which was originally developed to study patient screening behavior for tuberculosis and is commonly used to drive behaviors relating to health promotion and disease prevention. In this case, recipients received an SMS text message telling them that their health plan had put together "COVID-19 simple steps" to help them stay safe and healthy, and they were asked to click on a link to see more. Upon clicking, they viewed the mobile fotonovela, which consisted of 8 frames in story format showing different settings and how to stay safe around people in those settings. The various scenes in the story discussed high-risk individuals, washing hands, elbow bumps, elbow sneezes, using only trusted information sources on the internet, and avoiding crowds (see Figures 1 and 2).

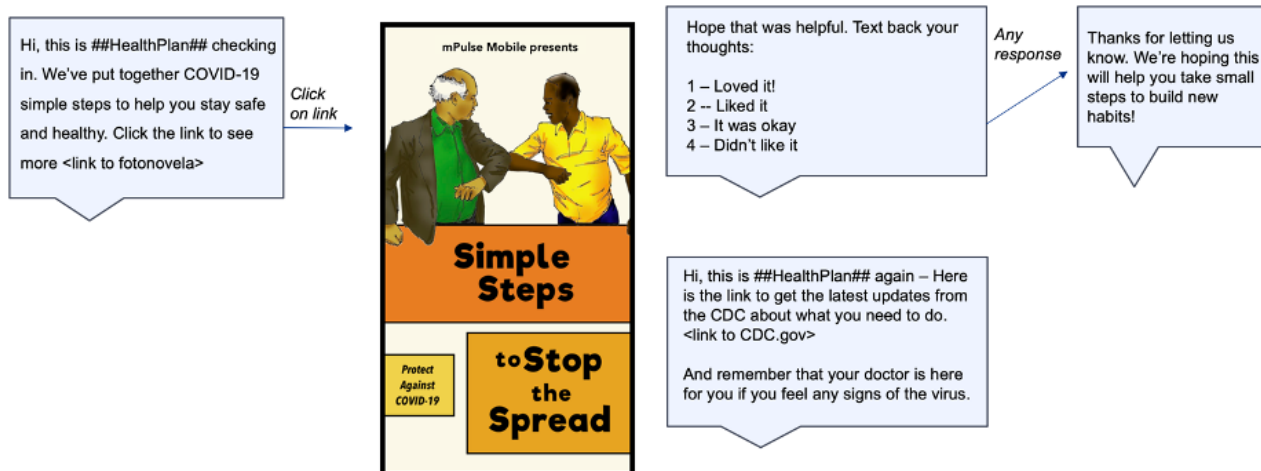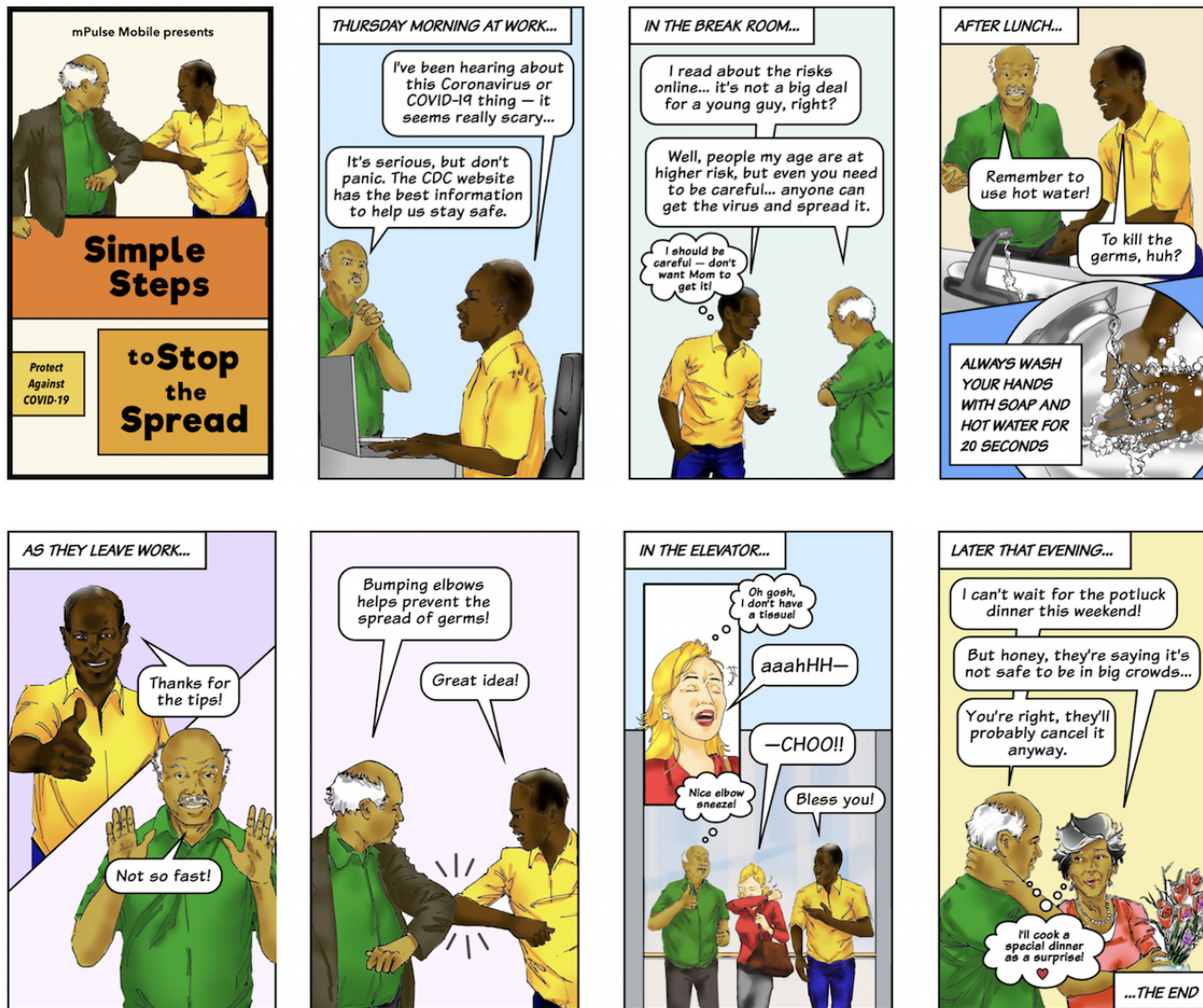**Figure 1.** Text workflow introducing Fotonovela 1.

Figure 2. Images from Fotonovela 1: Simple Steps to Stop the Spread.

We were addressing the various determinants of behavior or constructs that are outlined in the HBM and that are likely to influence preventive behaviors: perceived severity (Scene 1: how bad is the virus), perceived susceptibility (Scene 2: does it really apply to me and how do my actions impact others I care about), perceived benefits (Scene 3: washing hands will help keep me safe), perceived barriers to action (Scenes 4 and 5: but I have to change my behavior and stop shaking hands, be more careful about sneezing, and engage in social distancing), and exposure to factors that prompt action (Scene 6: influencing each other to avoid crowds and stay home). The goal of this type of mobile fotonovela is to build self-efficacy (the efficacy to influence events in one's life) [15] so that recipients feel empowered to pursue strategies and develop new habits that are likely to be successful in addressing the perceived threats or challenges posed by COVID-19. We also relied on social cognitive theory [16], which posits that understanding health risks and benefits will influence changes in health habits and behaviors if an individual believes that these new health behaviors can positively impact their health. As part of this process, there is an evaluation of possible benefits and losses and social approval or disapproval associated with new behaviors. By using a story format, we were able to (indirectly) share important messages, build empathy and understanding,

and avoid making the recipient feel defensive or detached. The focus on simple steps made the changes feel easier, and these small successes could build self-efficacy and confidence over time.

Fotonovela 2: Emphasizing the Importance of Staying Home to Save Lives

As the situation evolved, it became clear that we would need to provide support for extended periods of physical distancing (also called "social distancing") to slow the spread of COVID-19. An additional challenge was posed by social isolation as a result of restricted movement outside the home due to self-imposed or mandatory precautionary measures. There is much evidence that social isolation (or a lack of social interaction) can have adverse health effects and even reduce life expectancy when it is experienced as loneliness. Social isolation increases the risks of cardiovascular disease and cognitive decline, and it even weakens the immune system. A recent study suggested that the health effects of social isolation are as damaging as smoking 15 cigarettes a day, and this effect is particularly noticeable among older people [17]. Social

isolation in the era of COVID-19 adds an additional level of uncertainty, anxiety, and fear that can quickly influence quality of life and well-being and start to impact mental and physical health. In response, we created Fotonovela 2, which focuses on staying home whenever possible, maintaining healthy physical separation from others outside the home, developing strategies to remain balanced and positive despite limited social contact, and finding ways to remain emotionally connected with friends and family using technology. The messages in Fotonovela 2 address the effects of limited social contact over several weeks and the impact of these restrictions on personal well-being and perceived loneliness.

There is strong evidence from studies of psychology and behavioral economics [18] that people tend to engage in cognitive biases as part of the decision-making process, in which they assess the probability of uncertain events using heuristic principles that may contain severe errors. One such cognitive bias or illogical heuristic is *optimism bias* [19], where people underestimate risks when considering potential harm in the form of disease or catastrophe and also (erroneously) tend to believe that others are more likely to be impacted than themselves. This is coupled with *present bias* [20], which tends to overvalue immediate rewards in the present (seeing friends at a get-together in two days) at the expense of long-term benefits

(putting off the get-together so that everyone reduces their likelihood of contracting COVID-19). We believe that the prolonged nature and uncertain term of the stay-at-home requirements will only exacerbate these biases. To address both present bias and optimism bias, the consequences of not engaging in physical distancing are made particularly salient through the "Stay home, save lives" caption, which mirrors the messaging being promoted by the CDC.

Fotonovela 2 (see Figures 3 and 4) shows us four houses in a neighborhood and introduces us to the people in these homes. As the user scrolls down, a story unfolds behind closed doors and shares how people in the neighborhood are staying at home, coping with change, and finding ways to stay emotionally connected and positive. At the end of each story, there are specific suggestions to encourage users to explore new ways to stay actively engaged and develop a sense of routine even while being confined within their homes. By infusing the scenes with everyday examples of people making changes, expressing empathy, and staying positive, a secondary goal is established to build resilience and hope at a time of deep uncertainty. The primary goal is to overcome these biases, tap into prospect theory and loss aversion [21], and update risk assessment calculations so that individual decision-making and behavior more closely follow normative guidelines.

Figure 3. Text workflow introducing Fotonovela 2.

Figure 4. Images from Fotonovela 2: Stay Home, Save Lives.

For Fotonovela 2, we used photographs instead of illustrations, presenting images of real people against simple backgrounds with thought and speech bubbles. This visual presentation is more similar to that of traditional printed (or paper-based) fotonovelas and presented the stories within each vignette with elements of humor and lightheartedness (a dog that speaks or a toddler who has a design plan). The expectation is that the audience can relate to the situation presented in the first frame (the problem) but also can gain from the strategy offered in the subsequent frames (the solution). While each vignette ends with core points of the solution and related action suggestions, we rely on the visual story to achieve the communication objectives: to convey a sense of togetherness, address doubts and apprehensions, and provide concrete and positive ways to cope and build a sense of agency.

In addition to the fotonovelas, we created a series of check-ins via SMS text messages with advice, suggestions, and support on a range of topics, including cooking ideas, exercising at

home, importance of a daily routine, following a regular sleep schedule, suggestions to manage stress and increase mindfulness, and avoiding monitoring news reports throughout the day. The messages included links to helpful resources and were crafted to provide a lighter and more positive tone to help people reshape or reframe their day-to-day routines, even as they felt more cut off from social interactions and experienced high levels of uncertainty about the future.

While Fotonovela 2 and the related text messages do not explicitly discuss COVID-19, focusing instead on supportive and empathetic content, we built an extensive natural language understanding (NLU) system that can recognize and address member questions and concerns relating to COVID-19. For example, if a member asked "How do I know if I have coronavirus?", the NLU system would understand the question as "symptom-related" and automatically respond with a message pointing the member to the appropriate authoritative sources (eg, the CDC website, health plan website, or state website). This feedback system also allows us to add new content based on topics that were not included in the original program.

Early Feedback and Future Considerations

The mPulse Mobile platform delivers text messages to patients and members on behalf of health care companies. The platform consists of several components that collectively enable companies to interactively engage with their end users about appointments, refills, gaps in care, or other health-related topics [22]. This was our first use of a mobile fotonovela to share important health information to address health beliefs, build self-efficacy, and influence health behaviors. The characters in the visuals are culturally diverse, vary in age and gender, and communicate in English or Spanish. We were excited to find that this approach was effective in reaching over 100,000 health plan members across the age spectrum (as old as 97 years) and in providing value to Spanish speakers and people who are negatively impacted by social determinants of health (who may also be encountering health disparities in outcomes and health access issues relating to COVID-19). We are unable to present detailed results due to client confidentiality restrictions. Broadly, we measured the number of views of the content, the satisfaction survey responses, and the opt-out rates. These metrics suggest robust engagement with this material across audience groups.

Beyond COVID-19, our next step is to explore the ongoing use of these types of visual stories to build health literacy and awareness in other contexts. For example, is it necessary to get preventive screenings such as mammograms when you feel healthy and nobody in your family has a history of breast cancer? Similarly, how do you decide whether to use a nurse line, urgent care, or the emergency room when you feel sick after hours? We also want to consider tailoring the fotonovelas based on member demographics (age, gender, language, social determinants of health, health literacy) and psychographics (self-efficacy, health beliefs, stage of change) to build variations within story scenes based on these attributes, beliefs, and preferences. We expect that this type of data-driven and artificial intelligence–enabled segmentation will increase the relevance of the fotonovelas and further deepen engagement. Finally, we expect to build our dataset of member responses so that we can rely more heavily on machine learning–based natural language processing to improve recognition accuracy and response handling.

We were able to quickly develop and deploy a text messaging and fotonovela outreach in English and Spanish to address concerns and influence behavior relating to the COVID-19 crisis. This program is a cost-effective and convenient solution for building health literacy and engaging with underserved, under-resourced and hard-to-reach populations. Member responses and engagement insights can be used to improve the design of future text and mobile friendly visual story-based solutions.

Conflicts of Interest

None declared.

References

1. Get the Facts About Coronavirus. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/index.html [accessed 2020-08-05]
2. Coronavirus. World Health Organization. URL: https://www.who.int/health-topics/coronavirus#tab=tab_1 [accessed 2020-08-05]
3. U.S. Census Bureau Releases 2018 Families and Living Arrangements Tables. United States Census Bureau. 2018 Nov 14. URL: https://www.census.gov/newsroom/press-releases/2018/families.html [accessed 2020-08-05]
4. US Department of Health and Human Services. Healthy People 2010, 2nd edition. Washington, DC: US Government Printing Office; Nov 2000.

XSL•FO

RenderX

5. Berkman ND, Sheridan S, Donahue K, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107 [FREE Full text] [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]

6. Castro-Sánchez E, Chang PWS, Vila-Candel R, Escobedo AA, Holmes AH. Health literacy and infectious diseases: why does it matter? Int J Infect Dis 2016 Feb;43:103-110 [FREE Full text] [doi: 10.1016/j.ijid.2015.12.019] [Medline: 26751238]

7. Gold H. How governments are using text alerts to fight the coronavirus pandemic. CNN. 2020 Mar 27. URL: https://www.cnn.com/2020/03/27/tech/text-alert-coronavirus/index.html [accessed 2020-08-05]

8. Text message-based health interventions. Robert Wood Johnson Foundation. URL: https://www.countyhealthrankings.org/take-action-to-improve-health/what-works-for-health/strategies/text-message-based-health-interventions [accessed 2020-08-05]

9. Unger JB, Cabassa LJ, Molina GB, Contreras S, Baron M. Evaluation of a fotonovela to increase depression knowledge and reduce stigma among Hispanic adults. J Immigr Minor Health 2013 Apr 8;15(2):398-406 [FREE Full text] [doi: 10.1007/s10903-012-9623-5] [Medline: 22485012]

10. Increasing Health Literacy with Fotonovelas. USC School of Pharmacy. URL: https://pharmacyschool.usc.edu/fotonovelas/ [accessed 2020-08-05]

11. Koops van 't Jagt R, Hoeks JCJ, Duizer E, Baron M, Molina GB, Unger JB, et al. Sweet Temptations: How Does Reading a Fotonovela About Diabetes Affect Dutch Adults with Different Levels of Literacy? Health Commun 2018 Mar;33(3):284-290 [FREE Full text] [doi: 10.1080/10410236.2016.1258617] [Medline: 28094559]

12. Rosenstock IM. Historical Origins of the Health Belief Model. Health Educ Monogr 1974 Dec 01;2(4):328-335. [doi: 10.1177/109019817400200403]

13. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988 Sep 04;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]

14. Glanz K, Lewis FM, Rimer BK. Health Behavior and Health Education. Med Sci Sports Exerc 1991;23(12):1404. [doi: 10.1249/00005768-199112000-00016]

15. Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191]

16. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164 [FREE Full text] [doi: 10.1177/1090198104263660] [Medline: 15090118]

17. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237 [FREE Full text] [doi: 10.1177/1745691614568352] [Medline: 25910392]

18. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science 1974 Sep 27;185(4157):1124-1131. [doi: 10.1126/science.185.4157.1124] [Medline: 17835457]

19. Sharot T. The optimism bias. Curr Biol 2011 Dec 06;21(23):R941-R945 [FREE Full text] [doi: 10.1016/j.cub.2011.10.030] [Medline: 22153158]

20. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]

21. Kahneman D, Tversky A. Prospect Theory: An Analysis of Decision under Risk. Econometrica 1979 Mar;47(2):263-1979. [doi: 10.2307/1914185]

22. Brar Prayaga R, Agrawal R, Nguyen B, Jeong E, Noble H, Paster A, et al. Impact of Social Determinants of Health and Demographics on Refill Requests by Medicare Patients Using a Conversational Artificial Intelligence Text Messaging Solution: Cross-Sectional Study. JMIR Mhealth Uhealth 2019 Nov 18;7(11):e15771 [FREE Full text] [doi: 10.2196/15771] [Medline: 31738170]

Abbreviations

CDC: Centers for Disease Control and Prevention
COVID-19: coronavirus disease
HBM: health belief model
NLU: natural language understanding

XSL•FO
RenderX

Original Paper

A Tailored Motivational Messages Library for a Mobile Health Sleep Behavior Change Support System to Promote Continuous Positive Airway Pressure Use Among Patients With Obstructive Sleep Apnea: Development, Content Validation, and Testing

Sarah Alismail[1], MSc, MPH; Lorne Olfman[1], PhD

Claremont Graduate University, Claremont, CA, United States

Corresponding Author:
Sarah Alismail, MSc, MPH
Claremont Graduate University
150 E 10th St
Claremont, CA,
United States
Phone: 1 909 621 8000
Email: sarah.alismail@cgu.edu

Abstract

Background: Continuous positive airway pressure (CPAP) therapy is the most effective treatment for obstructive sleep apnea (OSA). Despite the reported benefits of CPAP therapy in treating OSA, its effectiveness is reduced by less-than-optimal adherence or use. Up to 50% of patients who accept CPAP therapy fail to adhere to it. As a lack of commitment to CPAP therapy is one of the most significant factors that hinder OSA treatment effectiveness, patient motivation and education are critical to help alleviate the problem of poor CPAP adherence or use.

Objective: This study aims to support the development of mobile health interventions or information systems solutions to promote CPAP adherence and use among patients with OSA through development, content validation, and testing of tailored motivational messages.

Methods: In phase 1, an initial library of 60 messages was developed to promote CPAP use among patients with OSA. In phase 2, draft messages were evaluated for content validation testing for relevance and clarity by research and clinical experts. In phase 3, patients with OSA (N=24) were recruited through a Qualtrics panel to rate the perceived persuasiveness of the messages in terms of threat and efficacy perceptions, as per their computed extended parallel process model (EPPM) response states. The average score of the ratings was calculated for each message. The messages were sorted according to their average (from highest to lowest) to select the best 12 messages for each tailored set based on the potential responses from the EPPM.

Results: In phase 1, 60 messages were developed based on the existing literature and a review of existing materials. In phase 2, the enumerated content validity of the messages was established through the use of the content validity index for items. A total of 57 messages were found to have acceptable content relevance and clarity. In phase 3, patients with OSA perceived the final library of 48 messages to be persuasive.

Conclusions: After the process of content validation and testing, the final library of messages met the criteria for clarity, relevance, and perceived persuasiveness. This study emphasizes the importance of developing and validating the content of motivational messages, grounded in EPPM theory, across the 4 possible response states in terms of high or low efficacy and threat perceptions.

XSL·FO
RenderX

Introduction

Background

Obstructive sleep apnea (OSA) is the most common breathing disorder that occurs during sleep in the United States [1], with 30 million adults affected by OSA, which is the second highest occurrence in the world [2,3]. Effective treatment of OSA is essential because untreated OSA has consequences for both individual health and society. These consequences include, but are not limited to, cognitive and behavioral deficits that affect a person's work performance, including excessive daytime sleepiness, as well as cardiovascular and metabolic dysfunction. Thus, OSA accounts for a significant portion of health care costs, either directly or by means of its associated illnesses and diseases [4]. In the United States alone, the estimated total societal-level cost attributed to OSA is reported to be US $160 billion annually [5].

Continuous positive airway pressure (CPAP) therapy is the most effective treatment for OSA [6,7]. CPAP therapy has been found to improve cognitive processing and daily functioning; reduce daytime sleepiness [8]; and reduce health care costs, physician visits, and work absences [5]. Despite the reported benefits of CPAP therapy in treating OSA, its effectiveness is reduced by less-than-optimal adherence/usage.

Up to 50% of patients who accept CPAP therapy fail to adhere to it [8]. Patients' failure to adhere to CPAP therapy is a major factor limiting its potential benefits [6,9]. Continued use of CPAP is fundamentally important for effective treatment of OSA [10]. As the lack of commitment to CPAP therapy is one of the most significant factors contributing to hindering the effectiveness of OSA treatment, patient motivation and education about their condition and treatment are critical to help alleviate the problem of poor CPAP adherence. Evidence suggests that behavioral interventions, along with education designed to improve patient self-efficacy, are promising strategies to improve adherence [8]. Thus, finding ways to improve CPAP usage is essential to improving the health and well-being of patients with OSA and might be a way to minimize the associated health care costs and adverse outcomes of the condition [3].

To support future studies on interventions for promoting CPAP use or adherence among patients with OSA, this study aimed to describe the process and results of developing, validating the content of, and testing a set of theory-based tailored motivational messages to be used in a mobile health (mHealth) sleep behavioral change support intervention that promotes CPAP use among patients with OSA.

Theoretical Framework

The extended parallel process model (EPPM) was developed by Witte [11] as a message design–related theoretical framework for effective health communication aimed at persuading target populations to engage in healthy behaviors [12]. The model considers both emotional and cognitive factors related to message processing and links these processes to the success or failure of a fear appeal [13].

On the basis of the EPPM, individuals are motivated to perform an action through their perception of a threat rather than the actual threat modeled [12]. Perceived threat encompasses 2 main dimensions: perceived susceptibility and perceived severity. The former refers to an individual's subjective perception and/or beliefs about the personal risk of experiencing a health condition, whereas the latter refers to one's beliefs regarding the seriousness of the health condition threatened [12,13].

When the perceived susceptibility and/or perceived severity is low (ie, *trivial or irrelevant*) [11], the recipient of the message will not be motivated to process the message, indicating that the efficacy will not be assessed and no response will take place to the fear appeal because the individual believes that the threat does not impact him or her and thus ignores it, as shown in Figure 1 [11,14]. However, when the processed threat message results in a moderate-to-high perceived threat, individuals become motivated and are likely to appraise efficacy (ie, an assessment of the efficacy of the recommended response) [11].

Figure 1. Extended Parallel Process Model, adapted from Witte.
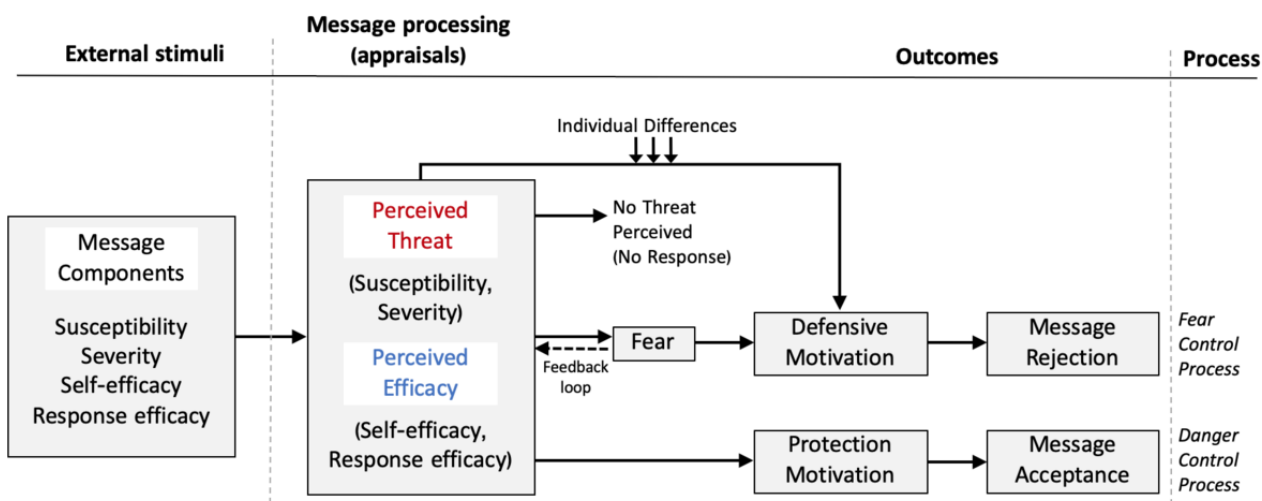
Perceived efficacy refers to the "cognition about the effectiveness, feasibility, and ease with which a recommended response alleviates or helps in avoiding a threat" [15]. The EPPM underlines 2 forms of efficacy under the perceived efficacy appraisal: perceived self-efficacy and perceived response efficacy. Perceived self-efficacy refers to one's perception that he or she can execute the recommended behavior to avert the threat [15]. Perceived response efficacy refers to one's belief about how effective the recommended response is in deterring the threat [15].

On the basis of the perceived efficacy level of the recommended behavior, individuals perform 1 of 2 responses: a danger control or a fear control response. If a message conveys both high threat and efficacy, it is predicted that it will lead individuals to control the danger, demonstrating protection motivation. Individuals will subsequently perform actions, typically the recommended message responses, to protect themselves against the threat. This means that the message succeeded in motivating people to take the recommended action (message acceptance). In contrast, when individuals perceive a high threat level but low efficacy in the message, they have a fear control response. Here, individuals will focus on alleviating the fear usually through psychological defense strategies (eg, defensive avoidance or denial) without taking action about the threat—a defense motivation outcome in the EPPM [14].

Objectives

This study had 3 phases related to the objectives of this study: (1) design and develop a tailored message library based on the potentially different response states of the EPPM, (2) have clinical and research experts validate the content of the messages, and (3) have members of the target audience evaluate the perceived persuasiveness of the messages. Phase 1 of this study involved designing and developing the message library per the EPPM. In phase 2, the content validity of the developed message library was tested by research and clinical experts. The content validation prioritized 2 aspects: perceived relevance and clarity. In phase 3, a sample of patients with OSA assessed the perceived persuasiveness of the messages. The findings of this study will inform the development of mHealth interventions or information systems (IS) solutions targeting improved use of CPAP therapy among patients with OSA.

Methods

Phase 1: Developing the Message Library

According to the EPPM, an individual is most likely to act on a recommended response when he/she is in a danger control state. A message library is designed to motivate participants to move in that direction (ie, high efficacy and high threat). If an individual is already in a danger control state, then the messages are designed and strive to reinforce his/her state.

A total of 60 messages were selected from existing sources or designed to make the message content relevant based on behavioral beliefs per the EPPM. These messages can be categorized into 4 sets, 1 for each potential response state of the EPPM (ie, responsive, proactive, avoidance, and indifference), as shown in Figure 2 (Phase 2 provides further details). Each quadrant is titled with descriptors used by Rimal and Real [16]. Messages were designed to address both threat and efficacy components that need to be altered, reinforced, or initiated. In all 60 messages, the threat is the health consequence of untreated OSA, and the recommended response is using CPAP at least four hours per night (using CPAP ≥4 hours per night was generally admitted as the clinical and empirical level of use required for adherence [17]).

Figure 2. Extended Parallel Process Model potential response states and message types.

Message set I targets those who are in a responsive state. They were designed to reinforce existing perceptions of both threat and efficacy (eg, Untreated obstructive sleep apnea increases your risk of heart problems. Use your CPAP for better health).

Message set II targets those who are in the avoidance state. The messages were designed to recognize the threat, but mainly to increase the efficacy of using CPAP (eg, Try adjusting your mask pads and straps for a better mask fit. This may help reduce air leaks. Use your CPAP to sleep well and be well!).

Message set III targets those who are in a proactive state. It was designed to increase the threat and to reinforce existing perceptions of efficacy (eg, when left untreated, obstructive sleep apnea can lead to chronic diseases. Use your CPAP for better health!).

Message set IV targets those who are in the indifference state, in which they do not feel the threat of OSA and have doubts about the effectiveness of CPAP and their ability to use CPAP therapy. Thus, this set of messages was designed to initiate the efficacy, but mainly to convince patients of the severity and seriousness of untreated OSA (eg, Sleep apnea could be hurting your heart. Use the CPAP to sleep well and be well!) [18].

The initial library of 60 messages consisted of 21 text-based messages, 20 video-based messages, and 19 image-based messages. The rationale for developing 60 messages was that this would allow experts and patients to select the ones that are most relevant, clear, and persuasive. The best messages would then be used in an mHealth intervention that could be evaluated through an experiment.

Message generation was carried out in an iterative manner. First, a review of existing materials was conducted to collect available existing videos, infographics, and images that can be added to the library. For the selected ready-made videos and images, permission was obtained from the source to use the materials in the study. Participants were referred to the ready-made materials via the source's web-based links. Text-, video-, and image-based messages were developed based on a review of the existing resources. The text, videos, and images provided in the messages were adapted from multiple sources.

Text messages were constructed by reviewing the following sources: the Mayo Clinic website [19], ResMed website [20,21], Breathe the Lung Association [22], the American Academy of Sleep Medicine [23], and the American Sleep Apnea Association [24]. The videos were YouTube-based videos provided through WebMD [25-27], the Cleveland Clinic [28,29], the American Academy of Sleep Medicine [30], Columbia Broadcasting System News [31], Lee Health [18], Johns Hopkins Medicine [32], and Mayo Clinic [33-35]. The length of the videos ranged from 43 seconds to approximately 5 min. Images and infographics were adopted from Adventist Health [36], National Jewish Health [37], Sleep Sherpa [38], the CPAP.com website [39], Sleep Well Respiratory Care [40], the American Sleep Apnea Association [41], Advanced Dental Sleep Treatment Center [42], Oral and Maxillofacial Surgeons [43], Very Well Health [44], Sleep Data [45], and Atlanta Headache temporomandibular joint Pain [46].

For the video- and image-based messages, the 12 principles of multimedia design outlined by Clark and Mayer [47] were used as a checklist to ensure that the messages address as many principles as possible. The length of the message was limited to no more than 160 characters to allow them to be delivered through a single mobile text message. The Gunning Fog test [48] was used to ensure that the text messages were at an eighth-grade reading level. The developed draft of the message library was maintained in a Microsoft Excel spreadsheet.

Phase 2: Content Validation of the Messages by Experts

After the message library was developed, it was subjected to content validation by clinical and research experts. The content validation prioritized 2 aspects: the relevance of the messages to the theoretical concepts and the clarity of the messages. The experts' suggestions were collected to improve the developed messages such that the content-validated messages could be tested by the target audience. The final message library was reviewed and improved based on the participants' input. The messages were filtered by the content validity index for items (I-CVI), first for relevance and then for clarity.

Perceived Relevance

We first tested the relevance of the developed messages to the theoretical concepts (ie, perceived self-efficacy, perceived response efficacy, perceived severity, and perceived susceptibility) through expert evaluation. An expert is defined as an individual with acquired knowledge, experience, and/or skills in his/her field of expertise (eg, research and/or clinical practice) [49]. The messages were reviewed and validated by experts in the fields of pulmonology/sleep medicine, psychology, health promotion, and health IS and technology. The experts were board-certified pulmonologists in the field of sleep medicine and faculty or PhD candidates from Claremont Graduate University programs in behavioral and social sciences, health promotion sciences, and IS and technology. To participate, an expert had to meet at least one of the following criteria per the definition of *expert* stated earlier: (1) knowledge and experience in treating patients with sleep apnea, (2) experience in the use of theoretical concepts, (3) methodological knowledge in designing intervention studies, and/or (4) experience in research in chronic diseases management. The experts validated the messages by determining whether the messages reflected their underlying theoretical concepts. Drafts of all 60 messages were grouped into 6 subsets for validation (10 messages per subset). Each subset had a combination of modes: text, video, and image. We sent an email to potential experts inviting them to provide a rating and give comments using a web-based survey developed via Qualtrics (Qualtrics LLC). In particular, experts rated the relevance of each message to its underlying theoretical concepts using a 4-point ordinal scale: 1 (not relevant), 2 (somewhat relevant), 3 (quite relevant), and 4 (highly relevant).

Perceived Clarity

After the messages were checked for content relevance, the text- and image-based messages were further evaluated for clarity. The Qualtrics web-based survey asked experts to rate the clarity of each message using a 4-point ordinal scale: 1 (not clear), 2 (somewhat clear), 3 (quite clear), and 4 (highly clear). For each message, experts were asked to provide suggestions for improvement, if any. Overall, 4 message sets—2 text based and 2 image based—were created for validation. Video-based messages were not included in clarity validation because they are based on ready-made materials (YouTube links) provided by health organizations and were therefore assumed to have been already reviewed and validated by domain experts.

The first author attended 2 days of a scientific conference targeted at pulmonologists and sleep specialists held in October 2019 to recruit experts by asking them if they would participate in the study. Those who agreed to participate completed the survey on a laptop provided by the researcher.

Phase 3: Message Testing by the Target Audience

After expert validation of the message library, the final version of the messages obtained from phase 2 was further validated by assessing the target audiences' reaction to the messages' persuasiveness. A cross-sectional survey was administered using a Qualtrics panel in February 2020. Qualtrics panels have been used extensively in health care research [50]. A Qualtrics panel is an opt-in research panel that offers participant recruitment services for researchers, where individuals who meet the study criteria are invited to participate. The Qualtrics-verified panel service provides access to potential participants who meet the

study criteria. It uses third-party sources to verify the potential participants' information and then sends a survey/study invitation to those who are eligible. This process is especially rigorous when potential participants indicate that they have been diagnosed with OSA. The survey's technical functionality was tested by the first author before launching the survey.

Each of the 4 message sets that were obtained from phase 2, tailored based on whether participants have high or low efficacy and threat perceptions, were split into 2 surveys, creating 8 subsets (Figure 3). This was performed to shorten the length of the survey.

Figure 3. Message sets for perceived persuasiveness testing by patients with obstructive sleep apnea. Some messages were matched with multiple sets based on the obtained I-CVI score. I-CVI: content validity index for items.
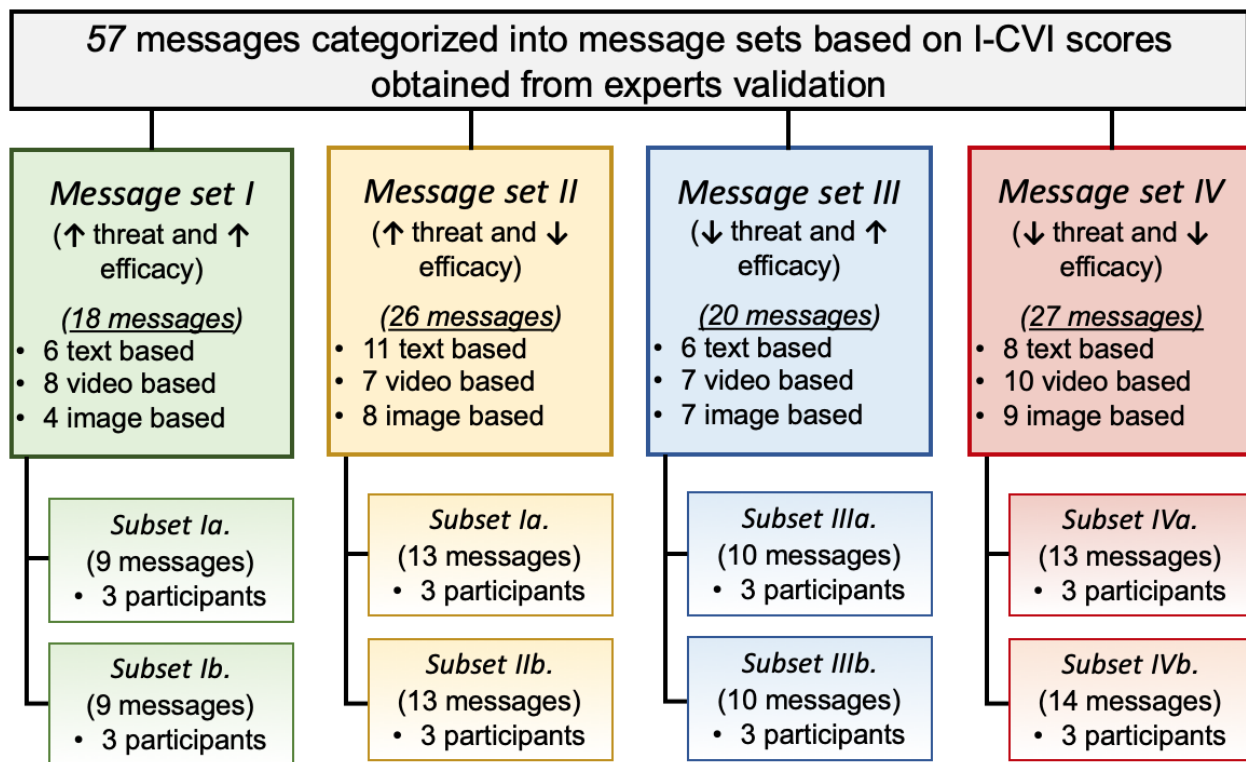
Participants

Web-based participants were recruited through a Qualtrics panel site (Qualtrics.com). Participants were a convenience sample of 24 adult patients with OSA across the United States who were selected based on their availability and willingness to participate in the study. Eligibility criteria included the following: (1) aged ≥18 years, (2) able to read and write in English, and (3) clinically diagnosed with OSA.

Procedure

The protocol for this study was approved by Claremont Graduate University's institutional review board. The survey questionnaire was developed using Qualtrics' web-based survey software. Recruitment was conducted by a Qualtrics project manager, where 676 potential participants in the United States, who indicated in their panel registration data that they have OSA, were invited via email to participate in the study. Qualtrics verified and confirmed their information. The invitation email communicated the opportunity to participate in an academic

survey, approximate length, purpose, and a web link to the anonymous and confidential survey. To minimize participant response bias, the population of interest was not acknowledged (ie, participants were not told that they were invited because they had OSA). People who were interested in participating and who volunteered to do so were first required to read an introductory page that welcomed them and described the study and its goals. After the participants read the welcome page and study goals, they indicated their agreement on an electronic consent form. After signing the consent form, they were asked a few eligibility screening questions before they were cleared to participate in the study. Eligibility screening was added to recruit individuals in the target audience of an mHealth intervention or an IS solution (ie, adults diagnosed with OSA). The eligibility screening questions included a question about age and a question about the kinds of ailments that affects the participant. A total of 634 of the 676 potential participants did not meet the inclusion criteria; hence, they were excluded from the study. Those who met the eligibility screening criteria

(N=42) were asked to report their demographic information, including ethnicity, educational level, and gender.

Next, they were asked to complete a survey to assess their efficacy and threat perceptions through the calculation of perceived self and response efficacy, perceived severity, and perceived susceptibility. Responses were then grouped into 4 quadrants as suggested by Popova [15]. To do this, the mean of each construct was calculated. Perceived efficacy was operationalized as the sum of the mean score of perceived self-efficacy and the mean score of perceived response efficacy as they are both considered as dimensions of the efficacy construct. Perceived threat was operationalized as the sum of the mean score of perceived severity and the mean score of perceived susceptibility as they are both considered as dimensions of the threat construct [11]. Sawyer and King [51] used the Self-Efficacy Measure for Sleep Apnea (SEMSA) with a score of 3 for each subscale as a critical indicator for tailoring an intervention. Thus, an average score of 3 was used as the critical indicator threshold to dichotomize the responses into high and low, with a score >3 being high and all others being classified as low. Thus, once the participants submitted their answers, a subset of messages from a set that matches their computed threat and efficacy perceptions appeared automatically on the screen, and they were asked to rate the extent to which each message is perceived persuasive on a 5-point Likert scale ranging from 1 (definitely not persuasive) to 5 (definitely persuasive). Once they were done, they hit the *submit* button.

Data were collected over 1 week from February 3 to 9, 2020. As monetary incentives have been found to be effective in enhancing the recruitment of participants [52], the participants were compensated upon the completion of the survey. Compensation was handled by Qualtrics. The average completion time for the survey was 25 min.

After obtaining all the participants' responses, the last step in this phase was to determine the best 12 messages in each set. Subsequently, the messages were selected based on the obtained averages, and the best 12 messages were added to the final library.

Measures and Instrumentation

In all, 4 constructs were measured in phase 3: perceived self- and response efficacy, perceived severity, and perceived susceptibility. The constructs are self-reported multi-item scales drawn from previously validated measures.

Self-Efficacy Measure for Sleep Apnea

Perceived self-efficacy, perceived response efficacy, and perceived susceptibility were adapted from the SEMSA [53]. The SEMSA was developed based on Bandura's social cognitive theory. It is divided into 3 subscales that measure self-efficacy, risk perception (perceived susceptibility), and outcome expectations (perceived response efficacy) [53]. The risk perception subscale is used for measuring perceived susceptibility as both terms refer to one's perception of the likelihood that they will be affected by the health consequences of untreated OSA. For instance, Olsen [54] used SEMSA's risk perception subscale to measure perceived susceptibility among patients clinically diagnosed with OSA. In addition, Witte et al

[14] noted that the term *outcome expectations* in Bandura's theory is *similar to response efficacy in fear appeals*. Therefore, the outcome expectations subscale was used to assess perceived response efficacy. Perceived self-efficacy, perceived response efficacy, and perceived susceptibility were assessed using a 4-point Likert scale ranging from 1 (not at all true) to 4 (very true). Higher scores denote greater perceived self-efficacy, greater perceived response efficacy, and higher perceived susceptibility.

Obstructive Sleep Apnea Knowledge and Beliefs

The OSA knowledge and beliefs (OSA-KAB) tool is a 16-item tool developed based on the health belief model to measure an individual's knowledge and beliefs in the context of OSA [55]. The perceived severity subscale was adapted from the OSA-KAB tool [55]. Perceived severity was assessed using 3 items based on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Scores range from 3 to 15.

Attention Checks and Quality Screens

Attention checks were employed to confirm that survey respondents were reading the survey questions carefully and thoroughly. The responses were checked for validity by checking the participants' answers to 2 embedded validity questions (ie, "please do not rate this statement so I know you are paying attention;" "please rate this statement at 'not persuasive' so I know you are paying attention"). One respondent failed one of the 2 embedded validity check questions, and 10 did not complete the entire survey; hence, they were excluded from the study.

Quality screens were also used to confirm that respondents spent an adequate amount of time completing the survey. The Qualtrics panel recommended excluding participants who completed the survey in one-third of the median time or less because that would indicate that the respondents were rushing through the survey. This criterion excluded 5 respondents.

Analysis

Phase 2: Content Validation of the Messages by Experts

The I-CVI [56] was used to analyze the content validation for relevance and clarity. This allowed us to decide which messages to retain, delete, or adjust.

Perceived Relevance

Content validation of perceived relevance was conducted to ensure that each message included the correct theoretical concepts. The I-CVI was calculated by summing the messages that received scores of 3 or 4 and then dividing the sum by the total number of responses. Messages with I-CVIs ≤0.78 should be eliminated and/or reformulated [56]. However, Lynn [56] recommended that when there are ≤5 experts, an I-CVI of 1.00 must be obtained for the item/message to have content validity. If an expert rated a message with 1 or 2, he/she was asked to provide suggestions for improvement. The experts' responses were analyzed and modifications were made to the first draft of the messages in accordance with the experts' feedback.

To analyze the experts' input, we first reviewed each expert's feedback. Then, the experts' ratings and suggestions were

transcribed into a table in a Microsoft Excel sheet. On the basis of the ratings of each message, the I-CVI was calculated to determine the extent to which each message related to each of the theoretical concepts. We then categorized each message into message sets based on the I-CVI scores. Some messages can be associated with more than one set. Specifically, we categorized each message dependent on whether an I-CVI of 1.00 was addressed to 0, 1, or 2 elements of threat and efficacy. As message set I targets those who are in the responsive state (high threat and high efficacy), it is designed to reinforce existing perceptions of both threat and efficacy; therefore, messages that received an I-CVI of 1.00 for at least one element of threat and at least one element of efficacy were matched to message set I. As message set II targets those who are in the avoidance state (high threat and low efficacy), the messages are designed mainly to increase efficacy. Thus, messages that received an I-CVI of 1.00 for at least one element of efficacy were matched to set II. As message set III targets those who are in the proactive state (low threat and high efficacy), it is designed to increase the threat and to reinforce existing perceptions of efficacy. Thus, messages that received an I-CVI of 1.00 for 1 or 2 elements of threat and 1 element of efficacy were matched to message set III. As message set IV targets those who are in the indifference state (low threat and low efficacy), it is designed to initiate the efficacy and mainly to convince patients of the threat of untreated OSA. Thus, messages with an I-CVI of 1.00 for at least one element of threat and at least one element of efficacy or 1 or 2 elements of threat were matched to set IV.

Perceived Clarity

The I-CVI was computed as the number of experts giving a rating of 3 (quite clear) or 4 (*highly clear*) divided by the number of experts. An I-CVI score of 1.00 denotes a 100% agreement among the experts. The experts' suggestions for improvement were also analyzed. Messages with low I-CVI scores (ie, <1.00) for clarity were reformulated based on the experts' written comments or deleted.

Phase 3: Message Testing by the Target Audience

Descriptive statistics were computed using frequency counts and percentages for categorical variables. To determine the best 12 messages in each set, participants' responses in each subset were analyzed as follows. Patient feedback on the survey was downloaded in Microsoft Excel. Each response state group (ie, responsive, proactive, avoidance, and indifference) was presented in a separate worksheet. The perceived persuasiveness ratings were provided next to each message and the average score of the ratings was calculated for each message. Then, the messages were sorted from highest to lowest based on the average score. This information enabled us to select the 6 best messages from each subset, yielding 12 from each message set.

Results

Phase 2: Content Validation of the Messages by Experts

Perceived Relevance

A total of 18 experts completed the survey, 3 for each subset (Table 1 provides the experts' characteristics). In the first draft of 60 messages, 1 message was eliminated ("Experiencing a return of symptoms or feeling like the CPAP is no longer working? Your mask may have an air leak") because of its low I-CVI score (<1.00) on each of the 4 theoretical concepts and because of an expert's comment that "there are acceptable leaks." As a result, the second draft included 59 messages.

The message library consisted of the following: 18 messages in set I (high threat and high efficacy: 6 text based, 4 image based, and 8 video based), 26 messages in set II (high threat and low efficacy: 11 text based, 8 image based, and 7 video based), 20 messages in set III (low threat and high efficacy: 6 text based, 7 image based, and 7 video based), and 27 messages in set IV (low threat and low efficacy: 8 text based, 9 image based, and 10 video based). It is important to note that some messages were matched with multiple sets based on the obtained I-CVI score of the messages' perceived relevance to efficacy and threat constructs.

Table 1. Experts' characteristics for message relevance validation.

Characteristics	Message subset 1	Message subset 2	Message subset 3	Message subset 4	Message subset 5	Message subset 6
Experts, n (%)	3 (16.6)	3 (16.6)	3 (16.6)	3 (16.6)	3 (16.6)	3 (16.6)
Gender, n (%)						
Female	2 (67)	0 (0)	1 (33)	2 (67)	2 (67)	2 (67)
Male	1 (33)	3 (100)	2 (67)	1 (33)	1 (33)	1 (33)
Area of expertise, n						
Pulmonology and sleep medicine	0	0	1	1	0	0
BSS[a]	1	0	2	0	1	1
IST[b]	1	2	0	1	0	1
HPS[c]	1	1	0	1	2	1
Current position, n						
Physician	0	0	1	1	0	0
Professor	1	1	1	0	1	1
Psychologist	0	0	1	0	0	0
PhD candidate	2	2	0	2	2	2

[a]BSS: behavioral and social science.

[b]IST: information systems and technology.

[c]HPS: health promotion science.

Perceived Clarity

A total of 16 experts completed the survey at the conference, and 3 experts provided their email addresses so that we could send them the survey link and they could complete the survey during their free time. As a result, a convenience sample of 19 experts (N=19) was used for clarity content validation. The experts were physicians and faculty members with expertise in the fields of pulmonology and/or sleep medicine.

The text-based messages were grouped into 2 sets for validation, and 3 experts evaluated the clarity of each set except for 1 set. That 1 set was evaluated by 4 experts as 1 expert had zero expertise in the field of pulmonology and/or sleep medicine. Then, we analyzed the data, and 8 messages (8/20) received an I-CVI for clarity <1.00. Modifications based on the experts' comments were made to these messages. We also modified 3 messages that received an I-CVI of 1.00 to address the experts' comments (eg, For the message "Do you feel sleepy or tired in the morning? Poor sleep may be a reason for sleepiness or tiredness. Use CPAP for at least 4 hours a night to feel better!" one expert commented, "Not only in the morning, probably just mention during the day"). As a result, 11 messages were reworded based on the experts' suggestions. We then created a survey to evaluate the clarity of the reworded messages. A total of 4 experts who were recruited from the conference completed the survey. The analysis of the reworded messages showed that 10 out of 11 messages received an I-CVI of 1.00 and 1 received an I-CVI of 0.75. However, 3 messages that received an I-CVI of 1.00 and 1 message that received an I-CVI of 0.75 were modified to address the experts' comments.

The image-based messages were also grouped into 2 sets for validation. Overall, 6 experts completed the survey (3 for each set of messages). Of the 19 image-based messages, 2 were eliminated because they received an I-CVI <1.00 as well as critical comments. In particular, 1 of the experts commented that 1 of the messages was "too wordy" and that the other message included inaccurate information (ie, the message should have indicated that the CPAP pumps air, not oxygen). Two messages were modified to address the experts' comments (eg, "the background color is too dark" and "perhaps changing the white font to black").

Finally, the modified messages were further tested for clarity by 3 additional experts. The modified messages consisted of 4 text-based messages that were modified based on the experts' comments and 2 image-based messages. Thus, 6 messages were included in *set 6* for clarity validation. The results show that all 6 messages received an I-CVI of 1.00. Therefore, a library of 57 messages (20 text-, 17 image-, and 20 video-based messages) was ready for use in the patient validation phase.

Phase 3: Message Testing by the Target Audience

Participants

A total of 676 potential participants were invited to participate in the web-based survey, 42 met the study criteria, and 30 participants completed the survey, of which 6 were excluded for not meeting the time quality screens or failed 1 of the 2 embedded validity questions (Figure 4).

Overall, 24 participants completed the web-based survey (7 men and 17 women). The average age of the participants was between 45 and 55 years. The demographic characteristics of

the participants are presented in Table 2. Overall, the participants (N=24) were predominantly women (n=17), white (n=22), and reported having a college degree (n=12).

Participants were categorized into 4 groups based on whether they had high or low efficacy and threat perceptions (ie, responsive, avoidance, proactive, and indifference). As stated

earlier, an average score of 3 was used as a critical indicator threshold to dichotomize the responses into high and low following Sawyer et al [51], with a score >3 being high and all others being classified as low. The indifference group had all women, only the responsive group had nonwhites, and only the proactive group had participants who were aged 65 years.

Figure 4. Patients with obstructive sleep apnea message testing process.
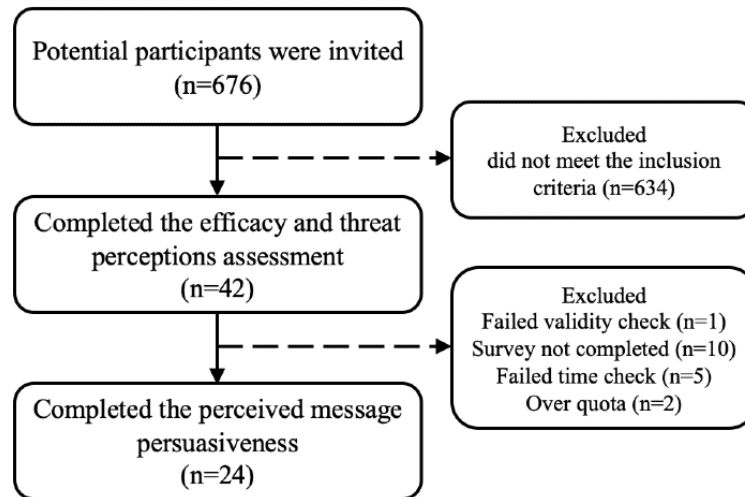
Table 2. Demographics of participants by response state.

Demographics[a]	Responsive (high threat and high efficacy)	Avoidance (high threat and low efficacy)	Proactive (low threat and high efficacy)	Indifference (low threat and low efficacy)	Combined
Participants, n (%)	6 (25)	6 (25)	6 (25)	6 (25)	24 (100)
Age (years), n (%)					
18-24	0 (0)	0 (0)	0 (0)	2 (33)	2 (8)
25-34	2 (33)	2 (33)	0 (0)	0 (0)	4 (17)
35-44	1 (17)	2 (33)	0 (0)	0 (0)	3 (13)
45-54	1 (17)	1 (17)	3 (50)	2 (33)	7 (29)
55-64	2 (33)	1 (17)	0 (0)	2 (33)	5 (21)
65-74	0 (0)	0 (0)	2 (33)	0 (0)	2 (8)
75-84	0 (0)	0 (0)	1 (17)	0 (0)	1 (4)
Gender, n (%)					
Female	4 (67)	3 (50)	4 (67)	6 (100)	17 (71)
Male	2 (33)	3 (50)	2 (33)	0 (0)	7 (29)
Ethnicity, n (%)					
White	4 (66)	6 (100)	6 (100)	6 (100)	22 (92)
Black	1 (17)	0 (0)	0 (0)	0 (0)	1 (4)
Hispanic	1 (17)	0 (0)	0 (0)	0 (0)	1 (4)
Education, n (%)					
High school	2 (33)	3 (50)	3 (50)	2 (33)	10 (42)
College	3 (50)	3 (50)	3 (50)	3 (50)	12 (50)
Master	0 (0)	0 (0)	0 (0)	1 (17)	1 (4)
Doctorate	1 (17)	0 (0)	0 (0)	0 (0)	1 (4)

[a]Response state refers to participants' categorization based on whether they have high/low threat and efficacy perceptions.

XSL•FO
RenderX

Message Set I

The message set I targets those who were in the responsive state (high threat and high efficacy), in which messages were designed to reinforce the patient's existing threat perception of untreated OSA and efficacy perception of using the CPAP device.

For set Ia, 7 messages received an average score of 4.67, 1 message had a mean score of 4.33, and 1 message had a mean score of 4.00. We intended to select the top 6 messages, sorted from highest to lowest based on the average score from each subset (ie, messages with the top high average scores), yielding 12 for set I. As 7 messages received an average score of 4.67, which represents the highest average score in this subset, a random number generator was used to select 6 of the 7 messages to be used in the final set. This was done to meet the predetermined message quantity.

For set Ib, 1 message received a mean score of 5.00, 1 message had a mean score of 4.67, 2 messages had a mean score of 4.33, and 3 messages had a mean score of 4.00. Among those that received a mean score of 4.00, 2 messages were randomly selected to be included in the final set to meet the predetermined message quantity. As a result, 12 messages were selected for set I, including 3 image-based, 6 video-based, and 3 text-based messages. Among the 18 messages that were under consideration, 12 messages were rated higher in perceived persuasiveness by patients with OSA who were in the responsive state and thus were added to the final library. Multimedia Appendix 1 provides the final 12 messages.

Message Set II

This message set targets those who were in the avoidance state (high threat and low efficacy), in which messages were designed mainly to increase patients' efficacy perception of using their CPAP device.

For set IIa, 2 messages received a mean score of 4.33 and 6 messages received a mean score of 4.00. Among the 6 messages that received a mean score of 4.00, 4 messages were selected randomly to be included in the final set II.

For set IIb, 1 message received a mean score of 4.67, 3 messages received a mean score of 4.33, and 3 messages received a mean score of 4.00. Among these messages with a mean score of 4.00, 2 messages were randomly selected to be included in the final set. Among the 26 messages that were under consideration, the selected 12 messages were perceived to be more persuasive than other messages in the set based on the feedback gained from patients with OSA who were in the avoidance state. As a result, 12 messages were added to the final message set II, including 6 image-based, 3 video-based, and 3 text-based messages (Multimedia Appendix 1).

Message Set III

This message set targets those who were in the proactive state (low threat and high efficacy), in which messages were designed to increase patient's threat perception of untreated OSA and reinforce existing efficacy perception of using their CPAP device.

For set IIIa, 1 message received a mean score of 5.00 and 5 messages received a mean score of 4.67, yielding 6 messages to be added to the final set. For set IIIb, 3 messages received a mean score of 4.67 and 6 messages received a mean score of 4.33. Three of the 6 messages that received a mean score of 4.33 were selected randomly to be included in the final set. As a result, 12 messages were included in the final set, including 3 image-based, 5 video-based, and 4 text-based messages (Multimedia Appendix 1).

Message Set IV

This message set targets those who were in the indifference state (low threat and low efficacy), in which messages were designed to initiate the patient's efficacy of using their CPAP device and to convince patients of the threat of untreated OSA.

For message set IVa, 2 messages received a mean score of 4.33 and 6 messages received a mean score of 4.00; thus 4 messages were selected randomly to be added to the final set. For message set IVb, 1 message received a mean score of 4.67, 4 messages received a mean score of 4.33, and 3 messages had a mean score of 4. One message was randomly selected from the 3 messages with a mean score of 4.00 and was added to the final set. As a result, 12 messages were selected for the final set IV, including 4 image-based, 4 video-based, and 4 text-based messages (Multimedia Appendix 1).

Discussion

This study aimed to develop, validate the content of, and test a set of messages for an mHealth behavior change support system that will target patients with OSA who are noncompliant with their CPAP therapy. The goal was to motivate them to use their CPAP device and become compliant with their CPAP therapy. The multiphase message development, content validation, and testing phases described herein offer a practical guide for researchers and health care practitioners to use for tailored motivational message development. The tailored messages were built based on the EPPM targeting patients with OSA to motivate them to use their CPAP devices.

In phase 1, messages were derived from the literature and a review of the existing materials. This phase was made easier due to the availability of appropriate videos and infographics; developing image- and video-based messages from scratch can be costly and time-consuming. Fertman and Allensworth [57] noted the importance of reviewing existing materials that can assist the researcher or practitioner in deciding whether to use them as they are, modify them, or develop new messages from scratch.

In phase 2, the enumerated content validity of the messages was established through the utilization of I-CVI [56]. A total of 57 messages were found to have acceptable content relevance and clarity. The experts' content testing and their comments enriched and improved the messages in terms of clarity and relevance to the theoretical concepts. The suggestions for modification of some sentences and/or the presentation of information provided guidance to develop and design messages. The expert ratings of the messages assisted us in the process of categorizing messages into high or low threat and efficacy perception. The

value of expert knowledge and feedback has been recognized in message design and content validation [58-60].

In phase 3, the messages were further tested using a convenience sample of the target audience for perceived persuasiveness. Pretesting potential messages to identify the ones that are more likely to be relatively effective is an important step to assist the message designer or researcher in the process of selecting messages to include in the intervention [61].

Acquiring feedback from the target audience and experts in the subject matter assisted us in revising and selecting a set of messages that was presumed to be more effective. The results showed that each of the 4 final message sets included all 3 modes (image-, video-, and text-based messages). Studies note the advantages of video-based messages that require less mental and cognitive effort compared with plain text messages, freeing resources to process the main messages, thus making them potentially more applicable for those with low educational levels [62,63]. However, text-based and image-based messages may be more applicable in situations such as being in a noisy place or having a limited amount of time. As a result, we believe that one size does not fit all. Having a variety of modes is advantageous for reaching and considering as many different individual characteristics and/or preferences as possible as well as the range of environmental situations that someone may experience in daily life.

The main contribution of this study was the development of a selection process for validating the content of and testing a set of messages to be used in an mHealth behavior change support system solution in the context of OSA. There is a need for future intervention-oriented research using these messages to examine and determine their impact on patients' short- and long-term CPAP usage. Future research should evaluate IS-/information technology–based solution features and strategies for implementation in delivering and presenting these messages as a means to support the motivation of noncompliant patients with OSA. It will also be important to assess the effectiveness of these features and strategies in promoting CPAP usage among patients with OSA. If the intervention demonstrates effectiveness, it could prove to be an efficient motivational tool for CPAP users. It can be delivered to a large number of patients with OSA.

Conclusions

Poor use of CPAP therapy among patients with OSA is a serious problem. Effective interventions are needed to improve treatment usage by patients with chronic health conditions such as OSA to enhance health outcomes and quality of life [64]. This study was designed to help promote CPAP usage among patients with OSA. This was a formative study for developing and validating the content of a set of messages to be used in an mHealth sleep behavior change support system.

After the process of content validation and testing, the final library of messages met the criteria for clarity, relevance, and persuasiveness. This study emphasizes the importance of developing and validating the content of motivational messages, grounded in theory, across the 4 possible response states in terms of high/low efficacy and threat perceptions. We expect the potential impact of these message sets will come from serving as a resource for different IS solutions or mHealth interventions targeted at increasing CPAP usage/adherence among patients with OSA. Through the evaluation of these IS solutions or mHealth interventions, more insights will be gained on the efficacy and effectiveness of messages on CPAP adherence.

This study has some limitations. First, there is a selection bias for the experts' responses in phase 2. All experts had clinical and/or research expertise, but most of those who participated in relevance content validation were research experts (ie, either professors or PhD candidates) from different disciplines. However, most of the experts who participated in the clarity content validation were clinical experts in the field of pulmonology and/or sleep medicine. As the experts were derived from different disciplines, and each message set was evaluated by different experts (3 experts evaluated each set), there is a potential for selection bias. Second, the majority of patients with OSA who participated in the perceived persuasiveness testing (phase 3) were white and women. Finally, the small sample size is based on a convenience sampling technique. Having a small sample size limited the variety of statistical techniques that could be used. However, this study was not intended to generalize the results to a larger population; thus, considering the available resources, using convenience purposeful sampling was applicable in this case. Purposively enrolling participants will encompass a range of perspectives and views regarding the constructed messages.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Final Message Library.
[DOCX File , 119 KB - mhealth_v8i8e18793_app1.docx]

References

1. Lattimore J, Celermajer DS, Wilcox I. Obstructive sleep apnea and cardiovascular disease. J Am Coll Cardiol 2003 May 15;41(9):1429-1437 [FREE Full text] [doi: 10.1016/s0735-1097(03)00184-0]

2. An E. Which Came First, Obstructive Sleep Apnea or Hypertension? A Retrospective Study of Electronic Records Over 10 Years, With Separation by Sex. medRxiv. 2020. URL: https://www.medrxiv.org/content/10.1101/2020.01.27.20019018v1 [accessed 2020-07-15]

3. Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MS, Morrell MJ, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. Lancet Respiratory Med 2019 Aug;7(8):687-698. [doi: 10.1016/s2213-2600(19)30198-5]

4. de Luca CG, Pachêco-Pereira C, Aydinoz S, Major PW, Flores-Mir C, Gozal D. Biomarkers associated with obstructive sleep apnea and morbidities: a scoping review. Sleep Med 2015 Mar;16(3):347-357. [doi: 10.1016/j.sleep.2014.12.007] [Medline: 25747333]

5. Wickwire EM, Albrecht JS, Towe MM, Abariga SA, Diaz-Abad M, Shipper AG, et al. The impact of treatments for OSA on monetized health economic outcomes: a systematic review. Chest 2019 May;155(5):947-961 [FREE Full text] [doi: 10.1016/j.chest.2019.01.009] [Medline: 30664857]

6. Weaver TE, Sawyer AM. Adherence to continuous positive airway pressure treatment for obstructive sleep apnoea: implications for future interventions. Indian J Med Res 2010 Feb;131:245-258 [FREE Full text] [Medline: 20308750]

7. Olsen S, Smith SS, Oei TP, Douglas J. Motivational interviewing (MINT) improves continuous positive airway pressure (CPAP) acceptance and adherence: a randomized controlled trial. J Consult Clin Psychol 2012 Feb;80(1):151-163. [doi: 10.1037/a0026302] [Medline: 22103957]

8. Weaver TE. Novel aspects of CPAP treatment and interventions to improve CPAP adherence. J Clin Med 2019 Dec 16;8(12):2220 [FREE Full text] [doi: 10.3390/jcm8122220] [Medline: 31888148]

9. Sawyer AM, Gooneratne NS, Marcus CL, Ofer D, Richards KC, Weaver TE. A systematic review of CPAP adherence across age groups: clinical and empiric insights for developing CPAP adherence interventions. Sleep Med Rev 2011 Dec;15(6):343-356 [FREE Full text] [doi: 10.1016/j.smrv.2011.01.003] [Medline: 21652236]

10. Kendzerska T, Wilton K, Bahar R, Ryan CM. Short- and long-term continuous positive airway pressure usage in the post-stroke population with obstructive sleep apnea. Sleep Breath 2019 Dec;23(4):1233-1244. [doi: 10.1007/s11325-019-01811-9] [Medline: 30820850]

11. Witte K. Putting the fear back into fear appeals: the extended parallel process model. Commun Monogr 1992 Dec;59(4):329-349. [doi: 10.1080/03637759209376276]

12. Maloney EK, Lapinski MK, Witte W. Fear appeals and persuasion: A review and update of the Extended Parallel Process Model. Soc Personal Psychol 2011;5(4):206-219. [doi: 10.1111/j.1751-9004.2011.00341.x]

13. Gore TD, Bracken CC. Testing the theoretical design of a health risk message: reexamining the major tenets of the extended parallel process model. Health Educ Behav 2005 Feb;32(1):27-41. [doi: 10.1177/1090198104266901] [Medline: 15642752]

14. Witte K, Meyer G, Martell D. Effective Health Risk Messages: A Step-By-Step Guide. Thousand Oaks, CA: Sage Publications; 2001.

15. Popova L. The extended parallel process model: illuminating the gaps in research. Health Educ Behav 2012 Aug;39(4):455-473. [doi: 10.1177/1090198111418108] [Medline: 22002250]

16. Rimal RN. Perceived risk and efficacy beliefs as motivators of change: use of the risk perception attitude (RPA) framework to understand health behaviors. Hum Commun Res 2003 Jul 1;29(3):370-399. [doi: 10.1093/hcr/29.3.370]

17. Haba-Rubio J, Vujica J, Franc Y, Michel P, Heinzer R. Effect of CPAP treatment of sleep apnea on clinical prognosis after ischemic stroke: an observational study. J Clin Sleep Med 2019 Jun 15;15(6):839-847 [FREE Full text] [doi: 10.5664/jcsm.7832] [Medline: 31138378]

18. Lee Health. Sleep Apnea Could be Hurting Your Heart. YouTube. 2015. URL: https://www.youtube.com/embed/vSxEvWjpqBo [accessed 2020-07-15]

19. Sleep Apnea - Symptoms And Causes. Mayo Clinic. 2020. URL: https://www.mayoclinic.org/diseases-conditions/sleep-apnea [accessed 2020-07-01]

20. Living With CPAP: 7 Tips for a Better Experience. ResMed. 2020. URL: https://www.resmed.com/us/en/blog/improving-therapy/living-with-cpap-7-tips-for-a-better-experience.html?b [accessed 2020-03-01]

21. CPAP Therapy. ResMed. 2020. URL: https://www.resmed.com.au/cpap [accessed 2020-03-01]

22. Sleep Apnea. Lung Cancer Canada. 2018. URL: https://www.lung.ca/lung-health/lung-disease/sleep-apnea/cpap [accessed 2020-02-15]

23. Risk of Motor Vehicle Accidents is Higher in People With Sleep Apnea. The American Academy of Sleep Medicine (AASM). 2018. URL: https://aasm.org/risk-of-motor-vehicle-accidents-is-higher-in-people-with-sleep-apnea/ [accessed 2020-02-15]

24. Troubleshooting Guide for CPAP Problem. American Sleep Apnea Association (ASAA). 2020. URL: https://www.sleepapnea.org/treat/cpap-therapy/troubleshooting-guide-for-cpap-problems/ [accessed 2020-06-15]

25. 3 Tips for Sleeping with a CPAP. YouTube. URL: https://www.youtube.com/embed/GzrJCqHCx8U [accessed 2020-02-15]

26. What Cause Sleep Apnea? YouTube. URL: https://www.youtube.com/embed/2JJExD2n39U [accessed 2020-06-15]

XSL•FO

RenderX

27.   How to Clean Your CPAP. YouTube. URL: https://www.youtube.com/embed/PgUeDkZvS4M [accessed 2020-02-15]
28.   Cleveland C. PAP Therapy: What to Expect. YouTube. URL: https://www.youtube.com/embed/tgSMlsoNDFg [accessed 2020-06-15]
29.   Harneet Walia, M.D. - 'Study: Common Sleep Apnea Treatment Reduces Drowsy Driving'. YouTube. URL: https://www.youtube.com/embed/d2UqGInReZ4 [accessed 2020-06-15]
30.   Sleep Apnea-Sleep Education. YouTube. URL: https://www.youtube.com/embed/6bqTOfwvWhk [accessed 2020-06-15]
31.   Sleep Apnea Increase Risk of Heart Attack, Study Finds. YouTube. URL: https://www.youtube.com/embed/vbTCWLnoK2A [accessed 2020-06-15]
32.   Obstructive Sleep Apnea. YouTube. URL: https://www.youtube.com/embed/VPwpJd8aBzM.Published [accessed 2020-02-18]
33.   Obstructive Sleep Apnea. YouTube. URL: https://www.youtube.com/embed/z12MEPiG4cg [accessed 2020-02-18]
34.   Mayo Clinic Minute: Signs of a Sleep Disorder. YouTube. URL: https://www.youtube.com/embed/lu-pMXDZ8IA [accessed 2020-02-20]
35.   Sleep Well with Positive Airway Pressure (PAP). YouTube. URL: https://www.youtube.com/embed/8_6CDuSiajA [accessed 2020-02-17]
36.   How Sleep Disorders Affect Your Heart. Adventist Health. URL: https://www.adventisthealth.org/images/import/wp-content/uploads/2016/03/sleep-disorders-infographic.png [accessed 2020-02-17]
37.   Top 10 CPAP Life Hacks. National Jewish Health. URL: https://www.nationaljewish.org/getattachment/Health-Insights/health-infographics/top-10-cpap-life-hacks/Top-10-CPAP-Life-Hacks-(1).jpg.aspx [accessed 2020-02-10]
38.   Sleep Apnea. Everything Birds. URL: https://2fgaqu3s48au2o452o28glwj-wpengine.netdna-ssl.com/wp-content/uploads/2019/01/Infographic_Sleep_Apnea-Symptoms-Risks-Effects-and-Treatments-150dpi.png [accessed 2020-06-20]
39.   Replacement Schedule. CPAP. URL: https://5faacd0c97cf0d23c6dc-064f3cd61bcd6e809053bf2fe3432ece.ssl.cf1.rackcdn.com/replacementschedulecpapdotcom.pdf [accessed 2020-06-19]
40.   What is Obstructive Sleep Apnea? WordPress. URL: https://pbs.twimg.com/media/DEksFuGVoAA8R9p.jpg [accessed 2020-06-20]
41.   How to Survive CPAP During Allergy Season. American Sleep Apnea Association. URL: https://www.sleepapnea.org/wp-content/uploads/2017/04/CPAP-and-Allergies-ASAA-infoG.png [accessed 2020-06-17]
42.   Hypertension Affects Your Whole Body. Omaha Sleep Apnea & Snoring Dentist: Dr Roger Roubal. URL: https://www.whywesnore.com/x/lc-content/uploads/2017/02/2016-sleep-apnea-infographic-8-2.png [accessed 2020-06-17]
43.   Obstructive Sleep Apnea. AAOMS: Experts in Face, Mouth & Jaw Surgery. URL: https://myoms.org/wp-content/uploads/2019/09/09_obstructive_sleep_apnea.png [accessed 2020-06-17]
44.   Risks of Untreated Sleep Apnea. Verywell Health - Know More. Feel Better. URL: https://www.verywellhealth.com/thmb/eyDOJ3PGeHv5x161dCOTZ34VpbE=/1500x1000/filters:no_upscale():max_bytes(150000):strip_icc()/sleep-apnea-overview-3014774-v1-5c5dbe9a46e0fb00017dd118.png [accessed 2020-06-17]
45.   Prevalence of Sleep Apnea in Associated Co-Morbidities. Squarespace. URL: https://static1.squarespace.com/static/591f6ccfb8a79bb177f17924/t/598b34ccd2b8578f996350f7/1506117076239/sleep-apnea-infographic-sleep-data.jpg?format=1500w [accessed 2020-06-17]
46.   Untreated Obstructive Sleep Apnea has Negative Effects on Medical Conditions. Craniofacial Pain and Dental Sleep Center of Georgia. URL: https://atlantaheadachetmjpain.com/wp-content/uploads/untreated-obstructive-sleep-apnea-has-negative-effects-on-medical-conditions..png [accessed 2020-02-17]
47.   Clark RC, Mayer RE. E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. Hoboken, New Jersey: John Wiley & Sons; 2016.
48.   Gunning R. The fog index after twenty years. Int J Bus Commun 2016 Sep 16;6(2):3-13. [doi: 10.1177/002194366900600202]
49.   Dey S, Sedera D. Information System Experts: Definition, Identification, and Application. In: Proceedings of the 19th European Conference on Information Systems. 2011 Presented at: ECIS'11; June 9-11, 2011; Helsinki, Finland.
50.   Williams J, Hadjistavropoulos T, Ghandehari OO, Malloy DC, Hunter PV, Martin RR. Resilience and organisational empowerment among long-term care nurses: effects on patient care and absenteeism. J Nurs Manag 2016 Apr;24(3):300-308. [doi: 10.1111/jonm.12311] [Medline: 26044745]
51.   Sawyer AM, King TS, Weaver TE, Sawyer DA, Varrasse M, Franks J, et al. A tailored intervention for PAP sdherence: the SCIP-PA trial. Behav Sleep Med 2019;17(1):49-69 [FREE Full text] [doi: 10.1080/15402002.2016.1276018] [Medline: 28128977]
52.   Grady C. Payment of clinical research subjects. J Clin Invest 2005 Jul;115(7):1681-1687 [FREE Full text] [doi: 10.1172/JCI25694] [Medline: 16007244]
53.   Weaver TE, Maislin G, Dinges DF, Younger J, Cantor C, McCloskey S, et al. Self-efficacy in sleep apnea: instrument development and patient perceptions of obstructive sleep apnea risk, treatment benefit, and volition to use continuous positive airway pressure. Sleep 2003 Sep;26(6):727-732. [doi: 10.1093/sleep/26.6.727] [Medline: 14572127]
54.   Olsen S, Smith S, Oei T, Douglas J. Health belief model predicts adherence to CPAP before experience with CPAP. Eur Respir J 2008 Sep;32(3):710-717 [FREE Full text] [doi: 10.1183/09031936.00127507] [Medline: 18417510]
55.   Homoud M. Adults' Knowledge and Beliefs Surrounding Obstructive Sleep Apnea. Seton Hall University. 2019. URL: https://scholarship.shu.edu/cgi/viewcontent.cgi?article=3697&context=dissertations [accessed 2020-07-15]

56. Lynn MR. Determination and quantification of content validity. Nurs Res 1986;35(6):382-385. [Medline: 3640358]

57. Fertman CI, Allensworth DD. Health Promotion Programs: From Theory to Practice. Hoboken, NJ: John Wiley and Sons; 2016.

58. Lima I, Galvão M, Pedrosa SC, Silva CA, Pereira ML. Validação de mensagens telefônicas para promoção da saúde de pessoas com HIV. Acta Paul Enferm 2017 May;30(3):227-232. [doi: 10.1590/1982-0194201700035]

59. Helitzer D, Hollis C, Cotner J, Oestreicher N. Health literacy demands of written health information materials: an assessment of cervical cancer prevention materials. Cancer Control 2009 Jan;16(1):70-78. [doi: 10.1177/107327480901600111] [Medline: 19078933]

60. Celik S, Cosansu G, Erdogan S, Kahraman A, Isik S, Bayrak G, et al. Using mobile phone text messages to improve insulin injection technique and glycaemic control in patients with diabetes mellitus: a multi-centre study in Turkey. J Clin Nurs 2015 Jun;24(11-12):1525-1533. [doi: 10.1111/jocn.12731] [Medline: 25422134]

61. O'Keefe DJ. Message pretesting using perceived persuasiveness measures: reconsidering the correlational evidence. Commun Methods Meas 2019 May 30;14(1):25-37. [doi: 10.1080/19312458.2019.1620711]

62. Stanczyk NE, Smit ES, Schulz DN, de Vries H, Bolman C, Muris JW, et al. An economic evaluation of a video- and text-based computer-tailored intervention for smoking cessation: a cost-effectiveness and cost-utility analysis of a randomized controlled trial. PLoS One 2014;9(10):e110117 [FREE Full text] [doi: 10.1371/journal.pone.0110117] [Medline: 25310007]

63. Walthouwer MJ, Oenema A, Soetens K, Lechner L, De Vries H. Systematic development of a text-driven and a video-driven web-based computer-tailored obesity prevention intervention. BMC Public Health 2013 Oct 20;13:978 [FREE Full text] [doi: 10.1186/1471-2458-13-978] [Medline: 24138937]

64. Long H, Bartlett YK, Farmer AJ, French DP. Identifying brief message content for interventions delivered via mobile devices to improve medication adherence in people with type 2 diabetes mellitus: a rapid systematic review. J Med Internet Res 2019 Jan 9;21(1):e10421 [FREE Full text] [doi: 10.2196/10421] [Medline: 30626562]

Abbreviations

CPAP: continuous positive airway pressure
EPPM: extended parallel process model
I-CVI: content validity index for items
IS: information systems
mHealth: mobile health
OSA: obstructive sleep apnea
OSA-KAB: obstructive sleep apnea knowledge and beliefs
SEMSA: Self-Efficacy Measure for Sleep Apnea

[XSL•FO]
RenderX

Original Paper

Two-Way Text Messaging to Support Self-Care and Delivery of an Online Sexual Health Service: Mixed Methods Evaluation

Sarah Shanks[1], BVSc, MPH; Alessandra Morelli[2], MSc; Elena Ardines[1], BSc; Gillian Holdsworth[1], MBChB, MPH, MSc; Paula Baraitser[1,3], MBBS, BSc, MA, MD, FFPH, FFSRH

[1]SH:24, London, United Kingdom

[2]Sexual Health Research Group, King's Centre for Global Health and Health Partnerships,, King's College London, London, United Kingdom

[3]Department of Sexual Health and HIV, Kings College Hospital National Health Service Foundation Trust, London, United Kingdom

Corresponding Author:
Sarah Shanks, BVSc, MPH
SH:24
35a Westminster Bridge Road
London, SE1 7JB
United Kingdom
Phone: 1 02076202250
Email: sshanks@nhs.net

Abstract

Background: Digital health care is increasingly used to improve health service accessibility and reduce costs. Remote health care requires a significant self-management role for service users, and this generates information provision and support needs that should be reflected in service planning. SMS text messaging offers a convenient and low-cost method of communication and is increasingly used across digital health care services to provide remote support.

Objective: The aim of this study was to quantify the number of messages generated through user interaction with a two-way SMS text messaging support service within an online sexual health service and to thematically explore the content of the messages and type of support required to facilitate self-management.

Methods: The content of all SMS text messages received by an online sexual health service was analyzed from April 4, 2018, to July 5, 2018. Messages were classified as being either administrative or clinical in nature and service or user initiated. For those messages that were both clinical and user initiated, a qualitative thematic analysis was completed to fully describe the content of the interactions.

Results: A total of 267 actionable messages were generated per 1000 orders requested through the service. Of the 8562 messages, 5447 (63.62%) messages were administrative and 3115 (36.38%) were clinical. Overall, 4306 of the 8562 messages (50.29%) responded to service-generated queries reflecting the public health and clinical responsibilities of an online provider, and 4256 (49.71%) were user-generated queries, demonstrating a willingness by users to proactively engage with a two-way SMS text messaging support service. Of the 3115 clinical messages, 968 (31.08%) clinical messages were user initiated and shared personal and complex clinical information, including requests for help with the self-testing process and personalized clinical advice relating to symptoms and treatment.

Conclusions: This study demonstrates the willingness of users of an online sexual health service to engage with two-way SMS text messaging and provides insight into the quantity and nature of the support required to facilitate service delivery and self-care. Further work is required to understand the range of clinical problems that can be managed within this medium.

Introduction

Digital health care is used to improve health service accessibility and reduce costs, and online health services are increasingly part of a "digital first" National Health Service (NHS) [1,2]. Sexual health services have been important innovators in this field, with online testing for sexually transmitted infections

(STIs) and online contraception now available in many parts of the United Kingdom.

Digital health services require users to take on new responsibilities for self-care, such as self-taken finger prick blood samples or interpreting results of tests completed at home, and these new roles require additional support [3,4]. This support may be delivered through a wide range of media, including telephone, video conversations, and messaging. Telephone and video conversations were used in early self-care services, possibly because the synchronous and voiced-based elements of these media reproduced face-to-face conversations more closely; however, health services are now progressively using text-based and asynchronous media for remote communications.

SMS text messaging is increasingly used within face-to-face sexual health services to provide test results, information, and support. One-way and two-way text messaging support is convenient, confidential, and anonymous. It is also low cost and accessible, especially to young people [5-9]. One-way messaging (usually service to user) is more commonly used than two-way messaging, and standardized messages appear to have less impact then tailored and customized messages [10]. There have been small-scale pilots of support via text messaging to facilitate self-care within online services [11-13]. However, there has been no evaluation of this type of support when delivered at scale. This gap in the evidence is important, as those providing online services need information about the number and type of messages generated to inform service planning, including staff time and expertise required.

This study reports an analysis of the use of two-way SMS text messaging support within a large online sexual health service. We used both quantitative and qualitative methods to understand (1) the volume of SMS text messages generated through user interaction with the service and (2) the nature of the messages sent by users and type of support required to facilitate service delivery and self-management.

Methods

Study Setting

We studied digital communication within a large digital sexual health service based in the United Kingdom. SH:24 is an online sexual health provider that offers self-sampling for STI tests, chlamydia treatment for users who test positive, and online contraception [14]. SH:24 is a community interest company that is commissioned by the NHS to provide services in 35 regions across the United Kingdom. The service was recently expanded to offer Fettle, a paid-for service for people living without a public sector–commissioned online service and who choose to have their care this way. SH:24 was designed to be integrated with face-to-face services and to work as part of the whole system of sexual health care [4], ensuring that users are appropriately referred between online- and clinic-based services according to clinical need.

The SH:24 and Fettle service is provided through an online portal. Users access the portal via a website to order sexual health tests, treatment, or oral contraception. Users are required to submit personal information as part of the ordering process when relevant to the service requested. For example, a sexual behavior history is required for sexual health testing and a medical history is required for contraception orders. Sexual health test kits are delivered to the user's home, and users take their own samples; this is usually a finger prick blood test, plus a self-taken vaginal swab for women or a urine sample for men. Samples are mailed by users directly to the laboratory and results are provided by SMS or a telephone call, depending on the infection identified. Treatments (for simple chlamydia infection and oral contraceptives) are prescribed based on test results and an online medical history. After prescription by a UK General Medical Council–registered doctor, treatments are dispensed and mailed to users by a UK-registered online pharmacy. The service is available 7 days a week and, at the time of this study, clinical support for the service was provided by a team of 3 specialist sexual health nurses and 1 clinical support worker. The clinical team has since expanded, as demand for the service has increased.

Once an order has been submitted, SMS text messaging is the primary medium for communication between service and user. SMS text messages have multiple functions within the service, from resolving logistical queries between the service and users to providing important health promotion information or guidance about self-management tasks, such as self-sampling or taking treatment. Two-way communication is encouraged and messages end with the option to contact the clinical team for additional information (see Textbox 1). Users can contact the clinical team at any time by replying to a message. Telephone consultations are used to provide reactive HIV results and are offered as part of the risk assessment process for safeguarding concerns.

Textbox 1. Examples of SMS text messages to support self-sampling and promote hepatitis B vaccination for men who have sex with men. GP: general practitioner.

Example self-sampling text message

"Hello. The Royal Mail should deliver your test kit within the next 72hrs. Infections may not show up in tests immediately after exposure, so it's important you test at the right time. If you are taking a blood test, we recommend taking it 4 weeks after your potential exposure - you can watch a 2 minute video to help you take a blood sample here: bit.ly/bloodtestSH24. If you are ONLY testing for chlamydia and gonorrhoea, we recommend taking the test 2 weeks after a potential exposure. Check the best time to take your test here: bit.ly/when2. Text back if you would like help. Thanks, SH:24"

Example hepatitis B vaccination text message

"Hello, you have told us that you have not been vaccinated against hepatitis B. Men who have sex with men can get vaccinated for free at any sexual health clinic or GP surgery. Hepatitis B is carried in blood so can be transmitted through having sex with someone who has the virus. Someone who has hepatitis B may appear well, whilst for some people, the infection can be long-lasting and lead to serious liver disease. Having a course of 3 hepatitis B vaccinations is the best way to protect yourself from the infection. Find your local sexual health clinic: https://www.nhs.uk/service-search/other-services/Sexual-health-information-and-support/LocationSearch/734. If you have any questions please reply to this message, your text will be read by a member of our clinical team. Thanks, SH:24"

Data Collection

We used routinely collected data to describe the demographics of users of SH:24 and Fettle between April 4, 2018, and July 5, 2018. These data are collected every time an order is placed with the service, including when multiple orders are generated by a single user. We extracted the text of all incoming SMS text messages received by SH:24 and Fettle during this time period. The data received contained the content of the message only; there were no details about the sender, their service use, or the time or date of sending. It was not possible to differentiate between text messages sent by users of SH:24 and text messages sent by users of Fettle unless this was specifically referred to in the content of the message.

Analysis

Demographic data were analyzed using Stata 15 (StataCorp LLC) in terms of age, sex, sexuality, and ethnicity. This analysis was done at the aggregate level to describe the population per order placed, not per individual user.

The SMS messages were analyzed by 3 members of the research team (SS, AM, PB). Messages were categorized by qualitative content and all messages were included in the initial analysis. Through initial reading and rereading of the messages, we developed a simple classification to describe them: (1) administrative or clinical and (2) user generated or service generated.

First, administrative messages were defined as those that required no clinical knowledge by the online support staff to reply to or act on. Clinical messages were defined as all messages that would require clinical knowledge by the online support staff to respond to or act on.

Second, service-generated messages were defined as messages that had been sent by users in response to a question asked by the service. User-generated messages were defined as those with content indicating they had been sent by the user without an initial prompt from the service.

The full data set was then coded according to this system. Messages containing content of no identifiable meaning or purpose were discarded at this stage and excluded from further analysis. The frequencies of different categories of text message were calculated to describe the volume of SMS text messages generated through user interaction with the service and the nature of the messages sent by users.

For messages that were categorized as both clinical and user generated, we completed a qualitative thematic analysis to fully describe the content of the interactions. This involved the research team reading and rereading all messages with an iterative process of theme generation, discussion, and development until a comprehensive structure for describing these data was developed and agreed upon by all the researchers involved in coding (SS, AM, PB). Initial coding frameworks were modified and refined during this process until the team agreed that the final framework effectively reflected the content of the SMS conversations analyzed. The final coding structure included 4 main themes and multiple subthemes. The main themes from the final coding analysis were (1) requests for personal support with the self-management process, (2) requests for help with personal assessment of risk of infection, (3) requests for help interpreting test results, and (4) requests for personalized clinical advice.

A final round of thematic analysis was completed within each category to describe and interpret the material within each main coding theme. When direct quotes were used to illustrate messages, we altered small details of the text to maintain confidentiality without changing the sense or tone of the message.

Ethics

Approval was obtained as a part of an MSc project from the Health Research Authority Research Ethics Committee (reference: 18/HRA/0128; Integrated Research Application System project ID: 224808).

Results

### Characteristics of Text Message Service Usage

Between April 4, 2018, and July 5, 2018, there were 38,033 orders requested through the service. Of the 38,033 orders, 34,494 (90.69%) were placed through SH:24 and 3539 (9.31%) through Fettle. The demographics of the users who created the orders are shown in Table 1.

Table 1. User demographics for orders requested between April 4, 2018, and July 5, 2018 (N=38,033).

Characteristic	Orders, n (%)
Age group (years)	
16-17	749 (1.97)
18-24	16,477 (43.32)
25-34	15,316 (40.27)
35-54	5077 (13.35)
55+	414 (1.09)
Gender	
Female	23,033 (60.56)
Male	15,000 (39.44)
Ethnicity	
White	24,402 (64.16)
Black/African/Caribbean/Black British	3929 (10.33)
Asian/Asian British	969 (2.55)
Mixed/multiple ethnicity	2172 (5.71)
Other	2560 (6.73)
Not known/prefer not to say	462 (1.21)
Not asked	3539 (9.31)
Sexual preference of female users	
Men and women	1046 (4.54)
Men	21,499 (93.34)
Women	488 (2.12)
Sexual preference of male users	
Men and women	750 (5.00)
Men	2584 (17.23)
Women	11,666 (77.77)

A total of 10,152 incoming SMS text messages were received during the 3-month time period. After the initial analysis, 1590 messages were discarded due to insufficient content. This included messages containing characters or words that did not make grammatical sense, for example ";;; I," and those that consisted only of generic conversational words or phrases, such as "thank you," "hello," or "ok." Of the remaining 8562 SMS text messages, 5447 (63.62%) were classified as administrative and 3115 (36.38%) were classified as clinical (Figure 1).

XSL•FO
RenderX

Figure 1. Results of quantitative analysis.
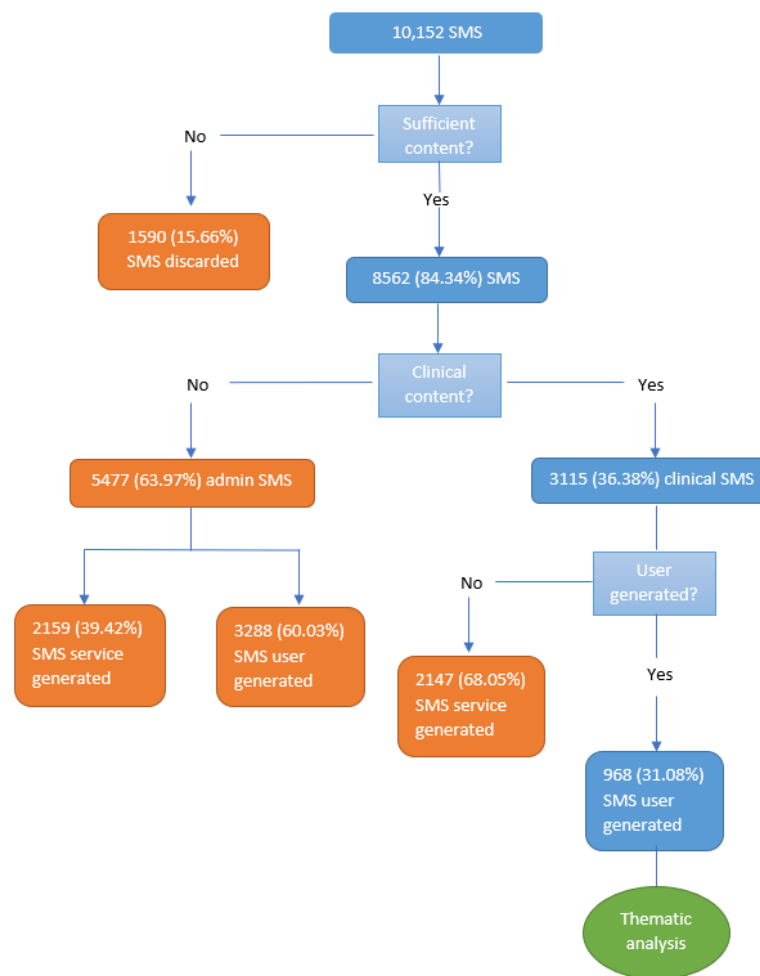
Of the 5447 administrative messages, 2159 (39.64%) were service generated to facilitate the running of the service, for example messages confirming user age or address and messages sent in reply to queries from the support team about unused test kits:

> *Yes that is the correct address.*
>
> *Yes sorry, hectic time recently!! Will send it off tomorrow, thanks.*

A total of 3288 of the 5447 administrative messages (60.36%) were user generated. The majority of these contained logistical queries about user orders, typical of those asked for any online order service:

> *HI I posted my kit today when do I get results?*
>
> *Can I track delivery and will I need to sign for it. And will "fettle" appear on packaging?*

There were also queries about the services provided at SH:24 and messages from users wanting help navigating sexual health services:

> *Do you do pregnancy tests?*
>
> *Can you tell my blood type?*
>
> *I visited the clinic before and they told me just to get a kit, that was couple of months ago now so I'm a bit*

hesitant about attending and being told the same thing again.

> *Hi there, could you please recommend a clinic that can do an instant HIV test?*

Of the 3115 messages that were classified as clinical messages, 2147 responded to service-generated questions that reflect the public health role and clinical function of the service. This included messages confirming that partner notification for STIs had been completed or that treatment had been received and taken and messages justifying a request to retest soon after a previous positive test result:

> *Hello, yes I confirm I have taken the treatment and experienced no vomiting or diarrhoea.*
>
> *Hi I have had one but am still getting the symptoms. Is it not possible to do one more? The wait to go to the clinic is weeks. Thank you.*

A total of 190 messages in this category contained responses to service-generated clinical questions prior to prescribing oral contraception or antibiotics:

> *Hi, I have been religiously taking my pill and am definitely not pregnant.*

Finally, there were a very small number of messages, a total of 24, sent in relation to a safeguarding risk assessment that is completed during the online order process as part of the service's

responsibility to identify and support vulnerable users, particularly those younger than 18 years [15]. In these messages, users were willing to disclose highly sensitive personal information and were open to discussions about support through messaging:

> I got forced to have sex.

> Hi. I'm not really sure what type of counselling I require, just feel sad and like I am losing interest in things.

Thematic Analysis of User-Initiated Clinical Messages

A total of 968 out of 3115 messages in the clinical category (31.08%) were user initiated, and an in-depth thematic analysis was completed on these data. The messages in this category had an informal, conversational tone and shared personal, often complex information. The thematic content analysis of these messages identified 4 main subjects of requests for personal support: help with the self-management of STI testing, help to assess risk of infection, requests for support with interpreting test results, and requests for personalized clinical advice.

Requests for Personal Support With the Self-Management Process

Messages in this theme were sent by users with questions about STI test self-sampling completion. These included procedural and technical questions about using the kit:

> How far do I insert the swab?

> Hi, I got a test kit at the beginning of the year but never got round to doing it, can I do it now?

The messages associated with blood test completion described the difficulty that some users had completing self-sampling and the new thinking that the process of a self-taken blood sample generated:

> Hi, I've just done the blood collection for the hiv sample, and it took me a long time to fill the tube…my blood is so thick and wouldn't flow out at all. I got a bit worried, isn't it supposed to be more fluid and liquid? Thx for letting me know something.

> Hi. Can I only send vaginal swap sample? I fainted when trying to take blood sample so I think it would be better to it done at the clinic. Thanks?

Requests for Personal Assessment of Risk of Infection

These messages reflected the level of concern generated by the possibility of STI transmission and the work that users did to predict their chances of infection, even during the process of testing. In these situations, complex descriptions were provided to support the anticipated clinical assessment of their risk:

> So I received oral sex from male that has chlamydia but he contracted it through genital sex. He didn't give oral sex to the person he contracted it from and when I received oral sex from him no fluids except saliva were exchanged. Is there any chance that I could receive chlamydia?

Help With Interpreting Test Results

Many users required reassurance about the accuracy and reliability of the tests when they received an unexpected result:

> How accurate are these tests…My boyfriend tested positive and got treatment. So i got tested with yourselfs and the test I received the results from showed negative?! This makes no sense?

> Thank you, is that 100%? Or could I just not have swabbed enough?

Other users needed support to understand the actual meaning of the test results:

> To clarify: my result was "negative". Does this mean I'm 100% sure about it being negative and there is no need to repeat any test? Thanks.

Requests for Personalized Clinical Advice

Messages in this theme contained clinical information that required a decision and response by a skilled health care professional. The messages were mainly related to the more complex services provided online (antibiotic treatment and contraception) or to users experiencing symptoms of infection:

> I have itching which I thought was thrush- would this need additional treatment?

> Hi I have white patches around my area down there, there was a lot more than there is right now but I've looked at them with a mirror and I'm really worried now.

The answers to these queries would not be available on general sexual health information websites and related to specific, often complex, clinical circumstances.

Discussion

Principal Findings

This study provides insight into how people interact with a two-way SMS text messaging support service within an online sexual health service and demonstrates the quantity and nature of support required to administer the service and facilitate self-care by users. During the study period, 8562 actionable SMS messages were received, corresponding to 267 messages generated per 1000 orders requested. A total of 63.62% (5447/8562) of the messages were administrative and 36.38% (3115/8562) required action by a member of staff with clinical knowledge.

Overall, 4306 of the 8562 actionable messages (50.29%) responded to service-generated queries that reflect the public health and clinical responsibilities of this service, and 4256 (49.71%) were user-initiated queries, demonstrating a willingness by users to proactively engage with a two-way SMS text messaging support service. In-depth analysis of the user-generated clinical messages was completed to explore the complexity of clinical questions asked by users over SMS. The results demonstrated the acceptability of providing personal information and discussing sensitive clinical matters using this platform. The queries highlighted the gaps in skills (for self-taken samples) and knowledge (to manage unanticipated

XSL•FO
RenderX

issues such as medication side effects) that are needed to complete the self-care process.

The use of SMS to support personalized clinical communication is part of a shift from remote communication by synchronous and voice-based communication media, such as telephone or video, to asynchronous and text-based media, such as SMS text messaging. This option may be less intrusive or susceptible to interruptions, more durable, and less likely to be subject to distortion than voice-based alternatives, and it reflects a change in the usage landscape of mobile devices [16,17]. The level of SMS support required is significant and needs adequate resourcing, and this should be built into new online service development. Since the time this study was completed, the combined SH:24 and Fettle service has expanded from processing 38,000 orders in a 3-month period to a current average of approximately 25,000 orders per month. The findings of the study were used to inform service planning, as the demand for online services increased. It is our experience that it takes an average of 2 minutes to act on an SMS text message. We calculate staff time per orders on the assumption that 25% of user orders (across STI kits, treatment, and contraception) will generate an incoming text message, based on the results of this study. For example, 25,000 orders will require staff to answer 6250 messages. We assume that 60% of the messages will require administrative action and 40% will require a response from a clinically trained member of staff. Therefore, 25,000 orders translates to 125 hours of administrator time and 83.3 hours of clinical time. These calculations reflect time spent on the dedicated SMS text messaging support service only and do not include the other clinical and administrative tasks required

in a service of this kind, such as telephone and email communications, prescribing, and supporting partner notification.

Text messaging solutions for health care have been successfully adopted to assist with remote clinical monitoring, information and education services, adherence to treatment or self-management, and consultation [18]. We conclude that the "digital first" plans for health care delivery would benefit from a better understanding of the value of this type of support. Further work is required to understand the range of clinical problems that can be managed within this medium outside of sexual health services.

Limitations

The main limitation of this study is that text messages could not be identified by service user. This information would determine what proportion of users engaged with the SMS support service and whether there were differences between users of a free or paid-for service. It would also enable patterns of communication between service providers and individual users to be examined. This information would not have altered the findings of the current study; the relative amounts of administrative and clinical support required for the online service is dependent on the content and number of messages received, not on the number of users responsible for sending them. However, further studies could provide additional useful insight into the volume and nature of user interactions with online health services. The generalizability of our findings will be variable across contexts depending on norms of SMS text messaging use and stigma surrounding the discussion of sexual health.

Conflicts of Interest

PB is a director of SH:24 and an employee at King's College London and King's College Hospital. GH is a director of SH:24. AM is a temporary online clinical support midwife at SH:24.

References

1. Wachter RM. Making IT work: harnessing the power of health information technology to improve care in England.: UK Department of Health; 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf [accessed 2019-11-25]
2. The NHS long term plan. London, United Kingdom: National Health Service; 2019. URL: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf [accessed 2019-11-25]
3. Baraitser P, Cribb A. "Putting people in charge of their own health and care?" Using meta-narrative review and the example of online sexual health services to re-think relationships between e-health and agency. Health Expect 2019 Oct;22(5):838-848 [FREE Full text] [doi: 10.1111/hex.12895] [Medline: 31054218]
4. Turner KM, Zienkiewicz AK, Syred J, Looker KJ, de Sa J, Brady M, et al. Web-Based Activity Within a Sexual Health Economy: Observational Study. J Med Internet Res 2018 Mar 07;20(3):e74 [FREE Full text] [doi: 10.2196/jmir.8101] [Medline: 29514776]
5. Perry RC, Kayekjian KC, Braun RA, Cantu M, Sheoran B, Chung PJ. Adolescents' perspectives on the use of a text messaging service for preventive sexual health promotion. J Adolesc Health 2012 Sep;51(3):220-225. [doi: 10.1016/j.jadohealth.2011.11.012] [Medline: 22921131]
6. Lim MSC, Hocking JS, Hellard ME, Aitken CK. SMS STI: a review of the uses of mobile phone text messaging in sexual health. Int J STD AIDS 2008 May;19(5):287-290. [doi: 10.1258/ijsa.2007.007264] [Medline: 18482956]
7. Lim MSC, Hocking JS, Aitken CK, Fairley CK, Jordan L, Lewis JA, et al. Impact of text and email messaging on the sexual health of young people: a randomised controlled trial. J Epidemiol Community Health 2012 Jan;66(1):69-74. [doi: 10.1136/jech.2009.100396] [Medline: 21415232]

XSL•FO

RenderX

8.   Willoughby JF, Jackson K. 'Can you get pregnant when u r in the pool?': young people's information seeking from a sexual health text line. Sex Education 2013 Jan;13(1):96-106. [doi: 10.1080/14681811.2012.677746]

9.   Willoughby JF, L'Engle KL. Influence of perceived interactivity of a sexual health text message service on young people's attitudes, satisfaction and repeat use. Health Educ Res 2015 Dec;30(6):996-1003. [doi: 10.1093/her/cyv056] [Medline: 26590247]

10. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]

11. Willoughby JF, Muldrow A. SMS for sexual health: A comparison of service types and recommendations for sexual health text message service providers. Health Education Journal 2016 Sep 27;76(2):231-243. [doi: 10.1177/0017896916661373]

12. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]

13. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]

14. Homepage. SH:24. URL: https://www.SH24.org.uk/ [accessed 2019-11-25]

15. Spencer-Hughes V, Syred J, Allison A, Holdsworth G, Baraitser P. Screening for Child Sexual Exploitation in Online Sexual Health Services: An Exploratory Study of Expert Views. J Med Internet Res 2017 Feb 14;19(2):e30 [FREE Full text] [doi: 10.2196/jmir.5911] [Medline: 28196790]

16. Horwitz LI, Detsky AS. Physician communication in the 21st century: to talk or to text? JAMA 2011 Mar 16;305(11):1128-1129. [doi: 10.1001/jama.2011.324] [Medline: 21406650]

17. Forgays DK, Hyman I, Schreiber J. Texting everywhere for everything: Gender and age differences in cell phone etiquette and use. Computers in Human Behavior 2014 Feb;31:314-321. [doi: 10.1016/j.chb.2013.10.053]

18. Bäck I, Mäkelä K. Mobile Phone Messaging in Health Care – Where are we Now? J Inform Tech Soft Engg 2012;01(02). [doi: 10.4172/2165-7866.1000106]

Abbreviations

NHS: National Health Service
STI: sexually transmitted infection

XSL•FO
RenderX

Original Paper

Characteristics of Neuropsychiatric Mobile Health Trials: Cross-Sectional Analysis of Studies Registered on ClinicalTrials.gov

Mia Tova Minen[1], MD, MPH; Julia Frederica Reichel[2], BA; Pallavi Pemmireddy[2], BA; Elizabeth Loder[3], MD, MPH; John Torous[4], MD, MBI

[1]NYU Langone Health, New York, NY, United States
[2]Barnard College of Columbia University, New York, NY, United States
[3]Brigham and Women's Hospital, Boston, MA, United States
[4]Beth Israel Deaconess Medical Center, Brookline, MA, United States

Corresponding Author:
Mia Tova Minen, MD, MPH
NYU Langone Health
222 E 41st Street,
New York, NY, 10017
United States
Phone: 1 2122637744
Email: minenmd@gmail.com

Abstract

Background: The development of mobile health (mHealth) technologies is progressing at a faster pace than that of the science to evaluate their validity and efficacy. Under the International Committee of Journal Medical Editors (ICMJE) guidelines, clinical trials that prospectively assign people to interventions should be registered with a database before the initiation of the study.

Objective: The aim of this study was to better understand the smartphone mHealth trials for high-burden neuropsychiatric conditions registered on ClinicalTrials.gov through November 2018, including the number, types, and characteristics of the studies being conducted; the frequency and timing of any outcome changes; and the reporting of results.

Methods: We conducted a systematic search of ClinicalTrials.gov for the top 10 most disabling neuropsychiatric conditions and prespecified terms related to mHealth. According to the 2016 World Health Organization Global Burden of Disease Study, the top 10 most disabling neuropsychiatric conditions are (1) stroke, (2) migraine, (3) major depressive disorder, (4) Alzheimer disease and other dementias, (5) anxiety disorders, (6) alcohol use disorders, (7) opioid use disorders, (8) epilepsy, (9) schizophrenia, and (10) other mental and substance use disorders. There were no date, location, or status restrictions.

Results: Our search identified 135 studies. A total of 28.9% (39/135) of studies evaluated interventions for major depressive disorder, 14.1% (19/135) of studies evaluated interventions for alcohol use disorders, 12.6% (17/135) of studies evaluated interventions for stroke, 11.1% (15/135) of studies evaluated interventions for schizophrenia, 8.1% (11/135) of studies evaluated interventions for anxiety disorders, 8.1% (11/135) of studies evaluated interventions for other mental and substance use disorders, 7.4% (10/135) of studies evaluated interventions for opioid use disorders, 3.7% (5/135) of studies evaluated interventions for Alzheimer disease or other dementias, 3.0% (4/135) of studies evaluated interventions for epilepsy, and 3.0% (4/135) of studies evaluated interventions for migraine. The studies were first registered in 2008; more than half of the studies were registered from 2016 to 2018. A total of 18.5% (25/135) of trials had results reported in some publicly accessible location. Across all the studies, the mean estimated enrollment (reported by the study) was 1078, although the median was only 100. In addition, across all the studies, the actual reported enrollment was lower, with a mean of 249 and a median of 80. Only about a quarter of the studies (35/135, 25.9%) were funded by the National Institutes of Health.

Conclusions: Despite the increasing use of health-based technologies, this analysis of ClinicalTrials.gov suggests that only a few apps for high-burden neuropsychiatric conditions are being clinically evaluated in trials.

KEYWORDS

smartphones; mobile phones; apps; mental health; regulation; stroke; migraine; major depressive disorder; Alzheimer disease; anxiety disorders; alcohol use disorders; opioid use disorders; epilepsy; schizophrenia

Introduction

The field of mobile health (mHealth), broadly defined as health care interventions that are delivered by mobile devices such as smartphones and tablets, is growing rapidly. Currently, there are over 325,000 mHealth apps [1], and the field of mHealth continues to attract new market entrants (28% of digital health practitioners have less than 2 years of industry experience) [1]. In addition, the global mHealth app market size is expected to hit 236 billion by 2026 [2]. mHealth apps claim to have varied purposes, from improving treatment adherence to increasing physical activity, to supplementing in-person counseling, and much more. The development of mHealth technologies is currently progressing at a much faster pace than that of the science to evaluate their validity and efficacy [3]. Thus, it is possible that ineffective or even harmful mHealth technologies might enter clinical practice without adequate evaluation [4]. For example, many apps make scientific claims on the app stores, but less than 2% of the apps can offer clinical evidence to back such claims [5]. The topic of digital health regulation by the Food and Drug Administration in the United States has been a news story since the fall of 2019 when it was discussed by a presidential candidate.

Currently, one path to increase scientific validity and transparency in the digital health space is through the mechanisms in place around publication. Since 2005, a condition of consideration for publication in the International Committee of Journal Medical Editors (ICJME) has been prospective registration of clinical trials. The ICJME defines a clinical trial as "any research project that prospectively assigns people or a group of people to an intervention, with or without concurrent comparison or control groups, to study the relationship between a health-related intervention and a health outcome." Trial registration is supposed to occur "in a public trials registry at or before the time of first patient enrollment." Registration requirements are intended to prevent well-documented problems that arise when the results of the trials are either unreported or are selectively reported [6]. A majority of the clinical trials are registered at ClinicalTrials.gov, a division of the US National Library of Medicine [7]. It is also possible to post trial results on ClinicalTrials.gov, although such posting is only required for certain studies. When reporting requirements apply, the results must generally be posted no later than 1 year after final data collection for the primary outcome [8].

Previous examinations of ClinicalTrials.gov have provided details on the types of studies being conducted in various fields and the various aspects of such studies, for example, location, number of study participants, interventions, and funders [9,10]. As is the case with clinical trials, in general, it is likely that many trials of mHealth interventions are not compliant with the requirements for trial registration. Selective reporting of results is of special concern in mHealth studies as they often collect vast amounts of data from surveys and sensors—often millions of data points from smartphone sensors and wearables. This creates many opportunities for both intentional and inadvertent systematic errors or biases in ongoing data collection, analysis, and interpretation. Prospective trial registration is an important protection against such mistakes, as well as a safeguard against intentional abuse.

However, the characteristics, number, and quality of registered trials of mHealth interventions for most therapeutic areas, including neuropsychiatric disorders, are unknown. According to the 2016 World Health Organization (WHO) Global Burden of Disease Study [11], the top 10 most disabling noninfectious neuropsychiatric conditions are, in the following order of ranking, (1) stroke, (2) migraine, (3) major depressive disorder, (4) Alzheimer disease and other dementias, (5) anxiety disorders, (6) opioid use disorders, (7) alcohol use disorders, (8) epilepsy, (9) schizophrenia, and (10) other mental and substance use disorders. For these conditions, we sought to evaluate the (1) number of mHealth trials of smartphone interventions registered on ClinicalTrials.gov; (2) study characteristics, for example, location of the study, type of interventions studied, number of study participants, length of study, and funder (National Institutes of Health [NIH] or other), among other characteristics; (3) frequency and timing of any outcome changes in the trial registry; and (4) the proportion of such studies that had reported results on ClinicalTrials.gov or in journal publications. Neuropsychiatric disorders comprise a broad range of medical conditions involving neurologic or psychiatric disturbances, including mental and behavioral disorders. Globally, they are the third leading cause of disability-adjusted life years (DALYs) [12,13]. Neuropsychiatric conditions are likely to be particularly amenable to mHealth interventions as their treatment may involve behavioral therapies and efforts to improve medication adherence, and there is a great need for scalable and accessible interventions [14-16].

Methods

According to the 2016 WHO Global Burden of Disease Study, [11] the following comprised the top 10 most disabling neuropsychiatric conditions worldwide: (1) stroke, (2) migraine, (3) major depressive disorder, (4) Alzheimer disease and other dementias, (5) anxiety disorders, (6) alcohol use disorders, (7) opioid use disorders, (8) epilepsy, (9) schizophrenia, and (10) other mental and substance use disorders (Multimedia Appendix 1 [17-149]). These conditions were searched for in ClinicalTrials.gov, along with smartphone-based keyword terms, and data were abstracted accordingly.

Data were summarized descriptively. Of note, we examined the reporting of results using two different methods. In the first method, for studies registered 3 or more years prior, we examined how many of these studies reported results either via one or both of the following: results reported on ClinicalTrials.gov or results automatically indexed to ClinicalTrials.gov. In the second method, we examined the studies marked completed on ClinicalTrials.gov by October

2017 (1 year before the date of data abstraction) and then examined which of those studies reported results either on ClinicalTrials.gov and/or had results automatically indexed to ClinicalTrials.gov.

We also analyzed whether there were any associations among (1) the number of study participants and study completion status, (2) the number of study participants and study results reporting status (3), the length of study intervention and completion status, and (4) the length of study intervention and results reporting status. We utilized *t* tests and regression models to analyze the results. A statistical analysis was conducted in the R programming environment (version R 3.6.1; The R Foundation).

Per the self-documentation form from the New York University (NYU) School of Medicine (SOM), the research did not involve human subjects. Thus, consistent with the NYU SOM institutional review board (IRB) policy and federal regulations governing human subject research, an IRB review was not required.

Results

A total of 135 studies on ClinicalTrials.gov met the search criteria. As shown in Multimedia Appendix 2 [17-64,66-82,84-114,116-133,135-141,143-153], the number of studies registered for each neuropsychiatric condition that used a smartphone for an intervention was as follows: major depressive disorder (n=39), alcohol use disorders (n=19), stroke (n=17), schizophrenia (n=15), anxiety disorders (n=11), other mental and substance use disorders (n=11), opioid use disorders (n=10), Alzheimer disease and other dementias (n=5), epilepsy (n=4), and migraine (n=4). Multimedia Appendix 1 shows the key findings for the various apps, including the myriad of purposes they served. They ranged from promoting rehabilitation and diet to medication adherence, symptom tracking, cognitive behavioral therapy, and more.

A breakdown of the studies is provided in Table 1. The altered outcomes have been described in Table 2.

Table 1. Breakdown of study status, results reporting, location of studies and National Institutes of Health funding.

Characteristics[a,b]	All conditions	Stroke	Migraine	Major depressive disorders	Alzheimer disease and other dementias	Anxiety disorders	Alcohol use disorders	Opioid use disorders	Epilepsy	Schizophrenia	Other mental and substance use disorders
Number of apps, n	135	17	4	39	5	11	19	10	4	15	11
Status of study, n											
Not yet recruiting	19	0	2	5	3	0	1	4	0	2	2
Recruiting	40	8	2	9	1	2	7	3	2	3	3
Enrolling by invitation	7	1	0	0	1	2	1	0	1	0	1
Active, not recruiting	7	1	0	4	0	0	1	0	0	1	0
Completed	59	7	0	21	0	8	7	3	1	9	4
Unknown	3	0	0	0	0	0	2	0	0	0	1
Number of participants estimated											
Mean (SD)	1078.64 (9088.35)	222.06 (470.64)	131 (52.57)	199.74 (347.83)	196 (175.23)	629 (1640.23)	439.59 (866.02)	144.8 (205.22)	25066.25 (49955.84)	102 (92.31)	234.10 (241.73)
Lowest; highest, n	8; 100,000	8; 2000	90; 200	15; 2000	52,448	8; 5000	30; 3600	9; 600	65; 100,000	36; 260	40; 800
Median (IQR)	100 (150)	80 (160)	117 (68)	103 (123)	142 (147)	70 (110)	105 (210)	64.5 (105.5)	100 (24983.75)	42.5 (113)	100 (303.5)
Number of participants actual											
Mean (SD)	242.19 (586.02)	301.55 (657.80)	N/A[c]	208.38 (426.21)	20 (N/A)	51.4 (41.52)	683.71 (1356.83)	67 (79.20)	95 (N/A)	92.3 (75.69)	322.33 (123.44)
Lowest; highest, n	10; 3702	20; 2274	N/A	11; 2010	20; 20	10; 105	15; 3702	11; 123	95; 95	27; 255	180; 400
Median (IQR)	80 (137.25)	100 (145)	N/A	84.5 (105)	20 (0)	40 (66)	58 (395.5)	67 (56)	95 (0)	59.5 (91.25)	387 (110)
Number of days of intervention											
Mean (SD)	149.26 (199.76)	155.88 (137.44)	173.75 (138.29)	104.44 (92.29)	202.2 (241.06)	198.91 (540.84)	125.90 (114.42)	158.6 (227.36)	112.25 (47.282)	189.93 (166.73)	205.18 (145.18)
Lowest; highest, n	0.0083; 1825	1; 456	56; 365	1; 395	28; 548	0.00; 1825	7; 365	7; 730	84; 183	30; 548	30; 548
Median (IQR)	84 (127)	91 (187)	137 (146.25)	84 (66)	42 (337)	14 (76)	84 (78.5)	56 (121.5)	91 (24.75)	168 (95.5)	183 (80.5)
Regional distribution, n											
Africa	1	1	0	0	0	0	0	0	0	0	0
Asia and Pacific	8	3	0	0	1	0	3	0	0	1	0
Central+South America	5	0	0	4	0	1	0	0	0	0	0
Europe	28	6	0	7	1	2	5	0	1	4	2
Middle East	2	0	0	0	0	1	0	0	0	1	0
North America	96	7	4	30	3	7	11	10	4	11	9

XSL•FO
RenderX

Characteristics[a,b]	All conditions	Stroke	Migraine	Major depressive disorders	Alzheimer disease and other dementias	Anxiety disorders	Alcohol use disorders	Opioid use disorders	Epilepsy	Schizophrenia	Other mental and substance use disorders
Study marked as National Institutes of Health funded, n	35	3	0	14	1	3	3	5	0	3	3

[a]Publication automatically indexed to the study by ClinicalTrials.gov identifier (NCT number) containing information pertaining to study without results.

[b]Publication automatically indexed to the study by ClinicalTrials.gov identifier (NCT number) containing information pertaining to study with results.

[c]N/A: not applicable.

Table 2. Altered outcomes.

Disorder and study title[a]	Altered outcomes
Stroke	
Effects of Patient-centered Stroke Educating System: A Randomized Controlled Trial [152]	• Original Primary Outcome Measures (submitted: October 28, 2015): Stroke Health Education Knowledge (time frame: 2 weeks) Changed August 16, 2018: Stroke Health Education Knowledge Questionnaire (time frame: 4 weeks)
Empowerment and Mobile Technology in the Control of Cardiovascular Risk Factors in Patients with Ischemic Stroke (CARDIOSTROKE) [24]	• Original Primary Outcome Measures (submitted: October 17, 2018) was Atrial Fibrillation (3 weeks) as detected in a 3-week ECG[b] monitoring period and Systolic/diastolic blood pressure (time frame: 12 months) measured as difference in systolic/diastolic blood pressure. • Current Primary Outcome Measures (submitted: October 18, 2018) are Number of Participants with New Atrial Fibrillation and Change in Blood Pressure. • Current Secondary Outcome Measures (submitted: October 18, 2018)-Number of participants was added before new cardiovascular events within 12 months and new cardiovascular events within 36 months.
Migraine	
RELAXaHEAD for Headache Patients [33]	• Changed December 29, 2017: Proportion of patients who enrolled in the study/were recruited for the study was eliminated as a primary outcome measure. Satisfaction using Likert scale questions on RELAXaHEAD usability, content, functionality was added as a primary outcome measure.
Major depressive disorder	
mHealth for Antenatal Mental Health [112]	• Submitted September 9, 2016: Adherence to sampling protocol (time frame: assessed after 6 months) was added as a primary outcome measure.
Text-Message-Based Depression for High-Risk Youth in the ED [41]	• Submitted: January 5, 2015: Δ in Depressive Symptoms was the primary outcome measure. Submitted November 7, 2017- Δ in Peer Violence Involvement was added as a primary outcome measures. • Submitted January 5, 2015-original secondary outcome measures was Δ in Peer Violence Involvement. • Submitted January 5, 2015: Acceptability/Feasibility: Follow Up Rate, Acceptability/Feasibility: Engagement of Intervention Group and Acceptability/Feasibility: Participant Questionnaire were the secondary outcome measures.
Behavioural Activation-Based Treatment Administered Through Smartphone [114]	• Original Primary Outcome Measures (submitted: October 31, 2011) was the Montgomery Asberg Depression Rating Scale-Self Rated (MADRS). Current Primary Outcome Measures (submitted: March 22, 2013) is the PHQ-9[c] and Beck Depression Inventory (BDI). • Original Secondary Outcome Measures (submitted: March 22, 2013) was the QOLI[d], AAQ[e], BAI[f] and TIC-P[g]. Current Secondary Outcome Measures (submitted March 22, 2013) eliminated the AAQ and BDI[h] as secondary outcome measures.
Mobile Technology to Engage and Link Patients and Providers in Antidepressant Treatment (MedLink) [116]	• Original Primary Outcome Measures (submitted: October 20, 2015) was Adherence to Antidepressant Medication measured as the number of days medication was taken when a dose was expected. Current Primary Outcome Measures (submitted: February 14, 2018) is Adherence to Antidepressant Medication measured through % of days adherent on Wisepill pillbox as well as self-reported adherence. • Original Secondary Outcome Measures (submitted: October 20, 2015) was changes in depression measured through self-report PHQ-9 and usability measured through Likert scale ratings. Current secondary outcome measures (submitted: February 14, 2018) is the PHQ-9 and Quick Inventory of Depressive Symptomology Clinician Rating (QIDS-C).

Disorder and study title[a]	Altered outcomes
Mobile Technology to Engage and Link Patients and Providers in Antidepressant Treatment [117]	• Original Primary Outcome Measures (submitted: July 24, 2013) was Adherence to Medication as measured by when the provided pill bottle is opened to remove a dose of medication and Δ in Depression Over Time through the PHQ-9. Current Primary Outcome Measures (submitted: January 12, 2015) is measured as Adherence to Antidepressant Medication measured as the frequency of medication usage from baseline to end of treatment. • Original Secondary Outcome Measures (submitted: July 24, 2013) was Presence of Side Effects and Δ s Over Time measured by the Patient Rated Inventory of Side Effects (PRISE) and Frequency, Intensity and Burden of Side Effects Rating (FIBSER). Current Secondary Outcome Measures (submitted: January, 12 2015) is changes in depression measured as the severity of depressive symptoms from baseline to end of treatment.
Lifestyle Intervention for Young Adults with Serious Mental Illness [57]	• Submitted: June 27, 2016 • Δ in serum lipids was added as a current secondary outcome measures.
Treating Depression on a Day-to-day Basis: Development of a Tool for Physicians Based on a Smartphone Application [122]	• Original Primary Outcome Measures (submitted: September 18, 2018) was comparison between the two groups of the number of participants with a decrease in HDRSS[i] scores of at least 50% at 8 weeks. Current Primary Outcome Measures (submitted: September 20, 2018) is a greater clinical response in the active group (smartphone application) comparatively to the comparator group (clinical response was defined as a decline in HDRS-17[j] score greater than 50%).
Using Mental Health Telemetry to Predict Relapse and Rehospitalization in Mood Disorders (PATH-MOD) [54]	• Submitted: November 21, 2014 • Quality-of-Life in Bipolar Disorder was added as a Primary Outcome Measures.
IntelliCare: Artificial Intelligence in a Mobile Intervention for Depression and Anxiety (AIM) [125]	• Original primary outcome measures (submitted: June 25, 2014) were changes in depression severity, adherence to mobile application intervention and changes in anxiety. Current Primary Outcome Measures (submitted: February 14, 2018) are the PHQ-9 and GAD-7[k]. Original Secondary Outcome Measures was participant satisfaction. Current Secondary Outcome Measures is the Mean Number of Treatment App Use Sessions by Study Week.
Enhancing Delivery of Problem Solving Therapy Using SmartPhone Technology [128]	• Original Primary Outcome Measures (submitted: June 28, 2013) was Depression, Anxiety and Stress, an instrument that measures clinical indices of depression and anxiety as well as acute stress. Current Primary Outcome Measures (submitted: July 27, 2016) is the Depression Anxiety and Stress Scale (DASS).
Technology Assisted Programs that Promote Mental Health for Teenagers (ProjectTECH) [48]	• Original Primary Outcome Measures (submitted: July 30, 2013) were Depression (Time Frame: Throughout participation, lasting up to 26 weeks) as measured by the CES-D[l] and MINIKid and Usability of the Program as measured by the Usefulness, Satisfaction and Ease of Use Questionnaire survey
Study of Technology-assisted Treatment of Adolescent Depression (iTAD) [49]	• Original Secondary Outcome Measures (submitted: April 19, 2012) was preliminary indicators of program efficacy (Time Frame: 12 weeks) measured by the 1. Depression Knowledge Test 2. Skill Self Efficacy Questionnaire 3. The Therapeutic Alliance Scale for Adolescents 4. The Cognitive Therapy Scale 5. Acceptance Questionnaire
Mobile Phone Sensing and Outreach as Adjuncts to Internet Based Behavior Intervention for Depression [129]	• Original Primary Outcome Measures (submitted: April 19, 2010) was Depression, as assessed by Quick Inventory of Depressive Symptoms, PHQ-9 and the Mini International Neuropsychiatric Interview Major Depressive Disorders Module (Time Frame: Measured as baseline, 4 and 8 weeks) • Original Secondary Outcome Measures (submitted: April 19, 2010) was Utilization-Adherence markers for the mobile phone (eg-number of responses to prompts for information) the website (Time frame: Measured from baseline to 8 weeks), Positive Affect, Anxiety (GAD-7) and Health-Related Quality of Life.
Effectiveness of a Technology Assisted Behavioral Intervention in Assisting People with Major Depressive Disorder [50]	

XSL•FO

RenderX

Disorder and study title[a]	Altered outcomes
	• Original Primary Outcome Measures (submitted: July 18, 2008) was Depression, as assessed by HRSD[m], PHQ-9 and SCID and Attrition (in arms with i-CBT[n]) • Original Secondary Outcome Measures was i-CBT utilization (eg, number of logins, average visit length, total time spent on the site, number of exercises completed), Health-Related Quality of Life (SF-36V[o]) and Patient Satisfaction (Satisfaction Index-Mental Health)
Online Peer Networked Collaborative Learning for Managing Depressive Symptoms (MoodTech) [51]	• Original Primary Outcome Measures (submitted: July 20, 2016) was Depression (Time Frame: Baseline to end of treatment) measured as the Δ in self-reported depressive symptom severity from baseline to end of treatment and adherence to the program measured as the number of times the program is accessed from start to last use from baseline to end of treatment
Anxiety	
Connection to Care: Pilot Study of a Mobile Health Tool for Patients with Depression and Anxiety [131]	• Original Primary Outcome Measures (submitted: July 13, 2015) were patient acceptability as determined by qualitative interview, care manager acceptability as determined by qualitative interview, patient report of usefulness as determined by qualitative interview and care manager report of usefulness as determined by qualitative interview. Current primary outcome measures (submitted: October 25, 2017) are app acceptability as measured by number of patient app users who rate app easy to use and time spent reasonable, app acceptability as measured by number of care manager dashboard users who rate dashboard easy to use and time spent reasonable, app usefulness as measured by number of patient app users who rate app easy to use and time spent reasonable and app usefulness as measured by number of care manager dashboard users who rate dashboard as useful.
Improving Medical Care With Electronic Interventions Based on Automated Text and Phone Messages [132]	• Current Secondary Outcome Measures (submitted: February 10, 2017)-Breast-feeding duration was added as a current secondary outcome measure.
Effect of Premedication Type on Preoperative Anxiety in Children [77]	• Current Primary Outcome Measures (submitted: August 13, 2018)-Mask acceptance (At anesthesia induction) was eliminated as a primary outcome measure and added as a secondary outcome measure.
Using Smartphones to Enhance the Treatment of Childhood Anxiety [80]	• Original Primary Outcome Measures (submitted: October 3, 2014) was the Clinical Global Impression Improvement (CGI-I) defined as treatment response at post treatment. Current Primary Outcome Measures (submitted: January 5, 2016) is the PARS[p] Treatment response. Original Secondary Outcome Measures was the PARS and Screen for Childhood Anxiety Related Emotional Disorders (SCARED). Current Secondary Outcome Measures is the Absence of diagnosis on K-SADS[q]
Youth Mayo Clinic Anxiety Coach Pilot Study [81]	• Current Primary Outcome Measures (submitted: July 30, 2014)-qualitative interview assessing subject safety and treatment adherence (Time frame: within 5 working days of treatment completion) was eliminated as a primary outcome measure
ACT-smart: Smartphone-supplemented iCBT for Social Phobia and/or Panic Disorder [82]	• Current Primary Outcome Measures and Current Secondary Outcome Measures (submitted: November 27, 2013) • All outcomes that listed month 24 after the treatment period in the time frame were changed to month 36 after the treatment period.
Alcohol use disorder	

XSL•FO
RenderX

Disorder and study title[a]	Altered outcomes
mWELLCARE: An Integrated mHealth System for the Prevention and Care of Chronic Disease (mWELLCARE [134]	• Changed September 12, 2016: no longer tracking 10-year risk of CHD[r] and added tracking for alcohol use, fasting blood sugar, total cholesterol, CVD[s] risk, and cost
Project Guard: Reducing Alcohol Misuse/Abuse in the National Guard [88]	• Updated August 8, 2016: No changes
Skills-Training for Reducing Risky Alcohol Use in App Form [137]	• Updated October 23, 2018: no longer looking for reduction in alcohol consumption
Usefulness of Supportive Text Messages in the Treatment of Depressed Alcoholics [96]	• Updated December 17, 2011: Becks Depression inventory Score was added, and global assessment of function score was added
Epilepsy	
Embrace: Seizure Characterization [39]	• Updated June 22, 2018: Original measures not given
Schizophrenia	
Development of a Mobile System for Self-Management of Schizophrenia (SOS) [60]	• Updated June 2, 2016: No longer tracking medication adherence.
A New Paradigm for Illness Monitoring and Relapse Prevention in Schizophrenia [63]	• Updated June 2, 2016: Now using BRPS to assess psychotic symptom severity instead of PANSS[t]
Other mental and substance use disorders	
Preventing HIV/STI in Urban Adolescents via an mHealth Primary Care Intervention [148]	• Updated February 12, 2018: No longer tracking Δ in adolescent STI[u] testing.
CopeSmart: Using Mobile Technology to Promote Positive Mental Health In Young People [149]	• Updated October 10, 2014: Changed to use Emotional Self-Awareness Scale.

[a]Publication automatically indexed to the study by ClinicalTrials.gov identifier (NCT number) without results.

[b]ECG: electrocardiogram.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]QOLI: Quality of Life Inventory.

[e]AAQ: Swiss Agency of Accreditation and quality assurance.

[f]BAI: Beck Anxiety Inventory.

[g]TIC-P: Trimbos and iMTA questionnaire on Costs associated with Psychiatric illness.

[h]BDI: Beck Depression Inventory.

[i]HDRSS: HDRS Hamilton Depressive Rating Scale.

[j]HDRS-17: HDRS Hamilton Depressive Rating Scale.

[k]GAD-7: General Anxiety Disorder-7.

[l]CES-D: Center for Epidemiological Studies-Depression.

[m]HRSD: Hamilton Rating Scale for Depression.

[n]i-CBT: internet-based Cognitive Behavioral Therapy.

[o]SF-36V: Satisfaction Index-Mental Health.

[p]PARS: Pediatric Anxiety Rating Scale.

[q]K-SADS: DSM 5 Diagnosis of Separation Anxiety, Social Anxiety, and Generalized Anxiety Disorder on the K-SADS interview.

[r]CHD: coronary heart disease.

[s]CVD: cardiovascular disease.

[t]PANSS: Positive and Negative Syndrome Scale.

[u]STI: sexually transmitted infection.

The second table shows that the studies were first registered in 2008, with more than half of the studies registered between 2016 and 2018. Across all 135 studies, the mean estimated enrollment was 1078, although the median was only 100. Across all 135 studies, the actual reported enrollment was lower, with a mean of 249 and median of 80. Only about a quarter (35/135, 25.9%) of the 135 studies were NIH funded.

Of the 135 studies included, only 9 (6.7%) studies reported their results. A total of 16.9% (23/135) of studies had publications automatically indexed to the ClinicalTrials.gov identifier (Multimedia Appendix 1). Moreover, 18.5% (25/135) of trials had results reported in some publicly accessible location (either the results section of ClinicalTrials.gov or in publications

XSL•FO

RenderX

indexed in databases, which were automatically associated with the ClinicalTrials.gov identifier; Multimedia Appendix 1).

Of those studies that were registered more than 3 years ago (a time frame the authors deemed reasonable given (1) the generally short study interventions, (2) the rapidly changing mHealth landscape, and (3) the 1-year time frame that ClinicalTrials.gov gives to post results after study completion), 14.5% (8/55) of the studies had results posted on ClinicalTrials.gov. The conditions and the number of studies for which the results were published are stroke (1/11), major depressive disorder (4/15), anxiety disorders (2/7), and schizophrenia (1/6). Of the 135 studies, 33 (24.4%) studies that altered their outcomes after the original outcome measures were posted (Multimedia Appendix 2). Of the 135 studies, there were 45 (33.3%) studies that were marked completed by October 2017, 1 year before the search date. Of the 45 studies, 20 (44%) had results reported in some publicly accessible location (either the results section of ClinicalTrials.gov or in publications indexed in databases, which were automatically associated with the ClinicalTrials.gov identifier). Moreover, 15% (7/45) of studies reported their results on ClinicalTrials.gov, and 30% (14/45) of studies had publications automatically indexed to the ClinicalTrials.gov identifier.

There was a statistically significant relationship, as determined by a two-sided t test, with studies reporting results having a shorter mean duration (121.7 days) compared with studies never reporting results (153.8 days; $P<.001$). There was also a statistically significant relationship with studies reporting results having lower actual enrollment (142.3 people) compared with studies never reporting results (295.6 people; $P=.01$). There was no statistically significant relationship between the estimated enrollment at the time of study registration and never reporting results. There was also no statistically significant relationship between the status of a study and its estimated or actual enrollment or length of intervention.

Discussion

Principal Findings

In this comprehensive analysis of registered mHealth studies of interventions for disabling neuropsychiatric conditions, there were 6 key findings. First, there has been a large increase in the number of clinical trial registrations in the past 2 years; almost half (44.5%) of the trial registrations were registered in the past 2 years. Thus, despite the increasing additions of health-based smartphone apps [1], this snapshot of ClinicalTrials.gov suggests that only a few such apps for high-burden neuropsychiatric conditions are being clinically evaluated in trials. Second, the studies were generally located in the United States, but a few of the studies were funded by the NIH. Third, the study characteristics were such that they would not generally be considered as high-quality evidence and for use in guideline recommendations because of small sample sizes and heterogeneous interventions. Fourth, the mean study duration has not changed with time, suggesting that long-term outcomes are still not the focus of research. Despite the myriad of ways in which results can be reported, a few trials had results reported either as entered on ClinicalTrials.gov or as study results

automatically indexed to ClinicalTrials.gov. As stated earlier, only 6.7% (9/135) of such trials that have been registered on ClinicalTrials.gov have reported results. Overall, 18.5% (25/135) of trials had results reported in some capacity (either under the results section of ClinicalTrials.gov or through publications indexed in databases, which were automatically associated with the ClinicalTrials.gov identifier). Fifth, initially specified outcomes were changed after trials commenced in a quarter of all trials. Finally, study duration and sample size of those enrolled are associated with the reporting results of studies.

More trials are being registered on ClinicalTrials.gov, given the requirements for prospective registration as a condition of publication. A previous study examined clinical trial registration for 3 groups of disorders (ie, cardiovascular disorders, mental health disorder, and oncologic disorders) that comprise the largest number of DALYs lost in the United States. The authors found that the number of trials submitted for registration between October 2004 and September 2007 and then October 2007 and September 2010 increased by about 140% from 28,881 to 40,970 [9]. A study assessing trial registration revealed that a few trials still reported on the trial registries [154]. However, more researchers are learning about this requirement, hence the expected increase in trial registrations.

Despite the increase in trial registrations, only 135 trials met the criteria for this study. Thus, a few apps for high-burden neuropsychiatric conditions are being clinically evaluated in the trials reported on ClinicalTrials.gov. A recent study of top-funded digital health companies examined the number of research studies collectively undertaken and found that of the top 10 disease categories examined, 7 were neuropsychiatric. However, none of these industry studies reported on the clinical effectiveness of the digital health tools for these high-burden conditions. Thus, although there is high interest in digital health toward neuropsychiatric conditions, there is little registered evidence that such apps on the market work [155].

A majority of the studies were based in the United States, followed by Europe. This is not because ClinicalTrials.gov is based in the United States, as ClinicalTrials.gov accounts for more than 80% of all the clinical studies in the WHO portal [9]. There may be fewer studies in Europe because of the newly implemented General Data Protection Regulation in 2018, which is a European Union (EU) law for the protection of data and privacy for all individuals within the EU and the European Economic Area. It also concerns the export of personal data outside the EU.

A few (26%) of the trials were NIH funded. This is not surprising as much of the research in digital health has occurred in the private sector. In 2016, 296 private digital health companies received venture funding that totaled to more than US $4.2 billion and approached US $6 billion in 2017 [155]. The 20 top-funded, private US-based digital health companies were studied to analyze their products and services, peer-reviewed evidence, and the potential for impact on patients with high-burden conditions. Less than one-third (27.9%) of the studies targeted patients with a high-burden condition. Only 16 (15%) studies assessed the clinical effectiveness of the product or service, and only 8 (8%) studies assessed the clinical

effectiveness of the product or service in a high-burden or high-risk factor population. Only a small number of studies published data, and interestingly, journals without impact factors were the most common (31%) source of publications [155].

There was tremendous heterogeneity in the purpose of the use of the apps not only across conditions but also within a given condition. The most common purpose across all conditions was symptom tracking and medication adherence. There were also some studies based on skills learning, for example, cognitive behavioral therapy or progressive muscle relaxation, and an app was used to help with the delivery of these skills. A small number of studies were designed to detect physical symptoms or signs, such as atrial fibrillation and seizure, or enhance communication with clinicians using apps connected to sensors and/or devices. This latter intervention was to provide support to patients, in some cases with the intent to prevent relapse or indicate the degree of symptoms to potentially lead to a change in medication management if warranted.

In terms of study design, the studies were generally small (samples sizes<100). This study's results were comparable with the study examining the ClinicalTrials.gov registration of interventional studies for cardiovascular disease, mental health, and oncology, which found that 62% of the trials had an anticipated enrollment mean of less than 100 [9]. This study's findings were also comparable with a study of trials conducted by the top 20 funded digital health companies, which found that 51% of the studies had less than 100 participants [155]. Thus, these studies will not be able to help in creating major guidelines with high-quality evidence related to these disabling neuropsychiatric conditions.

Although this study was not designed to assess why there is low registration of trials, the potential reasons may include lack of awareness and different prioritization in the app development industry. "Given the diversity of stakeholders involved in mHealth research, competing outcomes, priorities, funding, and publication requirements may potentially mean some studies are less likely to be registered" [156].

This study's results are in line with previous work showing that few studies reported published results on ClinicalTrials.gov. In a study examining trial registration compliance in publications related to headache, only 26% of all the studies that should have been registered were indeed registered [154]. A recent study of 556 trial registrations on ClinicalTrials.gov showed that out of all the trials in the study, 150 (27%) trials remained unpublished 5 years after the study completion dates [157]. There are a number of potential reasons for the low publication rates. First, the lack of reporting of negative results is a well-known phenomenon in academic medicine [6,158,159]. For example, some trials of triptans [160] and gabapentin [154,161] were never published. Second, in mHealth, there are likely specific challenges to publication, including high attrition rates, usability issues, and lack of sufficient previous formative research [162]. Third, previous research has shown that clinical trials with large sample sizes were more likely to be published [154,158,162,163]. Many of the mHealth clinical trial registrations in this study's sample sizes were considerably small.

As noted previously, the study outcomes were changed. In a systematic review in 2011 assessing the transparency of outcome reporting and trial registration of randomized controlled trials and top psychosomatic and behavioral health journals, of the 63 articles meeting the study criteria, only 25 (39.7%) articles had adequately declared primary or secondary outcomes [164]. Thus, this study's results are in line with previous research in the field of mental health. This is an especially prevalent concern in digital health, where researchers could easily change their outcomes and conduct selective analyses [165].

Finally, this study's findings suggest that study status is not associated with either estimated or actual sample size or length of the intervention. Study status may not be updated in real time; thus, it may not reflect the true status of the study. This study's findings that studies never reporting any results have a longer duration than those reporting results are logical in the sense that shorter studies may be easier to complete. Similarly, this study's findings that studies never reporting any results have a higher mean number of participants compared with studies that did report results make sense, as larger studies are more likely to be difficult to complete. In sum, these findings that shorter and smaller studies are associated with reporting results compared with longer and larger studies are intuitive and reflect that digital health studies have the same challenges that all clinical studies face in terms of reaching reporting status.

Limitations

One of the limitations of this study is that we may have failed to identify some mHealth studies due to the sampling methodology. The primary search terms that we used were "mHealth," "smartphone," "electronic diary," and "mobile technology," while specifying subcategories of the different neuropsychiatric conditions (ie, migraine, migraine with aura, migraine without aura, and migraine disorders). Other search terms such as "digital," "ecological momentary assessment," "experience sampling," or "log" may have captured additional studies. We only searched ClinicalTrials.gov, and other trial registration websites exist. Second, many of the studies are recent, and thus authors may not have completed their studies and posted results yet. Third, the potential usefulness of different behavioral interventions administered through smartphone apps is not listed on ClinicalTrials.gov and is likely unknown; it is generally the point of the studies to assess whether these interventions may be effective. Finally, the information on ClinicalTrials.gov will always be incomplete for two main reasons: (1) Individual studies may not be registered in the database. Second, information entered on ClinicalTrials.gov may be incomplete, for example, certain data elements may have had a different format or structure or may have been optional when the study information was initially entered. (2) There are few incentives to motivate responsible parties to update their studies [150] registered on ClinicalTrials.gov. One study found that 17% to 20% of the studies on ClinicalTrials.gov were observational and only 7% had posted results [150,151]. Inferential statistics were limited based on reports from individual studies.

Future Directions

The study information on ClinicalTrials.gov is helpful for understanding the landscape of smartphone-based studies for neuropsychiatric conditions. However, the information listed does not offer enough detail to fully understand the nature of smartphone apps and sensor data collection. Future solutions may include the posting of web-based demonstrations of apps being studied or links to the version of the app used in trials. In addition, it is clear that many of the health-based apps are being developed by the private sector. Efforts need to be made to encourage commercial companies to register their studies on ClinicalTrials.gov and to adhere to the trial registration guidelines, for example, report trial results in a timely manner, as described by ClinicalTrials.gov. For health-based apps that make claims of efficacy, there needs to be stringent oversight of the registered clinical trials.

Conclusions

Despite the increasing use of health-based smartphone apps by the general public, only a few such apps are rigorously evaluated in clinical settings. Similar to other research on the studies registered on ClinicalTrials.gov, studies of the top neuropsychiatric conditions involving mHealth, registered on ClinicalTrials.gov, tended to be small, and there was a large amount of heterogeneity in the methods (types of interventions), duration, and reporting methods [9]. Moreover, very few registered studies (6.7%) reported their results, raising the question of whether the burgeoning creation of mHealth-based interventions is efficacious, despite these apps being widely downloaded and used. There were few studies for the most disabling neuropsychiatric conditions that typically use electronic diaries for the self-management of migraine and epilepsy. Future work should focus on studying the efficacy of these mHealth interventions for neuropsychiatric conditions if they are to be used by patients with these disabling conditions. Such studies should be registered on ClinicalTrials.gov to ensure transparency and so that the public can also learn about the research being conducted using these interventions.

Authors' Contributions

Study conception was done by MM, EL, and JT. JR and PP conducted data analysis and performed descriptive analyses. Statistical analysis was conducted by JT, JR, and PP. Manuscript drafting was done by MM, JR, and PP, with all authors MM, JR, PP, EL, and JT revising it for intellectual content.

Conflicts of Interest

Both MM and JT have NIH-funded studies to conduct app-based research.

Multimedia Appendix 1
Study methodology, key findings, and data collected.
[DOC File , 837 KB - mhealth_v8i8e16180_app1.doc]

Multimedia Appendix 2
Mobile health studies, study criteria, and altered outcomes.
[DOC File , 237 KB - mhealth_v8i8e16180_app2.doc]

References

1. Pohl M. 325,000 Mobile Health Apps Available in 2017 – Android Now the Leading mHealth Platform. Research2Guidance. 2017. URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017/ [accessed 2020-06-17]
2. mHealth Apps Market Size Worth $236.0 Billion By 2026. Grand View Research. 2019. URL: https://www.grandviewresearch.com/press-release/global-mhealth-app-market [accessed 2020-06-17]
3. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
4. Schwartz JL. Real-world evidence, public participation, and the FDA. Hastings Cent Rep 2017 Nov;47(6):7-8. [doi: 10.1002/hast.779] [Medline: 29171057]
5. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: evaluation of mental health app store quality claims. NPJ Digit Med 2019;2:18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
6. Brassington I. The ethics of reporting all the results of clinical trials. Br Med Bull 2017 Jan 1;121(1):19-29. [doi: 10.1093/bmb/ldw058] [Medline: 28104631]
7. Registration of Clinical Trials on ClinicalTrials.gov. Johns Hopkins Medicine. 2019. URL: https://www.hopkinsmedicine.org/institutional_review_board/guidelines_policies/guidelines/clinical_trials.html [accessed 2020-06-17]

8. FDAAA 801 and The Final Rule. Clinical Trials. 2018. URL: https://clinicaltrials.gov/ct2/manage-recs/fdaaa#WhichTrialsMustHaveResults [accessed 2018-12-31]

9. Califf RM, Zarin DA, Kramer JM, Sherman RE, Aberle LH, Tasneem A. Characteristics of clinical trials registered in ClinicalTrials.gov, 2007-2010. J Am Med Assoc 2012 May 2;307(17):1838-1847. [doi: 10.1001/jama.2012.3424] [Medline: 22550198]

10. Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. Br Med J 2012 Jan 3;344:d7292 [FREE Full text] [doi: 10.1136/bmj.d7292] [Medline: 22214755]

11. GBD Results Tool. Global Health Data Exchange. 2019. URL: http://ghdx.healthdata.org/gbd-results-tool [accessed 2019-11-16]

12. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of diasease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. Lancet 2013 Nov 9;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]

13. Murray C, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the global burden of disease study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]

14. Minen M, Azarchi S, Sobolev R, Shallcross A, Halpern A, Berk T, et al. Factors related to migraine patients' decisions to initiate behavioral migraine treatment following a headache specialist's recommendation: a prospective observational study. Pain Med 2018 Nov 1;19(11):2274-2282 [FREE Full text] [doi: 10.1093/pm/pny028] [Medline: 29878178]

15. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]

16. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. Mhealth 2018;4:6 [FREE Full text] [doi: 10.21037/mhealth.2018.03.02] [Medline: 29682510]

17. Impact of M-health based Intervention on Adherence to Healthy Physical Activity After Stroke. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03507894?cond=Impact+of+M-health+based+Intervention+on+Adherence+to+Healthy+Physical+Activity+After+Stroke&rank=1 [accessed 2018-11-27]

18. Increasing Physical Activity in Stroke Survivors Using STARFISH. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT02494245?cond=Increasing+Physical+Activity+in+Stroke+Survivors+Using+STARFISH&rank=2 [accessed 2018-11-27]

19. The Stroke and Exercise Program. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT02701998?cond=The+Stroke+and+Exercise+Program&rank=1 [accessed 2018-11-27]

20. Stroke Inpatient Rehabilitation Reinforcement of ACTivity. Clinical Trials. 2010. URL: https://ClinicalTrials.gov/ct2/show/NCT01246882?cond=Stroke+Inpatient+Rehabilitation+Reinforcement+of+ACTivity&rank=1 [accessed 2018-11-27]

21. Survey of Lifelong Food and Nutrition Assistance (LIFANA) in Stroke Patients and Caregivers. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03635476?cond=LIFANA&rank=1 [accessed 2018-11-27]

22. mHealth Screening to Prevent Strokes (mSToPS). Clinical Trials. 2015. URL: https://ClinicalTrials.gov/ct2/show/NCT02506244?cond=mstops&rank=1 [accessed 2018-11-27]

23. VR-3D Movie-Based Education. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03104231?cond=vr+3d&rank=1 [accessed 2018-11-27]

24. Empowerment and Mobile Technology in the Control of Cardiovascular Risk Factors in Patients With Ischemic Stroke (CARDIOSTROKE). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03710902?term=NCT03710902&rank=1 [accessed 2018-11-27]

25. Hispanic Secondary Stroke Prevention Initiative (HISSPI). Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/NCT02251834?cond=hispanic+stroke&rank=1 [accessed 2018-11-27]

26. Phone-Based Intervention Under Nurse Guidance After Stroke (PINGS). Clinical Trials. 2015. URL: https://ClinicalTrials.gov/ct2/show/NCT02568137?cond=Phone-based+Intervention+Under+Nurse+Guidance+After+Stroke&rank=1 [accessed 2018-11-27]

27. Technology Intervention to Support Caregiving for Alzheimer's Disease (I-CARE) (I-CARE). Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03119259?cond=iCare-AD%3A+A+Mobile+Health+Application+for+Caregivers+of+Patients+With+Dementia&rank=1 [accessed 2018-11-27]

28. Diabetes as an Accelerator of Cognitive Impairment and Alzheimer's Disease (DIALCAT). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03578991 [accessed 2018-11-27]

29. Comparing Smartphone Technology and a Memory Strategy on Improving Prospective Memory in Alzheimer's Disease. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03384043?cond=Comparing+Smartphone+Technology+and+a+Memory+Strategy+on+Improving+Prospective+Memory+in+Alzheimer%27s+Disease&rank=1 [accessed 2018-11-27]

30. Therapeutic Efficacy of Categorical Language Fluency Smartphone Game Application. Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/show/

NCT02848404?cond=Therapeutic+Efficacy+of+Categorical+Language+Fluency+Smartphone+Game+Application&rank=1 [accessed 2018-11-27]

31.   Therapeutic Efficacy of Categorical Language Fluency Smartphone Game Application. Clinical Trials. 2016. URL: https:/ /ClinicalTrials.gov/ct2/show/ NCT02848404?cond=Therapeutic+Efficacy+of+Categorical+Language+Fluency+Smartphone+Game+Application&rank=1 [accessed 2018-11-27]

32.   Developing a Mobile Health Pain-Coping Skills Training Program for the Treatment of Chronic Migraine: AIM 4. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03465826 [accessed 2018-11-27]

33.   RELAXaHEAD for Headache Patients. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/ NCT03183791?cond=RELAXaHEAD+for+Headache+Patients&rank=1 [accessed 2018-11-27]

34.   Improving Health Outcomes of Migraine Patients Who Present to the Emergency Department. Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/show/ NCT02945839?cond=Improving+Health+Outcomes+of+Migraine+Patients+Who+Present+to+the+Emergency+Department&rank=1 [accessed 2018-11-27]

35.   A Study of the Effect of a Disease-Specific Migraine Smart Phone Application (App) on Participant Care. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03559088 [accessed 2018-11-27]

36.   Behavioral and Educational Tools to Improve Epilepsy Care. Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/ show/NCT02646631?cond=Behavioral+and+Educational+Tools+to+Improve+Epilepsy+Care&rank=1 [accessed 2018-11-27]

37.   Stress Management Intervention for Living With Epilepsy (SMILE). Clinical Trials. 2011. URL: https://ClinicalTrials.gov/ ct2/show/NCT01444183?cond=Stress+Management+Intervention+for+Living+with+Epilepsy+%28SMILE%29&rank=1 [accessed 2018-11-27]

38.   Generalized Seizure Detection And Alerting In The EMU With The Empatica Embrace Watch And Smartphone-Based Alert System. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03207685 [accessed 2018-11-27]

39.   Embrace: Seizure Characterization. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03206502?cond=. +Embrace%3A+Seizure+Characterization&rank=1 [accessed 2018-11-27]

40.   AniMovil mHealth Support for Depression Management in a Low-Income Country (AniMovil). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/ NCT03615118?cond=AniMovil+mHealth+Support+for+Depression+Management+in+Low-Income+Country&rank=1 [accessed 2018-11-27]

41.   Text-Message-Based Depression Prevention for High-Risk Youth in the ED (iDOVE). Clinical Trials. 2015. URL: https:/ /ClinicalTrials.gov/ct2/show/ NCT02332239?cond=Text-Message-Based+Depression+for+High-Risk+Youth+in+the+ED&rank=1 [accessed 2018-11-27]

42.   An Adaptive Intervention for Depression Among Latinos Living With HIV (Latino-SMART). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/ NCT03668379?cond=An+Adaptive+Intervention+for+Depression+Among+Latinos+Living+With+HIV&rank=1 [accessed 2018-11-27]

43.   Intervention to Prevent Peer Violence & Depressive Symptoms Among At-Risk Adolescents (iDOVE2). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03626103?term=NCT03626103&rank=1 [accessed 2018-11-27]

44.   Cognitive Behavioral Therapy Treatment of Depression With Smartphone Support. Clinical Trials. 2013. URL: https:/ /ClinicalTrials.gov/ct2/show/ NCT01819025?cond=Cognitive+Behavioral+Therapy+Treatment+of+Depression+With+Smartphone+Support&rank=1 [accessed 2018-11-27]

45.   Cognitive-Behavioral Intervention Via a Smartphone App for Depressive Symptoms in Caregivers (App Depression). Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/ NCT03110991?cond=Cognitive-behavioral+Intervention+via+a+Smartphone+App+for+Depressive+Symptoms+in+Caregivers&rank=1 [accessed 2018-11-27]

46.   Reducing Depressive Symptomatology With a Smartphone App. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/ show/NCT03060200?term=NCT03060200&rank=1 [accessed 2018-11-27]

47.   Smartphone-Enabled Health Coaching Intervention for Youth Diagnosed With Major Depressive Disorders. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03406052?term=NCT03406052&rank=1 [accessed 2018-11-27]

48.   Technology Assisted Programs That Promote Mental Health for Teenagers (ProjectTECH). Clinical Trials. 2013. URL: https://ClinicalTrials.gov/ct2/show/NCT01912729?term=NCT01912729&rank=1 [accessed 2018-11-27]

49.   Study of Technology-assisted Treatment of Adolescent Depression (iTAD). Clinical Trials. 2013. URL: https://ClinicalTrials. gov/ct2/show/NCT01582581?term=NCT01582581&rank=1 [accessed 2018-11-27]

50.   Effectiveness of a Technology Assisted Behavioral Intervention in Assisting People With Major Depressive Disorder. Clinical Trials. 2008. URL: https://ClinicalTrials.gov/ct2/show/NCT00719979?term=NCT00719979&rank=1 [accessed 2018-11-27]

51.   Online Peer Networked Collaborative Learning for Managing Depressive Symptoms (MoodTech). Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/show/NCT02841787?term=NCT02841787&rank=1 [accessed 2018-11-27]

52. New Technologies for Cognitive Behavior Therapy (CBT) Treatment of Adolescent Depression. Clinical Trials. 2013. URL: https://ClinicalTrials.gov/ct2/show/ NCT01868867?cond=New+Technologies+for+Cognitive+Behavior+Therapy+%28CBT%29+Treatment+of+Adolescent+Depression&rank=1 [accessed 2018-11-27]

53. Mental Health Telemetry for Self-Management in Major Depression (MHTV). Clinical Trials. 2013. URL: https:/ /ClinicalTrials.gov/ct2/show/NCT01999010?term=NCT01999010&rank=1 [accessed 2018-11-27]

54. Using Mental Health Telemetry to Predict Relapse and Re-hospitalization in Mood Disorders (PATH-MOD). Clinical Trials. 2013. URL: https://ClinicalTrials.gov/ct2/show/NCT01882608?term=NCT01882608&rank=1 [accessed 2018-11-27]

55. Training and Supervision Program for Depression Management. Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/ show/NCT02232854?cond=Training+and+Supervision+Program+for+Depression+Management&rank=1 [accessed 2018-11-27]

56. Effectiveness of a mHealth Intervention for the Treatment of Depression in People With Diabetes or Hypertension in Peru (LATIN-MHPeru). Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03026426 [accessed 2018-11-27]

57. Lifestyle Intervention for Young Adults With Serious Mental Illness. Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/show/NCT02815813?cond=Lifestyle+Intervention+for+Young+Adults+with+Serious+Mental+Illness&rank=1 [accessed 2018-11-27]

58. Comparing Mobile Health (mHealth) and Clinic-Based Self-Management Interventions for Serious Mental Illness. Clinical Trials. 2015. URL: https://ClinicalTrials.gov/ct2/show/NCT02421965?term=NCT02421965&rank=1 [accessed 2018-11-27]

59. Study of m-RESIST, an m-Health Program for Treatment-Resistant Schizophrenia (m-RESIST). Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/ NCT03064776?cond=Study+of+m-RESIST%2C+an+m-Health+Program+for+Treatment-resistant+Schizophrenia&rank=1 [accessed 2018-11-27]

60. Development of a Mobile System for Self-Management of Schizophrenia (SOS). Clinical Trials. 2013. URL: https:/ /ClinicalTrials.gov/ct2/show/ NCT01969500?cond=Development+of+a+Mobile+System+for+Self-Management+of+Schizophrenia+%28SOS%29&rank=1 [accessed 2018-11-27]

61. The Efficacy of Using a Smartphone App to Support Shared Decision Making in People With a Diagnosis of Schizophrenia. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03554655 [accessed 2018-11-27]

62. MedActive: A Smartphone Intervention to Improve Adherence to Antipsychotic Medications. Clinical Trials. 2013. URL: https://ClinicalTrials.gov/ct2/show/ NCT01953237?cond=MedActive%3A+A+Smartphone+Intervention+to+Improve+Adherence+to+AntiPsychotic+Medications&rank=1 [accessed 2018-11-27]

63. A New Paradigm for Illness Monitoring and Relapse Prevention in Schizophrenia. Clinical Trials. 2013. URL: https:/ /ClinicalTrials.gov/ct2/show/ NCT01952041?cond=A+New+Paradigm+for+Illness+Monitoring+and+Relapse+Prevention+in+Schizophrenia&rank=1 [accessed 2018-11-27]

64. Mobile Health Technology to Enhance Abstinence in Smokers With Schizophrenia. Clinical Trials. 2015. URL: https:/ /ClinicalTrials.gov/ct2/show/ NCT02420015?cond=Mobile+Health+Technology+to+Enhance+Abstinence+in+Smokers+with+Schizophrenia&rank=1 [accessed 2018-11-27]

65. Real-Time Mobile Cognitive Behavioral Intervention for Serious Mental Illness. Clinical Trials. 2014. URL: https:/ /ClinicalTrials.gov/ct2/show/ NCT02035202?cond=Real-Time+Mobile+Cognitive+Behavioral+Intervention+for+Serious+Mental+Illness&rank=1 [accessed 2018-11-27]

66. Feasibility and Outcomes of a Digital Health Support for the Schizophrenia Spectrum. Clinical Trials. 2018. URL: https:/ /ClinicalTrials.gov/ct2/show/ NCT03649815?cond=Feasibility+and+Outcomes+of+a+Digital+Health+Support+for+the+Schizophrenia+Spectrum&rank=1 [accessed 2018-11-27]

67. Motivation and Skills Support (MASS). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/ NCT03404219?cond=Motivation+and+Skills+Support+%28MASS%29&rank=1 [accessed 2018-11-27]

68. Smartphone Addiction Recovery Coach for Young Adults (SARC-YA) Experiment (SARC-YA). Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/ NCT03301012?cond=Smartphone+Addiction+Recovery+Coach+for+Adolescents+%28SARC-A%29+Experiment&rank=1 [accessed 2018-11-27]

69. The Cedar Project: Impact of mHealth for HIV Prevention Among Young Indigenous People Who Use Illicit Drugs. Clinical Trials. 2015. URL: https://ClinicalTrials.gov/ct2/show/NCT02437123 [accessed 2018-11-27]

70. Mobile Application to Improve Care Coordination Among HIV Clinic and Substance Use Treatment Providers. Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/show/

NCT02906215?cond=Mobile+Application+to+Improve+Care+Coordination+Among+HIV+Clinic+and+Substance+Use+Providers&rank=1 [accessed 2018-11-27]

71.   Mental Health Engagement Network (MHEN). Clinical Trials. 2011. URL: https://ClinicalTrials.gov/ct2/show/NCT01473550?cond=Mental+Health+Engagement+Network+%28MHEN%29&rank=1 [accessed 2018-11-27]

72.   Mobile Health Cognitive Stimulation in Heroin Users (Re@dict). Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/NCT02308878?cond=Mobile+Health+Cognitive+Stimulation+in+Heroin+Users&rank=1 [accessed 2018-11-27]

73.   Automated Recovery Line for Medication Assisted Treatment. Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/NCT02124980?cond=Automated+Recovery+Line+for+Medication+Assisted+Treatment&rank=1 [accessed 2018-11-27]

74.   Homeless Care Management App (Link2Care). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03399500?term=NCT03399500&rank=1 [accessed 2018-11-27]

75.   Augmenting Specialty Eating Disorder Clinical Treatment With a Smartphone Application. Clinical Trials. 2015. URL: https://ClinicalTrials.gov/ct2/show/NCT02484794?cond=Augmenting+Specialty+Eating+Disorder+Clinical+Treatment+with+a+Smartphone+Application&rank=1 [accessed 2018-11-27]

76.   Impact of Preanesthetic Information and Behavioral Intervention Using Smartphone on Anxiety of Children. Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/NCT02246062?cond=Impact+of+Preanesthetic+Information+and+Behavioral+Intervention+Using+Smartphone+on+Anxiety+of+Children&rank=1 [accessed 2018-11-27]

77.   Effect of Premedication Type on Preoperative Anxiety in Children (EPPA). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03530670?cond=Effect+of+Premedication+Type+on+Preoperative+Anxiety+in+Children&rank=1 [accessed 2018-11-27]

78.   Smartphone-Based Exposure Treatment for Dental Anxiety. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03461016?cond=Smartphone-Based+Exposure+Treatment+for+Dental+Anxiety&rank=1 [accessed 2018-11-27]

79.   Effects of Complementary Therapies Delivered Via Mobile Technologies. Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/NCT02236455?cond=Effects+of+Complementary+Therapies+Delivered+Via+Mobile+Technologies&rank=1 [accessed 2018-11-27]

80.   Using Smartphones to Enhance the Treatment of Childhood Anxiety (Smart-CAT). Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/NCT02259036?cond=Using+Smartphones+to+Enhance+the+Treatment+of+Childhood+Anxiety&rank=1 [accessed 2018-11-27]

81.   Youth Mayo Clinic Anxiety Coach Pilot Study. Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/NCT02205177 [accessed 2018-11-27]

82.   ACT-smart: Smartphone-supplemented iCBT for Social Phobia and/or Panic Disorder. Clinical Trials. 2013. URL: https://ClinicalTrials.gov/ct2/show/NCT01963806?cond=ACT-smart%3A+Smartphone-supplemented+iCBT+for+Social+Phobia+and%2For+Panic+Disorder&rank=1 [accessed 2018-11-27]

83.   Evaluating the Psychophysiological Effects of a Smartphone-Based Mindfulness Task. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03296007?cond=Evaluating+the+Psychophysiological+Effects+of+a+Smartphone-Based+Mindfulness+Task&rank=1 [accessed 2018-11-27]

84.   The Effectiveness of a Smartphone Application in the Treatment of Alcohol Use Disorder. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03396887?cond=The+Effectiveness+of+a+Smartphone+Application+in+the+Treatment+of+Alcohol+Use+Disorder&rank=1 [accessed 2018-11-27]

85.   AlcoChange: An Open Label Pilot Study of Smartphone Monitoring for Alcoholic Liver Disease. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03474328?cond=AlcoChange%3A+An+Open+Label+Pilot+Study+of+Smartphone+Monitoring+for+Alcoholic+Liver+Disease&rank=1 [accessed 2018-11-27]

86.   Health Mobile Cognitive Stimulation in Alcoholics. Clinical Trials. 2013. URL: https://ClinicalTrials.gov/ct2/show/NCT01942954?cond=Health+Mobile+Cognitive+Stimulation+in+Alcoholics&rank=1 [accessed 2018-11-27]

87.   Study of Mobile Phone Delivered Intervention to Reduce Alcohol Consumption (mROAD). Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/NCT02158949?term=NCT02158949&rank=1 [accessed 2018-11-27]

88.   Project Guard: Reducing Alcohol Misuse/Abuse in the National Guard. Clinical Trails. 2014. URL: https://ClinicalTrials.gov/ct2/show/NCT02860442?cond=Project+Guard%3A+Reducing+Alcohol+Misuse%2FAbuse+in+the+National+Guard&rank=1 [accessed 2018-11-27]

89.   A Text Message Behavioral Intervention to Reduce Alcohol Consumption in Young Adults (TRAC). Clinical Trials. 2012. URL: https://ClinicalTrials.gov/ct2/show/NCT01688245?term=NCT01688245&rank=1 [accessed 2018-12-27]

90.   A Tailored Physical Activity Smartphone App for Patients With Alcohol Dependence. Clinical Trials. 2016. URL: https:/
        /ClinicalTrials.gov/ct2/show/
        NCT02958280?cond=A+Tailored+Physical+Activity+Smartphone+App+for+Patients+With+Alcohol+Dependence&rank=1
        [accessed 2018-11-27]

91.   Young Adult Naturalistic Alcohol Study (YANAS) Using Smartphone Technology in a Simulated Laboratory Environment.
        Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/show/NCT02841735?term=NCT02841735&rank=1 [accessed
        2018-11-27]

92.   Smartphone Technology: Young Adult Drinking (STEADY). Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/
        show/NCT02963818?cond=Smartphone+Technology%3A+Young+Adult+Drinking+%28STEADY%29&rank=1 [accessed
        2018-11-27]

93.   Smartphone-Paired Breathalyzers and Loss- and Gain-Framed Texts for Reducing Drinking and Driving (BESAFE). Clinical
        Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03335735?term=NCT03335735&rank=1 [accessed 2018-11-27]

94.   Adaptive Goal-Directed Adherence Tracking and Enhancement (AGATE). Clinical Trial. 2011. URL: https://ClinicalTrials.
        gov/ct2/show/NCT01349985?term=NCT01349985&rank=1 [accessed 2018-11-27]

95.   The Efficacy of A Smartphone-Based Support System to Reinforce Alcohol Abstinence in Treatment-Seeking Patients.
        Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/
        NCT02385643?cond=The+Efficacy+of+A+Smartphone-based+Support+System+to+Reinforce+Alcohol+Abstinence+in+Treatment-seeking+Patients&rank=1
        [accessed 2018-11-27]

96.   Usefulness of Supportive Text Messages in the Treatment of Depressed Alcoholics. Clinical Trials. 2009. URL: https:/
        /ClinicalTrials.gov/ct2/show/
        NCT01037868?cond=Usefulness+of+Supportive+Text+Messages+in+the+Treatment+of+Depressed+Alcoholics&rank=1
        [accessed 2018-11-27]

97.   Text Messaging to Reduce Alcohol Relapse in Liver Transplant Patients. Clinical Trials. 2018. URL: https://ClinicalTrials.
        gov/ct2/show/NCT03402256?term=NCT03402256&rank=1 [accessed 2018-11-27]

98.   Lifestyle Physical Activity Intervention for Depressed Alcohol Dependent Women. Clinical Trials. 2016. URL: https:/
        /ClinicalTrials.gov/ct2/show/
        NCT02705898?cond=Lifestyle+Physical+Activity+Intervention+for+Depressed+Alcohol+Dependence&rank=1 [accessed
        2018-11-27]

99.   Reducing Non-Medical Opioid Use: An Automatically Adaptive mHealth Intervention. Clinical Trials. 2016. URL: https:/
        /ClinicalTrials.gov/ct2/show/
        NCT02990377?cond=Reducing+Non-Medical+Opioid+Use%3A+An+Automatically+Adaptive+mHealth+Intervention&rank=1
        [accessed 2018-11-27]

100.   mHealth for Patient Self-Management of Opioid Use Disorder. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/
        show/NCT03633929?cond=mHealth+for+Patient+Self-Management+of+Opioid+Use+Disorder&rank=1 [accessed
        2018-11-27]

101.   A Mobile Application for Post-Op Analgesic Consumption. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/
        NCT03197311?cond=A+Mobile+Application+for+Post-op+Analgesic+Consumption&rank=1 [accessed 2018-11-27]

102.   Using mHealth to Aid Opioid Medication Adherence Pilot Study. Clinical Trails. 2013. URL: https://ClinicalTrials.gov/
        ct2/show/NCT02017041?cond=Using+mHealth+to+Aid+Opioid+Medication+Adherence&rank=1 [accessed 2018-11-27]

103.   Using m-Health Tools to Reduce the Misuse of Opioid Pain Relievers. Clinical Trials. 2017. URL: https://ClinicalTrials.
        gov/ct2/show/NCT03012087?cond=Using+m-Health+Tools+to+Reduce+the+Misuse+of+Opioid+Pain+Relievers&rank=1
        [accessed 2018-11-27]

104.   Mobile Intervention for Young Opioid Users. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/
        NCT03610672?cond=Mobile+Intervention+for+Young+Opioid+Users&rank=1 [accessed 2018-11-27]

105.   Clinical Effect Size of an Educational Intervention in the Home and Compliance on People Who Suffer From Stroke.
        Clinical Trials. 2013. URL: https://ClinicalTrials.gov/ct2/show/NCT01980641 [accessed 2018-11-27]

106.   The Adherence and Knowledge Exchange Heart and Stroke Medicines Study (TAKEmeds). Clinical Trials. 2015. URL:
        https://ClinicalTrials.gov/ct2/show/NCT02597205?cond=TAKEmeds&rank=1 [accessed 2018-11-27]

107.   My Stroke Team (MYST): Stroke App Pilot Study. Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/
        NCT02230280?cond=myst+stroke&rank=1 [accessed 2018-11-27]

108.   iADAPT to Support Strategy Training After Stroke. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/
        NCT03253601?cond=iADAPTS+to+Support+Strategy+Training+After+Stroke&rank=1 [accessed 2018-11-27]

109.   Improving Medication Adherence Through SMS (Short Messaging Service) in Adult Stroke Patients: a Randomised
        Controlled Behaviour Intervention Trial. Clinical Trials. 2013. URL: https://ClinicalTrials.gov/ct2/show/NCT01986023
        [accessed 2018-11-27]

110.   TeleRehab for Stroke Patients Using Mobile Technology. Clinical Trials. 2015. URL: https://ClinicalTrials.gov/ct2/show/
        NCT02615132?cond=telerehab+stroke&rank=1 [accessed 2018-11-27]

111.   Developing Accessible mHealth Programs for Depression Management in Bolivia. Clinical Trials. 2016. URL: https:/
        /ClinicalTrials.gov/ct2/show/

NCT02765542?cond=Developing+Accessible+mHealth+Programs+for+Depression+Management+in+Bolivia&rank=1 [accessed 2018-11-27]

112. mHealth for Antenatal Mental Health (AMHS). Clinical Trials. 2019. URL: https://ClinicalTrials.gov/ct2/show/NCT02516982?cond=mHealth+for+Antenatal+Mental+Health&rank=1 [accessed 2018-11-27]

113. Scaling Up Science-Based Mental Health Interventions in Latin America (DIADA). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03392883?term=NCT03392883&rank=1 [accessed 2018-11-27]

114. Behavioural Activation-Based Treatment Administered Through Smartphone. Clinical Trials. 2011. URL: https://ClinicalTrials.gov/ct2/show/NCT01463020?cond=Behavioural+Activation-Based+Treatment+Administered+Through+Smartphone&rank=1 [accessed 2018-11-27]

115. Mobile Sensing and Support for Depression (MOSS). Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/show/NCT02776839?cond=Mobile+Sensing+and+Support+for+Depression&rank=1 [accessed 2018-11-27]

116. Mobile Technology to Engage and Link Patients and Providers in Antidepressant Treatment (MedLink). Clinical Trials. 2015. URL: https://ClinicalTrials.gov/ct2/show/NCT02583230?term=NCT02583230&rank=1 [accessed 2018-11-27]

117. Mobile Technology to Engage and Link Patients and Providers in Antidepressant Treatment (Medlink RCT). Clinical Trials. 2013. URL: https://ClinicalTrials.gov/ct2/show/NCT01909973?cond=Mobile+Technology+to+Engage+and+Link+Patients+and+Providers+in+Antidepressant+Treatment&rank=2 [accessed 2018-11-27]

118. Wellness Monitoring for Major Depressive Disorder (CBN-Well). Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/show/NCT02934334?term=NCT02934334&rank=1 [accessed 2018-11-27]

119. Patient Management of Depression Through Technology: A Study of Digitally Enabled Engagement. Clinical Trials. 2013. URL: https://ClinicalTrials.gov/ct2/show/NCT03242213?cond=Patient+Management+of+Depression+Through+Technology%3A+A+Study+of+Digitally+Enabled+Engagement&rank=1 [accessed 2018-11-27]

120. Effectiveness of a Mobile Texting Intervention for People With Serious Mental Illness. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03062267?cond=Effectiveness+of+a+Mobile+Texting+Intervention+for+People+With+Serious+Mental+Illness&rank=1 [accessed 2018-11-27]

121. Psychotherapeutic Text Messaging for Depression Pilot Study (TEXT4U). Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/show/NCT02872454?cond=Psychotherapeutic+Text+Messaging+for+Depression+Pilot+Study&rank=1 [accessed 2018-11-27]

122. Treating Depression on a Day-to-Day Basis: Development of a Tool for Physicians Based on a Smartphone Application (SMART). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03678194?cond=Treating+Depression+on+a+Day-to-day+Basis%3A+Development+of+a+Tool+for+Physicians+Based+on+a+Smartphone+Application&rank=1 [accessed 2018-11-27]

123. Testing the Value of Smartphone Assessments of People With Mood Disorders. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03429361?cond=Testing+the+Value+of+Smartphone+Assessments+of+People+with+Mood+Disorders&rank=1 [accessed 2018-11-27]

124. Evolution of Dark Ideas When Introducing or Switching an Antidepressant (DEPASSE). Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03327974?term=NCT03327974&rank=1 [accessed 2018-11-01]

125. IntelliCare: Artificial Intelligence in a Mobile Intervention for Depression and Anxiety (AIM). Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/NCT02176226?cond=IntelliCare%3A+Artificial+Intelligence+in+a+Mobile+Intervention+for+Depression+and+Anxiety+%28AIM%29&rank=1 [accessed 2018-11-27]

126. Augmenting Hospitalization for Serious Mental Illness: Cognitive Bias Modification. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03509181?cond=Augmenting+Hospitalization+for+Serious+Mental+Illness%3A+Cognitive+Bias+Modification&rank=1 [accessed 2018-11-27]

127. Evaluation of Text Message Engagement Support of Mindfulness Smartphone Applications. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03633682?cond=Evaluation+of+Text+Message+Engagement+Support+of+Mindfulness+Smartphone+Applications&rank=1 [accessed 2018-11-27]

128. Enhancing Delivery of Problem Solving Therapy Using SmartPhone Technology. Clinical Trials. 2013. URL: https://ClinicalTrials.gov/ct2/show/NCT01891734?cond=Enhancing+Delivery+of+Problem+Solving+Therapy+Using+SmartPhone+Technology&rank=1 [accessed 2018-11-27]

129. Mobile Phone Sensing and Outreach as Adjuncts to Internet-Based Behavioral Intervention for Depression. Clinical Trials. 2010. URL: https://ClinicalTrials.gov/ct2/show/NCT01107041?term=NCT01107041&rank=1 [accessed 2018-11-27]

130. Technology Enabled Mental Health Intervention for Individuals in the Criminal Justice System. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03105973?cond=Technology+Enabled+Mental+Health+Intervention+for+Individuals+in+the+Criminal+Justice+System&rank=1 [accessed 2018-11-27]

131. Therapeutic Efficacy of Categorical Language Fluency Smartphone Game Application. Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/show/NCT02848404?cond=Therapeutic+Efficacy+of+Categorical+Language+Fluency+Smartphone+Game+Application&rank=1 [accessed 2018-11-27]

132. Improving Medical Care With Electronic Interventions Based on Automated Text and Phone Messages. Clinical Trials. 2916. URL: https://ClinicalTrials.gov/ct2/show/NCT03002311?cond=Improving+Medical+Care+With+Electronic+Interventions+Based+on+Automated+Text+and+Phone+Messages&rank=1 [accessed 2018-11-27]

133. Youth Mayo Clinic Anxiety Coach Randomized Controlled Trial. Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/NCT02205203?cond=Youth+Mayo+Clinic+Anxiety+Coach+Randomized+Controlled+Trial&rank=1 [accessed 2018-11-27]

134. mWELLCARE:An Integrated mHealth System for the Prevention and Care of Chronic Disease (mWELLCARE). Clinical Trials. 2015. URL: https://ClinicalTrials.gov/ct2/show/NCT02480062 [accessed 2018-11-27]

135. Smartphone Based Continuing Care for Alcohol. Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/show/NCT02681406?cond=Smartphone+Based+Continuing+Care+for+Alcohol&rank=1 [accessed 2018-11-27]

136. Feasibility of a Smart-Phone Based Support System for Hazardous Drinkers (NZStepAway). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03553056?term=NCT03553056&rank=1 [accessed 2018-11-27]

137. Skills-Training for Reducing Risky Alcohol Use in App Form. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03696888?cond=Skills-Training+for+Reducing+Risky+Alcohol+Use+in+App+Form&rank=1 [accessed 2018-11-27]

138. Impact on Opioid Use of Bundling Medication-assisted Treatment With mHealth (Bundling). Clinical Trials. 2016. URL: https://ClinicalTrials.gov/ct2/show/NCT02712034?cond=.+Impact+on+Opioid+Use+of+Bundling+Medication-assisted+Treatment+with+mHealth+%28Bundling%29&rank=1 [accessed 2018-11-27]

139. Needle-X: Usability Testing of a Smartphone Application. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03665298?cond=Needle-X%3A+Usability+Testing+of+Smartphone+Application&rank=1 [accessed 2018-11-27]

140. Using mHealth to Aid Opioid Addicts. Clinical Trials. 2013. URL: https://ClinicalTrials.gov/ct2/show/NCT01955902?cond=Using+mHealth+to+Aid+Opioid+Addicts&rank=1 [accessed 2018-11-27]

141. Smartphone Technology to Alleviate Malignant Pain (STAMP). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03717402?cond=Smartphone+Technology+to+Alleviate+Malignant+Pain+%28STAMP%29&rank=1 [accessed 2018-11-27]

142. The Application of Wearable Technology to Improve the Physical Activity Level of People With Chronic Mental Illness. Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03408327?cond=The+Application+of+Wearable+Technology+to+Improve+the+Physical+Activity+Level+of+People+with+Chronic+Mental+Illness&rank=1 [accessed 2018-11-27]

143. Feasibility and Acceptability of a Smartphone App to Assess Early Warning Signs of Psychosis Relapse (ExPRESS:2). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03558529 [accessed 2018-11-27]

144. Using Mobile Technology to Enhance Early Psychosis Treatment Delivery (RWJFGinger). Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03303456?cond=Using+Mobile+Technology+to+Enhance+Early+Psychosis+Treatment+Delivery&rank=1 [accessed 2018-11-27]

145. Smartphone Applications Youth With Early Psychosis in Community Outpatient Settings (BHCOEMobi). Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03286595?cond=Smartphone+Applications+Youth+with+Early+Psychosis+in+Community+Outpatient+Settings&rank=1 [accessed 2018-11-27]

146. Mobile Enhancement of Motivation in Schizophrenia. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03059771?cond=Mobile+Enhancement+of+Motivation+in+Schizophrenia&rank=1 [accessed 2018-11-27]

147. Waitlist-Control Trial of Smartphone CBT for Body Dysmorphic Disorder (BDD). Clinical Trials. 2018. URL: https://ClinicalTrials.gov/ct2/show/NCT03673046?cond=Waitlist-Control+Trial+of+Smartphone+CBT+for+Body+Dysmorphic+Disorder+%28BDD%29&rank=1 [accessed 2018-11-27]

148. Preventing HIV/STI in Urban Adolescents Via an mHealth Primary Care Intervention. Clinical Trials. 2017. URL: https://ClinicalTrials.gov/ct2/show/NCT03368456?cond=Preventing+HIV%2FSTI+in+Urban+Adolescents+via+an+mHealth+Primary+Care+Intervention&rank=1 [accessed 2018-11-27]

149. CopeSmart: Using Mobile Technology to Promote Positive Mental Health in Young People. Clinical Trials. 2014. URL: https://ClinicalTrials.gov/ct2/show/ NCT02265978?cond=CopeSmart%3A+Using+Mobile+Technology+to+Promote+Positive+Mental+Health+In+Young+People&rank=1 [accessed 2018-11-27]

150. Tse T, Fain KM, Zarin DA. How to avoid common problems when using ClinicalTrials.gov in research: 10 issues to consider. Br Med J 2018 May 25;361:k1452 [FREE Full text] [doi: 10.1136/bmj.k1452] [Medline: 29802130]

151. Williams RJ, Tse T, Harlan WR, Zarin DA. Registration of observational studies: is it time? Can Med Assoc J 2010 Oct 19;182(15):1638-1642 [FREE Full text] [doi: 10.1503/cmaj.092252] [Medline: 20643833]

152. Effects of Patient-Centered Stroke Educating System: A Randomized Controlled Trial. Clinical Trials. 2018. URL: https:/ /ClinicalTrials.gov/ct2/show/NCT02591511?cond=stroke+educating+system&rank=1 [accessed 2018-11-27]

153. mWELLCARE:An Integrated mHealth System for the Prevention and Care of Chronic Disease (mWELLCARE). Clinical Trials. 2015. URL: https://ClinicalTrials.gov/ct2/show/NCT02480062 [accessed 2018-11-27]

154. Rayhill ML, Sharon R, Burch R, Loder E. Registration status and outcome reporting of trials published in core headache medicine journals. Neurology 2015 Nov 17;85(20):1789-1794. [doi: 10.1212/WNL.0000000000002127] [Medline: 26475691]

155. Safavi K, Mathews SC, Bates DW, Dorsey ER, Cohen AB. Top-funded digital health companies and their impact on high-burden, high-cost conditions. Health Aff (Millwood) 2019 Jan;38(1):115-123. [doi: 10.1377/hlthaff.2018.05081] [Medline: 30615535]

156. Calvo RA, Dinakar K, Picard R, Christensen H, Torous J. Toward impactful collaborations on computing and mental health. J Med Internet Res 2018 Feb 9;20(2):e49 [FREE Full text] [doi: 10.2196/jmir.9021] [Medline: 29426812]

157. Al-Durra M, Nolan RP, Seto E, Cafazzo JA, Eysenbach G. Nonpublication rates and characteristics of registered randomized clinical trials in digital health: cross-sectional analysis. J Med Internet Res 2018 Dec 18;20(12):e11924 [FREE Full text] [doi: 10.2196/11924] [Medline: 30485832]

158. Antes G, Chalmers I. Under-reporting of clinical trials is unethical. Lancet 2003 Mar 22;361(9362):978-979. [doi: 10.1016/S0140-6736(03)12838-3] [Medline: 12660049]

159. Hopewell S, Loudon K, Clarke M, Oxman A, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Library 2009:-. [doi: 10.1002/14651858.mr000006.pub3]

160. Tfelt-Hansen PC. Unpublished clinical trials with sumatriptan. Lancet 2009 Oct 31;374(9700):1501-1502. [doi: 10.1016/S0140-6736(09)61906-1] [Medline: 19880018]

161. Gabapentin for Pain: New Evidence From Hidden Data. Evidence-Based Prescription Drug Therapy Information. 2009. URL: https://www.ti.ubc.ca/2009/12/31/gabapentin-for-pain-new-evidence-from-hidden-data/ [accessed 2019-01-05]

162. 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]

163. Lee K, Bacchetti P, Sim I. Publication of clinical trials supporting successful new drug applications: a literature analysis. PLoS Med 2008 Sep 23;5(9):e191 [FREE Full text] [doi: 10.1371/journal.pmed.0050191] [Medline: 18816163]

164. Azar M, Riehm KE, McKay D, Thombs BD. Transparency of outcome reporting and trial registration of randomized controlled trials published in the journal of consulting and clinical psychology. PLoS One 2015;10(11):e0142894 [FREE Full text] [doi: 10.1371/journal.pone.0142894] [Medline: 26581079]

165. Barnett I, Torous J, Staples P, Keshavan M, Onnela J. Beyond smartphones and sensors: choosing appropriate statistical methods for the analysis of longitudinal data. J Am Med Inform Assoc 2018 Dec 1;25(12):1669-1674 [FREE Full text] [doi: 10.1093/jamia/ocy121] [Medline: 30272176]

Abbreviations

DALYs: disability-adjusted life years
EU: European Union
ICJME: International Committee of Journal Medical Editors
IRB: Institutional Review Board
mHealth: mobile health
NIH: National Institutes of Health
NYU: New York University
SOM: School of Medicine
WHO: World Health Organization

XSL•FO

RenderX

XSL•FO
RenderX

Original Paper

Developing a Mobile App (iGAM) to Promote Gingival Health by Professional Monitoring of Dental Selfies: User-Centered Design Approach

Guy Tobias[1*], MPH, DMD; Assaf B Spanier[2*], PhD

[1]Department of Community Dentistry, Faculty of Dental Medicine, The Hebrew University - Hadassah School of Dental Medicine, Jerusalem, Israel
[2]Department of Software Engineering, Azrieli College of Engineering, Jerusalem, Israel
[*]all authors contributed equally

Corresponding Author:
Guy Tobias, MPH, DMD
Department of Community Dentistry
Faculty of Dental Medicine
The Hebrew University - Hadassah School of Dental Medicine
Ein Kerem
Jerusalem, 91120
Israel
Phone: 972 527052333
Email: guy.tobias@mail.huji.ac.il

Related Articles:

Companion article: https://mhealth.jmir.org/2021/9/e24955

Companion article: https://formative.jmir.org/2021/10/e26125

Abstract

Background: Dental visits are unpleasant; sometimes, patients only seek treatment when they are in intolerable pain. Recently, the novel coronavirus (COVID-19) pandemic has highlighted the need for remote communication when patients and dentists cannot meet in person. Gingivitis is very common and characterized by red, swollen, bleeding gums. Gingivitis heals within 10 days of professional care and with daily, thorough oral hygiene practices. If left untreated, however, its progress may lead to teeth becoming mobile or lost. Of the many medical apps currently available, none monitor gingivitis.

Objective: This study aimed to present a characterization and development model of a mobile health (mHealth) app called iGAM, which focuses on periodontal health and improves the information flow between dentists and patients.

Methods: A focus group discussed the potential of an app to monitor gingivitis, and 3 semistructured in-depth interviews were conducted on the use of apps for monitoring gum infections. We used a qualitative design process based on the Agile approach, which incorporated the following 5 steps: (1) user story, (2) use cases, (3) functional requirements, (4) nonfunctional requirements, and (5) Agile software development cycles. In a pilot study with 18 participants aged 18-45 years and with different levels of health literacy, participants were given a toothbrush, toothpaste, mouthwash, toothpicks, and dental floss. After installing iGAM, they were asked to photograph their gums weekly for 4 weeks.

Results: All participants in the focus group believed in the potential of a mobile app to monitor gingivitis and reduce its severity. Concerns about security and privacy issues were discussed. From the interviews, 2 themes were derived: (1) "what's in it for me?" and (2) the need for a take-home message. The 5 cycles of development highlighted the importance of communication between dentists, app developers, and the pilot group. Qualitative analysis of the data from the pilot study showed difficulty with: (1) the camera, which was alleviated with the provision of mouth openers, and (2) the operation of the phone, which was alleviated by changing the app to be fully automated, with a weekly reminder and an instructions document. Final interviews showed satisfaction.

XSL•FO
RenderX

Conclusions: iGAM is the first mHealth app for monitoring gingivitis using self-photography. iGAM facilitates the information flow between dentists and patients between checkups and may be useful when face-to-face consultations are not possible (such as during the COVID-19 pandemic).

Introduction

Barriers to Dental Care

For most people, dental visits are unpleasant, take time, cost money, and are often accompanied by discomfort or pain [1,2]. In addition, the introduction of sharp instruments and splashing water in the mouth, the need to keep the mouth open for long periods, and the sounds and smells affiliated with dental procedures further intensify the negative experience [3]. Therefore, it is not surprising that many people only seek dental care when their pain becomes unbearable and they are out of options [4].

At the time of writing, we are in the midst of an urgent public health crisis. The coronavirus disease 2019 (COVID-19) global pandemic has necessitated social distancing [5]. One field that has been affected in particular is dentistry. In Israel, the Ministry of Health [6] has prohibited nonemergency dental treatment; as such, regular checkups and elective treatments have been postponed, and dentists cannot monitor their patients' oral health.

Dental Scientific Background

The 2 most prevalent diseases of the oral cavity in adults are caries and periodontal diseases. Caries involves the hard tissues in the mouth (the teeth); it is caused by bacteria that initially demineralizes tooth enamel and then penetrates and causes decay [7]. Periodontal diseases involve the tissues supporting the teeth, which include the alveolar bone, periodontal ligaments, cementum, and gingivae [8]. Periodontal diseases are divided according to severity, namely gingivitis and periodontitis. Gingivitis is a reversible condition; the lesions are restricted to the gums, which are typically red, swollen, and bleeding. Gingivitis heals within 10 days of professional care and with daily, thorough oral hygiene practices [9,10]. Untreated gingivitis, however, develops into periodontitis, the irreversible stage of periodontal disease, in which bacterial toxins and the immune response to them destroy the tissues that support the teeth. As the disease progresses, the teeth may eventually become mobile or lost [11]. A significant difference between caries and periodontal disease is that caries often causes pain in its early stages, whereas periodontal diseases usually remains asymptomatic until advanced disease is noted, and sometimes there is no pain at all [12]. These two diseases share the feature of being progressive, such that a delay in treatment may lead to the need for complex and expensive treatments or tooth loss [13,14].

Most people can recognize the early signs of gingival inflammation, which are bleeding while brushing or eating something hard, an unpleasant odor, or swollen gums. But because gingivitis is not painful, people tend to postpone dental appointments to when the disease is more advanced [15]. Epidemiologically, gingivitis is common among 18-year-olds, and over 80% of the global population suffers from gingivitis from time to time [16,17]. The treatment for gingivitis is relatively straightforward, primarily based on adequate oral hygiene practices that include twice-daily brushing and interproximal cleaning using dental floss, toothpicks, and mouthwash [18,19]. In the dental clinic, gingivitis is usually treated in a single cleaning session where plaque and calculus are removed [20]. Without appropriate oral hygiene, gingivitis may return and progress into periodontitis. Treatment for periodontitis is complex and requires multiple dental appointments, and sometimes surgical intervention. During routine dental checkups, the dentist surveys periodontal health and suggests the timing of the next appointment, often in 6 months' time. The interval between treatments is typically individualized according to variability in health behavior, lifestyle, and genetic diversity. In order to determine the optimal time interval between dental appointments, dental health should be monitored between them so that problems are detected early [21,22].

Regular checkups are essential for good oral health; however, the vast majority of people do not come for routine checkups [23]. Only 9% of people attend without an acute problem [24]. The results from a large cohort (n=608) telephone survey conducted by Sharabany et al [23] showed that 47.2% of respondents had a dental examination at least once a year, and the remaining individuals were examined once every two years or less frequently. They noted that the decision to schedule a routine dental examination was associated with socioeconomic status and personal beliefs about dental health.

Cellular Apps Background

Approximately 45% of people have at least one of the following devices: a desktop or laptop computer, a tablet, or a smartphone [25]. In the US, 56% of citizens own a smartphone capable of connecting to the internet and downloading content [26]. About a quarter of children use their smartphones more than 5 hours a day, and about 46% of teens aged 12-17 years surf the internet for 4 hours a day (browsing from their smartphones or computers) [27].

The Apple Store and Google Play app store contain more than 318,000 medical apps, with hundreds of new apps added daily [28,29]. EHealth (electronic health) is a broad term that describes the use of electronic devices to improve health. Mobile health (mHealth) is the part of eHealth that includes the use of mobile devices to gather data about an individual's health status and provides information to professionals and patients in

real-time. There are many mHealth apps, such as those that monitor medical conditions like blood glucose levels [30] (eg, SuCare, Sanofi-Aventis US LLC), blood pressure, and cholesterol [31] (eg, Vitadoc+, Medisana GmbH). In an app for type 1 diabetes, patients can submit information and get management instructions in real-time [32] (Gluci-Chek, Roche Diabetes Care Inc). There are also apps that provide exercise programs [33] (eg, Runkeeper, ASICS), weight loss plans [34] (eg, MyFitness Pal Calorie Counter and Diet Tracker, Under Armour Inc), and functions for family planning [35] (eg, Glow, Glow Inc). Some apps are able to alert a patient to when they need further professional assistance. Other apps assist in the diagnosis of specific pathologies, such as an app that uses the image processing of urine tests to diagnose kidney disease [36] (Healthy.io). In dentistry, there are apps that show clinical procedures using 3D imagery [37] (eg, Lexi-DENTAL COMPLETE, Lexicomp). However, following an extensive search, we did not find any apps that monitor gingivitis.

To the best of our knowledge, there are no available mHealth apps that monitor the periodontal status of patients and update the dentist or the patient about dental health during the interval between visits. The oral cavity remains unmonitored, and without proper oral hygiene, gingivitis can progress and worsen.

Purpose of the Study

The rationale for conducting this study was the understanding that between dental checkups, oral health in general, and periodontal health in particular, may decline; having an mHealth app that allows a dentist to monitor the periodontal status of patients with patient-generated photographs of the oral cavity (which we will refer to as dental selfies) should limit the deterioration that may occur between visits. Furthermore, a patient can be asked to come into the clinic when it is deemed necessary.

Study Objective

The aim of this study was to present an mHealth app to improve the flow of information between dentists and their patients during the intervals between checkups. We describe the characterization and development of an mHealth app called iGAM [38], which focuses on periodontal health. iGAM was developed as part of a quantitative and qualitative integrated research project to promote oral health with expert writing applications. iGAM can be downloaded from both the Apple Store and Google Play store.

Methods

Ethics Approval and Consent

The study was approved by the Hadassah research ethics committee (IRB, 0212-18-HMO), and informed consent was obtained from all participants.

Design and Development

The Agile approach [39] was used to develop iGAM and included the following steps: (1) user story, (2) use cases, (3) functional requirements, (4) nonfunctional requirements, and (5) the Agile software development life cycle.

The user story pertains to the conceptual scenarios of using a high-level app, aimed at defining the target audience and what is needed for the app to be self-contained for oral photography. Use cases are the specific actions that users (patients and dentists) will perform with the app. Functional requirements describe the functionality of the product, namely, which software tasks it must perform, the scope of the system, the boundaries of the product, and relationships to adjacent systems. Nonfunctional requirements describe the look and feel of the system, such as the visual characteristics [eg, the user interface (UI) and user experience (UX) of the system], its usability, and performance requirements (eg, how big, how fast). The Agile software development life cycle involves dividing the project into small increments to allow rapid changes to be made. The development life cycle includes customer satisfaction, delivering working software frequently (weeks rather than months), building around motivated individuals, using working software as the primary measure of progress, continuous attention to technical excellence and good design, and the maintenance of simplicity.

Results

User Story

The routine activity of 2 dental clinics, over the span of two 9-hour workdays, was monitored by the lead researcher (a dentist) and the app developer. Communication and patient management were observed in order to identify the digital dental technologies used and to assess the need for the iGAM app. We observed that dentists frequently employ administrative, communicative, clinical, and diagnostic technologies.

Then a focus group meeting was held with 10 participants: 2 dentists, 2 advanced-year dental students, and 6 patients who had just been treated. During the meeting, several questions were asked by the lead investigator using a guide. The conversation was recorded and transcribed with the prior approval of the participants. All participants believed in the potential of a mobile app to effectively monitor gingivitis and reduce the severity of gingivitis. The meeting also covered other issues, such as having access to easily understood information on gum diseases (including pictures). Some participants voiced concerns about information security, privacy issues, and the risk of personal oral health information being leaked.

Subsequently, we conducted 3 semistructured in-depth interviews on the use of cellphone apps for monitoring gum infections in the oral cavity. The mean duration of the interviews was 50 minutes. The interviewees were 2 men and 1 woman, which included a dentist, a nurse, and a dental student after dental treatment. The interviews were recorded and transcribed with permission. From the analysis of the interviews, 2 major themes emerged: (1) "What's in it for me?" and (2) the need for a take-home message. The first theme highlighted that users need to perceive a gain from the use of the app. For example, one participant said,

> I feel good. Occasionally I see some bleeding when I brush. So what? It doesn't mean anything to me, doesn't hurt me, and I don't have time to go to a

dentist. For me to use this app means there has to be a reason, and I don't understand what I will get out of it.

The second theme highlighted the user's need for a take-home message. For example, some participants said,

I would love to get answers to questions, and the doctor does not have time to answer me properly.

I would be happy if the dentist in the clinic would give me a leaflet with an explanation of how to brush my teeth; I'm not sure I brush right.

It's important for me to know that my mouth is healthy, that I'm doing things right.

Usage Scenarios

Once we understood what the users needed, we were able to describe the use cases of the app.

Architecture Diagrams

The iGAM mHealth app has 2 modes: (1) the patient mode, and (2) the dentist mode (Figure 1). In the first mode, the patient takes dental selfies (photographs of their gums) and sends them to the dentist through the app. In the second mode, the dentist determines the gingival status of the patient.

Figure 1. Interaction between the user, app administrator, and data storage server.
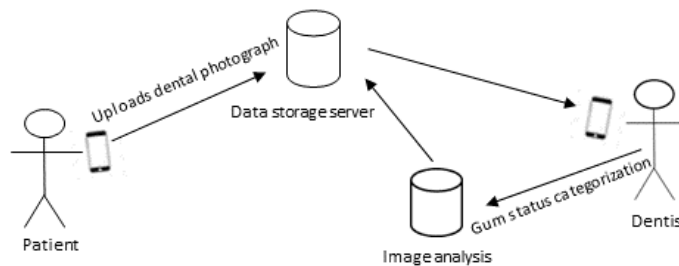
Use Cases: The Patient

The app's patient mode allows patients to: (1) login; (2) register and see general guidelines; (3) respond to questionnaires about health and dental behavior; (4) view tutorials with brushing instructions, information about periodontal problems, self-photography training, etc; (5) take photos; (6) and keep

track of responses to previously submitted photographs (Figure 2). All submitted patient data is saved in the database. The patient is instructed to take multiple dental selfies of their gums using the rear camera of their cellphone, once a week for 8 weeks. This time period was selected based on studies on gingivitis [40].

Figure 2. Patient use cases.
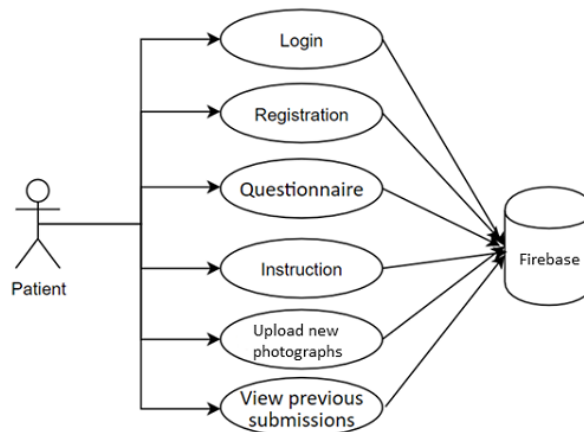
Use Cases: The Dentist

The app's dentist mode allows the dentist to: (1) login; (2) view the submitted dental selfies; (3) rate the image quality and

evaluate gingival health status using the modified gingival index (MGI; the MGI was introduced in the mid-1980s and was found to be reliable in the visual diagnosis of gingivitis) [41]; (4) add new patients; and (5) add and edit tutorials (Figure 3).

Figure 3. Dentist use cases.
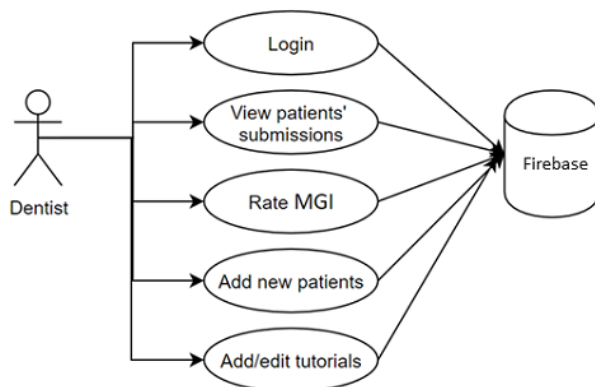
Functional Requirements

The following describes the intended patient user's experience with the iGAM app.

Following initial registration, a password known only to the patient is given to maintain anonymity. The first time the patient logs in, personal patient data (such as age, gender, origin, etc) is collected. In order to characterize the patient's interest in the app, the patient is prompted to fill out a questionnaire about dental knowledge and behaviors, including oral hygiene habits. Tutorials about oral hygiene and information on the dangers of poor oral hygiene maintenance are available on the app for the patient.

The patient is then instructed to photograph themselves once a week for 8 weeks. The dentist will assess inflammation using the MGI.

Nonfunctional Requirements

An intuitive UI was created to allow the user to perform editing tasks by clicking twice on the main screen. Data is synchronized to the cloud after each patient session. The app opens within 2 seconds. Communication with the text-to-speech (TTS) server is established in less than 30 seconds.

For interface design (UI/UX), the lead researcher and app developer sketched their ideas using pencil and paper (Figure 4).

Figure 4. Sketches of the essence of the app.
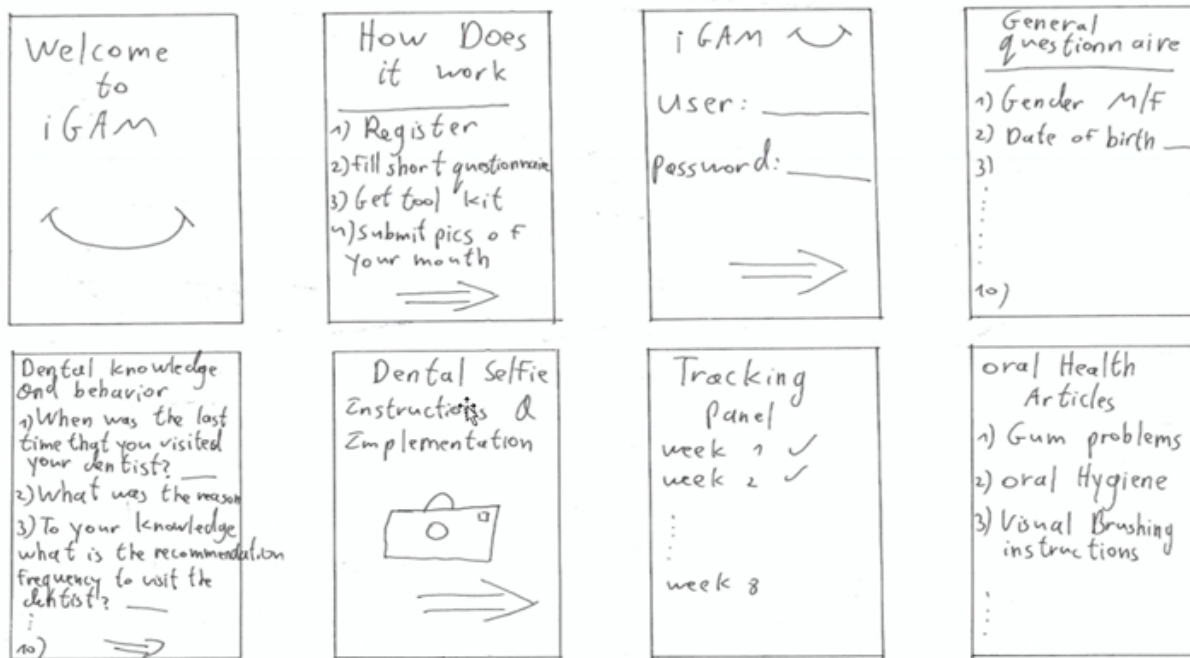
The prototype, which was designed using Adobe XD software (Adobe XD 21.0, Adobe Inc), included interface items with visual elements and colors, such as screens for the patient (Textbox 1, Figures 5-17) and screens for the dentist (Textbox 2, Figures 18-21). This facilitated the receipt of early feedback

in the design process. Hebrew was selected as the interface language as it is the primary language of the users; in the future, the app will also be available in English, Arabic, and Russian, the most commonly used languages in Israel.

Textbox 1. Prototype screens for patients in the iGAM app.

Initial Login Screen (Figure 5)

- Patient's name and password

General Explanation Screen (Figure 6)

Registration Questionnaire Screens

- General (Figure 7)

- General health (Figure 8)

- Oral hygiene (Figure 9)

- Oral hygiene, continued (Figure 10)

Tutorial Screens

- Main screen (Figure 11)

- Photo guidelines (Figure 12)

- Periodontal problems (Figure 13)

- Brushing your teeth (Figure 14)

Photo Panel Screens

- A button on each line allows a photograph to be taken at each prescribed time, and opens the camera (Figure 15).

- Clicking on the icon opens the camera (Figure 16).

- After a 10-second countdown, photographs are taken (Figure 17).

Figure 5. Initial login screen for the patient: patient name and password.
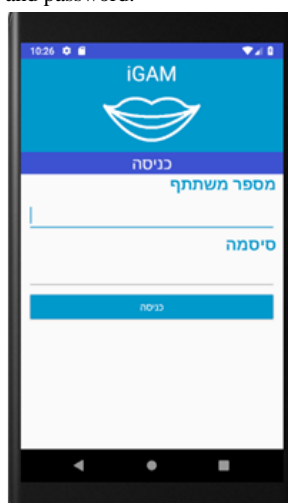
Figure 6. General explanation screen for the patient.

Figure 7. Registration questionnaire screen for the patient: general.

Figure 8. Registration questionnaire screen for the patient: general health.
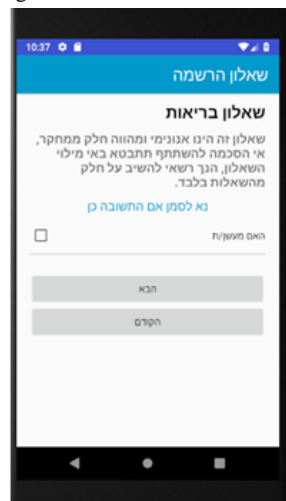
Figure 9. Registration questionnaire screen for the patient: oral hygiene.
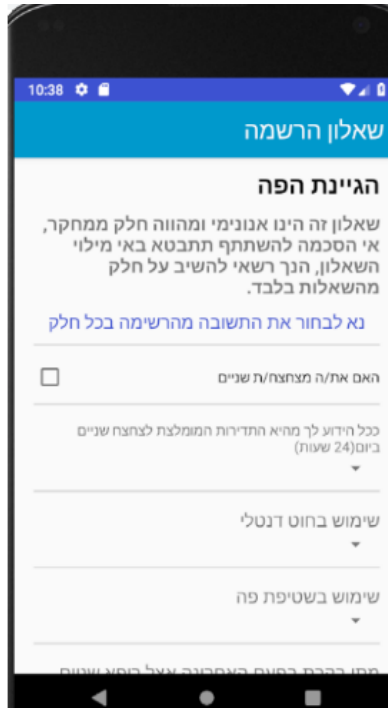
Figure 10. Registration questionnaire screen for the patient: oral hygiene, continued.

Figure 11. Tutorial screen for the patient: main.

Figure 12. Tutorial screen for the patient: photo guidelines.

Figure 13. Tutorial screen for the patient: periodontal problems.

XSL•FO
RenderX

Figure 14. Tutorial screen for the patient: brushing your teeth.

Figure 15. Photo panel screen for the patient: A button on each line allows a photograph to be taken at each prescribed time, and opens the camera.

Figure 16. Photo panel screen for the patient: Clicking on the icon opens the camera.

XSL•FO
RenderX

Figure 17. Photo panel screen for the patient: After a 10-second countdown, photographs are taken.

Textbox 2. Prototype screens for dentists in the iGAM app.

Initial Login Screen (Figure 18)

New Patient Screen (Figure 19)

- Add a new patient

Patient List Screen (Figure 20)

- The dentist can view the list of patients.

Viewing Screen (Figure 21)

- The dentist can view the dental selfie and rate the patient's condition.

Figure 18. Login screen for the dentist.
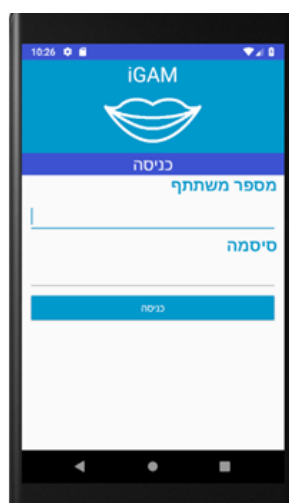
Figure 19. Patient screen for the dentist: the dentist may add a new patient.
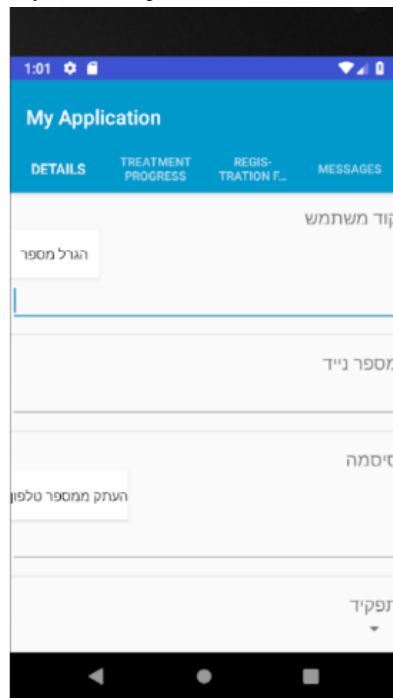
Figure 20. Patient list screen for the dentist: the dentist can view the list of patients.

XSL·FO
RenderX
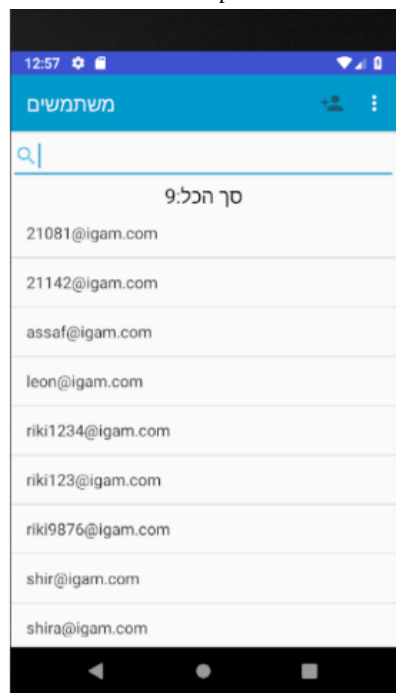
Figure 21. Viewing screen for the dentist: the dentist can view the dental selfie and rate the patient's condition.
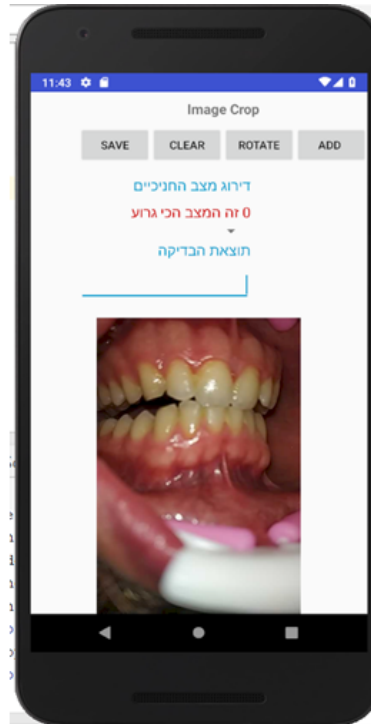
Design and Development

The sequence diagram in Figure 22 describes the processes involved in the use of the app by the patient.

After logging in for the first time, the patient fills out a form with their personal information and oral health information. The data is stored in the database (firebase) under a unique and anonymous identifier. All photographs and responses to the questionnaires are saved in the database, including the time and date of submission. After uploading a new dental selfie, the patient's status changes to "pending review."

The 4 major operations of the dentist (Figure 23) are (1) the verification of correct username and password, (2) viewing the list of new patient submissions and selecting the user to view, (3) formulating a rating using MGI, and (4) adding a new user to the trial.

Figure 22. Sequence diagram of the patient's use processes with the app; DB: database.
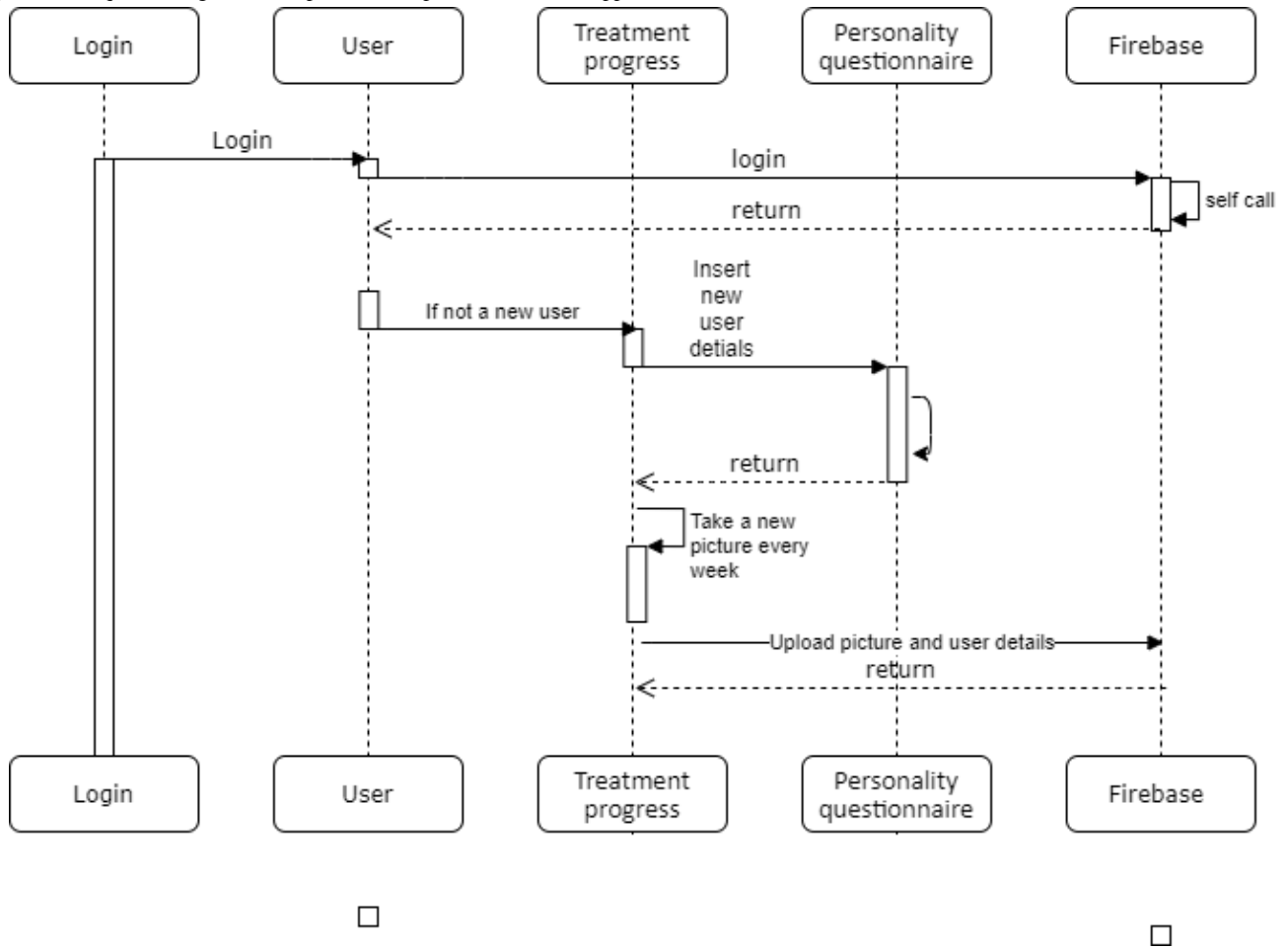
Figure 23. Sequence diagrams of the dentist side. After login the dentist can view patient submissions, rate gum status according to the MGI, add new patient and new tutorials.
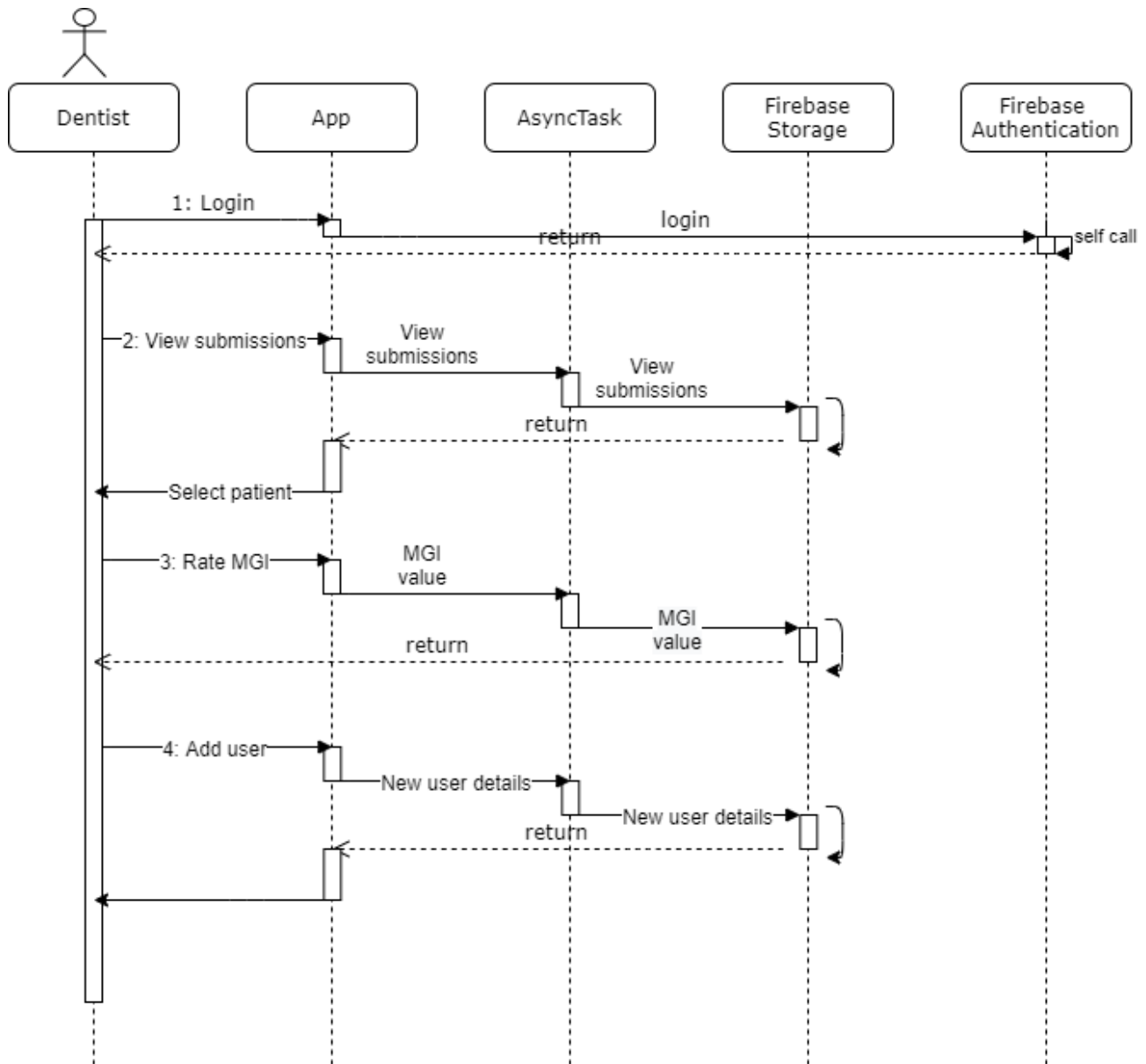
Implementation Technologies

Client Side

Two versions of the client side of iGAM were created. The first version was written for the Android operating system (ie, a native app, a method used for writing software in a programming language for certain types of mobile phones) in the Android Studio workspace. The second version was a hybrid app; development was done with web technologies (HTML, CSS, and mainly Javascript programming language) and was then packaged using tools like Phonegap/Cordova. We used an Ionic library to create the UI for this mobile app. The Ionic library runs on a wide range of Android iPhone devices. The technology environments we used were Android Studio for Android and Ionic Studio integrated development environment for Ionic. The development languages we used were Javascript for Android, and JavaScript, Angular, and TypeScript for Ionic.

Server Side

To store images and questionnaire results, Firebase storage was used. Firebase storage provides secure uploads and downloads for apps regardless of network quality, and the service is backed by Google Cloud storage.

The system data is managed by a database containing patient information. Specifically, 5 tables in the Firebase database were defined: (1) settings, (2) single shot, (3) submission reviews, (4) user labeling, and (5) user data. The settings table saves all app settings. The single-shot table saves the first image taken by each patient. The submission reviews table contains the dentist's reviews. The user labeling and user data tables contain patient information such as answers to the registration questionnaires, the number of photographs taken, the dates on which the photographs were taken, etc.

Software Tests

To test the possibilities of crashes or error cases, we tested the app on a number of Android and Apple mobile phones. We confirmed the operation of (1) system activation and loading by visualizing the home screen and the loading of links; (2) the existing user authentication; (3) the opening of the camera, the saving of images, and the uploading of the images to the database; (4) the dentist ratings using MGI, and the ability to add notes; and (5) new user additions, verifications, and unique patient identification assignment.

Agile Software Development Evaluation

The Agile software development life cycle approach was used for this research; the agile cycle involves dividing the work into small increments, allowing the product to adapt to changes quickly.

Following each photo session, several dental selfies are uploaded to the server (Firebase database). The dentist section of the app enables the dentist to view the patient's dental selfie, to rate its image quality, and to evaluate gingival health status.

Agile Software Development

After initial app development, a pilot study with a group of 18 participants aged 18-45 years (the inclusion criterion was 18 years of age and above) and with different levels of health literacy and education was conducted. The participants were given a kit containing toothpaste, a toothbrush, mouth wash, toothpicks, and dental floss to ensure they had the materials needed to improve their oral hygiene. After installing the app, the participants were asked to photograph their gums once a week for 4 weeks, and then to answer 3 open-ended questions: (1) Describe the feasibility of using the app and mouth opener; (2) Describe the usability of the photo feature; (3) Describe app acceptance over the time of the study.

Qualitative analysis of the answers involved 5 cycles. In cycle 1, we noted that some of the photographs were out of focus, so we reprogrammed the app to take 10 pictures each time to allow the researcher and dentist to analyze the best one. During cycle 2, 12 participants found that the physical positioning required to take the photograph (ie, holding the phone with one hand and aiming the camera at the gums with minimal movement while rolling down the lower lip with the other hand, as shown in Figure 24) to be clumsy and uncomfortable. Therefore, mouth openers (Figure 25) were added to the kit so that the user would not have to use their hand to retract their lip. The mouth opener was delivered to each participant, and as most participants found it difficult to use because of its rigidity, another, more comfortable mouth opener was tested and implemented (Figure 26), and the participants were satisfied.

Figure 24. Physical positioning for taking a dental selfie: without a mouth opener (suboptimal).

Figure 25. Physical positioning for taking a dental selfie: with a rigid mouth opener (suboptimal).

Figure 26. Physical positioning for taking a dental selfie: with a comfortable mouth opener (optimal).

In cycle 3, 5 participants reported difficulty with operating the phone, turning on the camera and the flash, and then pressing the photo button with one hand. Therefore, we reprogrammed the app so that a photo could be taken using the volume button, and the flash was set up to turn on automatically and to simultaneously initiate a 10-second timer, giving the participant time to position the camera. A sound indicated that the phone had started taking photographs, and another sound signaled the end of the photographing time.

During cycle 4, 13 participants reported difficulty remembering when to take another photo. Therefore, a weekly reminder feature (in the form of an SMS text message) was added.

In cycle 5, 8 participants reported difficultly placing the rear camera in the correct position. Therefore, users were instructed to stand in front of a mirror while taking the dental selfie, or to take pictures in the restroom behind a white wall, using natural light.

Discussion

Principal Findings

There are currently more than 318,000 medical apps that help diagnose and manage illness; for example, in recent years, many applications have been developed to monitor blood pressure or electrocardiography. Smartphones have very strong computing capabilities and have become involved in our daily lives [42,43]. Photography is a very valuable tool for documenting illness, supervising care, and educating patients. Most smartphones have cameras that can take and transfer high-quality dental selfies, and can also process and transmit audio and video files [44].

In this paper, we describe the characterization and development of a cellular app (iGAM) to monitor gingivitis, a prevalent condition of reversible periodontal inflammation. To the best of our knowledge, there are apps for improving oral hygiene [45,46], but no computerized platform exists to monitor gingivitis between dental appointments. This app was developed

in response to this need, and to examine whether oral health can be promoted using a mobile app and a feature we have termed a "dental selfie." In order to develop the app, we built a characterization protocol that included the following steps: (1) understanding user needs and perceptions toward the technology; (2) defining user identification and registration; (3) designing the UI/UX of the demo app screens; (4) creating user-administrator relationship chart and data retention systems; (5) connecting screens and user traffic on the app; (6) connecting screens and activities of the administrator.

As no apps of this nature are currently available, we went to dental clinics and observed dentist-patient interactions. Digital technology is widely used in dental clinics, and there is evidence that dentists appreciate the benefits of digital tools in their practices [47]. These tools enhance communication with peers and are perceived as useful in improving patient-dentist communication, management, and patient satisfaction [48,49]. After visiting the clinics, we conducted a focus group accompanied by semistructured, qualitative, in-depth interviews. Then we developed our first version of the app and tested it on 18 users. This pilot identified problems and enabled us to make the app more user-friendly. The app has 3 main characteristics: (1) It collects general information and information related to health behavior, with an emphasis on oral health; (2) it offers information about proper oral hygiene habits and the dangers of oral diseases, as well as visual tutorials demonstrating proper tooth brushing techniques; (3) dental selfie photographs are taken by the user and are evaluated by a dentist using the MGI.

The iGAM app was developed using the Agile approach, and the pilot study played an integral role in improving the features that users need. A randomized clinical trial in a large population and a qualitative study will follow and be reported separately.

Limitations

The inclusion criterion for the pilot study was individuals aged 18 years and older; however, the ages of the volunteers all fell into the range of 18-45 years. This type of app may not be applicable for older people or those with poor literacy. This study included a diverse set of research methods and was developed in Hebrew because of the limits of self-funding. We recognize that the cultural diversity of individuals speaking other languages may correspond with different needs that the app will need to meet. Future research may address these limitations.

Conclusions

The iGAM app is the first mHealth app for monitoring gingivitis to promote oral health using self-photography. iGAM presents a novel solution to improve the flow of information between dentists and their patients during the intervals between checkups. This app has tremendous potential for situations in which patients cannot meet their dentists in-person, such as during the COVID-19 pandemic. We believe that the collaboration between dentists, expert app developers, and typical dental patients—our target users—allowed us to make a quality and reliable app which will be updated regularly and improved.

Conflicts of Interest

None declared.

References

1. Pohjola V, Lahti S, Vehkalahti MM, Tolvanen M, Hausen H. Association between dental fear and dental attendance among adults in Finland. Acta Odontol Scand 2007 Aug;65(4):224-230. [doi: 10.1080/00016350701373558] [Medline: 17762985]

2. Armfield JM, Slade GD, Spencer AJ. Dental fear and adult oral health in Australia. Community Dent Oral Epidemiol 2009 Jun;37(3):220-230. [doi: 10.1111/j.1600-0528.2009.00468.x] [Medline: 19508270]

3. Berggren U, Meynert G. Dental fear and avoidance: causes, symptoms, and consequences. J Am Dent Assoc 1984 Aug;109(2):247-251. [doi: 10.14219/jada.archive.1984.0328] [Medline: 6590605]

4. Karibe H, Koeda M, Aoyagi-Naka K, Kato Y, Tateno A, Suzuki H, et al. Differences in the perception of dental sounds: a preliminary study. Patient Prefer Adherence 2019;13:1051-1056 [FREE Full text] [doi: 10.2147/PPA.S204465] [Medline: 31308637]

5. Coulthard P. Dentistry and coronavirus (COVID-19) - moral decision-making. Br Dent J 2020 Apr;228(7):503-505. [doi: 10.1038/s41415-020-1482-1] [Medline: 32277203]

6. https://www.health.gov.il/. 2020. URL: https://www.health.gov.il/English/Pages/HomePage.aspx [accessed 2020-03-27]

7. Zou J, Meng M, Law CS, Rao Y, Zhou X. Common dental diseases in children and malocclusion. Int J Oral Sci 2018 Mar 13;10(1):7 [FREE Full text] [doi: 10.1038/s41368-018-0012-3] [Medline: 29540669]

8. No authors listed. Periodontal Diseases of Children and Adolescents. Pediatr Dent 2017 Sep 15;39(6):431-439. [Medline: 29179386]

9. Hancock EB. Periodontal diseases: prevention. Ann Periodontol 1996 Nov;1(1):223-249. [doi: 10.1902/annals.1996.1.1.223] [Medline: 9118259]

10. Papapanou PN. Risk assessments in the diagnosis and treatment of periodontal diseases. J Dent Educ 1998 Oct;62(10):822-839. [Medline: 9847886]

11. No authors listed. Treatment of Plaque-induced Gingivitis, Chronic Periodontitis, and Other Clinical Conditions. Pediatr Dent 2017 Sep 15;39(6):445-454. [Medline: 29179388]

12. Page RC. Gingivitis. J Clin Periodontol 1986 May;13(5):345-359. [doi: 10.1111/j.1600-051x.1986.tb01471.x] [Medline: 3522644]

13. Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. J Periodontol 2018 Jun;89 Suppl 1:S159-S172. [doi: 10.1002/JPER.18-0006] [Medline: 29926952]

14. Michaud DS, Fu Z, Shi J, Chung M. Periodontal Disease, Tooth Loss, and Cancer Risk. Epidemiol Rev 2017 Jan 01;39(1):49-58 [FREE Full text] [doi: 10.1093/epirev/mxx006] [Medline: 28449041]

15. Chapple ILC, Van der Weijden F, Doerfer C, Herrera D, Shapira L, Polak D, et al. Primary prevention of periodontitis: managing gingivitis. J Clin Periodontol 2015 Apr;42 Suppl 16:S71-S76. [doi: 10.1111/jcpe.12366] [Medline: 25639826]

16. Jin LJ, Lamster IB, Greenspan JS, Pitts NB, Scully C, Warnakulasuriya S. Global burden of oral diseases: emerging concepts, management and interplay with systemic health. Oral Dis 2016 Oct;22(7):609-619. [doi: 10.1111/odi.12428] [Medline: 26704694]

17. Marcenes W, Kassebaum NJ, Bernabé E, Flaxman A, Naghavi M, Lopez A, et al. Global burden of oral conditions in 1990-2010: a systematic analysis. J Dent Res 2013 Jul;92(7):592-597 [FREE Full text] [doi: 10.1177/0022034513490168] [Medline: 23720570]

18. Caton J, Bouwsma O, Polson A, Espeland M. Effects of personal oral hygiene and subgingival scaling on bleeding interdental gingiva. J Periodontol 1989 Feb;60(2):84-90. [doi: 10.1902/jop.1989.60.2.84] [Medline: 2786069]

19. Kumar S. Evidence-Based Update on Diagnosis and Management of Gingivitis and Periodontitis. Dent Clin North Am 2019 Jan;63(1):69-81. [doi: 10.1016/j.cden.2018.08.005] [Medline: 30447793]

20. Kinane DF, Stathopoulou PG, Papapanou PN. Periodontal diseases. Nat Rev Dis Primers 2017 Jun 22;3:17038. [doi: 10.1038/nrdp.2017.38] [Medline: 28805207]

21. Clarkson JE, Pitts NB, Bonetti D, Boyers D, Braid H, Elford R, INTERVAL Trial Collaboration. INTERVAL (investigation of NICE technologies for enabling risk-variable-adjusted-length) dental recalls trial: a multicentre randomised controlled trial investigating the best dental recall interval for optimum, cost-effective maintenance of oral health in dentate adults attending dental primary care. BMC Oral Health 2018 Aug 07;18(1):135 [FREE Full text] [doi: 10.1186/s12903-018-0587-2] [Medline: 30086747]

22. Wang NJ, Petersen PE, Sveinsdóttir EG, Arnadóttir IB, Källestål C. Recall intervals and time used for examination and prevention by dentists in child dental care in Denmark, Iceland, Norway and Sweden in 1996 and 2014. Community Dent Health 2018 Mar 01;35(1):52-57. [doi: 10.1922/CDH_4186Wang06] [Medline: 29369547]

23. Shahrabani S. Factors affecting oral examinations and dental treatments among older adults in Israel. Isr J Health Policy Res 2019 May 10;8(1):43 [FREE Full text] [doi: 10.1186/s13584-019-0312-x] [Medline: 31077257]

24. Shahrabani S, Benzion U, Machnes Y, Gal A. The use of dental services for children: implications of the 2010 dental reform in Israel. Health Policy 2015 Feb;119(2):117-126. [doi: 10.1016/j.healthpol.2014.11.007] [Medline: 25465981]

25. pewresearch.org. 2020. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-03-27]

26. Panova T, Carbonell X. Is smartphone addiction really an addiction? J Behav Addict 2018 Jun 01;7(2):252-259 [FREE Full text] [doi: 10.1556/2006.7.2018.49] [Medline: 29895183]

27. Kaur N, Gupta M, Malhi P, Grover S. Screen Time in Under-five Children. Indian Pediatr 2019 Sep 15;56(9):773-788 [FREE Full text] [Medline: 31638012]

28. Statista. URL: https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/ [accessed 2020-03-27]

29. Statista. URL: https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/ [accessed 2020-03-27]

30. Goyal S, Morita P, Lewis GF, Yu C, Seto E, Cafazzo JA. The Systematic Design of a Behavioural Mobile Health Application for the Self-Management of Type 2 Diabetes. Can J Diabetes 2016 Feb;40(1):95-104. [doi: 10.1016/j.jcjd.2015.06.007] [Medline: 26455762]

31. medisana GmbH. URL: https://www.medisana.com/en/VitaDock-App-2-0.html [accessed 2020-03-27]

32. Roche Diabetes Care Inc. URL: http://www.pocket.md/1243/gluci-chek.html [accessed 2020-03-27]

33. https://runkeeper.com/cms/. URL: https://runkeeper.com/running-app [accessed 2020-03-27]

34. Good Housekeeping. URL: https://www.goodhousekeeping.com/health-products/g28246667/best-calorie-counting-apps/ [accessed 2020-03-27]

35. https://glowing.com/. URL: https://apps.apple.com/us/app/glow-period-fertility-tracker/id638021335 [accessed 2020-03-27]

36. Leddy J, Green JA, Yule C, Molecavage J, Coresh J, Chang AR. Improving proteinuria screening with mailed smartphone urinalysis testing in previously unscreened patients with hypertension: a randomized controlled trial. BMC Nephrol 2019 Apr 18;20(1):132 [FREE Full text] [doi: 10.1186/s12882-019-1324-z] [Medline: 30999886]

37. http://webstore.lexi.com/. URL: http://webstore.lexi.com/Lexi-DENTAL-COMPLETE [accessed 2020-03-16]

38. https://play.google.com/. 2016. URL: https://play.google.com/store/apps/details?id=com.igam.shira.myapplication&hl=en_US [accessed 2020-02-20]

39. Highsmith, James A. and Jim Highsmith. In: Agile software development ecosystems. Addison-Wesley Professional. Boston, USA: Addison-Wesley Professional; 2002.

40. Riley P, Lamont T. Triclosan/copolymer containing toothpastes for oral health. Cochrane Database Syst Rev 2013 Dec 05(12):CD010514 [FREE Full text] [doi: 10.1002/14651858.CD010514.pub2] [Medline: 24310847]

41. He T, Qu L, Chang J, Wang J. Gingivitis Models - Relevant Approaches to Assess Oral Hygiene Products. J Clin Dent 2018 Jun;29(2):45-51. [Medline: 30211990]

42. Martinez-Hurtado J, Yetisen A, Yun S. Multiplex Smartphone Diagnostics. Methods Mol Biol 2017;1546:302. [doi: 10.1007/978-1-4939-6730-8_26] [Medline: 27896779]

43. De Ridder B, Van Rompaey B, Kampen JK, Haine S, Dilles T. Smartphone Apps Using Photoplethysmography for Heart Rate Monitoring: Meta-Analysis. JMIR Cardio 2018 Feb 27;2(1):e4 [FREE Full text] [doi: 10.2196/cardio.8802] [Medline: 31758768]

44. Maurrasse SE, Schwanke TW, Tabaee A. Smartphone capture of flexible laryngoscopy: Optics, subsite visualization, and patient satisfaction. Laryngoscope 2019 Sep;129(9):2147-2152. [doi: 10.1002/lary.27803] [Medline: 30628089]

45. Lozoya CJS, Giblin-Scanlon L, Boyd LD, Nolen S, Vineyard J. Influence of a Smartphone Application on the Oral Health Practices and Behaviors of Parents of Preschool Children. J Dent Hyg 2019 Oct;93(5):6-14. [Medline: 31628171]

46. Kay E, Shou L. A randomised controlled trial of a smartphone application for improving oral hygiene. Br Dent J 2019 Apr;226(7):508-511. [doi: 10.1038/s41415-019-0202-1] [Medline: 30980006]

47. Jacox LA, Mihas P, Cho C, Lin F, Ko C. Understanding technology adoption by orthodontists: A qualitative study. Am J Orthod Dentofacial Orthop 2019 Mar;155(3):432-442. [doi: 10.1016/j.ajodo.2018.08.018] [Medline: 30826046]

48. van der Zande MM, Gorter RC, Aartman IHA, Wismeijer D. Adoption and use of digital technologies among general dental practitioners in the Netherlands. PLoS One 2015;10(3):e0120725 [FREE Full text] [doi: 10.1371/journal.pone.0120725] [Medline: 25811594]

49. Estai M, Kruger E, Tennant M. Perceptions of Australian dental practitioners about using telemedicine in dental practice. Br Dent J 2016 Jan 15;220(1):25-29. [doi: 10.1038/sj.bdj.2016.25] [Medline: 26768465]

Abbreviations

COVID-19: coronavirus disease 2019
CSS: cascading style sheets
eHealth: electronic health
MGI: modified gingival index
mHealth: mobile health
TTS: text to speech
UI: user interface
UX: user experience

XSL•FO
RenderX

Original Paper

Using Natural Language Processing and Sentiment Analysis to Augment Traditional User-Centered Design: Development and Usability Study

Curtis Lee Petersen[1,2], MSc, MPH; Ryan Halter[3], PhD; David Kotz[4], PhD; Lorie Loeb[4], MFA; Summer Cook[5], PhD; Dawna Pidgeon[6], PT; Brock C Christensen[2,7,8], PhD; John A Batsis[9,10,11], MD

[1]The Dartmouth Institute for Health Policy, Dartmouth, Hanover, NH, United States

[2]Quantitative Biomedical Sciences Program, Dartmouth, Hanover, NH, United States

[3]Thayer School of Engineering, Dartmouth, Hanover, NH, United States

[4]Computer Science, Dartmouth, Hanover, NH, United States

[5]Department of Kinesiology, University of New Hampshire, Durham, NH, United States

[6]Department of Physical Medicine and Rehabilitation, Dartmouth-Hitchcock Medical Center, Lebanon, NH, United States

[7]Department of Epidemiology, Dartmouth, Hanover, NH, United States

[8]Department of Molecular and Systems Biology at Dartmouth, Dartmouth, Hanover, NH, United States

[9]The Dartmouth Institute for Health Policy, Dartmouth, Lebanon, NH, United States

[10]Department of Medicine, Geisel School of Medicine, Dartmouth, Hanover, NH, United States

[11]Section of General Internal Medicine, Dartmouth-Hitchcock, Lebanon, NH, United States

Corresponding Author:
John A Batsis, MD
The Dartmouth Institute for Health Policy
Dartmouth
1 Medical Drive
Williamson Translational Research Building, Level 5
Lebanon, NH, 03756
United States
Phone: 1 603 653 9500
Email: john.batsis@gmail.com

Abstract

Background: Sarcopenia, defined as the age-associated loss of muscle mass and strength, can be effectively mitigated through resistance-based physical activity. With compliance at approximately 40% for home-based exercise prescriptions, implementing a remote sensing system would help patients and clinicians to better understand treatment progress and increase compliance. The inclusion of end users in the development of mobile apps for remote-sensing systems can ensure that they are both user friendly and facilitate compliance. With advancements in natural language processing (NLP), there is potential for these methods to be used with data collected through the user-centered design process.

Objective: This study aims to develop a mobile app for a novel device through a user-centered design process with both older adults and clinicians while exploring whether data collected through this process can be used in NLP and sentiment analysis

Methods: Through a user-centered design process, we conducted semistructured interviews during the development of a geriatric-friendly Bluetooth-connected resistance exercise band app. We interviewed patients and clinicians at weeks 0, 5, and 10 of the app development. Each semistructured interview consisted of heuristic evaluations, cognitive walkthroughs, and observations. We used the Bing sentiment library for a sentiment analysis of interview transcripts and then applied NLP-based latent Dirichlet allocation (LDA) topic modeling to identify differences and similarities in patient and clinician participant interviews. Sentiment was defined as the sum of positive and negative words (each word with a +1 or −1 value). To assess utility, we used quantitative assessment questionnaires—System Usability Scale (SUS) and Usefulness, Satisfaction, and Ease of use (USE). Finally, we used multivariate linear models—adjusting for age, sex, subject group (clinician vs patient), and development—to explore the association between sentiment analysis and SUS and USE outcomes.

Results: The mean age of the 22 participants was 68 (SD 14) years, and 17 (77%) were female. The overall mean SUS and USE scores were 66.4 (SD 13.6) and 41.3 (SD 15.2), respectively. Both patients and clinicians provided valuable insights into the needs of older adults when designing and building an app. The mean positive-negative sentiment per sentence was 0.19 (SD 0.21) and 0.47 (SD 0.21) for patient and clinician interviews, respectively. We found a positive association with positive sentiment in an interview and SUS score (ß=1.38; 95% CI 0.37 to 2.39; *P*=.01). There was no significant association between sentiment and the USE score. The LDA analysis found no overlap between patients and clinicians in the 8 identified topics.

Conclusions: Involving patients and clinicians allowed us to design and build an app that is user friendly for older adults while supporting compliance. This is the first analysis using NLP and usability questionnaires in the quantification of user-centered design of technology for older adults.

Introduction

Sarcopenia

Sarcopenia is the loss of muscle mass, strength, and function, which occurs with aging and is associated with serious health consequences such as disability [1], morbidity [2], and mortality [3]. Although the etiology is complex, mitigating its development is important in preserving long-term muscle function. Recommended treatments include resistance exercise programs [4] and protein supplementation, such as amino acids including creatine [5] or leucine [6], and vitamin D [7]. Strengthening exercises normally prescribed by physical therapists are cost-effective, safe, improve physical functioning [8], and prevent muscle loss [9]. Exercises to enhance muscle mass and strength conducted 2 to 3 times per week normally consist of using resistance exercise bands or weights—materials that are easily available and can be used in medical, community, or home-based settings.

Building Usable Solutions

Although the evidence base is clear for the efficacy of such treatment strategies, an estimated 40% of patients fail to adhere to their recommended regimens [10,11], making it difficult for clinicians to understand a patient's progression through therapy. Observation of activity provides a gold standard for therapists to tailor, evaluate, and encourage treatment regimens. However, direct observation is not practical or feasible in health systems and specifically for older adults facing barriers of travel and transportation to in-person visits [12]. Remote medical sensing and mobile health (mHealth) technologies have the potential to help patients and providers understand and track adherence and progression through therapy, and overcome some of the major barriers to attending in-person sessions. Technology has been used to track and communicate with patients with chronic diseases [13] and has demonstrated the ability to improve compliance with treatments [14].

As older adults have specific sensory needs and different perceptions of mHealth [15], it is important to employ user-centered design methods to ensure that the final device and app meet the needs and preferences of older adults [16]. A *user-centered design* incorporates the end user of a product in all phases of the design, ensuring that the result aligns with the users' needs. Qualitative methods of user-centered design incorporate interviews and the assessment of constructs identified through these conversations. Such methods are labor intensive and require specific training. Some of these barriers can be surmounted using quantitative techniques such as *natural language processing (NLP)*, which requires relatively little computational resources and leverages existing workflows and software pipelines. NLP has many applications, such as information retrieval; in the medical field, it is increasingly used to extract topics in electronic health record data [17]. In this context, *sentiment analysis* has been used to examine the perception of health care, drugs, treatment, or illnesses using social media data [18], although it has started to be used in a broad range of applications, from analyzing investor earnings calls [19] to interactions with chatbots [20]. These methods could be used beyond determining if people enjoy an investor call, health care system, or chatbot to assess how a person perceives a health product in development before a device or program has been finalized. Using them in the design and development of mHealth products for specific patient populations could lead to more rapid and accurate determination of how they feel about an mHealth product and how it could be improved without the burden of questionnaires.

We previously developed a Bluetooth-connected resistance exercise band that had the potential to provide feedback to both patients and providers on exercise compliance and treatment progress [21,22]. The addition of a mobile app permits real-time monitoring and has the ability to use cloud-computing resources to provide feedback, force, and detection of exercise repetitions to clinical or research teams. For patients with sarcopenia, connecting a resistance band to an app provides a platform for them to understand their progress through each exercise as they proceed through the regimen. With a suitable *dashboard*, the app allows clinicians to monitor not only compliance for each patient but also their entire patient population. As much of the user-centered design process is dependent on both interviewing users to assess their perception and quantifying it through questionnaires, we also sought to determine if the data generated through these conventional methods could also be analyzed through NLP and sentiment quantification methods. This exploratory work leverages the interviews and conversations that were collected during user development, quantifying them using new methods. A preliminary examination of correlation with questionnaires can also help shed light on how they may

be interpreted in the future. To our knowledge, this is the first study to apply NLP and sentiment analysis to interviews in a usability study for older adults.

Objectives

This study aims to create a mobile app for older adults to monitor their use of a Bluetooth-connected resistance band and to examine whether data collected through this process could be used for NLP and sentiment analysis.

Methods

Study Population

Participants were recruited through a primary care clinic at Dartmouth-Hitchcock in Lebanon, New Hampshire, a rural health care institution caring for 1.5 million patients in New Hampshire and Vermont. Word-of-mouth and study posters provided the main source of referrals. All study activities were conducted at the community-based Dartmouth-Hitchcock Aging Resource Center. Eligible participants were English-speaking, community-dwelling (eg, not residing in a nursing home or assisted living center) older adults aged 65 years or older without a self-reported diagnosis of dementia. There were no other specific inclusion or exclusion criteria. Clinicians were faculty members of the Section of General Internal Medicine at Dartmouth-Hitchcock. All participants were provided a research information sheet before the start of the study and were recruited as a convenience sample from Dartmouth-Hitchcock. The study was approved by the Committee for the Protection of Human Subjects at Dartmouth College and the Dartmouth-Hitchcock Institutional Review Board.

App and User-Centered Design Process

We built an app for Android, optimized for the Samsung Galaxy Tab A tablet. We chose this particular tablet for its large screen, Android operating system, and relatively low price; all factors were previously identified by patients. The app, written in JavaScript, allowed the user to sign up, connect to a Bluetooth-enabled resistance band, watch a video of an exercise, and record the data from the resistance band. Each exercise video consisted of an individual completing a specific exercise while verbally explaining it. The first exercise video was created without input from patients to provide an initial example of an exercise video.

The user-centered design process consisted of 3 rounds. Interviews with individual patients and clinicians were conducted in all rounds of the study by the same 1 or 2 research assistants. Participants were asked open-ended questions about their use of technology, their preferences on iterations of the app, and their reactions to design images and content, all of

which were documented by the interviewers as field notes. Field notes consisted of observations of the user discussing and using the app. After each round, the developers, designers, research assistants, and researchers would meet to review interview notes, discuss interviewees' perceptions, and determine issues discovered in the round that should be addressed. The team then determined potential solutions, deciding which were most feasible, and updated the design and content. Individuals only participated in a single round of development. We recorded the interviews and later used a commercial transcription service to produce a transcript for each interview.

Round 1 (predevelopment) consisted of 6 older adults. We evaluated their general perception of mHealth needs, physical activity, the Bluetooth-connected resistance exercise band, and the initial exercise video. As this round occurred before app development, the study team did not ask participants for feedback on a specific app; as such, they were not given System Usability Scale (SUS) and Usefulness, Satisfaction, and Ease of use (USE) questionnaires. Interviewees were asked about their exercise habits, their technology habits, their use of technology with exercise, and how they think they may use technology with exercise. Information from round 1 allowed the app developers to construct a prototype app for round 2 using initial designs with black and white mockups of the app.

During round 2 individual interviews, the team presented the updated exercise video to 3 clinicians and 4 patient participants and asked them to provide oral feedback. We then showed black and white wireframes of the app to the participants (Figure 1). Two versions of a weekly progress summary screen were shown in random order. Version 1 had toggles to switch between viewable summary data with a back arrow in the upper left-hand corner and vertical bar chart of repetition counts. Version 2 had buttons to switch between viewable summary data with a back button at the bottom of the workout summary screen and a horizontal bar chart of repetition counts (Figure 2). We also showed participants 2 approaches for displaying the number of completed repetitions in a workout: a vertical bar chart and a horizontal bar chart (Figure 2). Interviewers used think-aloud [23] and verbal prompting methods to encourage interviewees to share their thoughts about the function of each button and chart and their designs.

Round 3 included 3 clinicians and 6 patients. This round consisted of asking participants to start an exercise video and navigate through an exercise summary of videos in the prototype app. Each exercise video consisted of a person performing the exercise while describing how to position and move their bodies to complete the exercise successfully. Interviewers again used think-aloud and verbal prompting methods when asking the interviewees to explore the app and start a workout (Figure 3).

Figure 1. App versions for round 2 (round 1 did not use wireframes) consisted of black and white wireframes of the exercise selection screen and a preworkout screen.
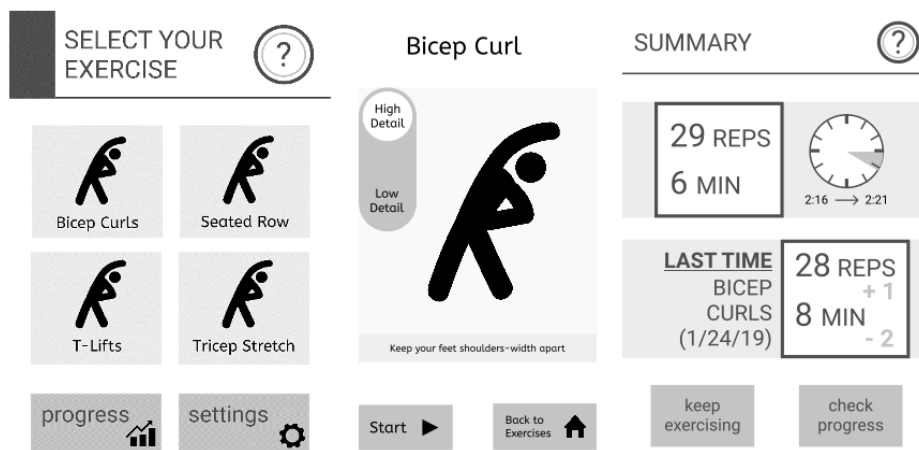
Figure 2. The App version for round 2 also had two different versions of the weekly progress screen shown to participants in random order.

Figure 3. The app versions for round 3 screens were in color, had the embedded video with a low-contrast background. Here, we present the home screen, the preworkout screen, the workout screen, the Borg Scale of Perceived Exertion questionnaire, and the progress screen.
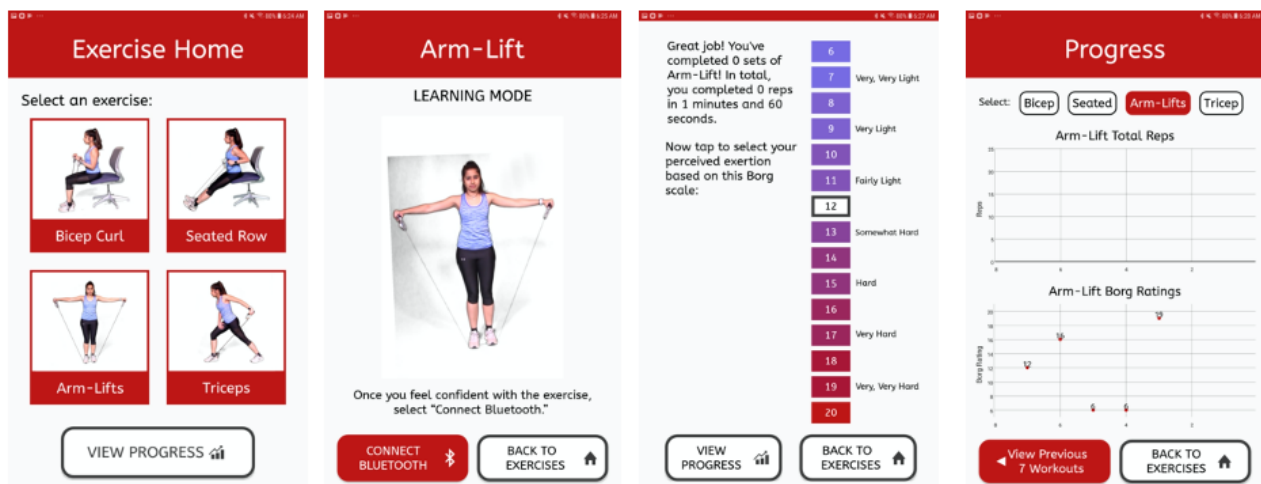
Usability and Acceptability Questionnaires

After each round 2 and round 3 interview, the study team asked participants to complete 2 validated questionnaires: the SUS [24] and the USE scale [25,26]. The SUS was developed for the global assessment of a system's usability and consisted of 10 questions, each on a 0- to 4-point Likert scale, with an overall score range of 0 to 100, following a scoring algorithm. The USE scale was developed for broad application in technology development in private industry and comprises 30 questions, each on a 1- to 7-point Likert scale and is broken down into 4 sections (usefulness, ease of use, ease of learning, and satisfaction). Each interviewee filled out both questionnaires after the interview session.

NLP and Sentiment Analysis

The study team subsequently used NLP methods to process the deidentified interview transcripts after all the rounds and interviews were completed. We separated utterances by speaker, removing interviewer data. We removed stop words and calculated the term frequency-inverse document frequency at the interviewee level. We then used latent Dirichlet allocation (LDA) [27] topic modeling to examine the content of the core concepts discussed in each interview. An LDA analysis is a generative probabilistic model that groups words based on how related they are—these clusters of related words are termed *topics*. LDA assumes that text is a mixture of topics that have a probability of occurring in each sentence and interview. LDA methods have been described in depth in several recent medical informatic manuscripts [28-30]. To examine the feature space of topic clusters from 2 to 60 using 4 different metrics [31-35] to calculate ideal cluster numbers, we used the mean number of ideal clusters (k=8) in the LDA analysis. We reviewed the 15 words most associated with each topic cluster and named them based on the concept that the authors believed they represented most accurately. This iterative process involved discussing how words could be used differently for an underlying concept or topic. Names were then decided on by consensus.

We used the Bing sentiment lexicon to define positive and negative words in our sentiment analysis [36]. The positive and negative sentiment words within each interview were then added, where positive sentiments had positive values (+1) and negative sentiments had negative values (−1). All other words had no value (0). The final sentiment value consisted of all the mean sentiments; thus, greater positive numbers indicate greater positive sentiment, whereas lower negative numbers indicate greater negative sentiment.

Statistical Analysis

We presented descriptive characteristics as mean (SD) and count (percentage). We determined differences in questionnaire values using the student *t* test. We used univariate and multivariate linear regression models to test the association between the total sentiment of the interview per individual (covariate of interest) and usability and acceptability questionnaire scores. Outcomes were the SUS score and each component of the USE questionnaire. All qualitative interview data were transcribed, managed, and coded in *Dedoose*, completed after categorizing data excerpts at each stage. Field notes were also obtained and aggregated. All codes were reviewed, and we identified positive or negative themes among the participants and juxtaposed them with the sentiment analysis. Multivariate linear models included age, sex, and subject group (clinicians vs patients). We conducted all analyses in R version 3.6.0 [37], and significance was defined as <.05.

Results

Study Population

We recruited 22 participants (6 clinicians and 16 patients) to review the app design and functionality. The age of the clinicians was significantly lower than that of patients with a mean age of 49 (SD 9) years; the mean age of the patients was 76 (SD 5) years (*P*<.001). The majority of the participants were female (17/22, 77%), and all were of non-Hispanic white ethnicity (Table 1).

Table 1. Study participant characteristics (N=22).

Participants	Clinicians (n=6)	Patients (n=16)
Age (years)		
Mean (SD)	49.0 (9.4)	75.56 (5.2)
Minimum to maximum	37 to 51	66 to 85
Male, n (%)	2 (33.3)	3 (18.8)
Non-Hispanic White, n (%)	6 (100)	16 (100)

Mobile App Design and Development

In round 1 (predevelopment), 6 patient participants were asked about their exercise behavior, how they used mobile devices, and their perception of the initial exercise video. All participants had experience with smartphones and computers and expressed that they were interested in using technology in a physical therapy program. Participants found it difficult to follow the exercise video with a low-contrast background and without each specific movement shown in close-up detail. They also found it difficult to hear the audio of the video or to distinguish the words being said. Participants expressed that they needed help in counting repetitions, knowing the exercise that they needed to do, and having clearly labeled buttons, as older adults may not be able to assume functionality as well as younger people. This led to designs that contained labeled buttons, such as back and home, along with some repetition counting functionality. Patients also hoped that technology can help provide feedback and guidance and thus impart confidence in completing exercises while at home. Finally, patients preferred the use of tablets over

XSL•FO
RenderX

phones because of their larger screen size as such designs were based on tablet screen layouts and not smartphones.

Round 2 focused on basic functionality using black and white wireframes. In total, 4 patients and 3 clinicians indicated that a vertical stacked bar chart was more interpretable than horizontal bar charts when reviewing the progress of previous exercise data. Patients found it difficult to follow the new video and considered the video instructions to be too fast and not detailed enough to understand the physical positioning required to complete the exercise. In addition, patients wanted to have large font text instructions along with the option to change the audio frequency of the video to help those with hearing impairment who may have difficulty hearing higher pitches. Clinicians had no issues with the videos and commented more about the content of the text on the screen. Specifically, they suggested changing the word associated with the videos from *recording* to *video* as one might think it will take a recording instead of playing one. Clinicians also suggested that we added the Borg Scale of Perceived Exertion as a measure of relative difficulty.

Round 3 involved 3 clinicians and 6 patient participants. Patients and clinicians reviewed the final videos and were asked to navigate through the app on the tablet in independent interviews. Some indicated that it would be helpful to include a separate video that taught the user how to complete the exercise and described in detail how the exercise was executed. This could be used along with the normal video that went through the exercise in real time to guide the user through it. Others indicated that some patients may have trouble with the technology and recommended that the app provide a means for the patient to call and talk with a live person.

Usability and Acceptability Questionnaires

All recruited clinicians and patients completed the usability and acceptability questionnaires. There was no statistically significant difference between patients and clinicians in their perceived use of, or satisfaction with, the app measured through either the SUS (mean 66.8, SD 16.5 vs mean 65.8, SD 7.7; $P=.90$) or the USE questionnaires (usefulness: mean 37.2, SD 18.0 vs mean 48.0, SD 5.2; $P=.18$; ease of use: mean 54.1, SD 16.4 vs mean 66.0, SD 9.1; $P=.13$; ease of learning: mean 22.4, SD 6.3 vs mean 27.5, SD 0.8; $P=.07$; and satisfaction: mean 33.3, SD 14.1 vs mean 44.3, SD 3.6; $P=.08$).

NLP and Sentiment Analysis

LDA identified 8 core topics from the transcribed interview corpus. The distribution of these topics was different between clinicians and patients (Figure 4). Clinicians' core topics included workouts over time, language of instructions, interaction with the app, and feature enhancement, whereas the core patient topics were improving fitness, help completing a workout, exercising with technology, and difficulty using the app. The total mean sentiment of each round increased from –0.17 (SD 15.92) in the predevelopment round 1 to 1.57 (SD 10.20) in round 2 and to 6.00 (SD 3.85) in round 3. Across all interviews, the maximum sentiment for a single statement was 6, whereas the minimum was –5 (Figure 5). The overall mean sentiment per phrase was 0.47 (SD 0.21) and 0.19 (SD 0.21) for clinician and participant interviews, respectively. When adjusting for age, sex, and subject group, the total sentiment value was significantly associated with the SUS, with each additional net positive statement associated with a 1.36 times increase in the SUS score ($P=.01$; Table 2). These results were not observed in any of the components of the USE questionnaire when adjusted for the same covariates (not shown). Table 3 presents representative statements that our sentiment analysis found positive and negative for both patients and clinicians.

Figure 4. LDA identified interview topics for clinicians and patients from transcripts. Each topic was named by examining the words within each topic and qualitatively determined the concept that summarized them. Here we display the proportion of the interview that was identified as each topic for both clinicians and patients. LDA: latent Dirichlet allocation.
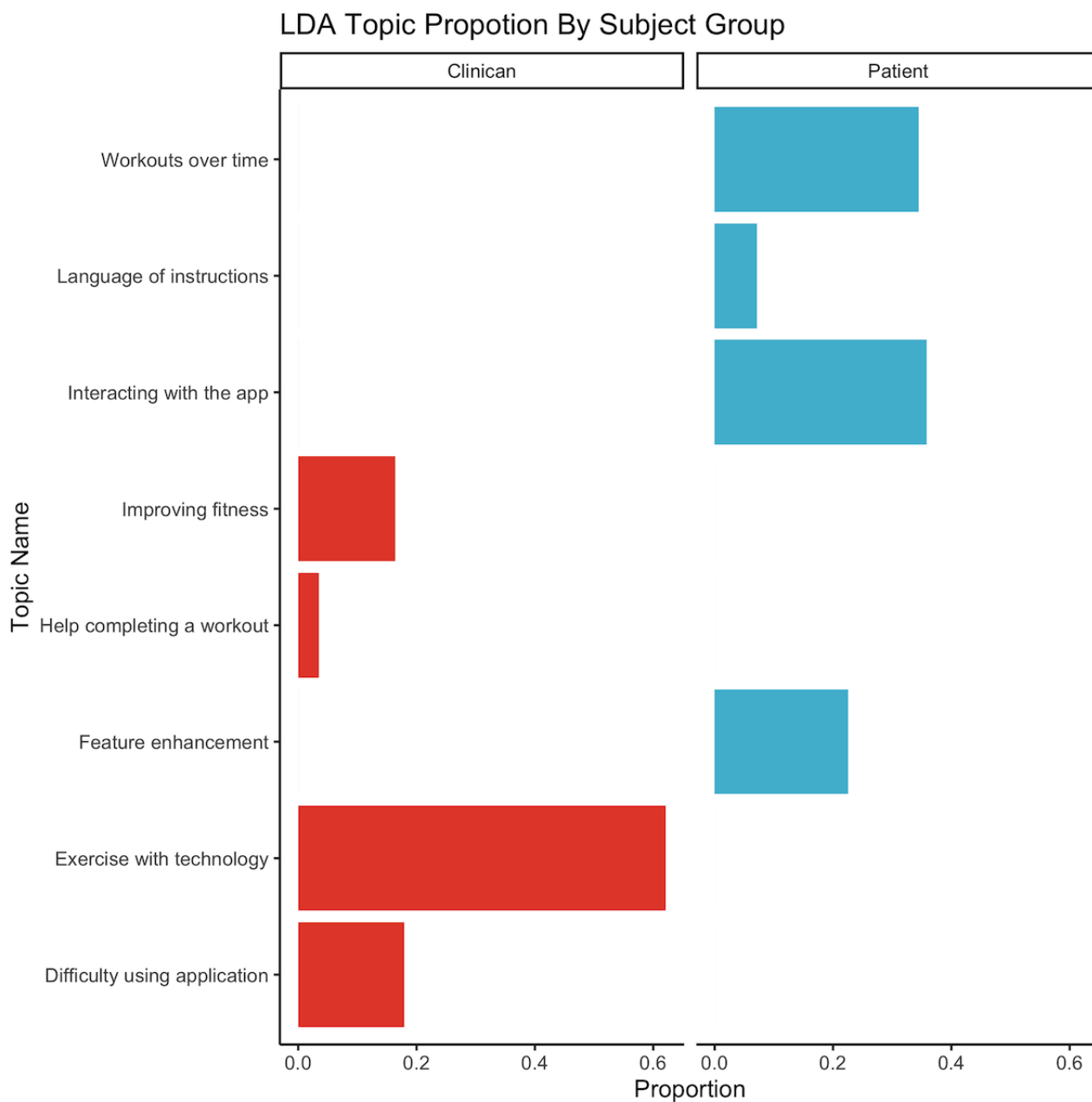
Figure 5. Sentiment over the course of the interview for clinicians and patients in each round of interviews. Negative sentiment values represent negative sentiments, while positive sentiment values represent positive sentiments.
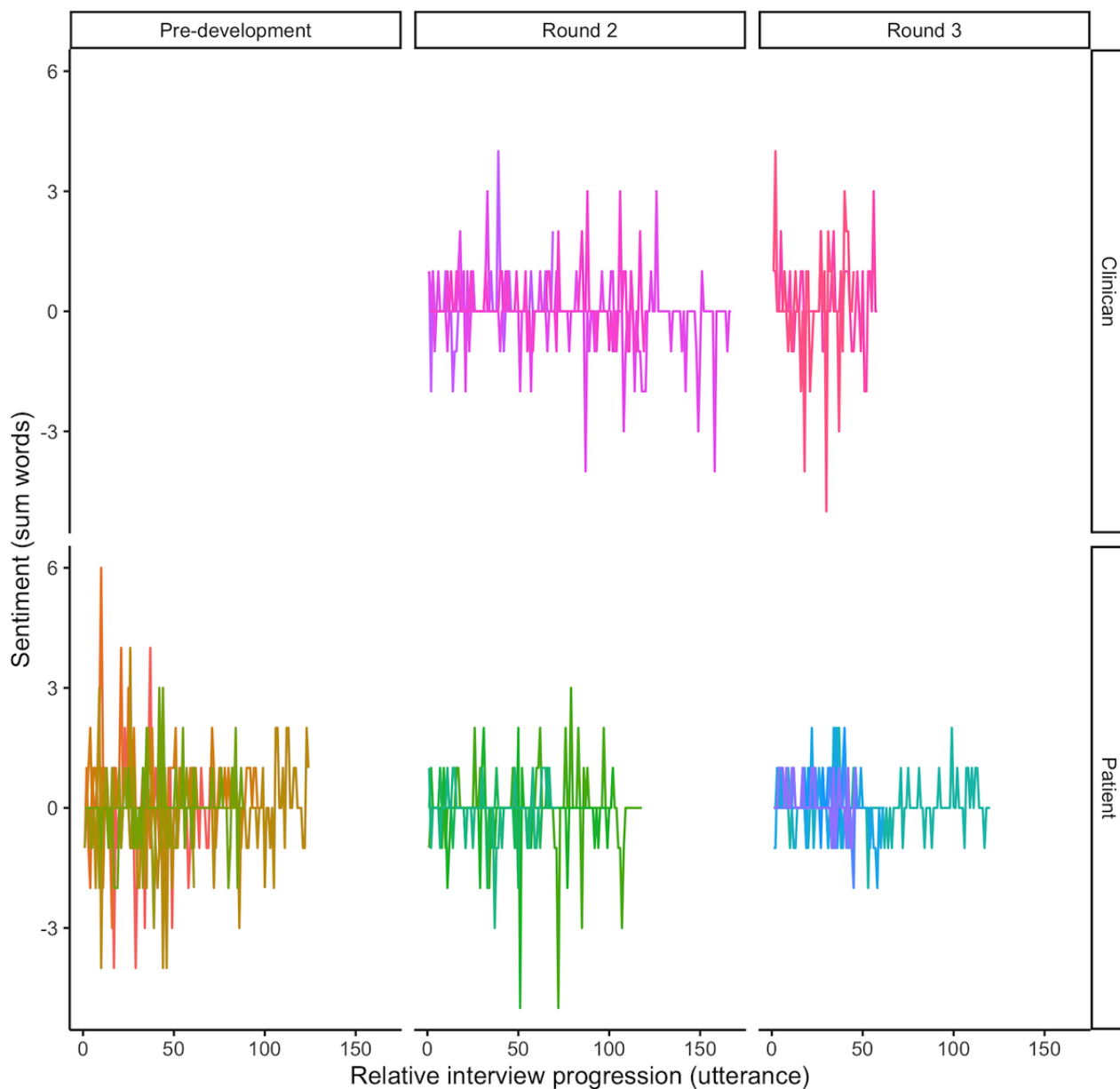
Table 2. Univariate and adjusted model for the System Usability Scale score with sentiment.

Characteristics	Univariate beta (95% CI)	P value	Adjusted beta (95% CI)	P value
Sentiment	1.11 (0.26 to 1.97)	.01	1.38 (0.37 to 2.39)	.01
Age (years)	0.03 (−0.51 to 0.58)	.89	0.59 (−0.72 to 1.96)	.25
Sex				
Female	Reference	N/A[a]	N/A	N/A
Male	−3.81 (−21.58 to 13.97)	.65	−2.59 (−18.45 to 13.26)	.59
Subject group				
Clinicians	Reference	N/A	N/A	N/A
Patients	−1.25 (−18.04 to 15.54)	.87	−8.48 (−42.74 to 25.89)	.72

[a]N/A: not applicable.

Table 3. Representative statements by patients and clinicians that were found to be positive or negative by sentiment analysis.

Participants, Statement types	Representative statements
Patient	
Positive	• "Oh, that's wonderful for the upper arms, in and out. That's wonderful. Yeah. I like that one."
	• "Honestly, it makes sense to be at the bottom. But I think most programs are at the top. I like the top."
	• "Technology. Well, I used to be quite good when I was working, because I had to keep up with the technology… [B]ut I use [MS] Word, I use email, and I text. I mean I'm not a real techie and I wouldn't say I'm up to date on all the latest gizmos that it's possible to do on my laptop or my smartphone. What I try to keep up, I can do more than is required of the life that I'm living."
Negative	• "Okay. Restraint, I missed a little bit. I had trouble with the palms up for a minute. Elbows straight I missed altogether, and slow return."
	• "Slow, slow. The key word with us elders is say it slow, slowly, key, and phone numbers too."
	• "You gotta go louder, I can't hear much. I've got my hearing aids on, but they're useless."
	• "Golly. The only thing that ever frustrates me is if something goes wrong technically and I don't know what to do about it. And usually it's something very simple. If I think about it, I very often can fix it myself. But I have a tech guy who comes in and helps me every once in a while."
Clinician	
Positive	• "I think overall the usability should be good, I think just these little tweaks here and there to make it just so they can ... I think in general it's fairly easy and at least for, you know, does show some basic ... you don't have to be super computer literate."
	• "No, I know what you guys are doing, it's cool. Okay, so it's really easy to discover the Bluetooth device; which is awesome. And same deal, back to ... It's really simple, it has a two layer architecture; which is really, really simple. Which I really like, actually."
Negative	• "[My] mindset is that I want it to go this pathway in order to start my freaking exercises, record them, and I want to press the minimal buttons to get to where I want to go. That's what anyone would want to do, and it's misleading with that video, because it makes me feel like, oh, now I can start my exercise by pressing play, but no, I'm not starting my exercise."
	• "Now, oh, I have to think. I have a working brain, but many people are slower, and they're tired, and this is just a lot of crap going on. I'm not happy with this. But yeah, I press something, and nothing happened, which is also annoying."

Discussion

Principal Findings

Through user-centered design practices, we developed a mobile app for a Bluetooth-connected resistance exercise band. For each round of development, the number of positive sentiments increased, although there was no significant difference in either patients' or clinicians' sentiment of the app in interviews. In the adjusted models, we demonstrated that increasing the total sentiment of an interview was positively associated with SUS scores. Finally, there are meaningful differences in the topics that were identified in the clinician and patient interviews through both qualitative and quantitative methods. Our exploratory results also suggest that NLP and sentiment analysis could be considered in future user-centered design processes.

Our experience with user-centered design aligns with previous insights involving older adults in the development of exercise games [38], active and assisted living [39] and wearable technologies [40]. Our results expand on these prior studies, as we found that patients and clinicians had different perspectives. Our approach allows additional methods for triangulation that are observed in formal mixed methods designs [41]. These results highlight the importance of seeking end user input and involving both clinicians and patients in the design of medical technology as they demonstrate measurable differences in their viewpoints.

The use of sentiment analysis in clinical research has typically been used to assess social media-based opinions around a specific topic such as hospital quality [42], palliative care [43], or tobacco use [44]. The methods used in the sentiment analysis vary—some build and train a model based on the study data, whereas others use open-source or private algorithms [18]. We used an open-source library, which allows others to use the same methodology. With preliminary correlation with the USE scale, the sentiment analysis suggests that it may assess similar domains in this context. Owing to the association between sentiment analysis and usability scores, we believe that a sentiment analysis can be actionable within the development cycle without the burden of a questionnaire. This can be accomplished with automated analytic scripts that parse, clean, and analyze interviews after they have been completed. However, future validation studies, similar to factor analyses for survey methods, should be considered.

We sought to increase external validity and reproducibility by using the public Bing sentiment library [45] because of its previous use and implementation simplicity. This lexicon enabled us to measure variations within and across interview texts, as in other sentiment analyses [18]. LDA methods assume that each text grouping is a mixture of topics and that such topics are a probabilistic mixture of words. Previous studies have applied these methods to clinical notes to identify topics predictive of specific diseases such as dementia [46] and heart failure [47]. In theory, these methods could be used to predict

such outcomes (through notes alone) before a diagnosis. The concept that a conversation or interview is also a mixture of topics that comprise probabilistic mixtures of words makes sense as the interviewee will express different topics that are important to them. Here, we did not use LDA-identified topics in prediction but in the identification of concepts that could assist in improving development and as a proof-of-concept that there are differences between the patients' and clinicians' perceptions of remote medical sensing. Analogs to naming discovered themes in the qualitative analysis (here the identified latent topic) must be entitled by the researchers themselves through interpretation, as it is not an automated process. Discovering the clear differences in topics between patient and clinician interviews validates the need to include both stakeholders in the design and development of remote medical sensing apps. This is an extension of the coproduction concept wherein patients and clinicians combine to help support patients contributing to the management of their own conditions [48,49]. As the use of sentiment analysis and LDA evolve, our results suggest that they can potentially be used alongside more traditional qualitative methods to examine the concepts and perceptions within the context of user-centered design and interviews.

Strengths

This study has several strengths. First, we were able to recruit both older adults and clinicians to participate in a user-centered design. Such an approach provides differing perspectives in addressing complex usability problems. Second, we demonstrated that researchers can use a sentiment analysis and LDA-based topic mapping for interviews with older adults and clinicians. Our evaluation goes beyond traditional qualitative methods that may not necessarily capture sentiments of developing specific apps. Finally, this study paired these evolving quantitative methods with known usability questionnaires to demonstrate the relationship between overall interview sentiment and perceived utility.

Limitations

Although our pilot study sample size was not very large, we were able to reach a total number of participants that allowed us to isolate conceptual themes within the suggested ranges for theoretical saturation [50]. The small sample size, magnitude, and significance of the regression results are less important; however, our goal was to demonstrate an important direction of effect for future research. The scope of the study is also limited to the development of the mobile app; we did not examine the efficacy of its implementation within a patient population of older adults or clinicians. We used a single sentiment library rather than building a sentiment library that worked for this data set. Our approach should be further tested in other medical usability studies to verify these results. As all NLP and sentiment analyses were conducted after development,

we did not examine how they could be used in the active development cycle and could potentially be considered in future design processes. As such, we did not have an *a priori* hypothesis of effect size, and this analysis serves as formative work for others interested in using NLP and sentiment analysis. In addition, LDA-identified topic clusters may be difficult to interpret; here, we qualitatively examined the words most associated with the topic and named them for their underlying concept. Finally, using sentiment analysis around spoken words in the medical and health setting might not be as accurate as its use in a written corpus [51].

Future Work

With the development of the mobile app, we will next focus on deploying it and the resistance exercise band in real at-home experiments. Although challenges still exist in mHealth's ability to impact health behavior [52], monitoring adherence and engagement is the first step. Not all of the insights identified by patients and clinicians could be implemented in this short study and should be added in future versions. Furthermore, new studies will examine the efficacy of the use of the app and its connected device.

The utility of NLP-based methods and sentiment analysis should be further examined in the context of interviews and usability questionnaires within the medical field with larger samples. Although we demonstrated correlation with a usability questionnaire, more work needs to be done to understand how sentiment analysis could be used in the context of conversations and how they may replace or augment questionnaire-based methods. These methods could be used within the development of technology used by patients and providers to assess how they are perceived. The use of NLP and sentiment analysis pipelines could be set up before interviews are conducted, allowing for analysis to occur immediately after the interview. Although not possible in this study, insights from NLP and sentiment analysis could then be incorporated into the design. Finally, additional research needs to examine how interviews could be structured to most effectively use NLP and sentiment analysis. Specifically, asking participants to describe how they felt during different uses or aspects of use may generate data that are actionable within the user design cycle. For user-centered design, many questionnaires are developed for assessing a fully functional product and not one that is partially functioning or disparate parts of a product, making it difficult to use such questionnaires in a pre-post fashion. Using sentiment analysis may allow researchers to more reliably compare the quantification of users' thoughts through the developmental rounds.

Conclusions

User-centered design with both patients and clinicians allowed us to build an app that older adults can use. This is the first analysis that used NLP and usability questionnaires to quantify the user-centered design of technology for older adults.

Acknowledgments

XSL•FO

RenderX

Conflicts of Interest

JB and RH submitted US patent application #62/672,827, titled Instrumented Resistance Exercise Device, and CP, JB, and RH submitted US patent application #63/033,022, titled Remote-sensing, Bluetooth-enabled Resistance Exercise Band.

References

1. Sayer AA, Syddall HE, Martin HJ, Dennison EM, Roberts HC, Cooper C. Is grip strength associated with health-related quality of life? Findings from the Hertfordshire cohort study. Age Ageing 2006 Jul;35(4):409-415. [doi: 10.1093/ageing/afl024] [Medline: 16690636]

2. Sayer AA, Dennison EM, Syddall HE, Gilbody HJ, Phillips DI, Cooper C. Type 2 diabetes, muscle strength, and impaired physical function: the tip of the iceberg? Diabetes Care 2005 Oct;28(10):2541-2542. [doi: 10.2337/diacare.28.10.2541] [Medline: 16186295]

3. Gale CR, Martyn CN, Cooper C, Sayer AA. Grip strength, body composition, and mortality. Int J Epidemiol 2007 Feb;36(1):228-235. [doi: 10.1093/ije/dyl224] [Medline: 17056604]

4. Marty E, Liu Y, Samuel A, Or O, Lane J. A review of sarcopenia: enhancing awareness of an increasingly prevalent disease. Bone 2017 Dec;105:276-286. [doi: 10.1016/j.bone.2017.09.008] [Medline: 28931495]

5. Candow DG, Chilibeck PD, Forbes SC. Creatine supplementation and aging musculoskeletal health. Endocrine 2014 Apr;45(3):354-361. [doi: 10.1007/s12020-013-0070-4] [Medline: 24190049]

6. Beaudart C, Buckinx F, Rabenda V, Gillain S, Cavalier E, Slomian J, et al. The effects of vitamin D on skeletal muscle strength, muscle mass, and muscle power: a systematic review and meta-analysis of randomized controlled trials. J Clin Endocrinol Metab 2014 Nov;99(11):4336-4345. [doi: 10.1210/jc.2014-1742] [Medline: 25033068]

7. Hickson M. Nutritional interventions in sarcopenia: a critical review. Proc Nutr Soc 2015 Nov;74(4):378-386. [doi: 10.1017/S0029665115002049] [Medline: 25923603]

8. Pahor M, Guralnik JM, Ambrosius WT, Blair S, Bonds DE, Church TS, LIFE Study Investigators. Effect of structured physical activity on prevention of major mobility disability in older adults: the LIFE study randomized clinical trial. J Am Med Assoc 2014 Jun 18;311(23):2387-2396 [FREE Full text] [doi: 10.1001/jama.2014.5616] [Medline: 24866862]

9. Batsis JA, Zagaria AB. Addressing obesity in aging patients. Med Clin North Am 2018 Jan;102(1):65-85 [FREE Full text] [doi: 10.1016/j.mcna.2017.08.007] [Medline: 29156188]

10. Hardage J, Peel C, Morris D, Graham C, Brown C, Foushee HR, et al. Adherence to exercise scale for older patients (AESOP): a measure for predicting exercise adherence in older adults after discharge from home health physical therapy. J Geriatr Phys Ther 2007;30(2):69-78. [doi: 10.1519/00139143-200708000-00006] [Medline: 18171490]

11. Martin LR, Williams SL, Haskard KB, Dimatteo MR. The challenge of patient adherence. Ther Clin Risk Manag 2005 Sep;1(3):189-199 [FREE Full text] [Medline: 18360559]

12. Fitzpatrick AL, Powe NR, Cooper LS, Ives DG, Robbins JA. Barriers to health care access among the elderly and who perceives them. Am J Public Health 2004 Oct;94(10):1788-1794. [doi: 10.2105/ajph.94.10.1788] [Medline: 15451751]

13. Petersen CL, Weeks WB, Norin O, Weinstein JN. Development and implementation of a person-centered, technology-enhanced care model for managing chronic conditions: cohort study. JMIR Mhealth Uhealth 2019 Mar 20;7(3):e11082 [FREE Full text] [doi: 10.2196/11082] [Medline: 30892274]

14. 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]

15. Batsis JA, Naslund JA, Zagaria AB, Kotz D, Dokko R, Bartels SJ, et al. Technology for behavioral change in rural older adults with obesity. J Nutr Gerontol Geriatr 2019;38(2):130-148 [FREE Full text] [doi: 10.1080/21551197.2019.1600097] [Medline: 30971189]

16. Swallow D, Petrie H, Power C, Lewis A, Edwards AD. Involving older adults in the technology design process: a case study on mobility and wellbeing in the built environment. Stud Health Technol Inform 2016;229:615-623. [Medline: 27534357]

17. Kim E, Rubinstein SM, Nead KT, Wojcieszynski AP, Gabriel PE, Warner JL. The evolving use of electronic health records (EHR) for research. Semin Radiat Oncol 2019 Oct;29(4):354-361. [doi: 10.1016/j.semradonc.2019.05.010] [Medline: 31472738]

18. Gohil S, Vuik S, Darzi A. Sentiment analysis of health care tweets: review of the methods used. JMIR Public Health Surveill 2018 Apr 23;4(2):e43 [FREE Full text] [doi: 10.2196/publichealth.5789] [Medline: 29685871]

19. Earnings Call Sentiment Analysis & Text NLP Tool. Amenity Analytics. 2020. URL: https://www.amenityanalytics.com/case-studies/earnings-call-transcript-analysis [accessed 2020-07-21]

20. Denecke K, May R, Deng Y. Towards emotion-sensitive conversational user interfaces in healthcare applications. Stud Health Technol Inform 2019 Aug 21;264:1164-1168. [doi: 10.3233/SHTI190409] [Medline: 31438108]

21. Batsis JA, Boateng GG, Seo LM, Petersen CL, Fortuna KL, Wechsler EV, et al. Development and usability assessment of a connected resistance exercise band application for strength-monitoring. World Acad Sci Eng Technol 2019;13(5):340-348 [FREE Full text] [doi: 10.5281/zenodo] [Medline: 31205628]

22. Petersen C, Wechsler E, Halter R, Boateng G, Proctor P, Kotz D. Detection and Monitoring of Repetitions Using an Mhealth-Enabled Resistance Band. In: Proceedings of the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies. 2018 Presented at: CHASE'18; May 14, 2018; Washington, DC. [doi: 10.1145/3278576.3278586]

23. Black AC, Serowik KL, Schensul JJ, Bowen AM, Rosen MI. Build a better mouse: directly-observed issues in computer use for adults with SMI. Psychiatr Q 2013 Mar;84(1):81-92 [FREE Full text] [doi: 10.1007/s11126-012-9229-z] [Medline: 22711454]

24. Brooke J. SUS-a quick and dirty usability scale. In: Jordan PW, Weerdmeester B, McClelland IL, Thomas B, editors. Usability Evaluation in Industry. Boca Raton, FL: CRC Press; 1996:189-194.

25. Gao M, Kortum P, Oswald F. Psychometric Evaluation of the USE (Usefulness, Satisfaction, and Ease of Use) Questionnaire for Reliability and Validity. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2018 Presented at: HFES'18; October 1-5, 2018; Los Angeles, CA p. 1414-1418 URL: https://doiorg/101177/1541931218621322 [doi: 10.1177/1541931218621322]

26. Lund A. Measuring usability with the USE questionnaire. Usab Interf 2001;8(2):3-6.

27. Blei D, Ng A, Jordan M. Latent dirichllocation. J Mach Learn Res 2003 Jan;3:993-1022.

28. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What patients can tell us: topic analysis for social media on breast cancer. JMIR Med Inform 2017 Jul 31;5(3):e23 [FREE Full text] [doi: 10.2196/medinform.7779] [Medline: 28760725]

29. Hwang Y, Kim HJ, Choi HJ, Lee J. Exploring abnormal behavior patterns of online users with emotional eating behavior: topic modeling study. J Med Internet Res 2020 Mar 31;22(3):e15700 [FREE Full text] [doi: 10.2196/15700] [Medline: 32229461]

30. Garcia-Rudolph A, Laxe S, Saurí J, Guitart MB. Stroke survivors on twitter: sentiment and topic analysis from a gender perspective. J Med Internet Res 2019 Aug 26;21(8):e14077 [FREE Full text] [doi: 10.2196/14077] [Medline: 31452514]

31. Arun R, Suresh V, Veni MC, Narasimha MM. On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. 2010 Presented at: PAKDD'10; June 21-24, 2010; Hyderabad, India. [doi: 10.1007/978-3-642-13657-3_43]

32. Cao J, Xia T, Li J, Zhang Y, Tang S. A density-based method for adaptive LDA model selection. Neurocomputing 2009 Mar;72(7-9):1775-1781 [FREE Full text] [doi: 10.1016/j.neucom.2008.06.011]

33. Deveaud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc information retrieval. Document Numérique 2014 Apr 30;17(1):61-84. [doi: 10.3166/dn.17.1.61-84]

34. Griffiths TL, Steyvers M. Finding scientific topics. Proc Natl Acad Sci U S A 2004 Apr 6;101(Suppl 1):5228-5235 [FREE Full text] [doi: 10.1073/pnas.0307752101] [Medline: 14872004]

35. Nikita M. ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters. The Comprehensive R Archive Network. 2019. URL: https://cran.r-project.org/web/packages/ldatuning/index.html [accessed 2020-07-23]

36. Hu M, Liu B. Mining and Summarizing Customer Reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2004 Presented at: KDD'04; August 22-25, 2004; Seattle, Washington. [doi: 10.1145/1014052.1014073]

37. R: A Language and Environment for Statistical Computing. The Global Biodiversity Information Facility (GBIF). 2019. URL: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing [accessed 2020-07-23]

38. Brox E, Konstantinidis ST, Evertsen G. User-centered design of serious games for older adults following 3 years of experience with exergames for seniors: a study design. JMIR Serious Games 2017 Jan 11;5(1):e2 [FREE Full text] [doi: 10.2196/games.6254] [Medline: 28077348]

39. Haslwanter JD, Neureiter K, Garschall M. User-centered design in AAL. Univ Access Inf Soc 2018 Jul 23;19(1):57-67. [doi: 10.1007/s10209-018-0626-4]

40. Batsis JA, Zagaria A, Kotz DF, Bartels SJ, Boateng GG, Proctor PO, et al. Usability evaluation for the amulet wearable device in rural older adults with obesity. Gerontechnology 2018 Sep;17(3):151-159 [FREE Full text] [doi: 10.4017/gt.2018.17.3.003.00] [Medline: 30631251]

41. Castro FG, Kellison JG, Boyd SJ, Kopak A. A methodology for conducting integrative mixed methods research and data analyses. J Mix Methods Res 2010 Sep 20;4(4):342-360 [FREE Full text] [doi: 10.1177/1558689810382916] [Medline: 22167325]

42. Hawkins JB, Brownstein JS, Tuli G, Runels T, Broecker K, Nsoesie EO, et al. Measuring patient-perceived quality of care in US hospitals using Twitter. BMJ Qual Saf 2016 Jun;25(6):404-413 [FREE Full text] [doi: 10.1136/bmjqs-2015-004309] [Medline: 26464518]

43. Nwosu AC, Debattista M, Rooney C, Mason S. Social media and palliative medicine: a retrospective 2-year analysis of global Twitter data to evaluate the use of technology to communicate about issues at the end of life. BMJ Support Palliat Care 2015 Jun;5(2):207-212. [doi: 10.1136/bmjspcare-2014-000701] [Medline: 25183713]

44. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013 Aug 29;15(8):e174. [doi: 10.2196/jmir.2534] [Medline: 23989137]

45. Liu B. Sentiment analysis and opinion mining. Synth Lect Hum Lang Technol 2012 May 23;5(1):1-167. [doi: 10.2200/s00416ed1v01y201204hlt016]

46. Shao Y, Zeng QT, Chen KK, Shutes-David A, Thielke SM, Tsuang DW. Detection of probable dementia cases in undiagnosed patients using structured and unstructured electronic health records. BMC Med Inform Decis Mak 2019 Jul 9;19(1):128 [FREE Full text] [doi: 10.1186/s12911-019-0846-4] [Medline: 31288818]

47. Shao Y, Mohanty AF, Ahmed A, Weir CR, Bray BE, Shah RU, et al. Identification and use of frailty indicators from text to examine associations with clinical outcomes among patients with heart failure. AMIA Annu Symp Proc 2016;2016:1110-1118 [FREE Full text] [Medline: 28269908]

48. Elwyn G, Nelson E, Hager A, Price A. Coproduction: when users define quality. BMJ Qual Saf 2019 Sep 5:- epub ahead of print [FREE Full text] [doi: 10.1136/bmjqs-2019-009830] [Medline: 31488570]

49. Batalden M, Batalden P, Margolis P, Seid M, Armstrong G, Opipari-Arrigan L, et al. Coproduction of healthcare service. BMJ Qual Saf 2016 Jul;25(7):509-517 [FREE Full text] [doi: 10.1136/bmjqs-2015-004315] [Medline: 26376674]

50. Gentles S, Charles C, Ploeg J, McKibbon K. Sampling in qualitative research: insights from an overview of the methods literature. Qual Res 2015;20(11):2015-2089 [FREE Full text]

51. He L, Zheng K. How do general-purpose sentiment analyzers perform when applied to health-related online social media data? Stud Health Technol Inform 2019 Aug 21;264:1208-1212. [doi: 10.3233/SHTI190418] [Medline: 31438117]

52. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. J Behav Med 2019 Feb;42(1):67-83 [FREE Full text] [doi: 10.1007/s10865-018-9966-z] [Medline: 30825090]

Abbreviations

LDA: latent Dirichlet allocation
mHealth: mobile health
NLP: natural language processing
SUS: System Usability Scale
USE: Usefulness, Satisfaction, and Ease of use scale

XSL•FO
RenderX

Original Paper

The Association Between App-Administered Depression Assessments and Suicidal Ideation in User Comments: Retrospective Observational Study

Shelly DeForte[1], PhD; Yungui Huang[1], PhD, MBA; Tran Bourgeois[1], MPH; Syed-Amad Hussain[1], BSc; Simon Lin[1], MD, MBA

Abigail Wexner Research Institute, Nationwide Children's Hospital, Columbus, OH, United States

Corresponding Author:
Simon Lin, MD, MBA
Abigail Wexner Research Institute
Nationwide Children's Hospital
700 Children's Drive
Columbus, OH, 43215
United States
Phone: 1 202 318 1627
Email: simon.lin@nationwidechildrens.org

Abstract

Background: Many people use apps to help understand and manage their depression symptoms. App-administered questionnaires for the symptoms of depression, such as the Patient Health Questionnaire-9, are easy to score and implement in an app, but may not be accompanied by essential resources and access needed to provide proper support and avoid potential harm.

Objective: Our primary goal was to evaluate the differences in risks and helpfulness associated with using an app to self-diagnose depression, comparing assessment-only apps with multifeatured apps. We also investigated whether, what, and how additional app features may mitigate potential risks.

Methods: In this retrospective observational study, we identified apps in the Google Play store that provided a depression assessment as a feature and had at least five user comments. We separated apps into two categories based on those having only a depression assessment versus those that offered additional supportive features. We conducted theoretical thematic analyses over the user reviews, with thematic coding indicating the helpfulness of the app, the presence of suicidal ideation, and how and why the apps were used. We compared the results across the two categories of apps and analyzed the differences using chi-square statistical tests.

Results: We evaluated 6 apps; 3 provided only a depression assessment (assessment only), and 3 provided features in addition to self-assessment (multifeatured). User comments for assessment-only apps indicated significantly more suicidal ideation or self-harm (n=31, 9.4%) compared to comments for multifeatured apps (n=48, 2.3%; X^2_1=43.88, P<.001). Users of multifeatured apps were over three times more likely than assessment-only app users to comment in favor of the app's helpfulness, likely due to features like mood tracking, journaling, and informational resources (n=56, 17% vs n=1223, 59% respectively; X^2_1=200.36, P<.001). The number of users under the age of 18 years was significantly higher among assessment-only app users (n=40, 12%) than multifeatured app users (n=9, 0.04%; X^2_1=189.09, P<.001).

Conclusions: Apps that diagnose depression by self-assessment without context or other supportive features are more likely to be used by those under 18 years of age and more likely to be associated with increased user distress and potential harm. Depression self-assessments in apps should be implemented with caution and accompanied by evidence-based capabilities that establish proper context, increase self-empowerment, and encourage users to seek clinical diagnostics and outside help.

KEYWORDS

mobile health; mHealth; depression; qualitative research; mental health

XSL•FO

RenderX

Introduction

Digital devices have become an essential part of our lives. People increasingly rely on digital content and functionalities delivered through mobile apps for information, entertainment, daily task management, work functions, and even for health-related activities, such as symptom tracking, diagnosis, and digital health treatment. Increased patient willingness and appetite to adopt mental health apps and share data [1,2] has resulted in a growing market for mental health apps. Mental health apps could play a critical role in addressing unmet needs in mental health disease screening, self-management, monitoring, and health education [3,4]. This role is especially needed, considering the severe shortage of mental health professionals and the long wait for mental health services.

However, most publicly available mental health apps have not been scientifically validated [5-7], provide inaccurate information, and do not follow clinical guidelines [8]. Often these apps also employ misleading marketing tactics [9] that may suggest to the end user an unearned medical or algorithmic authority. App-administered questionnaires for the symptoms of depression, such as the Patient Health Questionnaire-9 (PHQ-9) [10], are of particular concern because the results are often presented without proper context or access to additional resources. Although safety concerns have been raised about leaving patients to their own devices [11,12], it has not been established whether apps that administer a depression self-assessment are beneficial or dangerous, especially when an app solely offers a depression self-assessment. Given that people with mental health concerns are at high risk, it is essential to evaluate the safety of readily available mobile apps and the presence or absence of mitigating factors to reduce this risk.

In this study, we assessed the self-reported user experience using user comments publicly posted for app review. This study aimed to compare user experiences with apps that are assessment-only for depression with no supportive content to those that have additional resources and contextual information for depression. We hypothesized that consumers using assessment-only apps might report more distress without proper support, while those using multifeatured apps might fare better with additional supportive information and contexts. Our objective was to determine to what extent and how app-administered depression self-assessments may be associated with a perceived benefit versus self-reported distress or harm.

Methods

Due to the public availability of these comments, this study was exempt from the requirements for institutional review board approval.

Inclusion Criteria

We limited our analysis to apps available in Google Play store, as we were unable to extract comments automatically from the Apple App store. This study also included only those apps that offered a depression assessment feature and had more than five user comments.

Data Collection

Author SD searched the Google Play store on October 15, 2018, with Google Chrome in incognito mode using the term "depression test." SD read each app description and selected apps that met the inclusion criteria. Author SAH manually visited each Google Play app web page on October 19, 2018, selected "Read All Reviews," scrolled downward until the page had loaded all of the reviews, and then selected "view page source" from the Google Chrome browser menu. After saving these pages' source codes as files, author SAH developed a custom web scraping script to remove extraneous information (eg, HTML coding, rating), extract the user comments from each file, and compile them for analysis. This script used the Python library BeautifulSoup 4 (PyPi) to walk through the source page's HTML tree and then used regular expressions to extract the relevant information.

Data Analysis

The apps were categorized as those that provide only a depression assessment (assessment-only) and those having other features in addition to the assessment (multifeatured).

This observational study employed the theoretical thematic analysis approach defined by Braun and Clarke [13] to analyze the qualitative data of user comments. Guided by our research interests to evaluate whether the apps may be associated with perceived benefit or harm, we utilized a hybrid inductive and deductive framework. We adopted preconceived themes of benefit and distress but induced other themes by reading and coding the user comments. We first familiarized ourselves with the data, highlighting relevant words, phrases, and sentences related to helpfulness, risks, and functionalities. We began with free-form thematic tagging before grouping based on similar themes. Then we generated new themes, reviewed and finalized the definitions of themes, and finally coded all comments using the definitions and guideline. Coding for all themes was defined after reviewing the data multiple times. The full guideline and results can be found in the Results section. SD performed the initial manual thematic coding and grouping, and TB retagged the comments blind using the thematic scheme developed by SD. The final tagging was obtained based on discussion and consensus for any discrepancies, and a third independent coder (SAH) settled any disagreements. The prevalence-adjusted and bias-adjusted κ (PABAK) score determined inter-rater reliability.

Pearson chi-squared test with Yates continuity correction was used to analyze the difference between the assessment-only apps and multifeatured apps. The degree of freedom of the chi-squared test is 1. The significance level was set at $P<.05$.

Results

Apps Included in the Study

The apps that met the inclusion criteria, number of comments retrieved for each app, and the number of times each app had been rated are listed in Table 1. While ratings were not factored into the thematic coding, the number of ratings might reflect the relative user base of each app, as no additional data was available to indicate the actual number of users. Eight apps with

depression assessments were excluded due to having fewer than   five comments.

Table 1. The apps included in this analysis.

Features	Google Play app ID	In-app purchase	Ratings, n[a]	Comments, n[b]
Multifeatured[c] (proprietary assessment)[d]	de.moodpath.android	Yes	7062	1226
Multifeatured (PHQ-9[e])	com.moodtools.moodtools	Yes	2871	788
Multifeatured (modified PHQ-9)	com.williamalexander.android.depressiontracker	No	267	63
Assessment only[f] (PHQ-9)	nl.japps.android.depressiontest	No	1408	264
Assessment only (proprietary assessment)[d]	com.programming.advanced.depressiontest	No	436	43
Assessment only (PHQ-9)	com.moodtools.depressiontest	No	213	27

[a]The number of ratings for each app, some of which may not include a comment.

[b]The number of ratings that included a written comment.

[c]Apps that have other capabilities aside from a depression assessment.

[d]A nonstandard assessment.

[e]PHQ-9: Patient Health Questionnaire-9 (a 9-question survey to determine the severity of depression symptoms).

[f]Apps that have a depression assessment only.

Themes

Two tags, which capture benefit and harm for all apps selected, tested our hypothesis that more users of assessment-only apps than multifeatured apps reported distress and suicidal ideation (see Methods). In familiarizing ourselves with the comments and generating new themes, we were surprised to find that some comments indicated the users were probably minors. As pediatric users might be more vulnerable and should be considered separately, we created a youth (vs adult) tag to capture whether the post was apt to be from someone who was under 18 years of age. Three themes (tracking, report, and library) emerged to capture the useful app features in multifeatured apps, and three tags (management, self-knowledge, and therapy) were generated to capture the stated utility of additional features, when available. The full thematic coding guideline and finalized themes are described in Table 2.

XSL•FO
RenderX

Table 2. Thematic coding guidelines and finalized themes.

Theme	Description	Coding guideline
Youth	To capture whether the user was likely a youth or adult	• Comment specifically mentioned the user age was under 18, or self-reported as being a child or being young. • Comment indicated that they still needed their parents' help or approval (which suggests they were under 18 years of age).
Perception of helpfulness and harmfulness		
Helpful	To indicate that the app was affirmatively beneficial to the user.	• Marked "Y" if the user said the app is good (or similar) but also mentions a specific app feature, indicating that they are using the feature and think it is good. • Marked "N" if the user only said the app is "good" or complements a design aspect. • Marked "Y" if the comment used the words "helpful" or "useful" to describe the app (aside from this, most short comments were an "N." • Marked "Y" if the user stated that the app would be generically beneficial for depression, but mark "N" if the user also noted the app was not helpful for them. • Marked "Y" if the app prompted them to get professional help. • Marked "Y" if the user said they love the app or similar.
Distress	To capture suicidal ideation or self-harm	• Marked "Y" for any comments like "kill me now" or "I want to die" or "why am I alive?" • Marked "Y" for comments where they talked about someone else who is suicidal. • Marked "Y" for comments that talked about hospitalization for past suicide attempts. • Marked "Y" for comments that indicated the app 'saved their life' or app is a "life-saver." • Marked "N" if the comment only referenced the safety plan.
Features mentioned (for multifeatured only)		
Tracking	To capture whether this app was used as a tracking tool	• Comments mentioned "tracking" or used similar terminology. Mood tracking, journaling, and cognitive behavioral therapy were grouped under this tag. The key distinguisher here was activities over time. • Since the Moodtools app calls journaling a "Thought Diary," any mention of mood tracking, thought diary or journal features were included with this tag. • In the Moodpath app, a series of questions tracked mood, so references to a "questionnaire" were tagged with tracking.
Reporting	To capture any mention of the depression assessment or waiting for the depression assessment results.	• For Moodtools, the users may mention a "report" or a "doctor's note." Since the depression assessment was meant to be shared with the doctor, tagged comments with "Y" when a user mentioned sharing results with their doctor. • For Moodtools, since the assessment follows two weeks of mood tracking, we also marked references to waiting for two weeks as references to the depression assessment, even without explicitly mentioning the evaluation. • Marked "Y" any mentions of a "test," "diagnosis," "results," or "accuracy," which probably refer to the assessment
Library	To encapsulate features that can be used once and be useful; are not ongoing tracking but are also not references to the depression assessment.	• Terms like "resources," "educational," "informative," etc, suggest that this tag was appropriate. • This tag encompassed mentioning information, activities, videos, and the "Safety plan" in Moodtools.
How and why it helps (for multifeatured only)		
Management	To capture using the app for managing depression, moods, or other mental illness through activities done over time with broader interpretation.	• This tag overlapped with the "Tracking" tag frequently. The comment must indicate that the user is using the tracking feature to help their symptoms. "Great for tracking" would be an example of a comment that was tagged "Management" but not "Tracking." • Included comments that captured activities that would relieve symptoms of depression, such as "I can express myself," "it's like talking to a friend," "I like that it checks in on me." • Included comments in which the user expressed that the app helped them feel better. • Included comments that referenced depression or other mental illness and indicated that they were using the app to manage this state. • Included comments that indicated the app helped the user handle or manage things. • Included comments that referenced self-knowledge over time for symptom management. Note: this tag overlapped with the self-knowledge tag.

XSL•FO
RenderX

Theme	Description	Coding guideline
Self-knowledge	To capture comments that indicate that the user learned more about themselves through use of the app.	• Despite some overlap with the "Library" tag, we were looking for more personal comments that indicated the user had learned something about themselves. • This tag overlapped with the management tag when the user referenced that they have learned more about themselves and helped them to manage their symptoms. • Tagged comments that referred to learning about patterns or learning new skills. Words like "insightful" or "educational" frequently indicated use for self-knowledge.
Therapy	Infers if the user is in therapy, using the app with a therapist, or prompted by the app to go to a therapist.	• Included any comments by therapists. • Included comments that generically mention that the user could or might have used the app with a therapist or medical professional. • Included comments such as "you should use this app with a professional" because it indicates that using the app with a therapist is a priority for the user.

Coding Results

The PABAK scores for all tags, which were used to determine inter-rater reliability, 0-1 with 1.0 being the most reliable, were between 0.737 and 0.996 after the initial tagging (Multimedia Appendix 1), except for the "Help" tag for the multifeatured comments which had a PABAK score of 0.584. The two coders decided to treat comments that indicated general enthusiasm for the app as "helpful" comments, resolving the conceptual discrepancy that resulted in the low PABAK score with the updated PABAK score for Help of 0.682 and 95% CI 0.649-0.713.

The results of the thematic coding after a third coder adjudicated the final decisions on discordant coding are summarized in Table 3.

Table 3. Thematic coding results summary.

Theme	Assessment only (n=329), n (%)	Multifeatured (n=2069), n (%)	P value (assessment-only vs multifeatured)
Helpful	56 (17.02)	1223 (59.11)	<.001
Distress	31 (9.42)	47 (2.27)	<.001
Youth (vs adult)	40 (12.16)	9 (0.43)	<.001
Tracking	N/A[a]	509 (24.60)	N/A
Report	N/A	180 (8.70)	N/A
Library	N/A	253 (12.23)	N/A
Management	N/A	438 (21.17)	N/A
Self-knowledge	N/A	359 (17.35)	N/A
Therapy	N/A	118 (5.70)	N/A

[a]N/A: not applicable.

Thematic Analysis

The most common words in all app comments were words with the root "help*" (1007 occurrences) and "depress*" (443 occurrences). Of the multifeatured app comments, 59% (1223/2069) expressly indicated the app was beneficial in some way, which is statistically different from 17% (56/329) of the comments for the assessment-only apps (X^2_1=200.36, P<.001). Many comments from assessment-only apps simply reported the score the user received or expressed powerlessness or frustration over their self-diagnosed condition. Of the comments for multifeatured apps, 2.3% (31/329) indicated suicidal ideation or self-harm, compared to 9.4% (47/2069) of the comments for assessment-only apps, representing a statistically significant 4-fold increase (X^2_1=43.88, P<.001).

Of users who self-reported their age category (youth vs adult), 12% (40/329) of comments from assessment-only app users indicated they were under 18 years of age, compared with 0.04% (9/2069) for the multifeatured app users, a statistically significant difference (X^2_1=189.09, P<.001). The P values of the comparison tests between depression test-only and multifeatured apps are listed in Table 3.

As indicated by the coding guideline, additional app features for multifeatured apps were grouped into tracking, library (informational resources), and depression assessment (report theme) categories. Comments mentioning tracking or informational resources (Library) appear 664/2069 times (32%) versus 180/2069 (9%) for the depression assessment. Users who discussed how and why they used the multifeatured apps most often referenced self-management of moods and using the apps to understand triggers and improve their state of mind (438/2069, 21%; Table 3). Others referenced how the apps allowed them to increase self-awareness and validate misunderstood feelings (359/2069, 17%). Few people mentioned using the app with a clinical provider (118/2069, 6%). The numbers of comments reflecting the induced themes for the multifeatured apps and the numbers of comments covering multiple themes are illustrated in Figure 1.

Figure 1. Comments that indicate app features being used (left) and explanations of app use and why it is beneficial (right). 'Tracking' refers to mood tracking or journal features, 'Library' refers to informational resources, videos, or prompts for activities, and 'Assessment' refers to the depression assessment (left). 'Management' refers to comments that indicate applying the app for self-management of moods or mental health over time, 'Knowledge' refers to self-knowledge or insight gained from using the app, while 'Therapy' refers to app use in conjunction with some kind of clinical therapy.
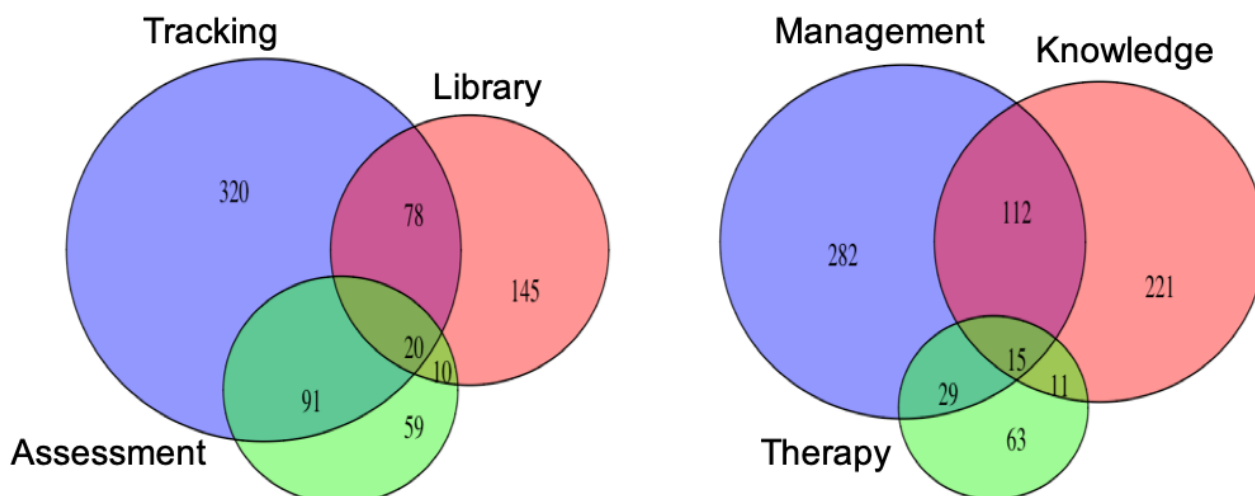
Discussion

Principal Findings

Our study supported our hypothesis that compared to users of multifeatured apps, assessment-only app users reported more distress, as seen in the significantly more substantial amount of comments indicating suicidal ideation. Additionally, users of assessment-only apps were more likely to be under 18 years of age than those using multifeatured apps. Compared to users of assessment-only apps, more users of multifeatured apps commented in favor of the apps' helpfulness, mood tracking, journaling, and informational resources.

Our findings were consistent with other studies that showed that people who used health apps valued the ongoing tracking features greatly [14] and suggested that standalone apps that only administered a depression assessment were less beneficial than multifeatured apps. Alarmingly, we observed a remarkable 4-fold increase in comments indicating suicidal ideation and self-harm in the assessment-only apps, suggesting these apps may be more associated with potential harm. In another study, users scored higher on suicidal ideation indicators when the PHQ-9 was given via an app versus a clinician [15], indicating the importance of clinical guidance and resources at the time of assessment. While some evidence suggested that the privacy afforded by an app might increase self-disclosure [16-18], our analysis, which only compared apps, did not provide sufficient data to derive this conclusion for depression self-assessments. Instead, our results suggested that self-assessments given without proper support and context may be associated with triggering and exacerbating suicidal ideation. An underexplored factor in the literature is whether the self-diagnosis of depression through an app may induce *demoralization*. Demoralization is characterized by a sense of disheartenment and disempowerment that can follow a severe diagnosis and which is significantly associated with suicidal ideation [16]. We think this potentially confounding factor merits further study. Activities that lead to

self-empowerment and self-awareness can decrease depressive symptoms [19,20], and an increase in perspective and perceived self-control can directly combat demoralization [21]. The additional features offered by the multifeatured apps might have provided the functionalities and engagement needed to enhance their self-awareness and sense of self-empowerment.

Limitation and Challenges

A fundamental limitation of this observational study is that participants were self-selecting, in addition to the small sample size of the three assessment-only apps and the three multifeatured apps. The expressed opinions through user comments might not be representative of the user population due to the selection bias. Also noted in this comparative analysis is the highly disparate number of comments on the two types of apps. The smaller number of comments in the assessment-only apps might not be representative of the underlying user base. The retrospective nature of this research precluded obtaining user assessments of all the different themes we are trying to analyze from all the users. Instead, we had to rely solely on the publicly available comments in the app store. Inferring user opinion was also a challenge. Even though we followed a clearly defined coding guideline for consistency, coders still had different interpretations. We recognize that despite our best efforts, we may have misrepresented what users tried to convey.

This observational study is insufficient to establish causality. The assessment-only apps that we studied appear to have a higher number and proportion of underage users. This more substantial proportion might be due to the app's simplicity and the likely associated ease of use, the free assessment, or possibly due to statistical variation resulting from a lack of a larger sample of data. Without a proper study design on a more extensive user base, we are unable to explain the reasons underlying the difference in the frequencies of underage users in the two types of apps. However, we feel that our results warrant caution and further study on the use and development

of mental health assessment apps. Especially concerning is the large proportion of adolescent users self-reporting suicidal ideation in the user comments of these apps. With the rising pediatric suicide rate [22], adolescents need more protective interventions against potential harms. All the best intentions of app creators to fill in the resource gaps before first mental health visits or between visits, the desire to give patients independence and privacy, and the goodwill to offer cheaper alternatives to clinical appointments might not be enough to keep a good balance on benefits and risks when implementing mental health apps. This caution is particularly important for younger users.

Finally, the landscape of health apps is extraordinarily dynamic and rapidly evolving [23]. We were only able to capture the data at one point and unable to validate our findings against more current data from 2019 to 2020, which is another limitation of this study. Nonetheless, we are confident our research provides an important data point in the continuous timeline, and our conclusions are expected to stand with the newer data.

Conclusions

In summary, we express our reservation about using mental health assessment-only apps without providing additional resources and functionalities. We also recommend that evidence-based, mitigating activities (eg, mood tracking, journaling, educational materials, and self-empowerment and self-awareness activities) should accompany any app that can lead to the self-diagnosis of depression or other mental health conditions. Assessment-only apps should firmly emphasize that assessment done in the app is not diagnostic, provide a clear recommendation to follow up with health professionals to conduct clinical diagnostic testing, and provide links to additional evidence-based information on depression [24,25]. If an assessment-only app can give the proper warnings and provide informational links, they can still be beneficial, especially to people who would want to an initial understanding of their mental health in private, or people with limited financial means. The advocated enhancement should be straightforward to implement and should not add much development costs to the apps. It might be more beneficial for the multifeatured app to offer the assessment module for free. These additions need to be balanced by engagement design factors [5] to avoid deterring use by younger users.

Even though multifeatured mental health apps seem to mitigate some risks of harm, additional larger-scale prospective research studies are needed to accurately assess the long-term benefits and risks of mental health apps [1,4]. All aspects of apps, including user experience, user engagement, age appropriateness of contents, values of different features, impacts attributed to apps, and others, should be investigated. We agree with many research studies that recommend apps incorporate evidence-based content [5], adopt a digital health app development standard [3,26], rely on clinician recommendations of validated apps versus social media, personal searches, or word-of-mouth to explore or adopt a health app [27,28] and require app stores to standardize reporting. We realize that implementation of these recommendations presents significant challenges requiring collaboration and development efforts across multiple fields, including clinical research, mobile health research, health IT industry, regulation, clinical practice, and others. Nevertheless, these standardizations and safeguards are essential to help patients find validated apps, prevent harm, and assure mental health apps create the value they claim to provide.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Thematic coding interrater reliability.
[DOCX File , 14 KB - mhealth_v8i8e18392_app1.docx]

References

1. BinDhim NF, Alanazi EM, Aljadhey H, Basyouni MH, Kowalski SR, Pont LG, et al. Does a Mobile Phone Depression-Screening App Motivate Mobile Phone Users With High Depressive Symptoms to Seek a Health Care Professional's Help? J Med Internet Res 2016 Jun 27;18(6):e156 [FREE Full text] [doi: 10.2196/jmir.5726] [Medline: 27349441]
2. Di Matteo D, Fine A, Fotinos K, Rose J, Katzman M. Patient Willingness to Consent to Mobile Phone Data Collection for Mental Health Apps: Structured Questionnaire. JMIR Ment Health 2018 Aug 29;5(3):e56 [FREE Full text] [doi: 10.2196/mental.9539] [Medline: 30158102]
3. BinDhim NF, Shaman AM, Trevena L, Basyouni MH, Pont LG, Alhawassi TM. Depression screening via a smartphone app: cross-country user characteristics and feasibility. J Am Med Inform Assoc 2015 Jan;22(1):29-34 [FREE Full text] [doi: 10.1136/amiajnl-2014-002840] [Medline: 25326599]

4. 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]

5. 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 Jun 06;20(6):e10120 [FREE Full text] [doi: 10.2196/10120] [Medline: 29875087]

6. Wasil AR, Venturo-Conerly KE, Shingleton RM, Weisz JR. A review of popular smartphone apps for depression and anxiety: Assessing the inclusion of evidence-based content. Behav Res Ther 2019 Dec;123:103498. [doi: 10.1016/j.brat.2019.103498] [Medline: 31707224]

7. Anthes E. Mental health: There's an app for that. Nature 2016 Apr 7;532(7597):20-23. [doi: 10.1038/532020a] [Medline: 27078548]

8. Martinengo L, Van Galen L, Lum E, Kowalski M, Subramaniam M, Car J. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. BMC Med 2019 Dec 19;17(1):231 [FREE Full text] [doi: 10.1186/s12916-019-1461-z] [Medline: 31852455]

9. Parker L, Bero L, Gillies D, Raven M, Mintzes B, Jureidini J, et al. Mental Health Messages in Prominent Mental Health Apps. Ann Fam Med 2018 Jul;16(4):338-342 [FREE Full text] [doi: 10.1370/afm.2260] [Medline: 29987082]

10. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]

11. Ho A, Quick O. Leaving patients to their own devices? Smart technology, safety and therapeutic relationships. BMC Med Ethics 2018 Mar 06;19(1):18 [FREE Full text] [doi: 10.1186/s12910-018-0255-8] [Medline: 29510750]

12. Lupton D, Jutel A. 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. Soc Sci Med 2015 May;133:128-135. [doi: 10.1016/j.socscimed.2015.04.004] [Medline: 25864149]

13. Braun V, Clarke V, Rance N. How to use thematic analysis with interview data Internet. In: The Counselling and Psychotherapy Research Handbook. 2600 Virginia Avenue NW Suite 600 Washington, DC 20037 United States: SAGE Publications Ltd; Dec 08, 2017:183-197.

14. Rubanovich CK, Mohr DC, Schueller SM. Health App Use Among Individuals With Symptoms of Depression and Anxiety: A Survey Study With Thematic Coding. JMIR Ment Health 2017 Jun 23;4(2):e22 [FREE Full text] [doi: 10.2196/mental.7603] [Medline: 28645891]

15. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a Personal Smartphone Custom App to Assess the Patient Health Questionnaire-9 (PHQ-9) Depressive Symptoms in Patients With Major Depressive Disorder. JMIR Ment Health 2015;2(1):e8 [FREE Full text] [doi: 10.2196/mental.3889] [Medline: 26543914]

16. Vehling S, Kissane DW, Lo C, Glaesmer H, Hartung TJ, Rodin G, et al. The association of demoralization with mental disorders and suicidal ideation in patients with cancer. Cancer 2017 Sep 01;123(17):3394-3401 [FREE Full text] [doi: 10.1002/cncr.30749] [Medline: 28472548]

17. Scott MA, Wilcox HC, Schonfeld IS, Davies M, Hicks RC, Turner JB, et al. School-based screening to identify at-risk students not already known to school professionals: the Columbia suicide screen. Am J Public Health 2009 Feb;99(2):334-339. [doi: 10.2105/AJPH.2007.127928] [Medline: 19059865]

18. Bradford S, Rickwood D. Acceptability and utility of an electronic psychosocial assessment (myAssessment) to increase self-disclosure in youth mental healthcare: a quasi-experimental study. BMC Psychiatry 2015 Dec 01;15:305 [FREE Full text] [doi: 10.1186/s12888-015-0694-4] [Medline: 26627041]

19. Kauer SD, Reid SC, Crooke AHD, Khor A, Hearps SJC, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]

20. Bastiaansen JA, Meurs M, Stelwagen R, Wunderink L, Schoevers RA, Wichers M, et al. Self-monitoring and personalized feedback based on the experiencing sampling method as a tool to boost depression treatment: a protocol of a pragmatic randomized controlled trial (ZELF-i). BMC Psychiatry 2018 Sep 03;18(1):276 [FREE Full text] [doi: 10.1186/s12888-018-1847-z] [Medline: 30176845]

21. Boscaglia N, Clarke DM. Sense of coherence as a protective factor for demoralisation in women with a recent diagnosis of gynaecological cancer. Psychooncology 2007 Mar;16(3):189-195. [doi: 10.1002/pon.1044] [Medline: 16869021]

22. Segu S, Tataria R. Paediatric suicidal burns: A growing concern. Burns 2016 Jun;42(4):825-829. [doi: 10.1016/j.burns.2015.12.012] [Medline: 26803366]

23. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6:27882 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]

24. Hsin H, Torous J. Creating boundaries to empower digital health technology. BJPsych Open 2018 Jul;4(4):235-237 [FREE Full text] [doi: 10.1192/bjo.2018.37] [Medline: 29998820]

25. Silow-Carroll S, Smith B. Clinical management apps: creating partnerships between providers and patients. Issue Brief (Commonw Fund) 2013 Nov;30:1-10. [Medline: 24312989]

26. Van Velthoven MH, Smith J, Wells G, Brindley D. Digital health app development standards: a systematic review protocol. BMJ Open 2018 Aug 17;8(8):e022969 [FREE Full text] [doi: 10.1136/bmjopen-2018-022969] [Medline: 30121614]

XSL•FO
RenderX

27.   Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and Interest in Health Apps Among Those With Mental Health Needs: Survey and Focus Group Study. J Med Internet Res 2018 Jun 11;20(6):e10141 [FREE Full text] [doi: 10.2196/10141] [Medline: 29891468]

28.   Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AAR, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015 Feb 16;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]

Abbreviations

PABAK: prevalence-adjusted and bias-adjusted κ
PHQ-9: Patient Health Questionnaire-9

Original Paper

Quality of Psychoeducational Apps for Military Members With Mild Traumatic Brain Injury: An Evaluation Utilizing the Mobile Application Rating Scale

Chelsea Jones[1,2], MSc; Kaitlin O'Toole[1], MSc; Kevin Jones[3], BSc; Suzette Brémault-Phillips[1], PhD

[1]Heroes in Mind, Advocacy and Research Consortium (HiMARC), Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada
[2]Canadian Forces Health Services, Department of National Defence, Government of Canada, Ottawa, ON, Canada
[3]Edmonton North Primary Care Network, Edmonton, AB, Canada

Corresponding Author:
Chelsea Jones, MSc
Heroes in Mind, Advocacy and Research Consortium (HiMARC)
Faculty of Rehabilitation Medicine
University of Alberta
8205 - 114 Street NW
Edmonton, AB, T6G 2G4
Canada
Phone: 1 780 492 0404
Email: cweiman@ualberta.ca

Abstract

Background: Military personnel have an elevated risk of sustaining mild traumatic brain injuries (mTBI) and postconcussion symptoms (PCS). Smartphone apps that provide psychoeducation may assist those with mTBI or PCS to overcome unique barriers that military personnel experience with stigma and access to health care resources.

Objective: This study aims to (1) use the Mobile Application Rating Scale (MARS) to evaluate smartphone apps purporting to provide psychoeducation for those who have sustained an mTBI or a PCS; (2) explore the relevance, utility, and effectiveness of these apps in facilitating symptom management and overall recovery from mTBI and PCS among military personnel; and (3) discuss considerations pertinent to health care professionals and patients with mTBI when considering the use of mobile health (mHealth), including apps for mTBI psychoeducation.

Methods: A five-step systematic search for smartphone apps for military members with mTBI or PCS was conducted on January 31, 2020. Cost-free apps meeting the inclusion criteria were evaluated using the MARS and compared with evidence-based best practice management protocols for mTBI and PCS.

Results: The search yielded a total of 347 smartphone apps. After applying the inclusion and exclusion criteria, 13 apps were subjected to evaluation. Two apps were endorsed by the US Department of Veterans Affairs and the US Department of Defense; all the others (n=11) were developed for civilians. When compared with evidence-based best practice resources, the apps provided various levels of psychoeducational content. There are multiple considerations that health care professionals and those who sustain an mTBI or a PCS have to consider when choosing to use mHealth and selecting a specific app for mTBI psychoeducation. These may include factors such as the app platform, developer, internet requirement, cost, frequency of updates, language, additional features, acknowledgment of mental health, accessibility, military specificity, and privacy and security of data.

Conclusions: Psychoeducational interventions have a good evidence base as a treatment for mTBI and PCS. The use of apps for this purpose may be clinically effective, cost-effective, confidential, user friendly, and accessible. However, more research is needed to explore the effectiveness, usability, safety, security, and accessibility of apps designed for mTBI management.

KEYWORDS

XSL•FO
RenderX

Introduction

Background

Mobile health (mHealth) is an emerging field, with health care professionals increasingly using apps as part of clinical practice [1]. Apps geared toward well-being and health promotion have drastically increased in number and availability and may offer new opportunities for individuals seeking immediate and confidential health care [2,3]. One population that has the potential to benefit from the use of such apps is public safety personnel and military members [4]. Canadian Armed Forces (CAF) service members (SMs), similar to other global militaries, have significant mental and physical health challenges associated with unique occupational stressors associated with military duties [4,5]. Military service commonly involves high-risk activities during physical training, daily trade-related tasks, overseas deployment, and responses to natural disasters among others. Their involvement in such activities can increase their likelihood of sustaining physical and mental health injuries, including mild traumatic brain injury (mTBI) [5,6].

mTBI

mTBI, also known as concussion, is defined as a temporary change in brain functioning caused by an insult to the head with a period of posttraumatic amnesia lasting less than a day [7,8]. In contrast, moderate and severe TBI include changes in brain functioning resulting from a head insult causing periods of posttraumatic amnesia lasting longer than a day and often a period of hospitalization in an acute care facility and/or tertiary rehabilitation [7,8]. Symptoms of mTBI may include headaches, fatigue, nausea, sensitivity to light and sound, visual disturbances, cognitive dysfunction, memory loss, sleep disturbances, balance or vestibular issues, emotional disturbances, seizures, and loss of consciousness to name a few [7-20]. Symptoms of mTBI generally resolve within 2 weeks when no additional physical or mental comorbidities and extenuating factors are present [7,8]. If 3 or more symptoms of mTBI persist for longer than 3 months, a diagnosis of postconcussion symptoms (PCS) may be made [5,7,8].

Incidence of mTBI and PCS Among Military Populations

The cause of an mTBI varies among CAF-SMs, with some occurring as a result of motor vehicle collisions, falls, sports, explosions, or other forces related to combat and military training [9-16,18]. Rates of mTBI prevalence and severity of symptoms vary by element (ie, Army, Navy, or Air Force), age, gender, trade or profession, and unit [18]. Military members who experience an mTBI in combat may be at risk of developing career-limiting medical conditions [16]. As of 2019, mTBI affected 1 in 25 CAF-SMs, with 5.7% female and 3.9% male CAF-SMs diagnosed with mTBI over a 5-year period (2012-2017) [18]. Notably, this was after the completion of CAF's involvement in Operation Iraqi Freedom (OIF) and Operation Enduring Freedom (OEF), indicating that these incidences of mTBI took place largely outside of combat zones [18]. Among CAF-SMs deployed to Afghanistan during OEF between 2009 and 2012, 5.2% self-reported experiencing an mTBI, 21% of whom noted PCS [5,7]. In comparison, US studies among military populations report mTBI rates of 12%

to 22.8% during OEF and OIF, with PCS rates of 15.8% to 35% [9-11,14]. The UK Armed Forces report a 4.4% mTBI prevalence among SMs deployed into these global conflicts [11]. Although numbers vary greatly between different global militaries, the evidence base consistently demonstrates higher rates of mTBI and PCS in military personnel than in civilian populations. Similarly, although the incidence of PCS among the global civilian population has been estimated at 15% [8], it is well documented that this rate is elevated among military populations, with global estimates ranging from 15.8% to 35% [5,9,12,14-21]. This is because of a host of factors that are more prevalent among military populations than civilians, including a higher incidence of mental health disorders, exposure to traumatic experiences, previous mTBI, stigma, and a general lack of knowledge about mTBI [9,12,14-16].

Military members experience a higher incidence of posttraumatic stress disorder (PTSD), anxiety, and depression, which can have significant functional implications when co-occurring with mTBI. mTBI or TBI and mental health disorders, such as PTSD, can co-occur from the same or separate traumatic incidents [6,7,9-16]. The presence of trauma, as well as previously diagnosed neurological or mental health disorders, has been demonstrated to exacerbate mTBI symptoms and may be a major factor in the presence, longevity, and severity of PCS [5,12,13]. Additionally, it is possible that military members have had multiple previous mTBI that may or may not have been formally diagnosed [18]. The compounding effects of subsequent mTBI have been researched in recent years; however, the severity, longevity, and specificity of these symptoms and subsequent dysfunction they may cause are still widely unknown [7,8].

In addition to symptoms and stressors directly attributed to mTBI, psychosocial stressors may also be experienced by military members. Such stressors may include social and geographical isolation as well as concerns regarding medical employment limitations. It is widely acknowledged that mTBI is underreported both in the CAF and other global militaries because of several factors including stigma, fear of career implications, and ignorance of the potential seriousness of mTBI [9,12,18]. Seeking medical care and receiving an mTBI diagnosis may result in time away from work or absences from courses, training exercises, or deployments [13,16,18]. Military members may not have awareness that resources and interventions available through primary care, physical rehabilitation, and mental health could assist with recovery from mTBI or PCS. Widespread education about mTBI and treatment options that reduce perceived or actual stigma and threats to careers may be effective in reducing the negative impact of mTBI and PCS.

Psychoeducation for mTBI Among Military Populations

Various interventions for treating mTBI symptoms have been studied among military populations [15-19,21]. A 2015 study reviewing the effectiveness of interventions for military members with mTBI, PCS, and mental health comorbidities isolated 4 categories of interventions: psychoeducation, psychotherapy, cognitive rehabilitation therapy, and integrated behavioral health interventions [15]. Psychoeducational

XSL•FO

RenderX

interventions have a strong evidence base both as treatments for mTBI and as supplements to therapies for mental health disorders in both the acute and chronic phases of the illness [8,14,15,17,19-21]. Access to appropriate and timely psychoeducation is important to facilitate timely recovery from mTBI [8,15,17,19-21]. Providing psychoeducation directly after sustaining a mTBI and during the chronic phase (ie, PCS) has been demonstrated to reduce the impact and longevity of somatic symptoms and the potential exacerbation of mental health distress [8,15,17-20].

Military members require psychoeducational interventions that are clinically effective, cost-effective, user friendly, available in multiple environments, secure, and confidential [3]. This is particularly important when in-person therapy is not possible, such as during deployment, natural disaster response, or a pandemic. At such times, it is essential that clinicians explore more novel interventions and modes of service delivery, which may include the use of smartphone apps. As apps have evolved with better accessibility, usability, and quality, the delivery of psychoeducational material for behavioral change and health improvement via this method has become more common. This paper will refer to *apps* as opposed to *applications* in accordance with recommendations Lewis et al [22] in 2014.

mHealth and mTBI

As with the civilian population, the use of health apps is becoming more widespread within military populations [3,23]. A 2018 scoping review of mental health mobile apps for use by military members reviewed the current literature aimed at determining whether or not mobile apps are perceived as an acceptable form of mental health support [23]. Studies included in the review addressed app utilization through an assessment of users' general attitudes of the app, perceived ease of use, and whether they would recommend the app to others [23]. Although the majority of the studies were conducted with the US military, the results of the review overwhelmingly indicated that military members were generally willing to use apps [23] and viewed mobile apps as being an ideal supplement to traditional health care [3,23].

Since the first appearance of an mTBI-based app in 2009, apps specific to mTBI have been rapidly produced and evolving [24]. A 2018 review searching a wide variety of available mTBI-focused apps found 5 general categories: (1) education and prevention, (2) diagnostic assessment, (3) head impact sensors, (4) symptom tracking, and (5) treatment [25]. The most common type of available mTBI apps, and arguably the most controversial in the evidence-based literature, are those in the diagnostic assessment category, which are designed and/or marketed to sports medicine professionals or the general public for use in the recognition and assessment of concussion [24]. Although diagnostic assessment apps may widen the opportunity to identify mTBI, they also provide the potential for less-qualified individuals to use such apps inappropriately [24]. To date, only a limited number of medical devices have been approved or cleared by the Food and Drug Administration to aid in the diagnosis, treatment, or management of head injury, and these do not yet include smartphone apps [26]. Research on apps specific to mTBI has largely focused on the diagnostic

abilities of apps and not on the quality of psychoeducational content that these apps provide. This gap in the literature relates especially to the military medical context.

There are both potential benefits and challenges associated with using apps for health and behavioral change. Benefits include decreased stigma and improved privacy, immediate access to psychoeducational content, reduced wait times to access resources, less administrative burden for appointment scheduling, and the ability to track symptoms and share information with health care providers [2,3,27]. There is also evidence to suggest that health care apps may increase help-seeking behaviors and engagement with health care services [27]. Specific to military personnel, the use of mobile technologies for mental health support and care may be a desirable option for many military members and veterans who fear stigmatization and the career implications of engaging with health care systems [23]. Along with the potential benefits of health care apps are potential challenges. Some apps may have a limited evidence base or health care professional involvement during their development phase. This is particularly problematic given that apps in development are not required to undergo any certification or regulatory process, and as many health apps do not utilize peer-reviewed research, their purported claims may not be backed by evidence and may be misleading [2,3,28,29]. Although multiple apps are available specifically for TBI, few guidelines exist to assist individuals in evaluating whether the value of such an app is supported by evidence [24,25].

Purpose

The purpose of this evidence-based app review was to (1) use the Mobile Application Rating Scale (MARS) [2] to evaluate smartphone apps that advertise the provision of psychoeducational support for those who have sustained an mTBI or a PCS, (2) explore the relevance, utility, and effectiveness of these apps to facilitate symptom management and overall recovery from mTBI among military personnel, and (3) discuss considerations pertinent to health care professionals and patients with mTBI when considering the use of mHealth, including apps for mTBI psychoeducation. It is hypothesized that multiple high-quality apps exist specific to mTBI in the military population that provide evidence-based, population-specific psychoeducation for mTBI.

Methods

An app search was conducted on January 31, 2020. The Google Search Engine, the Google Play Store (Canadian) and the Apple App Store (Canadian) were the 3 platforms used for the search. Initial search terms employed with the Google search engine included "military" and "mtbi" or "mild traumatic brain injury" or "concussion" and "apps" or "applications" or "mobile device applications." Google was selected as the database because of its familiarity, popularity as a search tool, and accessibility from CAF computers. A Google search also provides the user with peer recommendations, which can provide more information about the usefulness of the app for specific populations and may describe features in more detail than the description provided by the 2 app stores. Searches were then conducted on the Apple App Store and the Google Play Store using the terms

"concussion" and "mTBI." The Department of Defense Mobile Health Practice Guide (3rd ed) was also utilized as a starting point; however, it had not been recently updated and did not yield any apps that were not found via the Google search [3]. Review of the apps involved 5 rounds of elimination (Figure 1) and the evaluation of the selected apps using the MARS.

Figure 1. Summary of the 5 rounds of elimination to analyze the apps. mTBI: mild traumatic brain injury.
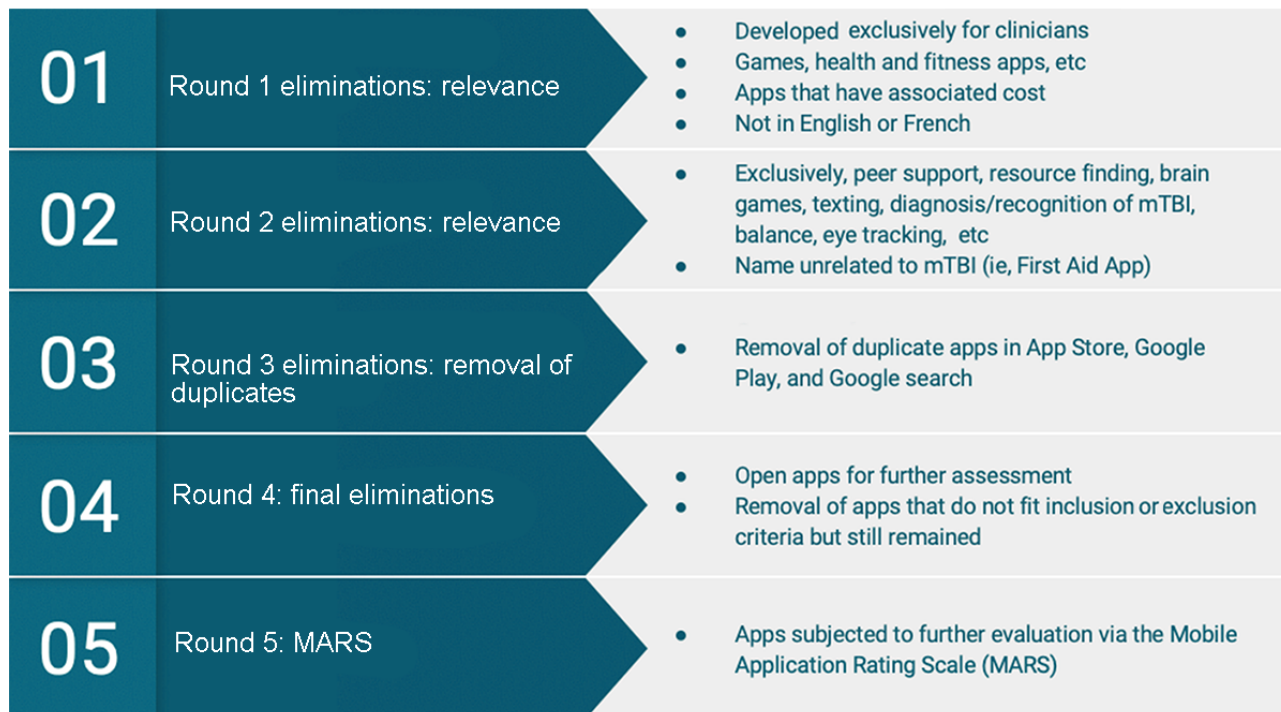
Rounds of Elimination

In Round 1, the writer screened descriptions from the app stores and excluded apps that were irrelevant, which included games and health and fitness apps unrelated to mTBI. Apps were also excluded if they had an associated cost, were specifically created for use by health care clinicians, were not in English or French, and were not accessible by the author on either the Apple App Store or the Google Play store. In Round 2, apps that were meant for peer support purposes, brain games, text messaging apps, for individuals aged 18 years or younger, and/or specific to 1 symptom of mTBI were also excluded. These included eye-tracking apps, balance apps, and apps aiming to diagnose an mTBI (such as sport-specific sideline apps) that did not have an associated psychoeducational component. Apps were excluded if they did not specifically address mTBI or concussion in the description or title and/or were designed for other purposes (ie, first aid apps). This was because of the assumption that if an app is not specifically designated for concussions or mTBI and does not list this in the name or description, it is unlikely that a person would recognize and choose the app. In Round 3, the remaining apps were compared from the Apple App Store with the Google Play Store to remove duplicates. In elimination Round 4, the writer opened the apps using an iPhone 8 iOS 12.3 (Apple Inc) and Google Pixel XL with Android 9.0. Apps were then excluded if they were intended to be used alongside an in-person therapist or sports coach, or if they otherwise did not include a psychoeducational component for

a patient recovering from an mTBI. If the app remained and its purpose was for mTBI detection but it had an educational component, it was included for further consideration. Round 5 involved the use of the MARS for evaluation. A literature search of the 13 apps that remained at Round 5 was conducted using the Scopus, Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica dataBASE (EMBASE), and Cumulative Index of Nursing and Allied Health Literature (CINAHL) Plus databases. The aim of this literature search was to determine whether studies using the apps had been published in the evidence-based literature.

MARS

Once the remaining 13 apps were screened and selected for inclusion in this review, they were evaluated by 3 raters using the MARS [2]. The MARS was created by Stoyanov et al [2] in 2015 for developing an mHealth app rating tool that was reliable, multidimensional, and which could provide a framework to trial, classify, and rate apps. Stoyanov et al [2] developed the MARS and found a high level of interrater reliability for overall MARS scores after evaluating 50 health and well-being apps. The components of the MARS were selected following a literature review to determine existing websites and app evaluation tools and after consulting a multidisciplinary advisory team that included psychologists and mHealth developers [2]. The resulting 3 MARS categories, app quality, app subjective quality, and app specific, were chosen to merge existing mHealth evaluation criteria into a format that

was not overly technical, difficult to use, or specific to any one health domain [2].

The app quality category is broken down into 4 subsections: (A) engagement, (B) functionality, (C) aesthetics, and (D) information (Table 1) [2]. Each subsection is averaged with a mean score out of 5. Engagement (A) is addressed using 5 questions and explores the ability of the apps to hold a user's attention by rating (1) how interesting it was to use, (2) how well-tailored it was to the targeted users, (3) how fun or entertaining it was, and (4) how well it could be customized for each user. Functionality (B) has 4 questions that rate how well the app works for the user. This includes how intuitive the gestures and icons are, how well the components worked when trialed, and how easy it is for a new user to learn. Aesthetics (C) is the shortest subsection, with 3 questions addressing the quality of the visual appeal and graphics. Information (D) is the last and largest subsection, with 7 questions exploring the credibility of the developer and content in the app and whether the app's efficacy has been tested and reported in published research. This section also addresses how the information within the app is presented to users (eg, is the information accessible to the target audience?) and if the information is of sound quality [2].

App-specific items are available to assess how effectively the app is perceived to address or impact a targeted health behavior [2]. In this study, the targeted health behavior was mTBI management. The 6 app-specific items indicated the rater's belief that the assessed app would be able to increase the knowledge of, attitude toward, and intention to change health management strategies of someone with an mTBI. The app-specific items also address how strongly the rater felt that the use of the app would actually result in a change in this targeted health behavior. These items were rated on a scale from 1 (strongly disagree) to 5 (strongly agree).

Use of the MARS to rate apps has numerous benefits. Its consideration of a health app based both on design elements (ie, color, graphic resolution, and layout) and content is essential for evaluating quality [2]. Newly published research exploring the quality of apps intended to equip individuals to self-manage their disease reports that apps that have been tested and rated highly on the MARS tool are trustworthy for clinicians to recommend to clients [30]. The MARS was selected to facilitate evaluation of apps in this research project for these features and because it is more comprehensive than the subjective star ratings featured on app stores [4] and is customizable for this specific research focus [2].

Three raters were trained in the MARS by watching the *MARS training video* posted by Stoyanov on YouTube [31]. These raters included a PhD candidate and a CAF Occupational Therapist, an MSc Occupational Therapist, and an American College of Sports Medicine Certified Clinical Exercise Physiologist. The finalized app selection was evaluated by the writers using the MARS January 30, 2020, with apps downloaded onto a Samsung Galaxy Tab A SM-T350 with Android version 7.1.1 software, Google Pixel XL with Android 9.0 software, and/or iPhone 8 iOS 12.3. The results from the raters were averaged to create the final MARS score. The Ontario Neurotrauma Foundations' Guideline for Concussion/Mild Traumatic Brain Injury and Persistent Symptoms: 3rd Edition, 5th International Conference on Concussion in Sport Concussion Consensus Statement and relevant military-specific evidence-based publications [15,18-20] were used to evaluate the quality and accuracy of the information (Information: Section D of MARS) provided within each app [7,8].

Table 1. The Mobile Application Rating Scale.

Sections and subsections	Question number and headings
App quality ratings	
A. Engagement	1. Entertainment 2. Interest 3. Customization 4. Interactivity 5. Target group
B. Functionality	1. Performance 2. Ease of use 3. Navigation 4. Gestural design
C. Aesthetics	1. Layout 2. Graphics 3. Visual appeal
D. Information	1. Accuracy of app description 2. Goals 3. Quality of information 4. Quantity of information 5. Visual information 6. Credibility 7. Evidence base
App subjective quality	
N/A[a]	1. Would you recommend this app to people who might benefit from it? 2. How many times do you think you would use this app in the next 12 months if it was relevant to you? 3. Would you pay for this app? 4. What is your overall star rating of the app?
App specific	
N/A	1. Awareness 2. Knowledge 3. Attitudes 4. Intention to change 5. Help seeking 6. Behavior change

[a]N/A: not applicable.

Results

The search yielded 347 apps that were subjected to 5 rounds of elimination (Figures 1 and 2). The results of the Apple App Store and Google Play Store searches conducted on January 31, 2020, resulted in 13 apps (refer to Figure 2 for a flowchart of the exclusion process). When the Round 1 exclusion criteria were applied, 254 apps from the 347 original apps were excluded as they did not meet the inclusion criteria regarding language, target audience, cost, or relevance. In Round 2, 60 apps were further eliminated because of the goal of the app not specifically targeting psychoeducation for individuals with mTBI. A total of 12 duplicate apps were removed in Round 3. After 20 apps were opened for additional screening, 7 apps were removed in Round 4. This resulted in a total of 13 apps that were evaluated by the writer using the MARS. Information about each of the 13 apps is available in Tables 2 and 3.

Scores for each individual section of the MARS are listed in Table 3. Scores ranged from 1 to 5 out of 5, with higher scores indicating higher engagement, function, information, and overall quality. The interrater reliability was 90% among the raters in all sections of the MARS.

Figure 2. App search results. MARS: Mobile Application Rating Scale.

XSL·FO
RenderX
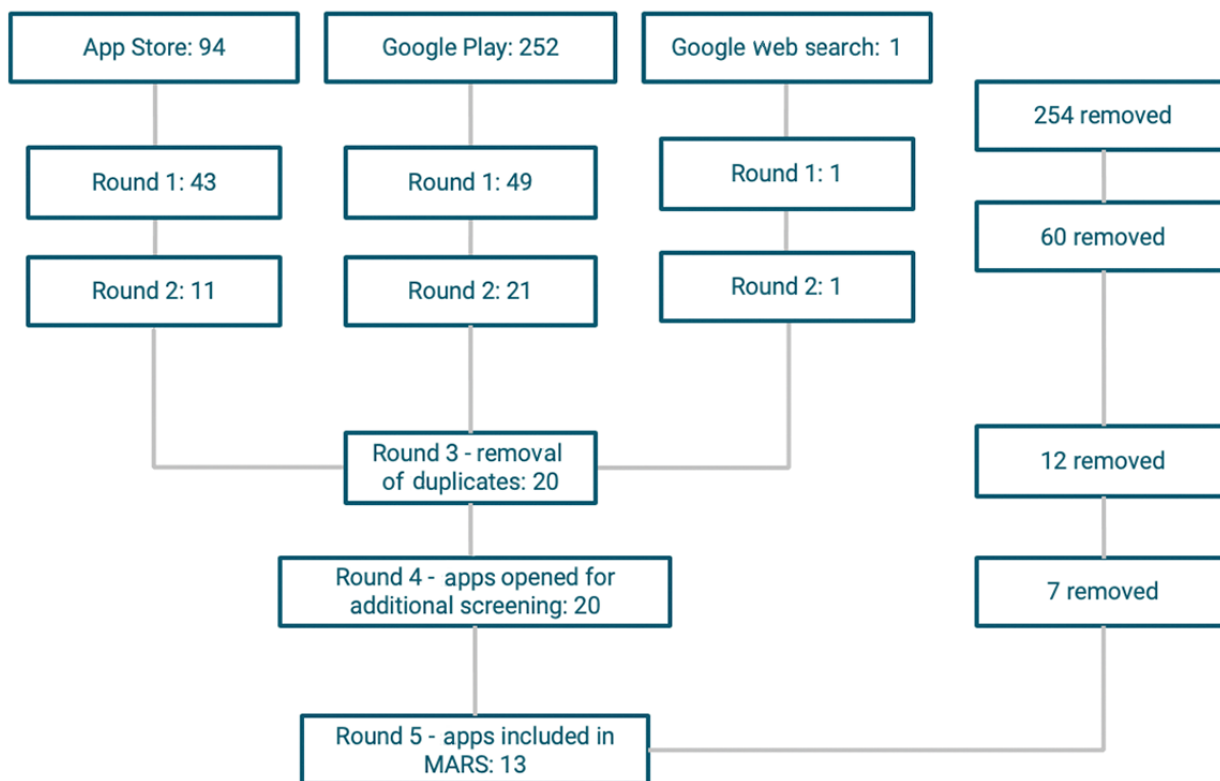
Table 2. Individual features of the 13 apps.

App names	Developers	Availability	Log-in	Internet	mTBI[a] Ax.	Mindfulness	CBT[b]	Goal setting	Symptom tracking	Return to activity	French	Military
Concussion Coach	US Department of Veterans Affairs	Google Play and Apple App Store	No	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes
World Rugby Concussion App	Mobanode	Google Play and Apple App Store	No	No	Yes	No	No	No	Yes	Yes	No	No
CDC Heads Up! Concussion and Helmet Safety App	Centers for Disease Control (CDC)	Google Play and Apple App Store	No	No	No	No	No	No	No	Yes	No	No
Concussion Awareness	Hockey Canada	Google Play and Apple App Store	No	Yes	No	No	No	No	No	Yes	No	No
Concussion Ed	Parachute	Google Play and Apple App Store	No	No	Yes	No	No	No	Yes	Yes	Yes	No
LifeArmor	National Centre for Telehealth and Technology	Google Play and Apple App Store	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes
CCMI Concussion Tracker	CCM Inc	Google Play and Apple App Store	Yes	Yes	Yes	No	No	No	Yes	No	No	No
Concussion Smart	ABI Ireland/Medtronic	Apple App Store	No	No	Yes	No	No	No	Yes	Yes	No	No
Concussion NI	Sport Northern Ireland	Google Play	No	Yes	No	No	No	No	No	Yes	No	No
Concussion Quick Check	American Academy of Neurology	Google Play and Apple App Store	No	No	Yes	No	No	No	No	Yes	No	No
Concussion Info	Programming is Fun	Google Play	No	No	No	No	No	No	No	No	No	No
How to Treat a Concussion	nermine_92	Google Play	No	No	No	No	No	No	No	No	No	No
PACE Concussion	PACE Concussion	Google Play	Yes	N/A[c]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

[a]mTBI: mild traumatic brain injury.

[b]CBT: cognitive behavioral therapy.

[c]N/A: not applicable.

XSL·FO
RenderX

Table 3. Mobile Application Rating Scale scores for each of the 13 apps.

App names	Section A: engagement	Section B: functionality	Section C: aesthetics	Section D: information	App quality	Section E: subjective	Section F: app specific
Concussion Coach	3.8	4.3	4.7	4.6	4.4	5.0	4.3
World Rugby Concussion App	4.4	5.0	4.8	4.3	4.6	4.5	4.0
CDC Heads Up! Concussion and Helmet Safety App	3.8	5.0	5.0	4.3	4.5	4.0	3.8
Concussion Awareness	2.4	2.8	2.3	2.7	2.6	1.8	2.7
Concussion Ed	3.6	4.8	4.3	4.0	4.2	3.5	4.0
LifeArmor	3.4	4.0	3.7	4.0	3.8	3.0	3.0
CCMI Concussion Tracker	3.4	4.5	4.7	4.1	4.2	3.3	3.7
Concussion Smart	3.2	4.8	4.3	3.4	3.9	3.0	3.7
Concussion NI	2.8	4.0	4.7	4.0	3.9	3.0	3.7
Concussion Quick Check	2.6	3.8	3.0	3.6	3.3	2.0	3.0
Concussion Info	2.4	2.3	2.3	2.3	2.3	1.8	2.7
How to Treat a Concussion	2.6	3.8	3.3	2.3	3.0	1.5	2.7
PACE Concussion	N/A[a]	1.0	3.0	1.3	1.8	1.0	1.0

[a]N/A: not applicable.

Discussion

The decision of whether to use mHealth apps, as well as selecting which apps are appropriate, can be complex even without the addition of military health care, culture, and contexts. There are multiple considerations that need to be taken into account by both the person with an mTBI and/or a clinician using an app to assist with a psychoeducational intervention. To date, most of the available literature surrounding the use of mHealth and smartphone apps for mTBI focuses on diagnostic assessment or concussion recognition, and few studies have been published regarding app quality for psychoeducation and mTBI management. On the basis of the results of this search and subsequent MARS evaluations, 5 apps demonstrated superior app quality domains, including engagement, functionality, aesthetics, and information. The 5 highest-scoring apps included *Concussion Coach, World Rugby Concussion, CDC Heads Up!, Concussion Ed, and LifeArmor* (Tables 2 and 3; Figure 3). For the app-specific component of evidence-based merit (Information: Section D) for mTBI psychoeducation, the highest-scoring apps included *Concussion Coach*, *World Rugby Concussion*, *CDC Heads Up!*, *Concussion Ed*, and *LifeArmor* (Table 3; Figure 3).

Rating apps using the MARS highlighted features of the apps that may be beneficial to individuals with mTBI. Of the top apps, *World Rugby Concussion*, *CDC Heads Up!*, and *Concussion Ed* addressed acute mTBI recognition and management by providing parents, coaches, and individuals with instructions to follow in the event of sustaining an mTBI in a sporting context (ie, *remove from play*); these instructions

are consistent with the best practices for sport-related mTBI [7,8]. All 5 apps provided a feature for symptom tracking. They also provided psychoeducational material for mTBI management and recovery, including information on mTBI acute management, recognition, and return to activity. *Concussion Coach* had the most robust psychoeducational resources of all the apps, with extensive information on sleep, headaches, goal setting, and cognition, in line with current best practice literature for mTBI and PCS psychoeducation for civilians and military populations [7,8,16,18]. Only 2 of the apps provided additional links to find community support in the United States (*Concussion Coach* and *LifeArmor*). Overall, the psychoeducational content provided was consistent with the current evidence base for mTBI recognition and management for both civilian and military populations. At no point was the information observed contraindicated for the recognition, management, and recovery of mTBI. The top-5 rated apps provided references or a link to a web resource where references could be found. References were evidence-based materials and seminal publications on mTBI management and best practices.

Most app users will be familiar with the 5-star rating system of the Google Play Store and the Apple App Store as a subjective manner for users to provide feedback on apps. Within this star rating system, the scores of the aforementioned top apps, as rated by the MARS, varied greatly. In the Apple App Store, *Concussion Coach, World Rugby Concussion*, and *ConcussionEd* each had 5 out of 5 stars, with *CDC Heads Up!* and *LifeArmor* not having enough ratings to formulate an average score. In the Google Play Store, *CDC Heads Up!* had 5 out of 5 stars, whereas *World Rugby Concussion* had 4.5 stars,

Concussion Coach and *LifeArmor* had 4 stars, and *Concussion Ed* had 3 stars. It was noted that many of these scores were deemed as having less than 10 reviews. Although the 5-star rating system of the platforms could be helpful for some users deciding on whether to use the app, it is clear that the MARS provides greater overall reliability and validity because of being less subjective and having defined scoring criteria.

Although scoring from the MARS was helpful in assessing these health apps and providing a numeric score and ranking, there are many additional factors that need to be considered. These factors will be further discussed, including app platform, developer, internet requirement, cost, frequency of updates, language, additional features, acknowledgment of mental health, accessibility, and military specificity.

Figure 3. The top 5 scoring apps based on the Mobile Application Rating Scale (MARS) evaluation. Concussion Coach, World Rugby Concussion, Concussion Ed, CDC Heads Up!, and LifeArmor.

App Platform

On searching both the Apple App Store and the Google Play Store, it was evident that the Google Play Store contained more apps; however, the Apple App Store produced a more refined and relevant search results (Figure 1). The requirements to publish apps vary between the Apple App Store and the Google Play Store, which influence the quality and volume of apps available. The Apple App Store has more rigorous review processes before allowing an app to be available to the public, which results in a more restrictive selection [32]. In addition, apps that score poorly in their customer five-star rating system risk being removed from the platform [32]. The Apple App Store may be a better option for finding adequate and usable apps for psychoeducation as a person with active symptoms of mTBI may struggle with the volume of results and the need to filter those that are not relevant [29]. Availability in both the Google Play Store and the Apple App Store will maximize the availability of the app to the target audience. It is evident that a clear title and description is key for finding and using an app [29].

Developer

Ideally, the developer of the app is an organization whose policies and protocols are based on evidence-based, peer-reviewed literature, such as a government or an academic institution. The most common evidence-based apps are those that were developed through nationally competitive government or research funding and have undergone rigorous research ideally with randomized control trials [2]. None of the results yielded apps that had been researched to this degree. Additionally, health care app developers do not always involve medical professionals in app development; therefore, apps may have limited referencing and make misleading claims [33]. When health providers recommend apps, they need to be competent in the evidence-based literature on the topic to evaluate the apps for the quality of the content and ensure that

it is in line with the desired therapeutic interventions and outcomes [3]. Although this expectation may be realistic for health care professionals who are clinically trained in a military health care system, it could be an unattainable expectation for military members in distress from an mTBI who are exploring mobile resources for themselves [4].

Internet Requirement and Log-In

Most of the apps reviewed did not require an internet connection to function once downloaded and installed. This may be an asset for those apps intended for the military population as access to the internet may not always be available, especially when on deployment, in rural areas, or on a military exercise. Additionally, many smartphone users have limited data and are reliant on Wi-Fi. The requirement of a log-in and an account may benefit users in that they can save more of their customizations and information within the app and virtually share their progress with health care professionals. Regular access to the internet, however, may be required as well as the sharing of more personal information within the app. The top-rated apps, *Concussion Coach, World Rugby Concussion, CDC Heads Up!, Concussion Ed, and LifeArmor* (Figure 3), did not require internet access after the app was initially downloaded to the device. In addition, they did not require a log-in or an account that negated the need for additional personal data to be stored and transmitted.

Cost

Cost is also a factor in selecting the right app for health care needs, and health care professionals recommending the use of apps should consider the benefits and drawbacks of free versus paid apps. Free apps are more accessible and more likely to be used but may come with intrusive advertisements that may be distracting and confusing for someone experiencing cognitive, vestibular, or visuospatial dysfunction from mTBI. Paid apps may have fewer or no advertisements but are less likely to be downloaded [34]. A 2015 study concluded that although 93%

of smartphone users download apps, only 35.8% will purchase apps that have an associated cost [34]. All 13 apps in this study were free to download and did not provide an option to pay for an upgraded version of the app. Advertisements were either absent or subtle enough to not disrupt the user experience.

Up-to-Date

Several apps were outdated and had not undergone a version update in several years. All apps encouraged periods of complete rest after an mTBI; however, the evidence base is evolving toward a more active approach to recovery [7,17]. Although no regulations or standards exist regarding the frequency with which a developer should update an app, these apps should ideally be updated regularly enough to reflect the rapidly changing evidence base and recommendations regarding psychoeducation and mTBI management [28]. The 5 top-rated apps had received updates within 24 months of the app search, except the French version of *Concussion Ed* (last updated version in 2017).

Language

Concussion Ed was the only app available in both of Canada's official languages, English and French. All other apps reviewed were only available in English. If the CAF wishes to use a standard app across all regions, there is currently only 1 available option that will meet the official language needs of the country.

Return to Activity

Of the 13 apps reviewed, 6 contained education specific to return to activities, such as sports, work, military duty, and/or school. Of the top 5 rated apps, the *World Rugby Concussion, CDC Heads Up!,* and *Concussion Ed* included this component and had concrete, specific examples for activity grading of frequency, volume, and intensity for both physical and cognitive tasks. The suggested return to play, school, or work guidance was consistent with evidence-based and best practice recommendations [7,8]. Guidance on return to activity is a common reason civilians and military members alike seek guidance from health care professionals [17]. This is also an area that is rapidly evolving in the literature as recent studies increasingly promote a more active recovery in contrast to previous recommendations that advocated for a longer period of complete rest immediately after sustaining an mTBI [7,8,17].

Symptom Tracking

Of the 13 apps, 6 had a component of interaction with the app that allowed for customized day-to-day symptom tracking. Symptom tracking features addressed symptoms such as dizziness, headaches, balance issues, cognitive dysfunction, nausea, visual disturbances, sensitivity to noise, hearing difficulties, fatigue, sleep disturbances, and mood changes. Consistent with best practice recommendations, some apps utilized evidence-based outcome measures such as the Neurobehavioral Symptom Inventory to guide questions regarding symptom occurrence and severity [8,35]. Symptom tracking may have benefits such as allowing a patient with an mTBI to visualize improvement and better report their symptom status to their health care team. This may assist with decisions to re-engage with returning to certain activities. Conversely,

frequent tracking could conceivably foster hypervigilance and preoccupation with symptoms, especially among anxious individuals [25].

Acknowledgment of Mental Health

To provide evidence-based best practice care for mTBI, it is important that a holistic multidisciplinary approach is encouraged because of the variable experiences of those who experience mTBI [10]. This requires that mental health distress and diagnoses, such as PTSD, anxiety, and depression, are acknowledged as potential contributors to mTBI and PCS, especially in the military context where evidence of these comorbidities is strong [15,17-20]. The majority of apps reviewed did not contain reference to mental health. Although mental health symptomology and its comorbid relationship with mTBI is a key piece of the psychoeducational strategies recommended for mTBI management and intervention, only 1 app, *Concussion Coach,* explicitly acknowledged the comorbidity of mTBI and mental health conditions [8,15]. Furthermore, only 2 apps, *Concussion Coach* and *LifeArmor*, provided resources for mindfulness, goal setting, and cognitive behavioral therapy, which have been identified as components of a psychoeducational or integrated behavioral health intervention for mTBI and comorbid mental health conditions [15].

Diagnostic Assessment or Recognition

Of the 13 apps, 7 had a component of acute mTBI assessment or recognition. Several were based on well-known and validated diagnostic tools utilized in a health care context such as the Glasgow Coma Scale and the Sport Concussion Assessment Tool 3 or 5 (SCAT3/SCAT5) [7,36,37]. The SCAT3/SCAT5 is a commonly used and evidence-based mTBI screen recommended for use during sport [7,37]. Of the top 5 rated apps, 3 (*World Rugby Concussion*, *CDC Heads Up!*, and *Concussion Ed*) included this component. As all the apps in this review are indicated for use by the general public, the mTBI assessment and recognition could carry significant legal liability, which is a concern that has been voiced in other reviews [24,25]. As stated earlier, none of the available smartphone apps in Canada and the United States have been federally regulated or approved for use in diagnosing mTBI [25,26]. The diagnostic utility of apps in mTBI detection is a controversial topic with emerging evidence-based literature [38-40].

Accessibility

Readability and function of a concussion app needs to be appropriate for an individual experiencing mTBI symptoms; however, most apps contain layouts and fonts that would be difficult to read and process for individuals with visuospatial or cognitive symptoms of mTBI. As symptoms of concussion may include light sensitivity, difficulty reading, and visual disturbances, larger fonts, high visibility text, pictures or diagrams, and an intuitive interface are very important for usability [41]. Only *Concussion Ed* had a customizable interface that allowed the user to easily enlarge the text. In the opinion of the authors, the apps that scored in the top 5 had the best visibility and usability of the 13 apps assessed. With mTBI and its symptomatology being multifaceted and complex,

intervention and rehabilitation will vary widely depending on the needs of the person [20].

Military Specificity

As mentioned earlier, the use of "military" as a search term confounded the app search with multiple games and other nonrelevant apps. It is unrealistic to expect someone to search through all the options, especially if they are experiencing distress and cognitive dysfunction, as may be the case with military members who sustained an mTBI or a PCS and/or have a mental health condition [42]. Given the rate at which new apps are entering the marketplace, the open access nature of app stores may impact the effectiveness of starting a search there; consequently, search results are not guaranteed to be consistent [42].

Only 2 apps, *Concussion Coach* and *LifeArmor*, provided military-specific resources. The majority of apps, as well as evidence-based literature on mTBI, focus on sport-related mTBI, which is common in the civilian population. Although military populations experience these types of mTBI, apps specific to military contexts would be an asset as there are multiple complexities within military organizations, cultures, occupational roles, and the environment to which civilians are not exposed [13].

Military members are exposed to a variety of physical and psychosocial variables, which either in isolation or in combination can exacerbate the severity, longevity, and dysfunctionality of mTBI symptoms [9-16,43]. Psychosocial factors that are prevalent at a higher rate in military populations include increased geographical isolation, alcohol consumption, mental health diagnoses (ie, depression, anxiety, and PTSD), chronic pain, TBI, and sleep disturbances, all of which can exacerbate mTBI symptoms [9-16]. In addition, military contexts necessitate higher levels of cognitive functioning. Cognitive dysfunction can potentially result in decreased efficiency and effectiveness, along with an increased risk of harm to self, the unit, and a mission [18]. Reduced physical, mental, and/or cognitive functioning may result in involuntary release from the CAF because of the inability to meet the Universality of Service criteria for employment, deployment, and fitness [44]. Moreover, as a CAF-SM transitions from the CAF to veteran status, PCS may continue to contribute to challenges within the transition processes, the family unit, civilian employment, leisure activities, and self-care [21].

Blast injuries are also more unique to military populations, with a portion of the mTBI sustained by military members during OEF and OIF being potentially attributable to members being in close proximity to explosions [5,9-13,43]. A blast mTBI is an injury to the brain leading to dysfunction that is the result of an explosion or a blast [13,43]. Despite differences among mechanisms of injury, no significant variations in mTBI symptoms and PCS caused by blast versus blunt force have been identified apart from a blast mTBI preceding more severe hearing loss [13]. Finally, repeated exposure to low-level head trauma, such as being in proximity to low-level blasts, use of a ram, or other weapon utilization, is another mechanism of possible head trauma that is unique to military and paramilitary personnel being explored in the literature [45]. More research

is needed to determine whether these mechanisms cause differences in the presentation of mTBI and PCS, and whether military-specific mTBI needs to be addressed as a unique subset compared with civilian mTBI.

Health Care Use and Future Direction

There is potential for psychoeducational apps for mTBI to be utilized in a clinical setting [29]. A health care provider could recommend a vetted app to assist with education, reassurance, and potentially behavioral change. This may reduce the need for multiple follow-up appointments and could assist with symptom reporting and monitoring [3]. Some apps have a feature that allows data to be electronically provided to the health care professional, which could reduce the administrative burden on both the health care professional and the patient as long as data sharing, privacy, and security are considered [26]. Although there could be some benefits, this area of mHealth remains novel and not without significant issues [1,27-29,33].

Currently, health-related mTBI apps, including those providing psychoeducation and included in this review, are not formally regulated by government agencies, although *Concussion Ed* is endorsed by Health Canada [46]. The purpose of such regulation would be to provide the consumer with confidence that the product can be safely used [26]. Until this is changed, the onus is on app developers and the app provider (ie, Apple or Google) to provide the consumer with a well-documented and described product and a clear indication as to the intended target group of the app [24]. The abundance of available health apps and the rapid rate at which they continue to be released indicate that it will take considerable time for agencies or regulatory bodies to monitor a virtual market for regulation [25,26].

None of the apps investigated were found to have rigorous peer-reviewed research published on the app-specific effectiveness of the psychoeducational or mTBI management components. The lack of empirical research to demonstrate effectiveness may be related to the short time frame during which mHealth apps have emerged, the speed at which their availability changes as well as the focus on diagnostic apps opposed to psychoeducational apps [25].

The type and volume of data gathered from an electronic device when an app is downloaded and the details of the electronically signed end user license vary [46]. App creation, app use, and data storage may all occur in different countries with different laws and regulations regarding data privacy and sharing of data collected through electronic means [47]. Data sharing and privacy is a consideration that requires attention from researchers, health care professionals, and the general public when deciding on which app to utilize or if app utilization is appropriate at all [33]. Future systems of app evaluation and research would benefit from adding a component that considers data sharing, storage, and privacy.

Technology acceptance and usability studies are also lacking for apps related to mTBI [24]. There are few early feasibility and compliance studies for mTBI apps published with small sample sizes that are not specific to the psychoeducational components of the apps [25]. Investigation into the feasibility, logistics, security, IT compatibility, and acceptance by health

care providers and military members with mTBI is critical to the implementation of standardized practices in military health care systems. The rapid rate at which mHealth is evolving contrasts with the slower pace of traditional evidence-based research practices [25]. Exhaustive evaluation for effectiveness, efficacy, and usability through research may not be practical and new approaches to evidence may have to be considered [25]. New methods of research with novel tools, such as the MARS, may also need further consideration and acceptance to assist both with the rapid need for mHealth research and the regulation of health apps.

Study Limitations and Strengths

This study has a number of limitations. First, the search and identification of apps was limited to 1 day, and given the fast-paced release of new apps, it may not have captured all apps available till date. Second, there are specific concerns regarding the use of the MARS as a rating tool for health apps. The MARS does not address data sharing, security, and privacy, which are important components to consider when making decisions regarding mHealth utilization. Furthermore, the MARS involves several potentially subjective responses by the raters, most notably in the area of app subjective quality. For instance, question 2 asks, "Is the app interesting to use? Does it use any strategies to increase engagement by presenting its content in an interesting way?" [2]. As such questions may be interpreted subjectively, mitigation strategies were implemented. These included rater participation in standardized video training before rating the apps, engagement of multiple raters in the rating process, and averaging ratings across raters [2,29].

The main strength of this study is the systematic, methodological evidence-based approach undertaken to evaluate the apps, including the use of an evidence-based tool and multiple rounds of elimination. In addition, the researchers engaged in this project have clinical experience, are employed in clinical settings, and routinely work with military and civilian populations who have sustained mTBI. They are also skilled at providing psychoeducational content to address mTBI and support subsequent recovery. The execution of the *a priori* process for app rating, coupled with the educational and clinical experience of the researchers, contributed to the validity and rigor of the app evaluation results and subsequent knowledge synthesis of findings. As mHealth is a rapidly evolving field, the brief time from app search and evaluation to manuscript preparation facilitated timely knowledge translation and integration into clinical practice.

Conclusions

As a component of mHealth, smartphone apps have become widely available in recent years as app technology rapidly improves. Similar to civilians, military populations have also embraced the use of health apps, which may have advantages specific to the challenges and barriers faced by military personnel, including geographical isolation and stigma. Health apps have the potential to be an engaging and accessible means of providing psychoeducational information for mTBI management. Of the 13 apps reviewed in this study, 5 (*Concussion Coach, World Rugby Concussion, CDC Heads Up!, Concussion Ed,* and *LifeArmor)* were well-suited to provide evidence-based psychoeducational information on mTBI management. Only 2 (*Concussion Coach* and *LifeArmor*) contained information specific to military populations and addressed mental health information and strategies, which are a critical component of mTBI and PCS management and recovery [8]. Further research should investigate the applicability, technology acceptance, and usability specific to the utilization of psychoeducational apps among military members with mTBI at the patient and health professional level within military contexts such as garrison, training, and deployed environments.

Conflicts of Interest

None declared.

References

1. Buijink AW, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med 2013 Jun;18(3):90-92. [doi: 10.1136/eb-2012-100885] [Medline: 22923708]
2. 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]
3. Armstrong C, Edwards-Stewart A, Ciulla R, Bush N, Cooper D, Kinn J, et al. Department of Defense Mobile Health Practice Guide. Military Health System. 2017. URL: https://health.mil/Reference-Center/Publications/2019/08/14/US-DoD-Mobile-Health-Practice-Guide-Fourth-Edition-Sept-2018 [accessed 2020-05-04]
4. O'Toole K, Brown C. Evaluating the Quality of Resilience Apps for Military Members and Public Safety Personnel. J Mil Veteran Fam Health 2020:- (forthcoming).
5. Garber BG, Rusu C, Zamorski MA. Deployment-related mild traumatic brain injury, mental health problems, and post-concussive symptoms in Canadian Armed Forces personnel. BMC Psychiatry 2014 Nov 20;14:325 [FREE Full text] [doi: 10.1186/s12888-014-0325-5] [Medline: 25410348]
6. Health and Lifestyle Information Survey of Canadian Forces Personnel 2013/2014. Government of Canada. 2016. URL: https://www.canada.ca/content/dam/dnd-mdn/documents/health/health-and-lifestyle-survey-2013-2014.pdf [accessed 2020-04-15]

7. McCrory P, Meeuwisse W, Dvořák J, Aubry M, Bailes J, Broglio S, et al. Consensus statement on concussion in sport-the 5 international conference on concussion in sport held in Berlin, October 2016. Br J Sports Med 2017 Jun;51(11):838-847. [doi: 10.1136/bjsports-2017-097699] [Medline: 28446457]

8. Marshall S, Bayley M, McCullagh S, Berrigan L, Fischer L, Ouchterlony D, et al. Guidelines for Concussion /Mild Traumatic Brain Injury and Prolonged Symptoms. Ontario Neurotrauma Foundation. 2018. URL: https://onf.org/3rd-edition-guidelines-for-concussion-mild-traumatic-brain-injury-and-persistent-symptoms/ [accessed 2020-04-01]

9. Armistead-Jehle P, Soble J, Cooper D, Belanger H. Unique aspects of traumatic brain injury in military and veteran populations. Phys Med Rehabil Clin N Am 2017 May;28(2):323-337. [doi: 10.1016/j.pmr.2016.12.008] [Medline: 28390516]

10. Schwab K, Terrio HP, Brenner LA, Pazdan RM, McMillan HP, MacDonald M, et al. Epidemiology and prognosis of mild traumatic brain injury in returning soldiers: a cohort study. Neurology 2017 Apr 18;88(16):1571-1579. [doi: 10.1212/WNL.0000000000003839] [Medline: 28314862]

11. Hoge CW, McGurk D, Thomas JL, Cox AL, Engel CC, Castro CA. Mild traumatic brain injury in US soldiers returning from Iraq. N Engl J Med 2008 Jan 31;358(5):453-463. [doi: 10.1056/NEJMoa072972] [Medline: 18234750]

12. Rona RJ, Jones M, Fear NT, Hull L, Murphy D, Machell L, et al. Mild traumatic brain injury in UK military personnel returning from Afghanistan and Iraq: cohort and cross-sectional analyses. J Head Trauma Rehabil 2012;27(1):33-44. [doi: 10.1097/HTR.0b013e318212f814] [Medline: 22241066]

13. Doneva SP. Mild traumatic brain injury in military service personnel: key issues and considerations. J Mil Veteran Fam Health 2018 Oct;4(2):121-135. [doi: 10.3138/jmvfh.2017-0012]

14. Janak JC, Cooper DB, Bowles AO, Alamgir AH, Cooper SP, Gabriel KP, et al. Completion of multidisciplinary treatment for persistent postconcussive symptoms is associated with reduced symptom burden. J Head Trauma Rehabil 2017;32(1):1-15. [doi: 10.1097/HTR.0000000000000202] [Medline: 26709579]

15. Cooper DB, Bunner AE, Kennedy JE, Balldin V, Tate DF, Eapen BC, et al. Treatment of persistent post-concussive symptoms after mild traumatic brain injury: a systematic review of cognitive rehabilitation and behavioral health interventions in military service members and veterans. Brain Imaging Behav 2015 Sep;9(3):403-420. [doi: 10.1007/s11682-015-9440-2] [Medline: 26330376]

16. Garber BG, Rusu C, Zamorski MA, Boulos D. Occupational outcomes following mild traumatic brain injury in Canadian military personnel deployed in support of the mission in Afghanistan: a retrospective cohort study. BMJ Open 2016 May 4;6(5):e010780 [FREE Full text] [doi: 10.1136/bmjopen-2015-010780] [Medline: 27147386]

17. Quatman-Yates CC, Hunter-Giordano A, Shimamura KK, Landel R, Alsalaheen BA, Hanke TA, et al. Physical therapy evaluation and treatment after concussion/mild traumatic brain injury. J Orthop Sports Phys Ther 2020 Apr;50(4):CPG1-CP73. [doi: 10.2519/jospt.2020.0301] [Medline: 32241234]

18. Besemann M, Godsell P, Mahoney N, Hawes R. Traumatic Brain Injury in the Canadian Armed Forces: Epidemiology, Management, and Rehabilitation. In: Canadian Institute for Military and Veteran Health Research Forum. 2019 Presented at: MVHR'19; October 21-23, 2019; Gatineau, Quebec.

19. Radomski M, Davidson L, Voydetich D, Erickson M. Occupational therapy for service members with mild traumatic brain injury. Am J Occup Ther 2009;63(5):646-655. [doi: 10.5014/ajot.63.5.646] [Medline: 19785265]

20. Radomski M, Weightman M, Davidson L. Mild Traumatic Brain Injury Rehabilitation Toolkit. US Army MEDCoE. 2019. URL: https://www.cs.amedd.army.mil/Portlet.aspx?ID=065de2f7-81c4-4f9d-9c85-75fe59dbae13#Mild%20Traumatic%20Brain%20Injury%20Rehabilitation%20Toolkit [accessed 2020-05-06]

21. Jones C, Pike A, Brémault-Phillips S. Brain Bootcamp: pre–post comparison findings of an integrated behavioural health intervention for military members with reduced executive cognitive functioning. J Mil Veteran Fam Health 2019 Apr;5(1):131-140. [doi: 10.3138/jmvfh.2018-0005]

22. Lewis TL, Boissaud-Cooke MA, Aungst TD, Eysenbach G. Consensus on use of the term 'app' versus 'application' for reporting of mHealth research. J Med Internet Res 2014 Jul 17;16(7):e174; discussion e174 [FREE Full text] [doi: 10.2196/jmir.3460] [Medline: 25033233]

23. Tam-Seto L, Wood VM, Linden B, Stuart H. A scoping review of mental health mobile apps for use by the military community. Mhealth 2018;4:57 [FREE Full text] [doi: 10.21037/mhealth.2018.12.01] [Medline: 30701175]

24. Lee H, Sullivan SJ, Schneiders AG, Ahmed OH, Balasundaram AP, Williams D, et al. Smartphone and tablet apps for concussion road warriors (team clinicians): a systematic review for practical users. Br J Sports Med 2015 Apr;49(8):499-505. [doi: 10.1136/bjsports-2013-092930] [Medline: 24668048]

25. Kwan V, Bihelek N, Anderson V, Yeates K. A review of smartphone applications for persons with traumatic brain injury: what is available and what is the evidence? J Head Trauma Rehabil 2019;34(2):E45-E51. [doi: 10.1097/HTR.0000000000000425] [Medline: 30045225]

26. The FDA Recommends Only Using Cleared or Approved Medical Devices to Help Assess or Diagnose a Head Injury, Including Concussion. FDA Safety Communication. 2019. URL: https://www.fda.gov/medical-devices/safety-communications/fda-recommends-only-using-cleared-or-approved-medical-devices-help-assess-or-diagnose-head-injury [accessed 2020-05-05]

XSL•FO

RenderX

27.   O'Neill S, Brady RR. Colorectal smartphone apps: opportunities and risks. Colorectal Dis 2012 Sep;14(9):e530-e534. [doi: 10.1111/j.1463-1318.2012.03088.x] [Medline: 22646729]

28.   Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 7;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]

29.   Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? BMC Med 2018 Dec 3;16(1):225 [FREE Full text] [doi: 10.1186/s12916-018-1211-7] [Medline: 30501638]

30.   Stec M, Arbour MW. Wellness and disease self-management mobile health apps evaluated by the mobile application rating scale. Adv Fam Pract Nurs 2020 May;2:87-102. [doi: 10.1016/j.yfpn.2020.01.003]

31.   Stoyanov S. MARS Training Video. YouTube. 2016. URL: https://www.youtube.com/watch?v=25vBwJQIOcE&t=3s [accessed 2020-01-28]

32.   App Store Review Guidelines. Apple Developer. 2020. URL: https://developer.apple.com/app-store/review/guidelines/ [accessed 2020-07-28]

33.   Balebako R, Cranor L. Improving app privacy: nudging app developers to protect user privacy. IEEE Secur Privacy 2014 Jul;12(4):55-58. [doi: 10.1109/msp.2014.70]

34.   Abroms LC, Westmaas JL, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]

35.   Cicerone KD, Kalmar K. Persistent postconcussion syndrome: the structure of subjective complaints after mild traumatic brain injury. J Head Trauma Rehab 1995;10(3):1-17. [doi: 10.1097/00001199-199506000-00002]

36.   Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet 1974 Jul 13;2(7872):81-84. [doi: 10.1016/s0140-6736(74)91639-0] [Medline: 4136544]

37.   McCrory P, Meeuwisse W, Johnston K, Dvorak J, Aubry M, Molloy M, et al. Consensus statement on concussion in sport: the 3rd international conference on concussion in sport held in Zurich, November 2008. J Athl Train 2009;44(4):434-448 [FREE Full text] [doi: 10.4085/1062-6050-44.4.434] [Medline: 19593427]

38.   Eghdam A, Bartfai A, Oldenburg C, Koch S. How do persons with mild acquired cognitive impairment use information and communication technology and e-services? Results from a Swedish national survey. PLoS One 2016;11(7):e0159362 [FREE Full text] [doi: 10.1371/journal.pone.0159362] [Medline: 27427947]

39.   Fitzgerald S. Not so fast, federal regulators warn. There is no approved device to diagnose concussion on the spot. Neurology 2019;19(10):5-6. [doi: 10.1097/01.nt.0000559865.92231.19]

40.   Kutcher JS, McCrory P, Davis G, Ptito A, Meeuwisse WH, Broglio SP. What evidence exists for new strategies or technologies in the diagnosis of sports concussion and assessment of recovery? Br J Sports Med 2013 Apr;47(5):299-303. [doi: 10.1136/bjsports-2013-092257] [Medline: 23479488]

41.   Crowle C. Low Vision Rehabilitation: A Practical Guide for Occupational Therapists. Memphis, TN: Slack Incorporated; 2010.

42.   Owings-Fonner N. A Review of Meditation Apps for Adults. Let's Get Technical: A Review of the Latest Apps and Tools for Practicing Psychologists. American Psychological Assocaition. 2019. URL: https://www.apaservices.org/practice/business/technology/tech-column/apps-tools-psychologists [accessed 2020-05-09]

43.   Bryden D, Tilghman J, Hinds S. Blast-related traumatic brain injury: current concepts and research considerations. J Exp Neurosci 2019;13:1179069519872213 [FREE Full text] [doi: 10.1177/1179069519872213] [Medline: 31548796]

44.   DAOD 5023-0, Universality of Service. Government of Canada. 2018. URL: https://www.canada.ca/en/department-national-defence/corporate/policies-standards/defence-administrative-orders-directives/5000-series/5023/5023-0-universality-of-service.html [accessed 2020-01-28]

45.   Rhea CK, Kuznetsov NA, Ross SE, Long B, Jakiela JT, Bailie JM, et al. Development of a portable tool for screening neuromotor sequelae from repetitive low-level blast exposure. Mil Med 2017 Mar;182(S1):147-154. [doi: 10.7205/MILMED-D-16-00140] [Medline: 28291466]

46.   Concussion Management Resources. Government of Canada. 2019. URL: https://www.canada.ca/en/public-health/services/diseases/concussion-sign-symptoms/concussion-prevention-risks.html#_Concussion_management_resources [accessed 2020-01-28]

47.   Understanding Mobile Apps. Federal Trade Commission. 2017. URL: https://www.consumer.ftc.gov/articles/0018-understanding-mobile-apps [accessed 2020-01-28]

Abbreviations

CAF: Canadian Armed Forces
MARS: Mobile Application Rating Scale
mHealth: mobile health
mTBI: mild traumatic brain injury
OEF: Operation Enduring Freedom

OIF: Operation Iraqi Freedom
PCS: postconcussion symptom
PTSD: posttraumatic stress disorder
SCAT3: Sport Concussion Assessment Tool 3
SCAT5: Sport Concussion Assessment Tool 5
SM: service member
TBI: traumatic brain injury

XSL•FO
RenderX

Original Paper

Smartphone and Mobile Health Apps for Tinnitus: Systematic Identification, Analysis, and Assessment

Muntazir Mehdi[1], MSc; Michael Stach[2], MSc; Constanze Riha[3], MSc; Patrick Neff[4,5], PhD; Albi Dode[2], MSc; Rüdiger Pryss[6], Prof Dr; Winfried Schlee[4], PhD; Manfred Reichert[2], Prof Dr; Franz J Hauck[1], Prof Dr

[1]Institute of Distributed Systems, Ulm University, Ulm, Germany

[2]Institute of Databases and Information Systems, Ulm University, Ulm, Germany

[3]Department of Psychology, University of Zurich, Zurich, Switzerland

[4]Clinic and Policlinic for Psychiatry and Psychotherapy, Regensburg, Germany

[5]URPP Dynamics of Healthy Aging, University of Zurich, Zurich, Switzerland

[6]Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany

Corresponding Author:
Franz J Hauck, Prof Dr
Institute of Distributed Systems
Ulm University
Albert-Einstein-Allee 11
Ulm, 89081
Germany
Phone: 49 731 50 24140
Email: franz.hauck@uni-ulm.de

Abstract

Background: Modern smartphones contain sophisticated high-end hardware features, offering high computational capabilities at extremely manageable costs and have undoubtedly become an integral part in users' daily life. Additionally, smartphones offer a well-established ecosystem that is easily discoverable and accessible via the marketplaces of differing mobile platforms, thus encouraging the development of many smartphone apps. Such apps are not exclusively used for entertainment purposes but are also commonplace in health care and medical use. A variety of those health and medical apps exist within the context of tinnitus, a phantom sound perception in the absence of any physical external source.

Objective: In this paper, we shed light on existing smartphone apps addressing tinnitus by providing an up-to-date overview.

Methods: Based on PRISMA guidelines, we systematically searched and identified existing smartphone apps on the most prominent app markets, namely Google Play Store and Apple App Store. In addition, we applied the Mobile App Rating Scale (MARS) to evaluate and assess the apps in terms of their general quality and in-depth user experience.

Results: Our systematic search and screening of smartphone apps yielded a total of 34 apps (34 Android apps, 26 iOS apps). The mean MARS scores (out of 5) ranged between 2.65-4.60. The Tinnitus Peace smartphone app had the lowest score (mean 2.65, SD 0.20), and Sanvello—Stress and Anxiety Help had the highest MARS score (mean 4.60, SD 0.10). The interrater agreement was substantial (Fleiss κ=0.74), the internal consistency was excellent (Cronbach α=.95), and the interrater reliability was found to be both high and excellent—Guttman $\lambda6$=0.94 and intraclass correlation, ICC(2,k) 0.94 (95% CI 0.91-0.97), respectively.

Conclusions: This work demonstrated that there exists a plethora of smartphone apps for tinnitus. All of the apps received MARS scores higher than 2, suggesting that they all have some technical functional value. However, nearly all identified apps were lacking in terms of scientific evidence, suggesting the need for stringent clinical validation of smartphone apps in future. To the best of our knowledge, this work is the first to systematically identify and evaluate smartphone apps within the context of tinnitus.

KEYWORDS

XSL·FO
RenderX

Introduction

Tinnitus is a condition mainly associated with the perception of a continuous ringing noise in the ears in the absence of any external sound source. The direct causative factors of the perception of subjective tinnitus are manifold and are not fully understood. However, tinnitus is often associated with underlying damage in the inner ear, such as the loss of cochlear hair cells. Worldwide, roughly 15% of the population suffers from tinnitus; among them 2% experience a substantial decrease in quality of life due to the phantom percept [1]. At present, tinnitus is considered to be a condition that involves changes at different levels of the auditory pathway, the auditory cortex, and nonauditory areas such as the limbic system. These changes may additionally be influenced by psychosocial stress (eg, negative thoughts, stress at home, increased workload, etc), which affects not only the emotional status of the patient but also the auditory system [2,3]. Consequently, people with tinnitus often report that their perception of it varies, including its loudness or related distress [4]. This moment-to-moment variability can be captured by utilizing ecological momentary assessments [5]. Herein, we consider smartphone-based solutions and apps that can be employed to better understand tinnitus, or to offer assistance to patients with tinnitus in managing tinnitus-related distress.

Interestingly in recent years, smartphones, smartphone apps, and auxiliary health devices such as heart monitors and smart wristbands have gained significant popularity by helping patients to monitor and treat their health problems [6-8]. Specifically, smartphones provide an app ecosystem that can easily be extended with new apps. For instance, on one hand, smartphone-based app solutions can be applied to monitor the ecological or environmental surroundings of patients to better understand health phenomena [9,10]. On the other hand, these solutions can also easily be designed or tailored to assist patients in managing or mitigating the symptoms of their health problems [11]. Within the scope of this paper, such solutions can be directly applied within the context of tinnitus and other closely related health complaints, such as stress [12-15], Ménière disease [16,17], hearing loss [18-20], vertigo [21-23], and dementia [24,25]. Consequently, developers frequently push new apps to the markets to capitalize on the growing interest in health care–related smartphone apps in both academia and industry. Thus, timely assessment and evaluation of these smartphone apps are critically important, particularly due to the sensitive nature of their target domain.

Several mobile health (mHealth) app assessment tools and models are available [26-29], yet they either are focused on a particular health domain or are too time consuming and complex to employ in research. Furthermore, it is often hard to establish an objective score based on these rating and assessment tools. Nonetheless, among them the Mobile Application Rating Scale (MARS) [30] is a reliable and valid instrument for the quality assessment of smartphone-based medical apps or mHealth apps. It offers a straightforward multidimensional rating process to objectively assess health care apps without requiring excessive training for the rater. In addition, MARS has been diversely and widely used in evaluating smartphone-based health apps, for instance, pain management apps [31], weight management apps [32], diabetes management apps [33], and rheumatology apps [34,35]. Additionally, MARS can be employed for rating smartphone apps for complex psychological or physiological conditions such as depression [36], hypertension [37], or tinnitus [38]. In the field of tinnitus, Sereda and colleagues [38] gathered tinnitus management apps based on patient opinion via a web-based survey. Features of the patients' most cited apps were then analyzed with MARS. We again apply MARS to evaluate tinnitus management apps, yet in this paper, we emphasized systematic search and exploration of the most prominent commercial mobile app platforms to identify the relevant tinnitus apps. Furthermore, we aimed to identify and evaluate recent apps (those newly added since 2019 or not previously identified) in comparison to those previously reviewed [38].

Methods

Overview

Our work offers 3 major contributions. First, based on the PRISMA guidelines [39], we systematically searched, screened, and identified smartphone apps aimed at assisting patients with tinnitus. Second, with respect to the objective quality of the smartphone apps and the user experience ratings, all identified apps were critically evaluated and assessed based on MARS. As an added step, we compared our MARS ratings to ratings from other sources and computed interrating agreements. Last, we gathered information for quality ratings of the health apps from various established information platforms, as well as the star ratings from the Google Play Store and the Apple App Store.

Finding Relevant Apps

In order to generate an exhaustive overview of relevant tinnitus apps, we employed PRISMA guidelines for a systematic search, screening, and identification of the apps.

We performed an open keyword search (Textbox 1) on 2 of the most prominent app markets, namely Google Play and Apple App stores to cover both major mobile platforms (ie, Android and iOS, respectively). Due to the device-specific limitations of apps from different app stores, we did not include app stores such as Amazon App Store, Sony Apps, Samsung Galaxy Apps, Huawei App Store, and LG SmartWorld in our app search workflow. Furthermore, third-party app providers such as Aptoide and F-Droid were not taken into account as they are not considered to be reliable sources because of security issues and their reliance on rooted devices. Rooting is the process of acquiring full system access or administrative control of mobile devices. This process is highly discouraged by device manufacturers and app developers as it may introduce security vulnerabilities [40].

The overall workflow to systematically identify relevant apps was based on PRISMA guidelines and is illustrated in Figure 1. The search yielded a total of 675 apps from both app markets (Google Play Store: 334; Apple App Store: 341 apps); 311 apps were identified after removing duplicates. These were screened

based on the title and description resulting in 29 apps that satisfied the required criteria.

Textbox 1. Search summary.

Sources: Google Play Store and Apple App Store

Keywords: *tinnitus*, *hearing*, *noise*, *cognitive behavior therapy (CBT)*, *self-help*

**Strategy:** manual investigation of app title, description, and top 10 comments (if available)

Inclusion criteria:

1. Smartphone apps with English title and description

2. Availability of smartphone app on either Google Play Store or Apple App Store

3. App title or description clearly addressing tinnitus, CBT, or self-help as the main subject matter

Exclusion criteria:

1. Apps with missing description

2. Misleading apps claiming to address tinnitus, self-help, or CBT.

Figure 1. PRISMA workflow for systematic identification of smartphone apps.
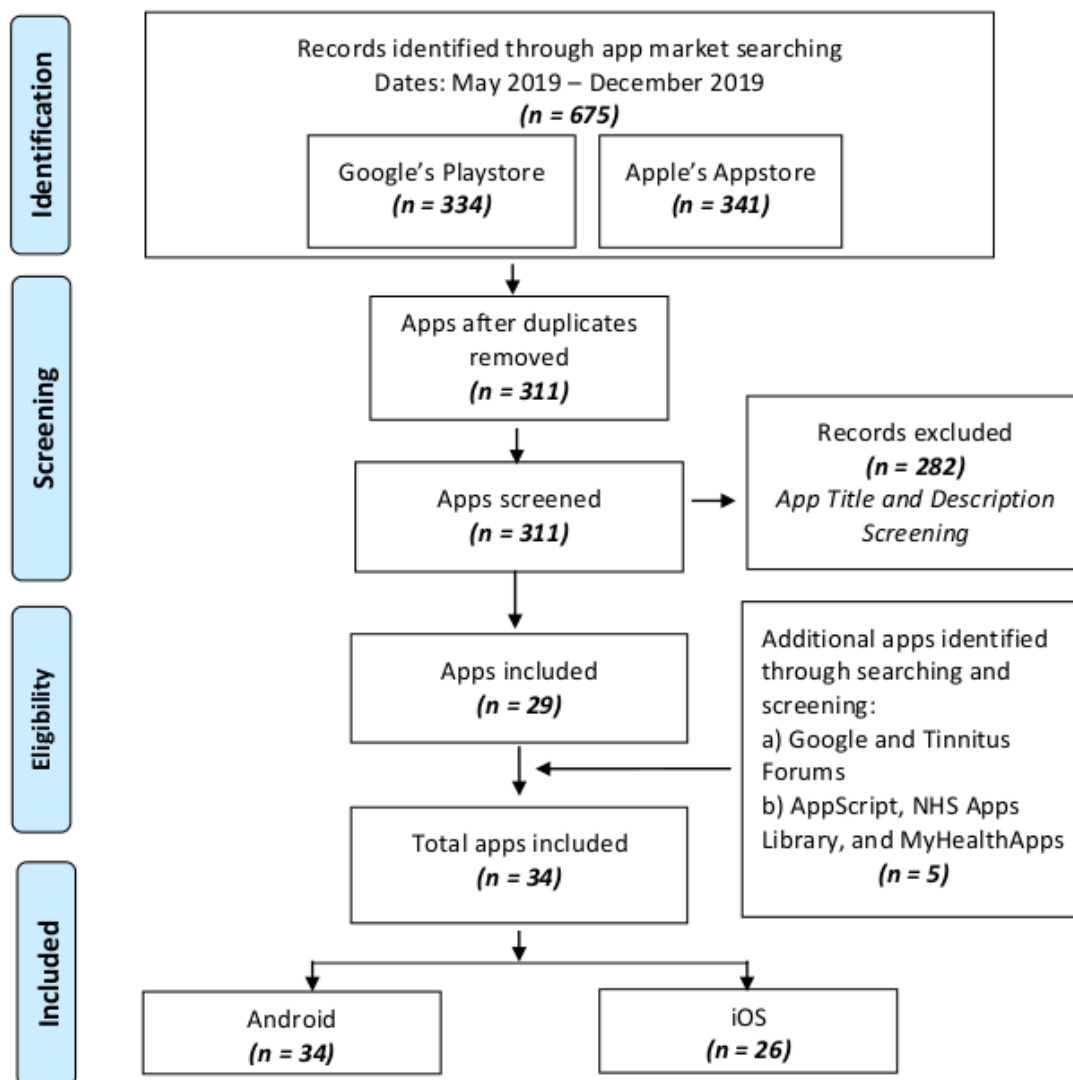
Using the same keywords (with the keyword *app* appended), Google searches were performed to find any missing or additional apps (in May 2019 and in December 2019). The Google search yielded multiple webpages and forum posts. The contents of these webpages and forum posts were manually investigated to identify additional potentially relevant apps. We also performed searches on 3 independent third-party mHealth app libraries: (1) the government-funded National Health Service Apps Library [41], (2) the privately funded AppScript Library [42], and (3) the privately funded MyHealthApps Cochrane

Library [43]. These third-party mHealth app libraries are web portals that curate smartphone apps [44]. This ancillary search of smartphone apps on webpages, tinnitus forums, and third-party mHealth app libraries resulted in the identification of 5 additional apps; therefore, a total of 34 apps were identified for assessment and evaluation. All were available for Android, whereas only 26 apps could also be used on the iOS platform.

Apps Assessments and Evaluations

The smartphone apps that were identified in the app-store search based on PRISMA guidelines were evaluated using MARS scoring guidelines. To rate the smartphone apps, 4 raters (2 raters from a tinnitus domain, and 2 raters from a mobile app–development domain with background in tinnitus research) were recruited. According to the recommendations of the MARS developers, the raters were instructed to watch a video presentation [45] to familiarize themselves with the scoring process. In addition and to further facilitate the familiarization, the raters were requested to read the MARS scoring paper [30]. Next, all raters rated a sample app for training purposes, and the results were discussed briefly to ensure that all raters had the necessary understanding of the MARS scoring process as well as the individual items of the MARS scale. Finally, all of the raters were assigned all of the identified apps for rating purposes. Note that the quality rating of the MARS is based on a scale ranging from 1 to 5 points: 1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent. MARS further includes 19 items that are divided into 4 subscales, namely, *engagement*, *functionality*, *aesthetics*, and *information quality*. Additionally, MARS includes a fifth category, namely *subjective*, which is not included in the calculation of the final MARS score.

In order to evaluate MARS scores from the 4 raters, we calculated the interrater agreement based on Fleiss κ [46], the internal consistency was based on Cronbach α [47], and the interrater reliability was based on Guttman λ6 [48] as well as intraclass correlation—ICC(2,k) with 95% CI [49].

Results

Tinnitus Relief Using Smartphones

A comprehensive list of apps that assist patients (ie, for tinnitus-related relief) and that were identified through PRISMA is shown in Table 1, with their respective properties. Among the app properties, the *downloads* property provides insight into the apps' usage, however, in the case of the iOS platform, the number of users is not publicly provided by the app store; *rating* provides a general understanding of the quality of the app based on the user opinion and is according to the app store's rating system; *update* reports the last recorded update for the corresponding store; and for *pricing*, the app price is given, or if the app was free, further information is given.

A categorical distribution of the smartphone apps that were identified is depicted in Figure 2. Two main categories were identified based on the app descriptions: tinnitus therapy (24 apps) and cognitive behavior therapy (10 apps). In the tinnitus-therapy category, 18 apps had the main focus of providing of sound therapy, including sounds for tinnitus masking (7 apps), sound habituation (4 apps), neuromodulation (4 apps), and distraction (3 apps). The remaining 6 apps of the tinnitus-therapy category were almost evenly distributed among zen therapy (1 app), notch therapy (1 app), game-based therapy (1 app), individual therapy (1 app), and tinnitus management (2 apps). Cognitive behavior therapy for tinnitus made up the other main category (self-help: 6 app; chatbots: 3 app; acceptance and commitment therapy: 1 app).

Table 1. Tinnitus relief–related apps.

App[a]	Description	Platform	Downloads	Rating (out of 5)	Update	Pricing, €(US $)[b]
Beltone Tinnitus Calmer[c]	Combination of relaxation exercise and sound therapy to avoid tinnitus	Android	>1000	4.7	September 2019	In-app purchases
		iOS	—	5.0	September 2019	In-app purchases
CBT[d] Companion	Employs visual tools to learn and practice CBT techniques	Android	>50,000	4.6	February 2019	In-app purchases
		iOS	—	4.7	February 2019	Free
Diapason for tinnitus[c]	Game-based digital therapeutic providing app for tinnitus relief	Android	>5000	3.1	May 2019	In-app purchases
		iOS	—	—	May 2019	In-app purchases
H & T Sound Therapy	Noise player (pink noise, white noise or brown noise) for masking tinnitus	Android	>10,000	4.3	October 2019	Free
Kalmeda mynoise[c]	Offers medically based individual tinnitus therapy	Android	>1000	3.0	July 2019	In-app purchases
		iOS	—	3.6	July 2019	In-app purchases
MindShift CBT[c]	CBT tools to manage and control anxiety	Android	>100,000	3.9	October 2019	Free
		iOS	—	4.2	October 2019	Free
Moodfit—Stress & Anxiety	Stress and anxiety management and tracking, and offers CBT exercises	Android	>5000	4.4	August 2019	Free
myNoise[c]	Controlling tinnitus via combination of different sounds and noises	Android	>100,000	4.4	March 2018	In-app purchases
		iOS	—	4.6	April 2019	In-app purchases
Quirk CBT	Self-help CBT companion based on 3-column technique	Android	>10,000	3.6	July 2019	In-app purchases
		iOS	—	4.7	September 2019	In-app purchases
Relax Melodies[c]	Sleep assisting app that combines sounds and melodies	Android	>10,000,000	4.6	May 2019	Ad-supported, in-app purchases
		iOS	—	4.8	May 2019	In-app purchases
Relax Noise 3[c]	Masking tinnitus by using red, white, or pink noise	Android	>100,000	4.2	March 2015	Free
ReSound Relief[c]	Avoiding tinnitus using combination of sound therapy and relaxation exercise	Android	>100,000	4.5	February 2019	In-app purchases
		iOS	—	4.7	January 2019	In-app purchases
Sanvello—Stress & Anxiety Help	Audio and video CBT exercises, anxiety tracking and management	Android	>1,000,000	4.6	February 2019	In-app purchases
		iOS	—	4.8	Nov 2019	In-app purchases
SimplyNoise[c]	Controlling and managing stress and tinnitus using white and brown noise	Android	>50,000	3.7	June 2012	Free
		iOS		4.4	May 2018	In-app purchases
Stress & Anxiety Companion	CBT based visual exercises to manage stress and anxiety	Android	>10,000	4.2	July 2019	In-app purchases
		iOS	—	4.6	June 2019	In-app purchases
Starkey Relax[c]	Tinnitus masking, self-management, and education app	Android	>10,000	4.3	October 2017	Free
		iOS	—	3.9	October 2017	Free
StopTinnitus[c]	Masking tinnitus using customized tones	Android	>100	2.7	January 2015	7.95 (9.38)
		iOS	—	1.3	January 2015	8.03 (9.48)
Tinnitracks[c]	Controlling and managing tinnitus by filtering out music for sound therapy	Android	>10,000	3.8	April 2019	In-app purchases
		iOS	—	3.6	February 2019	In-app purchases
Tinnitus Balance App[c]	Controlling annoying tinnitus using customized sounds or music	Android	>50,000	3.7	March 2016	Free
		iOS	—	2.3	March 2016	Free

XSL•FO
RenderX

App[a]	Description	Platform	Downloads	Rating (out of 5)	Update	Pricing, €(US $)[b]
Tinnitus Help[c]	Tinnitus masking using natural sounds or music	Android	>500	3.0	November 2015	9.90 (11.68)
		iOS	—	4.4	January 2019	17.99 (21.23)
Tinnitus Notch	Provided custom tailored notch therapy for tinnitus relief	Android	>1000	2.7	September 2016	In-app purchases
Tinnitus Peace	Offers melodies to match the frequency of tinnitus to reduce its effects	Android	>5000	3.8	November 2015	Free
Tinnitus Relief	Tinnitus masking using head-phones	Android	>10,000	4.2	December 2019	In-app purchases
Tinnitus Sound Therapy	Sound/acoustic therapy for masking tinnitus	Android	>10,000	3.9	June 2019	Free
Tinnitus Therapy (Lite)[c]	Avoiding tinnitus with sound masking and therapy	Android	>500	3.6	February 2019	6.49 (7.66)
		iOS	—	5.0	March 2019	5.36 (6.32)
Tonal Tinnitus Therapy[c]	Helps to mitigate symptoms of tonal tinnitus based on acoustic neuromodulation	Android	>10,000	4.0	July 2018	In-app purchases
Track Your Tinnitus[c]	Managing tinnitus by tracking tinnitus patterns in daily activity	Android	>1000	2.1	October 2018	Free
		iOS	—	5.0	June 2017	Free
What's Up? A Mental Health App	Offers CBT and ACT[e] methods to manage stress, anxiety as well as depression	Android	>50,000	4.4	June 2019	In-app purchases
		iOS	—	4.6	December 2016	In-app purchases
Whist[c]	Controlling tinnitus using sounds with adjusted volume, pitch, etc	Android	>1000	4.2	March 2017	2.18 (2.57)
		iOS	—	3.7	January 2019	1.78 (2.10)
White Noise (Lite)[c]	Masking and controlling tinnitus using environmental sounds	Android	>5000	4.6	September 2018	3.19 (3.76)
		iOS	—	4.8	April 2019	2.67 (3.15)
Widex Zen[c]	Avoiding tinnitus using relaxing zen sounds and exercises to manage tinnitus	Android	>10,000	3.8	May 2017	Free
		iOS	—	5.0	November 2017	Free
Woebot—Your Self-Care Expert[c]	A chatbot for guided-CBT to manage stress and anxiety	Android	>100,000	4.8	November 2019	Free
		iOS	—	4.7	November 2019	Free
Wysa: Mental Health Therapy[c]	A chatbot offering CBT and DBT[f] techniques	Android	>1,000,000	4.7	November 2019	In-app purchases
		iOS	—	4.7	December 2019	In-app purchases
Youper—Emotional Health[c]	A chatbot based on CBT and ACT techniques monitoring and tracking mood changes	Android	>1,000,000	4.7	December 2019	In-app purchases
		iOS	—	4.9	December 2019	In-app purchases

[a]Retrieved December 15, 2019.

[b]An exchange rate of €1 to US $1.18 is applicable.

[c]Apps reported in literature.

[d]CBT: cognitive behavior therapy.

[e]ACT: acceptance and commitment therapy.

[f]DBT: dialectical behavior therapy.

XSL•FO
RenderX

Figure 2. Categorical distribution of identified apps. CBT: cognitive behavior therapy; ACT: acceptance and commitment therapy.
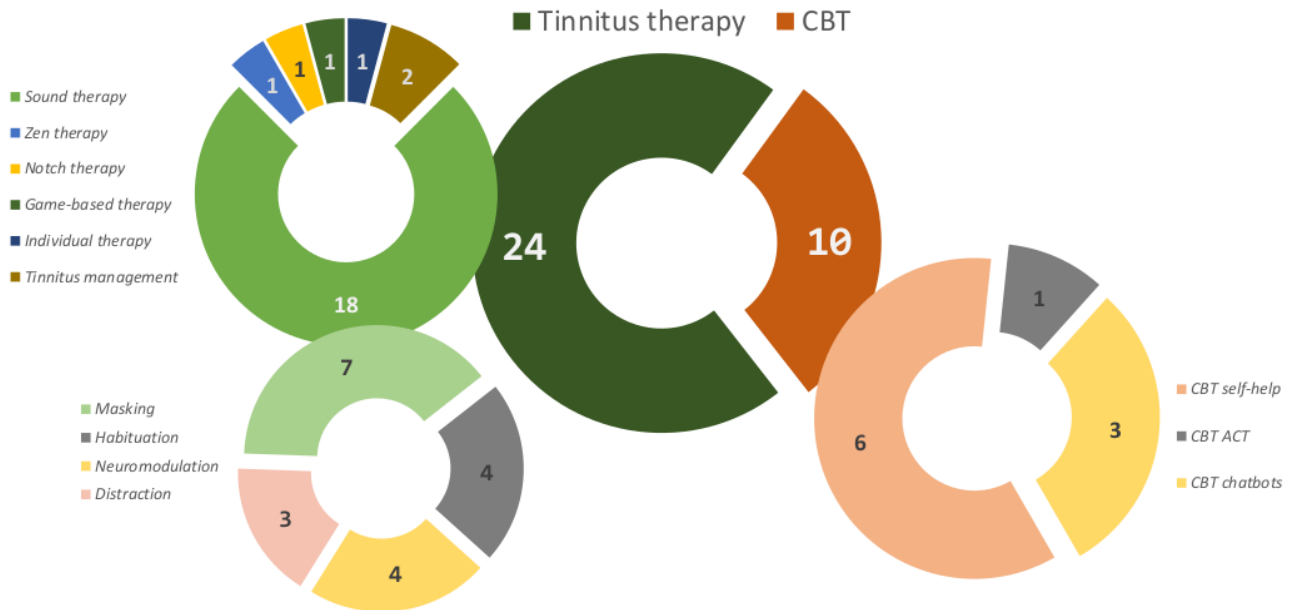
Evaluation of Tinnitus Relief Apps

The ratings of each individual rater and the mean of all 4 raters are depicted in the Figure 3; it can be seen that evaluations are rather consistent between the 4 raters. Note, that MARS ratings range from 1 (inadequate) to 5 (excellent); however, none of the apps in our evaluation process scored less than 2. To ensure consistency between raters and internal consistency, as well as

reliability, we performed statistical psychometric analyses (Table 2).

In addition to the objective MARS scores calculated using the arithmetic mean of 4 categories (engagement, functionality, aesthetics, and information quality), MARS guidelines also allow subjective scoring of the smartphone apps, reflecting individual rater opinion. In Figure 4, the results of the subjective criteria of the MARS questionnaire are shown.

Figure 3. MARS scores.
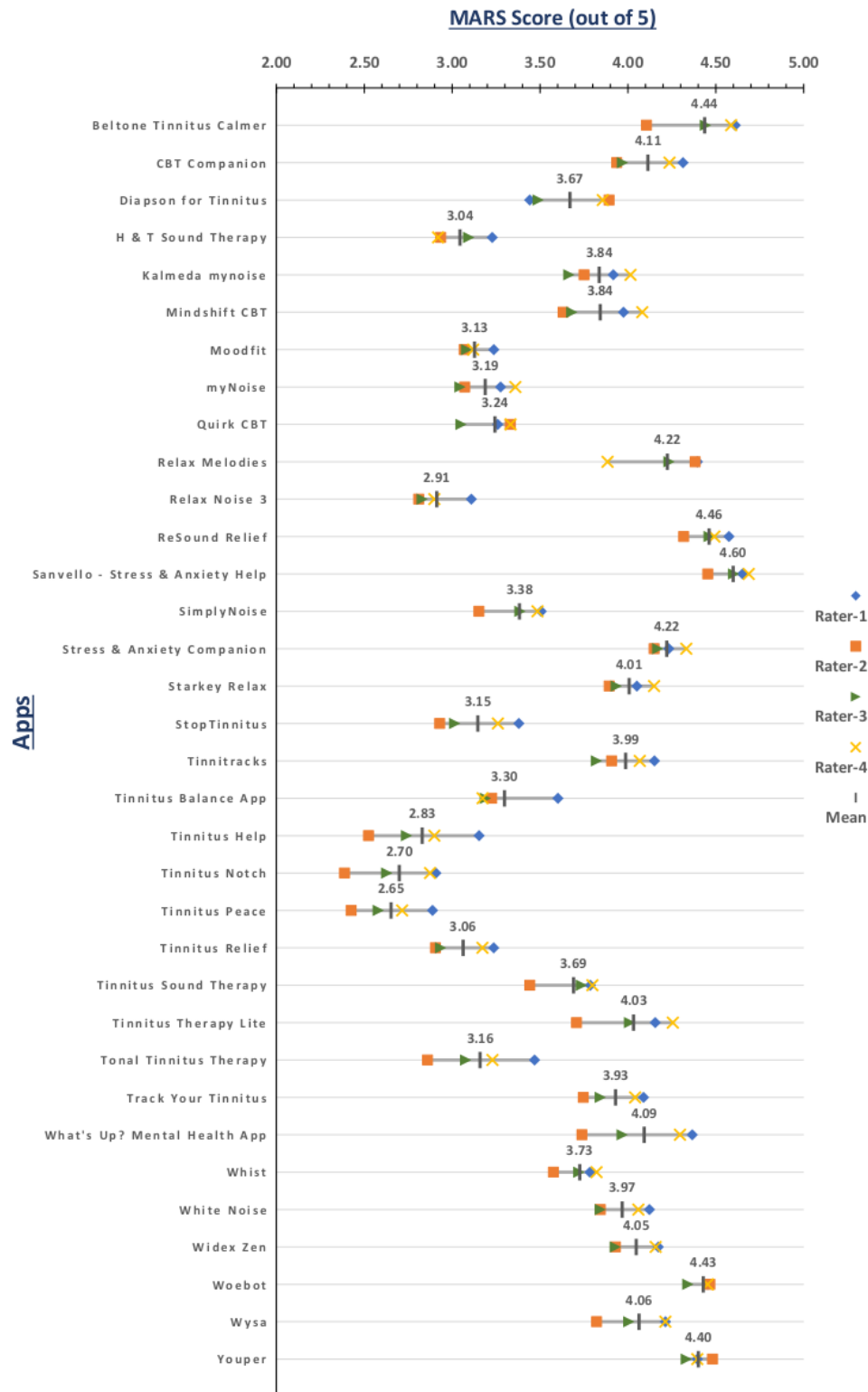
Table 2. Interrater agreement, internal consistency, and reliability with interpretations [46-49].

Rating category	Fleiss κ	Cronbach α	Guttman λ6	ICC(2,k)[a] (95% CI)
Engagement	0.62 (substantial)	.93 (excellent)	0.93 (high)	0.92 (0.88-0.95) (excellent)
Functionality	0.52 (moderate)	.90 (excellent)	0.89 (high)	0.89 (0.82-0.93) (good)
Aesthetics	0.49 (moderate)	.96 (excellent)	0.97 (high)	0.94 (0.88-0.97) (good)
Information	0.78 (substantial)	.96 (excellent)	0.96 (high)	0.96 (0.94-0.98) (excellent)
MARS[b] score	0.74 (substantial)	.95 (excellent)	0.94 (high)	0.94 (0.91-0.97) (excellent)
Subjective score	0.45 (moderate)	.96 (excellent)	0.96 (high)	0.95 (0.92-0.97) (excellent)

[a]ICC: intraclass correlation.

[b]MARS: Mobile Application Rating Scale.

Figure 4. Mean MARS subjective score (with range).
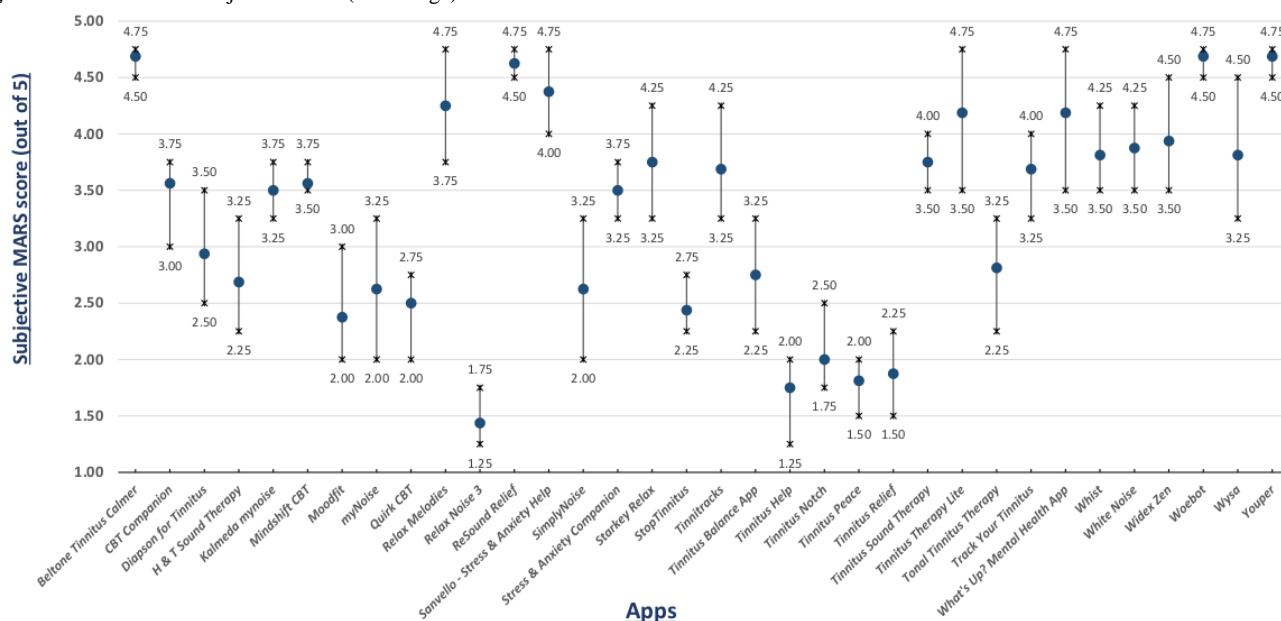
Additionally, we gathered information on the quality of the apps from various repositories. We used existing ratings both from established information platforms for health app quality ratings as well as from the Google Play Store and the iOS App Store (star ratings). The first information platform that we chose was PsyberGuide [50], which is a nonprofit website that is funded by One Mind and operated by Northwestern University. PsyberGuide's app reviews consist of a *credibility* score that represents the research support, a *user experience* score that is based on MARS, and a *transparency* score that represents the app developer's privacy information transparency. The second information platform that we chose was ORCHA [51], an organization that offers evaluations of health apps and advice for governments, health, and social care organizations. ORCHA app reviews consist of a score that is a calculated mean of the

3 domains: *data privacy*, *clinical assurance*, and *user experience* plus a level that classifies the app in 1 of 5 levels based on their focus and functionality. This quality information is given in Table 3.

Since the objective MARS scores calculated in this paper and PsyberGuide and ORCHA ratings (apart from PsyberGuide's user experience) are incomparable, a separate chart (Figure 5) depicts a comparison of MARS scores from our study with those from 2 papers from literature [38,52] and with PsyberGuide's user experience score. The MARS scores in [38] are for general tinnitus apps such as Beltone Tinnitus Calmer, Relax Noise 3, ReSound Relief, myNoise, Tinnitus Therapy (Lite), and White Noise (Lite), while MARS scores in [52] are given for mindfulness and cognitive behavior therapy apps such as Relax Melodies and MindShift CBT.

Table 3. Rating scores comparison (higher numbers are better for all criteria).

App	PsyberGuide			ORCHA		Star rating	
	Credibility	User experience	Transparency	Score, %	Level	Android	iOS
MindShift CBT	2.14	3.70	Questionable	69	3	3.9	4.2
Sanvello—Stress & Anxiety Help	4.29	4.79	Acceptable	N/A[a]	N/A	4.6	4.8
Stress & Anxiety Companion	1.80	N/A	N/A	N/A	N/A	4.2	4.6
What's Up? A Mental Health App	1.43	3.83	Unacceptable	51	3	4.4	4.6
Woebot—Your Self-Care Expert	4.64	4.38	Acceptable	N/A	N/A	4.8	4.7
Wysa: Mental Health Therapy	3.86	3.86	Acceptable	N/A	N/A	4.7	4.7
Youper—Emotional Health	2.50	4.33	Questionable	N/A	N/A	4.7	4.9
Widex Zen	N/A	N/A	N/A	77	3	3.8	5.0
Whist	N/A	N/A	N/A	71	3	4.2	3.7
Tinnitus Therapy (Lite)	N/A	N/A	N/A	68	3	3.6	5.0
White Noise (Lite)	N/A	N/A	N/A	59	3	4.6	4.8
Tinnitus Relief	N/A	N/A	N/A	56	3	4.2	N/A
Relax Melodies	N/A	N/A	N/A	55	3	4.6	4.8
Beltone Tinnitus Calmer	N/A	N/A	N/A	54	3	4.7	5.0
Resound Relief	N/A	N/A	N/A	54	3	4.5	4.7

[a]N/A: not available.

Figure 5. MARS score comparison.
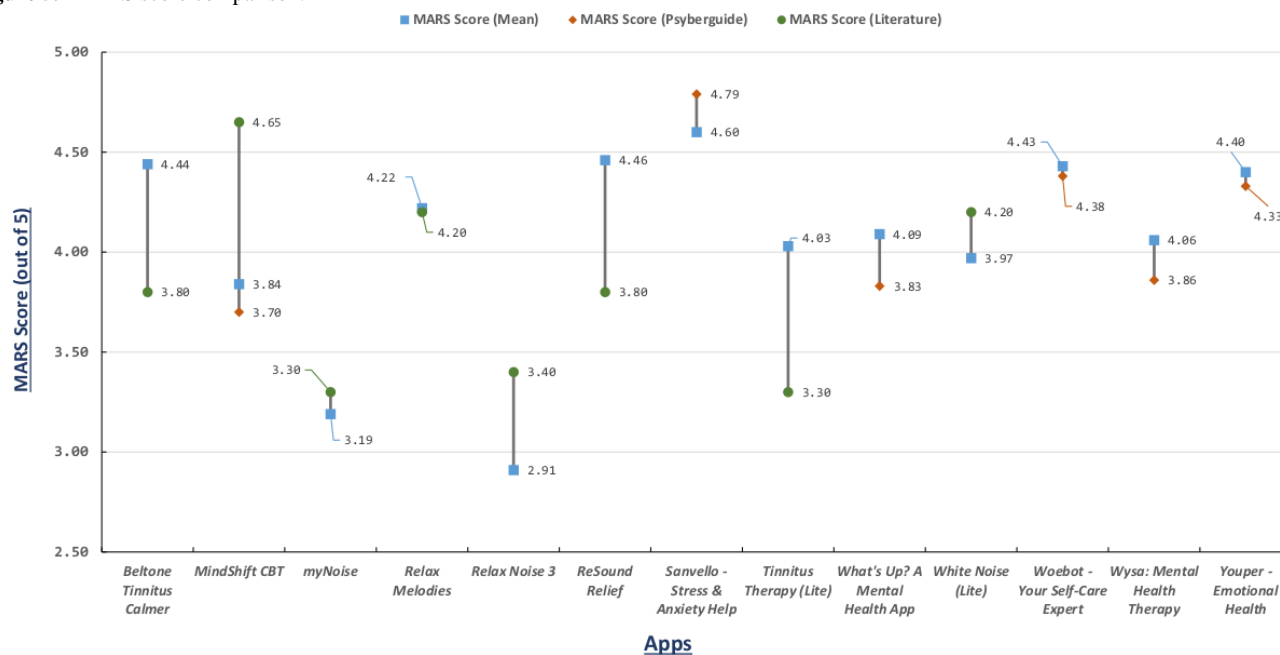
Discussion

Literature

A significant portion of tinnitus scientific literature [53-57] reports on different smartphone and mobile crowdsensing apps to support clinicians in better understanding tinnitus, ranging from data collection to mitigating tinnitus symptoms via therapeutic interventions. These apps are specifically designed to assist patients, clinicians, and researchers alike. From the perspective of patients, these apps aim to provide the necessary means to mask, control, mitigate, or manage tinnitus symptoms. For example, the TrackYourTinnitus smartphone app systematically records data about fluctuations of tinnitus symptoms over time from patients, thus providing information about patient's tinnitus variability [53]. Similarly, Henry et al [54] delve into the development of an app, based on progressive tinnitus management, to support patients in learning and using coping skills for tinnitus. From the perspective of clinicians and researchers, these apps support a better understanding of tinnitus, particularly in identifying symptom severity and tinnitus

characteristics in different patients. For instance, the data collected from the TrackYourTinnitus app can be used to associate tinnitus with daily routines or activities [55] or to shape recruiting strategies for tinnitus-related studies [56]. Similarly, TrackYourTinnitus app was also used to better understand tinnitus variability and tinnitus loudness and stress associations [57].

Generally, health care apps have been exhaustively discussed and reviewed in the literature [58]; however, the number of papers that focus on reviewing and evaluating tinnitus-related smartphone apps is underwhelming. Nevertheless, among pre-existing literature, some papers have discussed the role of tinnitus smartphone apps in clinical scenarios. For instance, internet- and smartphone-based solutions for treatment and management of tinnitus have been reviewed in [59], similarly, the review by Kalle et al [60] discusses internet- or smartphone-delivered cognitive behavior therapy with particular focus on self-help for tinnitus. Both of these papers demonstrate the role of several approaches and technologies involved in advancing tinnitus clinical practice but focus less on current and available apps for patients. Furthermore, the review by Lui et al [61] addresses efficacy or effectiveness of therapeutic solutions provided by mental health apps and hearing health care apps have been discussed in [62]. Comparatively, both [61,62] list a limited number of apps and many are no longer commercially available on the app stores. Specifically, in terms of the assessment of tinnitus-related smartphone apps using MARS, the reviews by Sereda et al [38], which were further extended and repeated by Smith et al [63], are the only closely related works in relation to this work. Both of these reviews [38,63] list tinnitus management apps based on patient opinions, gathered via a web-based survey. The added value of our review was primarily the exploration of smartphone app markets to reveal relevant apps as opposed to using a survey. Additionally, our proposed work also compares the star ratings and MARS scores with quality information gathered by third-party app assessment platforms.

Limitations

A noticeable limitation of our work was the restricted search of relevant smartphone apps to only 2 app stores. Although the restriction was justified in the paper, it might be possible that there would be benefit in exploring other app stores, such as Amazon and Samsung app stores. Another possible limitation lay in the inclusion criteria for the apps. To include an app, we inspected the app description and a few top-rated comments from the users. Despite being effective and straight-forward, this approach is subjective and highly relies on the knowledge of the inspector about the domain. This limitation can be overcome or can be further improved by gathering additional opinions from domain experts.

Future Work

In future research, the study will be extended and developed in 2 directions. First, additional app evaluation and assessment instruments (currently under development or newly developed) will be used to repeat the study. For example, recently, the THESIS app evaluation instrument was presented [64]. Therefore, we intend to extend our current work by evaluating

the identified apps using THESIS and comparing the results with those of MARS. This will include updates on the available and relevant apps in the app stores. Second, although the already developed instruments systematically and objectively measure the quality of mHealth apps from a user-experience perspective, the lack of instruments to clinically validate smartphone apps is undeniable. Therefore, we intend to further invest our efforts in this research direction. Additionally, as a consequence of this study, we learned that there exist 7 evidence-based tinnitus apps; therefore, we are currently working on a review paper detailing the scientific evidence of these 7 apps.

Principal Findings

The aim of this study was to systematically identify smartphone apps within the context of tinnitus. The identification process yielded a total of 34 commercially available tinnitus smartphone apps, which were divided in 2 categories: tinnitus therapy (24 apps) and cognitive behavior therapy (10 apps). In an added step, we evaluated the identified apps using MARS. From MARS objective scores (Figure 3), first, we can see that all 34 identified apps have MARS scores higher than 2, indicating that most of these apps provide some level of user experience and that they all have some functional value. Furthermore, the MARS rating process discovered that only 7 apps—Tinnitus Therapy (Lite), ReSound Relief, SimplyNoise, Audio Notch, Wysa, Woebot, and MindShift—out of 34 apps scored in the evidence-based subitem of the information category, suggesting a lack of clinical validation for most of the apps. Furthermore, the mean MARS scores (Figure 3) of all 4 raters ranged from 2.65-4.60, with Tinnitus Peace having the lowest mean score (mean 2.65, SD 0.20) and Sanvello—Stress and Anxiety Help having the highest mean score (mean 4.60, SD 0.10). On the individual-rater level, Tinnitus Notch had the lowest (score: 2.39), while, Sanvello—Stress and Anxiety Help had the highest (score: 4.69). Furthermore, some apps were rated better in comparison to the others. For instance, Beltone Tinnitus Calmer, Resound Relief, Sanvello—Stress and Anxiety Help, as well as Woebot and Youper received very good ratings, whereas Tinnitus Help, Tinnitus Notch, and Tinnitus Peace were the worst of all. The mean subjective scores presented in Figure 4 ranged from 1.44-4.69, with Relax Noise 3 having the lowest scores (mean 1.44) and Beltone Tinnitus Calmer, Woebot, and Youper having the highest scores (mean 4.69). From Figure 4, naturally, the range of the subjective scores was wider than those of the objective scores; however, it is notable that the mean subjective results were more or less in line with the mean objective results.

Values from the 4 psychometric measures (Table 2) confirm the reliability and validity of our MARS rating procedures, especially for measures testing internal consistency and reliability (>0.89). The interrater agreement, as demonstrated by Fleiss κ, was merely moderate for subjective, functionality, and aesthetics scores. This is noteworthy and may be related to individual differences in the raters with regards to their backgrounds. In any case, from Figure 3, Figure 4, and Figure 5, it is evident that our rating procedures generally produced valid, reliable, and thus viable results.

Interestingly, the expert ratings from both information platforms (PsyberGuide and ORCHA) and the user ratings from the app stores varied, sometimes considerably. For example, What's Up? A Mental Health App received high ratings from its users in the app stores (Google Play Store: 4.4; Apple App Store: 4.6), whereas both expert ratings were considerably lower (PsyberGuide: credibility: 1.43, transparency: unacceptable, user experience: 3.38; ORCHA: 51%). Another app whose app store rating differed significantly from its ORCHA rating is Beltone Tinnitus Calmer (Google Play Store: 4.7; Apple App Store: 5.0; ORCHA: 54%). The ORCHA score was moderate as a result of a moderate rating for data privacy and clinical assurance; the apps hadn't been rated by PsyberGuide. These examples illustrate that an independent rating with validated instruments is crucial for the informed selection of health apps.

From Figure 5, it can be seen that the MARS scores in our work were more in line with those of PsyberGuide's user experience score. Similarly, the differences between MARS scores from our paper and those from literature were evidently higher for all apps except myNoice, Relax Melodies, and White Noise (Lite). These differences were as a result of version changes in the apps. In our case, similar to PsyberGuide's case, the MARS scores present the contemporary assessment of the most recent version of the apps, thus validating the need of an up-to-date MARS assessment of tinnitus apps.

All 34 identified apps obtained a MARS objective score higher than 2 (ranging between 2.65-4.60), indicating that they provide some level of user experience and at least some technical functional value for the user. Furthermore, in addition to presenting the objective MARS scores, the subjective MARS scores (ranging between the values of 1.44-4.69) were also discussed. The 4 psychometric measurements—Fleiss κ, Cronbach α, Guttman λ6, and ICC(2,k)—confirmed and depicted positive interrater agreement, internal consistency as well as reliability between the raters. The only exception was noticed in Fleiss κ with moderate values for subjective score, functionality, and aesthetics. The quality information of identified apps from PsyberGuide and ORCHA as well as the star ratings from the Google Play Store and the Apple App Store were compared. The quality information comparison exhibited incongruity. Finally, the comparison between MARS scores from this work and MARS scores of smartphone apps reported in previously published papers depicted a high coherence. Through these steps, we were able to comprehensively capture the wide array of heterogeneous apps for tinnitus and present an up-to-date assessment of identified apps.

Conclusions

This work highlighted the impact of smartphone apps, specifically within the context of tinnitus research. As a consequence, we demonstrated that there exists a plethora of smartphone apps utilized in supporting and controlling tinnitus symptoms, understanding tinnitus, and monitoring patients with tinnitus. Among the 34 identified apps, only 7 apps were evidence-based suggesting that the majority were in need of more stringent clinical validation.

Authors' Contributions

MM undertook the database searches, cataloguing, conception, drafting, and revision of the paper. MS and AD helped with writing the Methods section, providing the app store scraper and app information database, and revising the manuscript. CR helped with Results section. PN and RP helped with writing Discussion section. CR, PN, WS, and MR helped by critically revising the manuscript. FH contributed by helping with writing the Introduction section, critical revisions, and final approval of the manuscript, as well as supervision of the entire work. All authors approved the submitted manuscript.

Conflicts of Interest

None declared.

References

1. Baguley D, McFerran D, Hall D. Tinnitus. Lancet 2013 Nov 09;382(9904):1600-1607 [FREE Full text] [doi: 10.1016/S0140-6736(13)60142-7] [Medline: 23827090]
2. Mazurek B, Szczepek AJ, Hebert S. Stress and tinnitus. HNO 2015 Apr;63(4):258-265. [doi: 10.1007/s00106-014-2973-7] [Medline: 25862619]
3. Jastreboff P, Jastreboff M. Tinnitus Retraining Therapy (TRT) as a method for treatment of tinnitus and hyperacusis patients. J Am Acad Audiol 2000 Mar;11(3):162-177. [Medline: 10755812]
4. Probst T, Pryss R, Langguth B, Schlee W. Emotion dynamics and tinnitus: Daily life data from the. Sci Rep 2016 Aug 04;6:31166 [FREE Full text] [doi: 10.1038/srep31166] [Medline: 27488227]
5. Schlee W, Pryss RC, Probst T, Schobel J, Bachmeier A, Reichert M, et al. Measuring the Moment-to-Moment Variability of Tinnitus: The TrackYourTinnitus Smart Phone App. Front Aging Neurosci 2016;8:294 [FREE Full text] [doi: 10.3389/fnagi.2016.00294] [Medline: 28018210]

6. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]

7. Naslund JA, Aschbrenner KA, Barre LK, Bartels SJ. Feasibility of popular m-health technologies for activity tracking among individuals with serious mental illness. Telemed J E Health 2015 Mar;21(3):213-216. [doi: 10.1089/tmj.2014.0105] [Medline: 25536190]

8. Nelson EC, Verhagen T, Noordzij ML. Health empowerment through activity trackers: An empirical smart wristband study. Computers in Human Behavior 2016 Sep;62:364-374. [doi: 10.1016/j.chb.2016.03.065]

9. Mehdi M. Smart mobile crowdsensing for tinnitus research: student research abstract. In: SAC '19: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing. In: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing. ACM; 2019 Presented at: SAC '19: The 34th ACM/SIGAPP Symposium on Applied Computing; April, 2019; Limassol, Cyprus p. 1220. [doi: 10.1145/3297280.3297654]

10. Mehdi M, Schwager D, Pryss R, Schlee W, Reichert M, Hauck F. Towards Automated Smart Mobile Crowdsensing for Tinnitus Research. In: Proceedings of the IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS).: IEEE; 2019 Jun Presented at: IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS); June, 2019; Cordoba, Spain p. 32. [doi: 10.1109/cbms.2019.00026]

11. Mehdi M, Mühlmeier G, Agrawal K, Pryss R, Reichert M, Hauck FJ. Referenceable mobile crowdsensing architecture: A healthcare use case. Procedia Computer Science 2018;134:445-451. [doi: 10.1016/j.procs.2018.07.185]

12. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. Mil Med 2014 Jan;179(1):12-18. [doi: 10.7205/MILMED-D-13-00271] [Medline: 24402979]

13. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]

14. Economides M, Martman J, Bell MJ, Sanderson B. Improvements in Stress, Affect, and Irritability Following Brief Use of a Mindfulness-based Smartphone App: A Randomized Controlled Trial. Mindfulness (N Y) 2018;9(5):1584-1593 [FREE Full text] [doi: 10.1007/s12671-018-0905-4] [Medline: 30294390]

15. Van AM, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. Depress Anxiety 2017 Jun;34(6):526-539. [doi: 10.1002/da.22657] [Medline: 28569409]

16. Schmidt W, Sarran C, Ronan N, Barrett G, Whinney DJ, Fleming LE, et al. The Weather and Ménière's Disease: A Longitudinal Analysis in the UK. Otol Neurotol 2017 Feb;38(2):225-233 [FREE Full text] [doi: 10.1097/MAO.0000000000001270] [Medline: 27861300]

17. Delbaere K, Valenzuela T, Woodbury A, Davies T, Yeong J, Steffens D, et al. Evaluating the effectiveness of a home-based exercise programme delivered through a tablet computer for preventing falls in older community-dwelling people over 2 years: study protocol for the Standing Tall randomised controlled trial. BMJ Open 2015 Oct 22;5(10):e009173 [FREE Full text] [doi: 10.1136/bmjopen-2015-009173] [Medline: 26493461]

18. Bright T, Pallawela D. Validated Smartphone-Based Apps for Ear and Hearing Assessments: A Review. JMIR Rehabil Assist Technol 2016 Dec 23;3(2):e13 [FREE Full text] [doi: 10.2196/rehab.6074] [Medline: 28582261]

19. Masalski M, Grysiński T, Kręcicki T. Hearing Tests Based on Biologically Calibrated Mobile Devices: Comparison With Pure-Tone Audiometry. JMIR Mhealth Uhealth 2018 Jan 10;6(1):e10 [FREE Full text] [doi: 10.2196/mhealth.7800] [Medline: 29321124]

20. Kardous CA, Shaw PB. Evaluation of smartphone sound measurement applications. J Acoust Soc Am 2014 Apr;135(4):EL186-EL192 [FREE Full text] [doi: 10.1121/1.4865269] [Medline: 25236152]

21. Dlugaiczyk J, Thiemer M, Neubert C, Schorn BA, Schick B. The aVOR App Increases Medical Students' Competence in Treating Benign Paroxysmal Positional Vertigo (BPPV). Otol Neurotol 2018 Jun;39(5):e401-e406. [doi: 10.1097/MAO.0000000000001778] [Medline: 29579015]

22. Bromwich MA, Parnes LS. The DizzyFix: initial results of a new dynamic visual device for the home treatment of benign paroxysmal positional vertigo. J Otolaryngol Head Neck Surg 2008 Jun;37(3):380-387. [Medline: 19128643]

23. Organ B, Liu H, Bromwich M. An iPhone-assisted particle repositioning maneuver for benign paroxysmal positional vertigo (BPPV): a prospective randomized study. J Am Board Fam Med 2015;28(1):118-120 [FREE Full text] [doi: 10.3122/jabfm.2015.01.120295] [Medline: 25567831]

24. Beauvais B, Rialle V, Sablier J. Myvigi: An android application to detect fall and wandering. 2012 Presented at: In: Proceedings of the sixth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, UBICOMM; 2012; Barcelona, Spain p. 156-160.

25. Sposaro F, Danielson J, Tyson G. iWander: An Android application for dementia patients. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology.: IEEE; 2010 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology; 2010; Buenos Aires, Argentina p. 3875-3878. [doi: 10.1109/iembs.2010.5627669]

26. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: A Comprehensive Quality and Therapeutic Potential Evaluation Tool for Mobile and Web-Based eHealth Interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]

27. 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]

28. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. Telemed J E Health 2015 Feb;21(2):97-104. [doi: 10.1089/tmj.2014.0088] [Medline: 25469795]

29. 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]

30. 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]

31. 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]

32. Bardus M, van BSB, 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]

33. Chavez S, Fedele D, Guo Y, Bernier A, Smith M, Warnick J, et al. Mobile Apps for the Management of Diabetes. Diabetes Care 2017 Oct;40(10):e145-e146. [doi: 10.2337/dc17-0853] [Medline: 28774944]

34. Grainger R, Townsley H, White B, Langlotz T, Taylor WJ. Apps for People With Rheumatoid Arthritis to Monitor Their Disease Activity: A Review of Apps for Best Practice and Quality. JMIR Mhealth Uhealth 2017 Feb 21;5(2):e7 [FREE Full text] [doi: 10.2196/mhealth.6956] [Medline: 28223263]

35. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German Mobile Apps in Rheumatology: Review and Analysis Using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]

36. Qu C, Sas C, Daudén Roquet C, Doherty G. Functionality of Top-Rated Mobile Apps for Depression: Systematic Search and Evaluation. JMIR Ment Health 2020 Jan 24;7(1):e15321 [FREE Full text] [doi: 10.2196/15321] [Medline: 32012079]

37. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone Apps to Support Self-Management of Hypertension: Review and Content Analysis. JMIR Mhealth Uhealth 2019 May 28;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]

38. Sereda M, Smith S, Newton K, Stockdale D. Mobile Apps for Management of Tinnitus: Users' Survey, Quality Assessment, and Content Analysis. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e10353 [FREE Full text] [doi: 10.2196/10353] [Medline: 30672739]

39. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]

40. Zhang H, She D, Qian Z. Android Root and its Providers: A Double-Edged Sword. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security.: ACM; 2015 Presented at: CCS'15: The 22nd ACM Conference on Computer and Communications Security; October 2015; Denver, Colorado, USA p. 1093-1104. [doi: 10.1145/2810103.2813714]

41. National Health Service (NHS) Apps Library Internet. NHS Apps Library - NHS. URL: https://www.nhs.uk/apps-library/ [accessed 2020-05-15]

42. AppScript Internet. AppScript | Discover, Deliver & Track Digital Health. URL: https://www.appscript.net/ [accessed 2020-05-15]

43. PatientView. MyHealthApps Internet. MyHealthApps - apps tried and tested by people like you. URL: https://myhealthapps.net/ [accessed 2020-05-15]

44. Baxter C, Carroll JA, Keogh B, Vandelanotte C. Assessment of Mobile Health Apps Using Built-In Smartphone Sensors for Diagnosis and Treatment: Systematic Survey of Apps Listed in International Curated Health App Libraries. JMIR Mhealth Uhealth 2020 Feb 03;8(2):e16741 [FREE Full text] [doi: 10.2196/16741] [Medline: 32012102]

45. Vimeo, Inc. Vimeo Internet. Leanne Hides on Vimeo. URL: https://vimeo.com/244067710 [accessed 2020-05-15]

46. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]

47. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]

48. A basis for analyzing test-retest reliability. Psychometrika 1945;10:255-282. [doi: 10.1007/BF02288892] [Medline: 21007983]

49. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]

50. PsyberGuide | A project of One Mind | A mental health App guide. One Mind. URL: https://onemindpsyberguide.org/ [accessed 2020-05-15]

51. Rated Health Apps | Best Health Apps | ORCHA. ORCHA. URL: https://www.orcha.co.uk/ [accessed 2020-05-15]

52. Weekly T, Walker N, Beck J, Akers S, Weaver M. A Review of Apps for Calming, Relaxation, and Mindfulness Interventions for Pediatric Palliative Care Patients. Children (Basel) 2018 Jan 26;5(2) [FREE Full text] [doi: 10.3390/children5020016] [Medline: 29373515]

53. Pryss R, Schlee W, Langguth B, Reichert M. Mobile Crowdsensing Services for Tinnitus Assessment and Patient Feedback. In: Proceedings of the 2017 IEEE International Conference on AI & Mobile Services (AIMS).: IEEE; 2017 Presented at: IEEE International Conference on AI & Mobile Services (AIMS); 2017; Honolulu, HI, USA p. 22-29. [doi: 10.1109/aims.2017.12]

54. Henry JA, Thielman E, Zaugg T, Kaelin C, Choma C, Chang B, et al. Development and field testing of a smartphone "App" for tinnitus management. Int J Audiol 2017 Oct;56(10):784-792. [doi: 10.1080/14992027.2017.1338762] [Medline: 28669224]

55. Probst T, Pryss R, Langguth B, Schlee W. Emotional states as mediators between tinnitus loudness and tinnitus distress in daily life: Results from the "TrackYourTinnitus" application. Sci Rep 2016 Feb 08;6:20382 [FREE Full text] [doi: 10.1038/srep20382] [Medline: 26853815]

56. Probst T, Pryss RC, Langguth B, Spiliopoulou M, Landgrebe M, Vesala M, et al. Outpatient Tinnitus Clinic, Self-Help Web Platform, or Mobile Application to Recruit Tinnitus Study Samples? Front Aging Neurosci 2017;9:113 [FREE Full text] [doi: 10.3389/fnagi.2017.00113] [Medline: 28484389]

57. Pryss R, Probst T, Schlee W, Schobel J, Langguth B, Neff P, et al. Prospective crowdsensing versus retrospective ratings of tinnitus variability and tinnitus–stress associations based on the TrackYourTinnitus mobile platform. Int J Data Sci Anal 2018 Mar 12;8(4):327-338. [doi: 10.1007/s41060-018-0111-4]

58. Mosa ASM, 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]

59. Nagaraj MK, Prabhu P. Internet/smartphone-based applications for the treatment of tinnitus: a systematic review. Eur Arch Otorhinolaryngol 2020 Mar;277(3):649-657. [doi: 10.1007/s00405-019-05743-8] [Medline: 31807891]

60. Kalle S, Schlee W, Pryss RC, Probst T, Reichert M, Langguth B, et al. Review of Smart Services for Tinnitus Self-Help, Diagnostics and Treatments. Front Neurosci 2018;12:541 [FREE Full text] [doi: 10.3389/fnins.2018.00541] [Medline: 30177869]

61. Lui JHL, Marcus DK, Barry CT. Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. Professional Psychology: Research and Practice 2017 Jun;48(3):199-210. [doi: 10.1037/pro0000122]

62. Paglialonga A, Tognola G, Pinciroli F. Apps for Hearing Science and Care. Am J Audiol 2015 Sep;24(3):293-298. [doi: 10.1044/2015_AJA-14-0093] [Medline: 26649533]

63. Smith S, Sereda M. Smartphone Apps for Tinnitus Management. The Hearing Journal 2020;73(2):20. [doi: 10.1097/01.hj.0000654928.79548.08]

64. Levine DM, Co Z, Newmark LP, Groisser AR, Holmgren AJ, Haas JS, et al. Design and testing of a mobile health application rating tool. NPJ Digit Med 2020;3:74 [FREE Full text] [doi: 10.1038/s41746-020-0268-9] [Medline: 32509971]

Abbreviations

ICC: intraclass correlation
MARS: Mobile Application Rating Scale
mHealth: mobile health

Review

mHealth Interventions to Promote Anti-Retroviral Adherence in HIV: Narrative Review

Stephen B Lee[1,2], MD, FRCPC; Joanne Valerius[2], PhD

[1]Department of Medicine, Division of Infectious Diseases, University of Saskatchewan College of Medicine, Regina, SK, Canada
[2]Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR, United States

Corresponding Author:
Stephen B Lee, MD, FRCPC
Department of Medicine, Division of Infectious Diseases
University of Saskatchewan College of Medicine
Regina General Hospital, Unit 4E
Regina, SK, S4P 0W5
Canada
Phone: 1 306 766 4247
Email: leestephenz@gmail.com

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2020/9/e24250/

Abstract

Background: Antiretrovirals (ARVs) are key in the management of HIV. Although no cure exists, ARVs help patients live healthy lives and prevent transmission to others. Adherence to complex regimens is paramount to outcomes and in avoiding the emergence of drug-resistant viruses. The goal of therapy is to reach an undetectable viral load. However, adherence is a common problem, stemming from issues such as mental health, chaotic home situations, and busy work schedules. Mobile health (mHealth) represents a new approach in improving medication adherence, and multiple studies have been performed in this area.

Objective: This study aims to review the current implementation of mHealth in the management of HIV among different groups of patients.

Methods: We used PubMed, Academic Search Elite, and 1 journal database with various search terms to review the current implementation of mHealth in HIV care.

Results: Titles and abstracts were screened, and 61 papers were identified and fully reviewed. The literature was divided into lower- and higher-income nations, as defined by the United Nations. A total of 20 studies with quantitative results were identified, with 10 being text- and SMS-based interventions (the majority of these being in lower-income countries) and 8 being smartphone-based apps (primarily in higher-income countries). The majority of these studies determined whether there was an effect on adherence or biochemical parameters (viral load and CD4 count). Various qualitative studies have also been conducted, and many have focused on determining the specific design of interventions that were successful (frequency of messaging, types of messages, etc) as well as priorities for patients with regard to mHealth interventions.

Conclusions: There seems to be a role of mHealth in the management of HIV in lower-income nations; however, the optimal design of an intervention needs to be delineated. In higher-income countries, where the 2 significant risk factors were injection drugs and men who have sex with men, the benefit was less clear, and more research is needed.

(JMIR Mhealth Uhealth 2020;8(8):e14739)   doi:10.2196/14739

KEYWORDS

mHealth; HIV; antiretroviral; adherence; mobile phone

XSL•FO
RenderX

Introduction

Background

HIV remains to be a significant public health concern, causing an estimated 770,000 deaths worldwide, with significant mortality from poorly managed disease, even in developed nations [1,2]. Particularly in the developing world, there is significant morbidity with 79.1% of HIV prevalence centered in Africa (25,700,000 people) and Southeast Asia (3,500,000 people) [3,4].

The burden of this disease is not only health related but also impacts economics and development. A large proportion of deaths from HIV occurs in young and middle-aged adults; the same group of people who raise children, are caretakers, work, and teach future generations. Especially in lower-income countries, families who have had a death from HIV cope by pulling children out of school, reducing food intake, or using up agricultural resources prematurely [5]. In addition, although first-line antiretrovirals (ARVs) cost US $80 per year, the cost of second- and third-line regiments (used with drug resistance) may be as high as US $2200 per year [6].

Despite all these bleak consequences, studies have shown that with adherence to ARV therapy, a patient aged 20 years could expect an additional 43.3 years of life [7]. In contrast, those who are not adherent and develop AIDS have a median survival of 12.1 months [8]. Adherence to ARVs is key to prolonging life, improving immune function and quality of life, and preventing transmission [9-14].

Without proper adherence, there is a risk of disease progression or drug resistance. Despite increasing access to ARVs worldwide, adherence is far from perfect [15,16]. One meta-analysis reported adherence at 62%, and studies on subgroups, such as illicit drug users, are as low as 27% [17,18].

Factors such as psychiatric disorders, cognitive impairment, social stigma, substance abuse, and volatile housing can prevent optimal adherence [19,20]. Burch et al [21] found clear associations between markers of the social determinants of health and HIV outcomes. Rachlis et al [22] found that substance abuse and unemployment were associated with suboptimal care. The risk of transmission and resistance make it crucial to develop methods for improving adherence to medications. Indeed, international organizations have called for strategies to improve adherence to medication and its importance [23-25].

Consumer informatics, particularly mobile health (mHealth), has shown promise in improving ARV adherence. Indeed, research has shown that patients living with HIV are interested in apps to support HIV self-management, and much past research has been done to support disease management [26,27].

Given the widespread adoption of cellphones (95% of the population) and smartphones (77%) in different parts of the world, mHealth is becoming an important tool to improve care [28]. Furthermore, in North America, access to mobile data and the internet ranges close to 80%, and with expanding mobile computing capabilities, this represents further opportunities [29].

Objectives

We believe that the major benefits that mHealth could offer would be helping with medication and appointment adherence, improving HIV education, and increasing engagement in care. Some studies have shown promise in improving viral suppression and CD4 count. However, there are existing gaps in current reviews [30]. First, as technology is a rapidly growing field, constantly updated work is needed to summarize developments. Second, and most importantly, people who live with HIV are a diverse group of patients with a variety of challenges, including living in developing countries with poor health infrastructure and having comorbid substance use issues [31-34]. Each of these patient populations has unique characteristics and concerns that may influence care. For instance, access to the internet and smartphones, health care, income, and psychiatric comorbidities could be vastly different in various groups. The majority of existing reviews homogenize and summarize the existing literature as a whole and do not consider the differences in patient populations [35-39]. An intervention may have varying levels of success in each group and likely needs to be tailored to fit unique needs. For instance, a smartphone app for disease education may be successful in college-educated, smartphone-owning patients; however, it may have a vastly different effect on a patient in the developing world without a smartphone or an advanced education. The underlying differences in patient groups need to be considered, and our review groups studies into subgroups and discusses the implications of such differences.

Methods

Search Strategy

Two separate databases with peer-reviewed, reputable articles were used, PubMed and Academic Search Elite. All relevant articles reviewed were within the time period of 2010 to 2019.

For PubMed, we used the following search terms: "mHealth AND HIV AND medication adherence," "app AND HIV AND adherence," "smartphone AND app AND HIV AND anti-retroviral adherence," and "mobile phone AND anti-retroviral adherence AND HIV." On reviewing the literature using these search terms, we added the search terms "WelTel" and "CAMPS and retroviral" to capture relevant studies found in the references of the articles retrieved from the initial search.

We used the search terms "mHealth AND HIV AND medication adherence," "mobile phone AND anti-retroviral adherence AND HIV," and "app AND HIV AND adherence" for the Academic Search Elite database search.

Finally, we used the *Journal of Medical Internet Research* database, a group of journals that focuses on informatics implementations in health care, to identify articles listed under "mHealth for Treatment Adherence."

Inclusion and Exclusion Criteria

A total of 386 search results were screened using the title or abstracts for relevance, and 61 articles were fully reviewed and are discussed in this study.

Articles were included if they were experimental in nature (encompassing qualitative or quantitative studies as well as randomized controlled trials [RCTs] or other studies). Study designs included RCTs, case studies, cohort studies, cross-sectional studies, and qualitative studies (focus groups, interviews, observations, and surveys). No age restrictions were set, although the majority of the literature focused on adults and studies were required to have all participants diagnosed with HIV. A significant number of articles were excluded (273/325, 84%). Reasons for exclusion were irrelevance to the topic discussed, inclusion of patients not diagnosed with HIV, articles written in a language other than English, reviews rather than original research on this topic, categorization as nonexperimental irrespective of whether qualitative or quantitative (for instance, a commentary of mHealth), or if they were protocols for experiments.

We defined mHealth in this study as any intervention that involved the use of mobile phones, smartphones, or other wireless devices (such as smart pill bottles). Although no formal definition of mHealth exists, this description is in agreement with the World Health Organization [40]. Pilot trials found on literature review were included in our review.

Results

Overview

A total of 61 papers were identified from the period of 2010 to 2019. In accordance with the mHealth categories previously described by the United Nations, we categorized apps according to type and found that these apps were focused on either treatment support (reminders) or education [41]. In addition, we categorized apps as either push or pull and as one-way versus dual communication. Interventions that were primarily reminders tended to use push, whereas those that were education used pull (Table 1).

As factors such as technology accessibility, socioeconomic background, and internet or cellular connectivity vary based on geographical location, the authors chose to separate the discussion into studies performed in lower- and higher-income countries. Studies were divided into categories (*upper middle and upper* vs *lower middle and low*) as per the definition by the United Nations [42].

The risk factor for HIV acquisition in upper middle- and upper-income countries was primarily men who have sex with men (MSM) or intravenous drug users (IVDU). The majority of these studies were performed in North America (Canada and the United States).

For lower middle- and low-income countries, the majority of these studies were based on the African continent, with some conducted in Asia.

Interventions were text or SMS based, voice message based, or smartphone apps. All work in the lower-income countries consisted of simple text or SMS or voice interventions (7/7, 100% of quantitative studies) [43-49]. In contrast, in higher-income countries, more of the work included advanced smartphone apps (8/13, 62%; Table 1) [50-58]. Although we have summarized the quantitative studies on efficacy in Table 1, other studies discussed the optimal designs or considerations in intervention development and are reviewed in the following sections.

On an evaluation, using the Oxford Centre for Evidence-Based Medicine's Levels of Evidence, there was a large amount of high-level evidence in developing countries, with many RCTs [59]. On the basis of this, we believe that conclusions regarding efficacy are possible. There is also a high level of evidence (with a large number of RCTs) in higher-income countries; however, there remain some limitations to those studies (Table 1). The authors believe that further work may need to be performed in higher-income countries.

Table 1. Quantitative studies on the effects of mobile health and ubiquitous Health interventions in HIV management and antiretroviral adherence.

Modes of intervention	Patient population	Designs	Effects	Push (central) vs pull (client) vs other	One-way vs dual communication	App type
Text message/SMS based (n=10)						
Resource-limited settings (n=5)						
Lester et al [43]; WelTel Kenya1	Low income, primarily urban population in Nairobi who were: 1. Starting ARVs[a] for the first time 2. >18 years old 3. Daily access to mobile phone. 538 randomized: 273 SMS and 265 control in primary analysis	Individually randomized multisite to SMS or control; intervention was weekly SMS with response required within 48 hours; primarily zidovudine or stavudine+lamivudine+efavirenz or nevirapine	Improved rate of viral suppression (57% vs 48%, $P=.04$); improved self-reported adherence (62% vs 50%; $P=.006$)	Push	Dual	Diagnostic and treatment support (reminder)
Pop-Eleches et al [44]	Rural government clinic in Kenya. 1. Patients were >18 years old 2. Patients were started ARV 3 months earlier	720 patients randomized to short daily (70), long daily (72), short weekly (73), and long weekly (74) messages. Data gathered from MEMS[b] caps	Improved adherence of >90% (53% vs 40%; $P=.03$); decreased lapses of >48 hours (in the weekly message subgroup; 81% vs 90%; $P=.03$)	Push	One way	Diagnostic and treatment support (reminder)
Maduka et al [45]; SMS and counseling	Tertiary care Nigerian hospital, patients who had the following: 1. ARVs for ≥3 months 2. History of nonadherence (<95%) 3. Access to mobile SMS	104 randomized to intervention of monthly adherence counseling and twice weekly ARV (52) or control with no SMS or counseling (52)	Improved self-reported adherence (76.9% vs 55.8%; $P=.02$); improved CD4 (193-575 cells/mL vs 131-361.5 cells/mL; $P=.007$)	Push	Dual	Diagnostic and treatment support (reminder)
Haberer et al [46]	Rural Southwestern Uganda. 1. Patients were >18 years old 2. Patients were initiating ARVs 3. Patients had an own operational cell 4. Patients had 1-2 social supporters 5. Patients were close to hospital	63 patients who were randomized to scheduled SMS and RTAM[c] (21), triggered SMS from RTAM (20), or RTAM only (21); the scheduled SMS group was daily for 1 month and weekly for 2 months, followed by reminders triggered by late or missed doses. SMS was also sent to *social supporters* if there was no signal >48 hours; triggered SMS was sent only in response to late or missed doses	Improved adherence (91% vs 79% vs 79%; $P=.04$); decreased lapses (7 vs 11 vs 11 for 48 hours and 1 vs 3 vs 4 at 96 hours; $P=.02$); no significant differences in VL[d] suppression	Other (one arm was *push* with scheduled messages, another arm was triggered by smart pill bottle)	One way	Diagnostic and treatment support (reminder)
Mbuagbaw et al [47]	Hospital in Yaounde, Cameroon with patients who: 1. Are >21 years old 2. Owned mobile phone, could read, and text Had >1 month of ARV	200 patients randomized to weekly SMS developed from focus groups that were varied, contemporary (eg, Season's Greetings), with a call back number (101) vs usual care (99)	No significant difference in adherence	Push	Dual	Diagnostic and treatment support (reminder)
Higher-income countries (n=5)						
Guo et al [60]	Large metropolitan South Chinese hospital. 1. Patients were >18 years old 2. Patients had >1 month of ARVs 3. Patients were able to read or write	62 primarily nonheterosexual males who were randomized to weekly SMS and reminders for ARVs and exercise with WeChat educational materials sent 3 times per month (31) or control (31)	No significant difference in CD4 counts or missed medications	Push	Dual	Diagnostic and treatment support (reminder)

Modes of intervention	Patient population	Designs	Effects	Push (central) vs pull (client) vs other	One-way vs dual communication	App type
Moore et al [61]; SMS for adherence and drug use, iTAB	Pilot at UCSD[e] whose patients: 1. Were >18 years old 2. Had DSM-IV-TR[f] diagnosis of methamphetamine abuse or dependence or self-reported use within 45 days 3. Were willing to participate in study components	75 randomized to iTAB, which were daily personalized texts built from focus groups and focused on responsibility to others, self-esteem, nonadherence risks, harm reduction, reminders, spirituality, celebration of health, and disease control (50) vs control (25). Also assessed methamphetamine use with daily texts	No significant difference in adherence (measured by MEMS caps); fewer methamphetamine use days (P=.05)	Push	Dual	Diagnostic and treatment support (reminder)
Ingersoll et al [51]; SMS including adherence, mood, and substance use	Pilot study from 2 outpatient clinics in Virginia serving mainly nonurban and rural, primarily male and white, patients who: 1. Were >18 years old 2. Had an active prescription for ARVs 3. Had <95% adherence in the past 2 weeks 4. Consumed illicit drugs or had a risky level of drinking within 30 days 5. Had good command of English	63 patients randomized to daily texts on medications, twice daily mood texts, and daily substance use texts eliciting responses from patients (33) or usual care (30)	No significant difference in proportion of missed visits (P=.12); no difference in substance use days; improved adherence (66%-85% vs 62%-71%; P=.04)	Push	Dual	Diagnostic and treatment support (reminder)
King et al [62]	Repeated measures study where patients recruited from a clinic for women and families living with HIV in Vancouver, British Columbia, who: 1. Attended a clinic for at least 1 year 2. Had CD4<500 cells/mm^3 3. Were detectable VL 1 year prior 4. Were ≥14 5. Had a high risk for disengagement	85 enrolled with 5 lost and administered intervention, which was modeled after WelTelKenya1 with SMS or texts every Monday asking "How are you?" requiring a response from patients within 1 day. If no response, this was followed-up by further texts and calls	Mean VL decreased from 1098 copies/mL to 439 copies/mL (P=.004); adherence to ARV (OR 1.14; P<.001); decreased appointment adherence (OR 0.81; P=.03)	Push	Dual	Diagnostic and treatment support (reminder)
Rana et al [63]	Patients recruited from Rhode Island clinic providing HIV services, primarily white, and MSM[i] as a risk factor with patients who: 1. Were ≥18 years 2. Had a cell phone capable of texting 3. Were newly engaged in care within 1 year or re-engaging after a lapse of >1 year or at risk for nonadherence	32 patients enrolled, with 20 completing study (exclusions were due to death, incarceration, transferal of care, or no response) given a daily bidirectional texting intervention	Improved VL suppression (P=.002)	Push	Dual	Diagnostic and treatment support (reminder)

Smartphone or computer apps (n=8)

Higher-income countries (n=8)

Modes of intervention	Patient population	Designs	Effects	Push (central) vs pull (client) vs other	One-way vs dual communication	App type
Venter et al [54]; lab results, appointment reminders, disease information	Multisite RCT[h] in the inner city of Johannesburg, South Africa (1 community health center, 3 clinics, 1 tertiary care hospital) with patients who: 1. Were >18 years old 2. Read English or Zulu 3. Were residents of the area	353 randomized to SmartLink, which provided appointment reminders, information about lab tests, ARV adherence and HIV info, and CD4 and VL results (181) as well as control (172)	No significant difference in adherence; improved linkage to care in the 18- to 30-year-old subgroup (20% increase, P=.02)	Push	One way	Diagnostic and treatment support (reminder) and education and awareness
Himelhoch t al [50]	Pilot study from a Baltimore urban outpatient HIV clinic whose patients: 1. Were aged 18-64 years 2. Were patients of the clinic's adherence program. The adherence counselor determined difficulty in adherence. 3. Had self-reported history of drug or alcohol use 4. Carried mobile phones	30 patients randomized to the Heart2HAART app giving medication reminders; information regarding adherence; ecological momentary for side effects, depression, and cravings for drug use and tailored education; recommendation; and encouragement (20) vs control of smartphone only (10)	No significant difference in adherence	Push	One way	Diagnostic and treatment support (reminder) and education and awareness
Dillingham et al [52]	Study from a clinic in Virginia whose patients were: 1. New (within 90 days) HIV diagnosis, returning to care after lapse, or considered elevated risk of nonadherence 2. Fourth grade reading level or better	77 enrolled in study primarily male, slightly less than half MSM given PositiveLinks iteratively designed user-centered app with tailored resources, queries of mood, stress, adherence; appointment reminders, and community forum	Improved retention in care at 12 months (P<.001); improved constant visits (P<.001); improved CD4 count (P<.001); improved VL suppression (P≤.001)	Pull	Dual	Diagnostic and treatment support (reminder) and education and awareness
Horvath et al [53]	Pilot study recruited from ads on Grindr from Miami, Orlando, Washington DC, Charlotte, Houston, and New Orleans as well as flyers or cards from organizations and clinics serving the targeted population. Patients were: 1. Male and MSM within 5 years 2. A US resident 3. Currently taking ARVs 4. Reporting suboptimal adherence 5. The use of illicit stimulants (eg, methamphetamines, cocaine, ecstasy, amphetamines) within 6 months 6. Owning an iPhone or Android phone	90 patients randomized to APP+ for MSM who use stimulants, giving informational material, a playable storyline of a fictional HIV+ character, and a tool to track personal adherence (45) vs control (45)	Temporary significant improvement in self-reported adherence (P=.04); no significant difference in substance use	Pull	One way	Education and awareness
Dworkin et al [55]	Patients in Chicago who were: 1. Aged 18-34 years old 2. On ARVs for >3 months 3. Owning an Android smartphone 4. Scheduled for a blood draw or visit during 3 months after baseline assessment 5. African American MSM	43 patients enrolled, with 11 lost to follow-up, given avatar conversational agent intervention on mobile phones	Improved adherence (via pill count adherence >80%; P=.05)	Pull	One way	Education and awareness
Whiteley et al [56]				Pull	One way	Education and awareness

XSL•FO
RenderX

Modes of intervention	Patient population	Designs	Effects	Push (central) vs pull (client) vs other	One-way vs dual communication	App type
	Patients recruited from the greater Jackson, MS area who were: 1. 14-26 years of age 2. Receiving or starting ARVs 3. Aware of status 4. Detectable VL within 1 month of screening 5. Speaking English 6. Able to consent	66 patients who were primarily male, black, and nonheterosexual enrolled, with 5 patients lost receiving BattleViro mobile game designed from qualitative user feedback that helped ARV adherence, viral load, and other knowledge regarding HIV	Greater decrease in VL (0.96 log greater decrease in intervention; P=.04); improved adherence (71% vs 48%; P=.05)			
Perera et al [57]	Patients who were primarily male and homosexual, recruited from the Auckland City Hospital Infectious Diseases Clinic and local HIV support organization who were: 1. On ARVs for at least 6 months 2. Using Android smartphones	28 randomized to augmented apps, including graphical representations of ARV concentrations and simulation of immune activity (17) vs normal apps with medication clock only (11)	Increased self-reported adherence via MARS[i] score (48.93 vs 47.09; P=.03); lower VL at 3 months (1.30 log10 vs 1.70 log10; P=.02)	Pull	One way	Education and awareness
Ownby et al [58]	Recruitment of patients in Broward County, FL with a large proportion of MSM[g]	124 participants recruited in educational intervention, with a final 109 patients analyzed for results	Improved adherence in subgroups with lower baseline adherence rates (in <85% adherence, P=.04; in <80% adherence P=.09)	Pull	One way	Education and awareness
Voice calls (n=2)						
Resource-limited settings (n=2)						
Shet et al [48]; HIVIND, interactive automated voice all and reminder pictorial messages	1 ambulatory and 1 private HIV clinics in India whose patients: 1. Were aged 18-60 years 2. Were ARV naïve 3. Had first-line ARV: (zidovudine or stavudine or tenofovir+lamivudine+nevirapine or efavirenz)	631 educated patients randomized to motivational call with inquiry of adherence requiring response and pictorial SMS (315) or control (316)	No significant difference in viral load; no significant difference in adherence	Push	Dual	Diagnostic and treatment support (reminder)
Rodrigues et al [49]	Bangalore, South Indian tertiary, nonprofit private facilities outpatients who: 1. Had access to mobile phones 2. Had >1 month on ARVs 3. Had first-line regimens (zidovudine or stavudine+lamivudine+nevirapine or efavirenz)	Quasi experimental, where 150 primarily male patients were all given automated interactive voice response call and noninteractive neutral picture SMS every week for 6 months and studied at the end of the intervention	Improved adherence (85%-91% during intervention and 6 months after, P=.02)	Push	Dual	Diagnostic and treatment support (reminder)

[a]ARV: antiretroviral.

[b]MEMS: medication event monitoring system

[c]RTAM: real-time adherence monitoring.

[d]VL: viral load.

[e]UCSD: University of California at San Diego.

[f]DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders – IV – Text Revision.

[g]MSM: men who have sex with men.

[h]RCT: randomized controlled trial.

[i]MARS: medication adherence report scale.

XSL•FO
RenderX

Lower-Income Countries Group

Given the rising rates of cellular access in the developing world, mHealth represents a significant opportunity to improve care [64].

A variety of studies have been conducted on mHealth interventions in the developing world, with a wide range of study methodologies. The quantitative experimental designs (including cohort studies and RCTs) are summarized in Table 1 for ease of comparison. Seven studies with quantitative data were available, with the majority being text- or SMS-based modalities (5 studies) and a minority being voice call based (2 studies) [43-49].

Other studies investigated user acceptability and surveyed patient concerns regarding the use of mHealth. A pilot study in rural Uganda by Musiimenta et al [65] enrolled 63 patients and randomized them to a scheduled SMS, triggered SMS (by missed or delayed doses), or no SMS. The feedback received was positive, citing usefulness for motivation, reminders, and social support. However, confidentiality, shared phones, ability to use the phone, and availability of electricity were concerns [65]. Similarly, an explorative study by Lepere et al [66] demonstrated that mHealth was acceptable among people who live with HIV in Cote d'Ivoire, Togo, and Burkina Faso and advocated for its use. Mbuagbaw et al [67] also demonstrated, in a qualitative arm of the CAMPS trial, that these interventions were acceptable, although content, timing, technical challenges, and privacy need to be considered.

Higher-Income Countries Group

Countries classified as high or high middle income were analyzed. The primary risk factors in these studies were either IVDU or MSM, in agreement with the previous literature [68,69].

Quantitative experimental studies are described in Table 1 [50-58,60-63,70]. In addition, a number of nonquantitative or nonexperimental studies were conducted to investigate the nature of optimal interventions. These are summarized in Table 2 [71-84].

Different studies have investigated aspects of mHealth apps other than efficacy. WelTel was studied in Vancouver, British Columbia, on patients involved with substance abuse using both quantitative and qualitative methodologies [62,84]. One such study was done by Campbell et al [85], who studied WelTelBC1 from an economic perspective on 85 viral load–positive vulnerable patients (defined as substance use and other psychosocial factors), finding that the cost of caring for a highly vulnerable patient was US $347.74, whereas the overall median cost was US $36.72.

Multiple quantitative studies investigated the acceptability of mHealth interventions. In Seattle, 224 patients with HIV were randomized to receive two-way pagers with personalized adherence messages. Of these patients, 55% identified as gay

or lesbian, and the majority were white males. Participants were surveyed for the usefulness of the pager and asked whether they adhered to their medications; 73% reported liking the pager, 51% began finding messages annoying, and 48% believed that the pager did not improve their adherence. The overall response rate was only 42.8% but declined over 3 months. The self-reported pager adherence was 90.8% (SD 33%) compared with the medication event monitoring system showing 53.6% (SD 37%) adherence [86].

Similarly, a Peruvian study by Krishnan et al [87] with 359 transgender women or MSM patients asked about mHealth preferences using a 5-point Likert scale. They found an overall positive uptake, with a mean of 3.21 for technological medication reminders and 3.56 for anonymous internet chats to discuss HIV issues.

Other studies, using focus groups and other quantitative study methods, were used to determine how best to deliver mHealth interventions. For instance, Krishnan et al [87] also found that daily messages were preferred in terms of frequency within their study population. In Bangkok, Anand et al [88] interviewed 16 MSM and 2 transgender women. They gathered information on preferences and how to address needs with electronic health. The group found that 39% (7/18) of patients preferred instant messaging and 11% (2/18) preferred phone calls. Of these, 39% (7/18) wanted a private website chat room and 11% (2/18) preferred Skype (Skype Communications) video chats. All patients desired personalized reminder messages; 50% (9/18) wanted reminders on an instant message app, 17% (3/18) preferred a stand-alone app, and 33% (6/18) wanted it as an SMS. The theme that emerged was web based, accessible, and reliable disease information [88].

Work performed in higher-income countries also had diversity in intervention type. For instance, Skrajner et al [89] enrolled a patient with cognitive impairment in a video-conferenced, cognitive-psychosocial program that improved the pill count adherence rate from 75% to 97.9% in 1 month (80% in 2-3 months). Similarly, the weCare team, which has done non-HIV work on social media, is currently developing a tailored Facebook, text, and social and sexual networking app–based intervention to help engage care [90]. Hwayoung et al [91] combined an electronic pill bottle, fitness tracker, and phone alerts and found that the pill bottle encouraged adherence and self-management of medications. However, they mentioned that the smart pill bottle was too small to fit all the pills, although they liked how it was easy to open and discrete [91].

Finally, Schnall et al [92] created the mVIP app, which attempted to provide self-care strategies for self-management of different HIV-related symptoms. They found improvements in a variety of symptoms, including anxiety (P=.001), depression (P=.001), neuropathy (P=.002), fevers or chills or sweats (P=.04), and weight loss (P=.02). Results for adherence were mixed, showing improvement only when measuring it via the center for adherence support evaluation adherence index [92].

Table 2. Qualitative studies on the priorities and important themes in mobile health interventions.

Study	Designs	Themes reported by patients
Senn et al [71]	22 African American MSM[a] with HIV in Rochester completed surveys and qualitative interviews	1. Importance of social support 2. Convenience 3. Anonymous and confidential
LeGrand et al [72,73]	EPIC Allies (University of North Carolina, Duke University) undergoing study now. Game-like app that uses game mechanics and social networking to improve HIV care. Development study looked at ARV[b] adherence needs, preferences, and usability testing in young MSM and transgendered women who have sex with men	1. Information about side effects 2. Discrete medication reminders 3. Interactive 4. Engaging 5. Social 6. Informational 7. Customizable
Dworkin et al [74]	5 different focus groups composed of African American MSM with HIV in Chicago iteratively designed a talking avatar app providing information about disease, improved adherence, and helped with appointment attendance	1. Positive impression 2. Importance of confidentiality 3. Motivational over negative messages 4. Customizable
Castel et al [75]	Focus groups and surveys on patients with HIV aged between 13 and 24 year (no delineation as to acquisition) for 3 different game prototypes that were linked with Wisepill dispensers	1. Game was feasible and acceptable 2. Customizable, challenging, and user-friendly games
Morano et al [76,77]	Primarily, 132 African American males enrolled to the Care4Today app that helped with medication management and appointments and tracked health, wellness, and goals	1. Higher education levels, staff support, and possession of smart phone predicted use of app 2. 70.2% of patients interested in medication reminders
Cook et al [78]	37 patients in Colorado studied to determine if messages matched to psychological states were useful. Patients were primarily ethnic minorities and nonheterosexual	1. Changing messages improved adherence (potentially indicating importance of novelty)
Olalla et al [79]	Spanish study on 30 patients (15 randomized to app group) who were >60 years old. The app offered medical news, reports about HIV, and an anonymous chat between patients	1. No clinical or analytical parameter changes (likely due to short intervention period and small sample) 2. Patients enjoyed the social aspect of the app 3. Analysis of the chat logs showed interested in general health over HIV specifically 4. Most patients seemed to have positive emotions and often mentioned happiness
Westergaard et al [80]	Recruited patients with HIV who were absent from appointments, had substance use, and had an unsuppressed viral load. Investigated acceptability and uptake of a 2-part intervention that included a smartphone app (delivering tailored interventions and communication) and a peer navigator (psychosocial and logistical support)	1. The participants universally commented on the app's usefulness as medication and appointment reminders 2. Some requested alternate apps at the end of the trial
Przybyla et al [81]	Used app for adult patients with HIV who had been on ARVs for >3 months with history of alcohol and a history of at least one recent day of nonadherence. Evaluated feasibility and acceptability of the app	1. Found high report completion rates demonstrating feasibility and acceptability of the app 2. Challenges include those with limited smartphone experience
Horvath et al [82]	Focus groups for stimulant using MSM with HIV (San Francisco, Minneapolis) and explored app components that were important for continued use and engagement	1. Important features included: low cost, customizable, integrate well into their lives, be engaging, credible, private, and provide appointment and medication reminders
Rosen et al [83]	Focus groups of patients with HIV with a history of substance abuse using the *iHAART* app (which used visual representations of adherence, VL[c], and CD4)	1. Need a balance of requested and provided information 2. Emotions provoked by the app can affect adherence
Smillie et al [84]	Pilot WelTel study in Vancouver, British Columbia, on patients with substance abuse using a qualitative methodology of semistructured interviews	1. Participants thought app was more useful as means of accessing psychosocial support and health care providers rather than as a reminder or source of information

[a]MSM: men who have sex with men.

[b]ARV: antiretroviral.

[c]VL: viral load.

XSL•FO
RenderX

Discussion

Lower-Income Countries Group

Many studies have been performed using mHealth to improve adherence, with most in sub-Saharan Africa. Some of the studies were rigorous, large RCTs, such as Kenya WelTel [43]. There were some limitations in other studies where mHealth was not studied independently. For instance, in the study by Maduka et al [45], the effects of adherence counseling and SMS messages were not separated, and in the trial by Haberer et al [46], the role of social supporters was unclear. However, most of the studies indicate an improvement in adherence and VL (Table 1) [43-46,48,49]. Furthermore, it seems that mHealth represents a cost-effective intervention compared with the alternatives, with 1 study showing that a nurse could manage 1000 patients per week [43].

Although most studies found benefits, there is a gap in the literature regarding the design of a successful intervention. It was found that daily messages were not optimal, habituation likely being a significant factor [43-47]. Pop-Eleches et al [44] found that weekly interventions were more effective than daily interventions and long versus short messages made no difference, potentially reflecting that the messages were used more as reminders.

Most other successful interventions agreed with tapering schedules, twice weekly messaging in Nigeria, weekly WelTel Kenya1 messages, and the weekly South Indian study [43,45,46,49]. Another variable was message interactivity and content. On the basis of the existing literature, it is unclear if the interactivity of messages is important. In terms of content, Pop-Eleches et al [44] showed that a motivational message "This is your reminder. Be strong and courageous, we care about you" was not better than a generic impersonal message. These findings potentially indicate that an interactive or customized message does not negate the effects of habituation. Further work is necessary to determine the exact type of message that generates the maximal benefit.

Cultural contexts likely play a role in feasibility, and caution should be exercised when extrapolating results to different locations. Different views on health care, HIV stigma, baseline demographics, and cultural or religious views could be important when designing an intervention. The HIVIND study illustrates this, with high baseline adherence further boosted by the Hawthorne effect [48].

These studies collectively show a role for mHealth in improving HIV care in low-income or lower middle-income nations; however, we still need to delineate the best design. A less-than-daily message frequency seems to be optimal; however, interactivity and content need to be considered. Geographical and cultural differences may also affect efficacy.

Higher-Income Countries Group

The major risk factor for HIV in higher-income countries was either IVDU or MSM [93-98]. Although in higher-income countries different factors were identified in HIV management compared with lower-income countries, optimizing adherence is still the key.

In those with IVDU as a risk factor, compliance is important as it has implications for disease transmission. Finding meaningful interventions to improve adherence is crucial as there are higher rates of drug use in patients with HIV+, and these patients often face barriers to adherence [20,99].

However, there are a series of unique challenges for these patients that have implications in implementation and interpretation of studies. An inherent difficulty in studies performed on patients with a history of substance misuse is their heterogeneity, with a variety of substances used, including alcohol, marijuana, cocaine, amphetamines, and opioids. Prolonged use of drugs such as methamphetamines can cause psychotic episodes, resulting in more chaotic lives than other addictive substances [100,101]. Opioids have substitution agents that represent an opportunity for frequent, recurrent health care interaction in dispensing [102]. An intervention efficacious in one group may not be efficacious in another.

In addition, a variety of barriers exist that confound the ability of mHealth interventions to help with adherence. These include unstable financial and housing situations as well as psychiatric comorbidities [103]. Some of these barriers may limit the ability for any intervention to be beneficial without first addressing them.

The majority of the interventions studied in patients with substance misuse were text or SMS based, whereas those focusing on populations with MSM as the major risk factor had more advanced smartphone apps (Table 1) [50-58,60-63]. A potential explanation is the financial resources required for a smartphone and an accompanying plan, which supports the need for stable financial situations before a successful intervention.

Overall, across risk groups in higher-income countries, there seemed to be a theme of advanced mHealth interventions (using apps, social media, and game-like interventions). Examples in our review included EPIC Allies, the talking avatar, and the weCare interventions that contrasted mainly to call and text-related interventions in lower-income countries [72-74,90]. This could potentially be explained by the availability of smartphones and internet connectivity in higher-income countries [28,29].

Regardless of risk factors, a limitation of many of these individual studies was the small sample size. For instance, the iTAB study only showed an effect on methamphetamine use, but as this factor could contribute to nonadherence, a larger sample size may have had an effect on adherence [61].

In addition, studies outside North America should be interpreted with caution if implemented in North America, as many different considerations exist. For instance, different regions of the world preferentially use different messaging apps [104]. These may have different functions and capabilities. Certainly, further work is required [105,106].

Design-related studies indicated the themes that interventions should be customizable, provide disease information, and maintain confidentiality [71-84]. A reminder intervention seems to be acceptable in the patients studied [76,77,80,82]. The role of applications as a conduit for health care support was also emphasized [71,84].

A reasonable conclusion from the totality of the data is that there is a role in improving adherence in higher-income countries, but specific nuances need to be considered on a case-by-case basis. There are limitations affecting the ability to make conclusions, including small sample sizes, heterogeneous populations, and heterogeneous interventions [53,54,60,62,80-82,88].

Conclusions

This study attempted to review the existing literature on mHealth in HIV adherence, divided the literature to the largest subgroups for which evidence was found, and discussed existing studies with each of these groups as specific contexts.

On the basis of an evaluation of both the results and the quality of evidence available, we believe a clear role for mHealth interventions in developing countries exists; however, further work is needed before a final conclusion can be made in the higher-income countries, where the 2 main subgroup risk factors were MSM and IVDU. In addition, a gap in the literature in all 3 groups is the exact nature of interventions (optimal message frequency, content, and intervention themes).

In addition, as demonstrated by the variety and scope of interventions in our review, there is a difference between digital and mobile access, with some groups having the hardware and connectivity to access advanced apps such as smartphone apps, whereas other groups simply have mobile access and can only access text or SMS or voice call interventions.

Our review of the literature provides an optimistic outlook on the role of mHealth in improving HIV care; however, there are limitations to our work. Our narrative, compared with a systematic methodology, provides a broad and comprehensive overview of the subject area but does not have the stringent inclusion and exclusion criteria of a systematic review. However, our review provides a broad view of the subject and identifies specific focus areas for future systematic reviews. Specific areas of future work include better delineation of the efficacy of different types of interventions in higher-income countries and in specific risk groups within higher-income countries.

Conflicts of Interest

None declared.

References

1. Global Health Observatory (GHO) Data: Number of Deaths Due to HIV/AIDS. World Health Organization. 2019. URL: https://www.who.int/gho/hiv/epidemic_status/deaths_text/en/ [accessed 2020-01-11]

2. Croxford S, Kitching A, Desai S, Kall M, Edelstein M, Skingsley A, et al. Mortality and causes of death in people diagnosed with HIV in the era of highly active antiretroviral therapy compared with the general population: an analysis of a national observational cohort. Lancet Public Health 2017 Jan;2(1):e35-e46 [FREE Full text] [doi: 10.1016/S2468-2667(16)30020-2] [Medline: 29249478]

3. Prevalence of HIV Among Adults Aged 15 to 49, 2017 by WHO Region. World Health Organization. 2017. URL: https://www.who.int/gho/hiv/hiv_013.png?ua=1 [accessed 2019-03-10]

4. Number of Deaths Due to HIV/AIDS: Estimates by WHO Region. World Health Organization. 2019. URL: https://www.who.int/gho/hiv/epidemic_status/deaths_text/en/ [accessed 2019-05-16] [WebCite Cache ID 78QBZ5HZn]

5. Preparing for the Future of HIV/AIDS in Africa: A Shared Responsibility. The National Academies Press. 2011. URL: https://www.nap.edu/catalog/12991/preparing-for-the-future-of-hivaids-in-africa-a-shared [accessed 2020-01-20]

6. Confronting the HIV/AIDS Deadly Pandemic. Médecins Sans Frontières (MSF) International. 2019. URL: https://www.msf.org/hivaids [accessed 2019-05-16] [WebCite Cache ID 78QBnp636]

7. Teeraananchai S, Kerr S, Amin J, Ruxrungtham K, Law M. Life expectancy of HIV-positive people after starting combination antiretroviral therapy: a meta-analysis. HIV Med 2017 Apr;18(4):256-266 [FREE Full text] [doi: 10.1111/hiv.12421] [Medline: 27578404]

8. Yarchoan R, Venzon DJ, Pluda JM, Lietzau J, Wyvill KM, Tsiatis AA, et al. CD4 count and the risk for death in patients infected with HIV receiving antiretroviral therapy. Ann Intern Med 1991 Aug 1;115(3):184-189. [doi: 10.7326/0003-4819-115-3-184] [Medline: 1676252]

9. Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents with HIV. AIDSinfo. 2019. URL: https://aidsinfo.nih.gov/guidelines/html/1/adult-and-adolescent-arv/10/initiation-of-antiretroviral-therapy [accessed 2019-05-16] [WebCite Cache ID 78QAAmKdR]

10. Le T, Wright EJ, Smith DM, He W, Catano G, Okulicz JF, et al. Enhanced CD4+ T-cell recovery with earlier HIV-1 antiretroviral therapy. N Engl J Med 2013 Jan 17;368(3):218-230 [FREE Full text] [doi: 10.1056/NEJMoa1110187] [Medline: 23323898]

11. INSIGHT START Study Group, Lundgren JD, Babiker AG, Gordin F, Emery S, Grund B, et al. Initiation of antiretroviral therapy in early asymptomatic HIV infection. N Engl J Med 2015 Aug 27;373(9):795-807 [FREE Full text] [doi: 10.1056/NEJMoa1506816] [Medline: 26192873]

12. Langford SE, Ananworanich J, Cooper DA. Predictors of disease progression in HIV infection: a review. AIDS Res Ther 2007 May 14;4:11. [doi: 10.1186/1742-6405-4-11] [Medline: 17502001]

13. Samji H, Cescon A, Hogg RS, Modur SP, Althoff KN, Buchacz K, North American AIDS Cohort Collaboration on Research and Design (NA-ACCORD) of IeDEA. Closing the gap: increases in life expectancy among treated HIV-positive individuals in the United States and Canada. PLoS One 2013;8(12):e81355 [FREE Full text] [doi: 10.1371/journal.pone.0081355] [Medline: 24367482]

14. Lohse N, Obel N. Update of survival for persons with HIV infection in Denmark. Ann Intern Med 2016 Nov 15;165(10):749-750. [doi: 10.7326/L16-0091] [Medline: 27842400]

15. Yoong D, Bayoumi AM, Robinson L, Rachlis B, Antoniou T. Public prescription drug plan coverage for antiretrovirals and the potential cost to people living with HIV in Canada: a descriptive study. CMAJ Open 2018;6(4):E551-E560 [FREE Full text] [doi: 10.9778/cmajo.20180058] [Medline: 30482757]

16. World Health Organization. Access to Antiretroviral Drugs in Low- and Middle-income Countries: Technical Report July 2014. Geneva, Switzerland: World Health Organization; 2014.

17. Ortego C, Huedo-Medina TB, Llorca J, Sevilla L, Santos P, Rodríguez E, et al. Adherence to highly active antiretroviral therapy (HAART): a meta-analysis. AIDS Behav 2011 Oct;15(7):1381-1396. [doi: 10.1007/s10461-011-9942-x] [Medline: 21468660]

18. Iacob SA, Iacob DG, Jugulete G. Improving the adherence to antiretroviral therapy, a difficult but essential task for a successful HIV treatment-clinical points of view and practical considerations. Front Pharmacol 2017;8:831 [FREE Full text] [doi: 10.3389/fphar.2017.00831] [Medline: 29218008]

19. Saracino A, Zaccarelli M, Lorenzini P, Bandera A, Marchetti G, Castelli F, Icona Foundation Study Group. Impact of social determinants on antiretroviral therapy access and outcomes entering the era of universal treatment for people living with HIV in Italy. BMC Public Health 2018 Jul 13;18(1):870 [FREE Full text] [doi: 10.1186/s12889-018-5804-z] [Medline: 30005709]

20. Feldman MB, Kepler KL, Irvine MK, Thomas JA. Associations between drug use patterns and viral load suppression among HIV-positive individuals who use support services in New York City. Drug Alcohol Depend 2019 Apr 1;197:15-21. [doi: 10.1016/j.drugalcdep.2018.12.015] [Medline: 30743195]

21. Burch LS, Smith CJ, Anderson J, Sherr L, Rodger AJ, O'Connell R, et al. Socioeconomic status and treatment outcomes for individuals with HIV on antiretroviral treatment in the UK: cross-sectional and longitudinal analyses. Lancet Public Health 2016 Nov;1(1):e26-e36 [FREE Full text] [doi: 10.1016/S2468-2667(16)30002-0] [Medline: 28299369]

22. Rachlis B, Burchell AN, Gardner S, Light L, Raboud J, Antoniou T, Ontario HIV Treatment Network Cohort Study. Social determinants of health and retention in HIV care in a clinical cohort in Ontario, Canada. AIDS Care 2017 Jul;29(7):828-837. [doi: 10.1080/09540121.2016.1271389] [Medline: 28027668]

23. Chaiyachati KH, Ogbuoji O, Price M, Suthar AB, Negussie EK, Bärnighausen T. Interventions to improve adherence to antiretroviral therapy: a rapid systematic review. AIDS 2014 Mar;28(Suppl 2):S187-S204. [doi: 10.1097/QAD.0000000000000252] [Medline: 24849479]

24. Koester KA, Johnson MO, Wood T, Fredericksen R, Neilands TB, Sauceda J, et al. The influence of the 'good' patient ideal on engagement in HIV care. PLoS One 2019;14(3):e0214636 [FREE Full text] [doi: 10.1371/journal.pone.0214636] [Medline: 30921440]

25. Cheever LW. Engaging HIV-infected patients in care: their lives depend on it. Clin Infect Dis 2007 Jun 1;44(11):1500-1502. [doi: 10.1086/517534] [Medline: 17479949]

26. Lucero RJ, Frimpong JA, Fehlberg EA, Bjarnadottir RI, Weaver MT, Cook C, et al. The relationship between individual characteristics and interest in using a mobile phone app for HIV self-management: observational cohort study of people living with HIV. JMIR Mhealth Uhealth 2017 Jul 27;5(7):e100 [FREE Full text] [doi: 10.2196/mhealth.7853] [Medline: 28751298]

27. Yi JY, Kim Y, Cho Y, Kim H. Self-management of chronic conditions using mhealth interventions in Korea: a systematic review. Healthc Inform Res 2018 Jul;24(3):187-197 [FREE Full text] [doi: 10.4258/hir.2018.24.3.187] [Medline: 30109152]

28. Mobile Fact Sheet. Pew Research Center. 2019. URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-05-16] [WebCite Cache ID 78QCxCcj9]

29. Bauman K. New Survey Questions Do a Better Job Capturing Mobile Use. United States Census. 2019. URL: https://www.census.gov/library/stories/2018/08/internet-access.html [accessed 2019-05-16] [WebCite Cache ID 78QChW6sa]

30. Catalani C, Philbrick W, Fraser H, Mechael P, Israelski DM. Mhealth for HIV treatment & prevention: a systematic review of the literature. Open AIDS J 2013;7:17-41 [FREE Full text] [doi: 10.2174/1874613620130812003] [Medline: 24133558]

31. HIV Prevention and Control Report. Government of Saskatchewan. 2017. URL: https://www.saskatchewan.ca/government/government-structure/ministries/health/other-reports/annual-report-archive [accessed 2020-07-24]

32. HIV Surveillance Reports. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2020-01-06]

33. Joseph HA, Pan Y, Mendoza M, Harawa NT, Lauby J, Hosek SG, et al. HIV acquisition and transmission potential among African American men who have sex with men and women in three US cities. Arch Sex Behav 2018 Jan;47(1):183-194. [doi: 10.1007/s10508-017-1052-z] [Medline: 29124541]

XSL•FO

RenderX

34.   Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV
        infection in men who have sex with men. Lancet 2012 Jul 28;380(9839):367-377 [FREE Full text] [doi:
        10.1016/S0140-6736(12)60821-6] [Medline: 22819660]

35.   Cooper V, Clatworthy J, Whetham J, Consortium E. Mhealth interventions to support self-management in HIV: a systematic
        review. Open AIDS J 2017;11:119-132 [FREE Full text] [doi: 10.2174/1874613601711010119] [Medline: 29290888]

36.   Daher J, Vijh R, Linthwaite B, Dave S, Kim J, Dheda K, et al. Do digital innovations for HIV and sexually transmitted
        infections work? Results from a systematic review (1996-2017). BMJ Open 2017 Nov 3;7(11):e017604 [FREE Full text]
        [doi: 10.1136/bmjopen-2017-017604] [Medline: 29101138]

37.   Quintana Y, Gonzalez Martorell EA, Fahy D, Safran C. A systematic review on promoting adherence to antiretroviral
        therapy in HIV-infected patients using mobile phone technology. Appl Clin Inform 2018 Apr;9(2):450-466 [FREE Full
        text] [doi: 10.1055/s-0038-1660516] [Medline: 29925099]

38.   Wang Z, Zhu Y, Cui L, Qu B. Electronic health interventions to improve adherence to antiretroviral therapy in people living
        with HIV: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Oct 16;7(10):e14404 [FREE Full text] [doi:
        10.2196/14404] [Medline: 31621641]

39.   Mulawa MI, LeGrand S, Hightow-Weidman LB. Ehealth to enhance treatment adherence among youth living with HIV.
        Curr HIV/AIDS Rep 2018 Aug;15(4):336-349 [FREE Full text] [doi: 10.1007/s11904-018-0407-y] [Medline: 29959649]

40.   World Health Organization. MHealth: New Horizons for Health Through Mobile Technologies. Geneva, Switzerland:
        World Health Organization; 2011.

41.   mHealth for Development: The Opportunity of Mobile Technology for Healthcare in the Developing World. Web Archive.
        2011. URL: https://web.archive.org/web/20121203014521/http://vitalwaveconsulting.com/pdf/2011/mHealth.pdf [accessed
        2020-05-04]

42.   World Economic Situation and Prospects. United Nations. 2020. URL: https://www.un.org/development/desa/dpad/
        wp-content/uploads/sites/45/WESP2020_FullReport.pdf [accessed 2020-06-11]

43.   Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on
        antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845.
        [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]

44.   Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de Walque D, et al. Mobile phone technologies
        improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message
        reminders. AIDS 2011 Mar 27;25(6):825-834 [FREE Full text] [doi: 10.1097/QAD.0b013e32834380c1] [Medline: 21252632]

45.   Maduka O, Tobin-West CI. Adherence counseling and reminder text messages improve uptake of antiretroviral therapy in
        a tertiary hospital in Nigeria. Niger J Clin Pract 2013;16(3):302-308 [FREE Full text] [doi: 10.4103/1119-3077.113451]
        [Medline: 23771450]

46.   Haberer JE, Musiimenta A, Atukunda EC, Musinguzi N, Wyatt MA, Ware NC, et al. Short message service (SMS) reminders
        and real-time adherence monitoring improve antiretroviral therapy adherence in rural Uganda. AIDS 2016 May
        15;30(8):1295-1300 [FREE Full text] [doi: 10.1097/QAD.0000000000001021] [Medline: 26760452]

47.   Mbuagbaw L, Thabane L, Ongolo-Zogo P, Lester RT, Mills E, Volmink J, et al. The Cameroon mobile phone SMS (CAMPS)
        trial: a protocol for a randomized controlled trial of mobile phone text messaging versus usual care for improving adherence
        to highly active anti-retroviral therapy. Trials 2011 Jan 7;12:5 [FREE Full text] [doi: 10.1186/1745-6215-12-5] [Medline:
        21211064]

48.   Shet A, de Costa A, Kumarasamy N, Rodrigues R, Rewari BB, Ashorn P, HIVIND Study Team. Effect of mobile telephone
        reminders on treatment outcome in HIV: evidence from a randomised controlled trial in India. Br Med J 2014 Oct
        24;349:g5978 [FREE Full text] [doi: 10.1136/bmj.g5978] [Medline: 25742320]

49.   Rodrigues R, Shet A, Antony J, Sidney K, Arumugam K, Krishnamurthy S, et al. Supporting adherence to antiretroviral
        therapy with mobile phone reminders: results from a cohort in South India. PLoS One 2012;7(8):e40723 [FREE Full text]
        [doi: 10.1371/journal.pone.0040723] [Medline: 22952574]

50.   Himelhoch S, Kreyenbuhl J, Palmer-Bacon J, Chu M, Brown C, Potts W. Pilot feasibility study of Heart2HAART: a
        smartphone application to assist with adherence among substance users living with HIV. AIDS Care 2017 Jul;29(7):898-904.
        [doi: 10.1080/09540121.2016.1259454] [Medline: 28043176]

51.   Ingersoll KS, Dillingham RA, Hettema JE, Conaway M, Freeman J, Reynolds G, et al. Pilot RCT of bidirectional text
        messaging for ART adherence among nonurban substance users with HIV. Health Psychol 2015 Dec;34S:1305-1315 [FREE
        Full text] [doi: 10.1037/hea0000295] [Medline: 26651472]

52.   Dillingham R, Ingersoll K, Flickinger TE, Waldman AL, Grabowski M, Laurence C, et al. PositiveLinks: a mobile health
        intervention for retention in HIV care and clinical outcomes with 12-month follow-up. AIDS Patient Care STDS 2018
        Jun;32(6):241-250 [FREE Full text] [doi: 10.1089/apc.2017.0303] [Medline: 29851504]

53.   Horvath KJ, Lammert S, MacLehose RF, Danh T, Baker JV, Carrico AW. A pilot study of a mobile app to support HIV
        antiretroviral therapy adherence among men who have sex with men who use stimulants. AIDS Behav 2019
        Nov;23(11):3184-3198. [doi: 10.1007/s10461-019-02597-3] [Medline: 31309348]

54. Venter WD, Fischer A, Lalla-Edward ST, Coleman J, Lau Chan V, Shubber Z, et al. Improving linkage to and retention in care in newly diagnosed hiv-positive patients using smartphones in South Africa: randomized controlled trial. JMIR Mhealth Uhealth 2019 Apr 2;7(4):e12652 [FREE Full text] [doi: 10.2196/12652] [Medline: 30938681]

55. Dworkin MS, Lee S, Chakraborty A, Monahan C, Hightow-Weidman L, Garofalo R, et al. Acceptability, feasibility, and preliminary efficacy of a theory-based relational embodied conversational agent mobile phone intervention to promote HIV medication adherence in young HIV-positive African American MSM. AIDS Educ Prev 2019 Feb;31(1):17-37. [doi: 10.1521/aeap.2019.31.1.17] [Medline: 30742481]

56. Whiteley L, Brown LK, Mena L, Craker L, Arnold T. Enhancing health among youth living with HIV using an iPhone game. AIDS Care 2018;30(sup4):21-33 [FREE Full text] [doi: 10.1080/09540121.2018.1503224] [Medline: 30626196]

57. 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]

58. Ownby RL, Waldrop-Valverde D, Caballero J, Jacobs RJ. Baseline medication adherence and response to an electronically delivered health literacy intervention targeting adherence. Neurobehav HIV Med 2012 Oct 18;4:113-121 [FREE Full text] [doi: 10.2147/NBHIV.S36549] [Medline: 23293544]

59. OCEBM Levels of Evidence. Centre for Evidence-Based Medicine. 2016. URL: https://www.cebm.net/2016/05/ocebm-levels-of-evidence/ [accessed 2020-05-07]

60. Guo Y, Xu Z, Qiao J, Hong YA, Zhang H, Zeng C, et al. Development and feasibility testing of an mhealth (text message and WeChat) intervention to improve the medication adherence and quality of life of people living with HIV in China: pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 Sep 4;6(9):e10274 [FREE Full text] [doi: 10.2196/10274] [Medline: 30181109]

61. Moore DJ, Pasipanodya EC, Umlauf A, Rooney AS, Gouaux B, Depp CA, et al. Individualized texting for adherence building (iTAB) for methamphetamine users living with HIV: a pilot randomized clinical trial. Drug Alcohol Depend 2018 Aug 1;189:154-160 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.05.013] [Medline: 29958127]

62. King E, Kinvig K, Steif J, Qiu AQ, Maan EJ, Albert AY, et al. Mobile text messaging to improve medication adherence and viral load in a vulnerable canadian population living with human immunodeficiency virus: a repeated measures study. J Med Internet Res 2017 Jun 1;19(6):e190 [FREE Full text] [doi: 10.2196/jmir.6631] [Medline: 28572079]

63. Rana AI, van den Berg JJ, Lamy E, Beckwith CG. Using a mobile health intervention to support HIV treatment adherence and retention among patients at risk for disengaging with care. AIDS Patient Care STDS 2016 Apr;30(4):178-184 [FREE Full text] [doi: 10.1089/apc.2016.0025] [Medline: 27028183]

64. Measuring Digital Development: Facts and Figures 2019. International Telecommunication Union. 2019. URL: https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx [accessed 2019-05-16] [WebCite Cache ID 78Q4g5E0P]

65. Musiimenta A, Atukunda EC, Tumuhimbise W, Pisarski EE, Tam M, Wyatt MA, et al. Acceptability and feasibility of real-time antiretroviral therapy adherence interventions in rural Uganda: mixed-method pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 May 17;6(5):e122 [FREE Full text] [doi: 10.2196/mhealth.9031] [Medline: 29773527]

66. Lepère P, Touré Y, Bitty-Anderson AM, Boni SP, Anago G, Tchounga B, et al. Exploring the patterns of use and acceptability of mobile phones among people living with HIV to improve care and treatment: cross-sectional study in three francophone west African countries. JMIR Mhealth Uhealth 2019 Nov 13;7(11):e13741 [FREE Full text] [doi: 10.2196/13741] [Medline: 31719023]

67. Mbuagbaw L, Bonono-Momnougui RC, Thabane L. Considerations in using text messages to improve adherence to highly active antiretroviral therapy: a qualitative study among clients in Yaoundé, Cameroon. HIV AIDS (Auckl) 2012;4:45-50 [FREE Full text] [doi: 10.2147/HIV.S29954] [Medline: 22570574]

68. HIV - HIV/AIDS Annual Reports. Publications Saskatchewan. 2019. URL: http://www.publications.gov.sk.ca/details.cfm?p=64628 [accessed 2019-05-16] [WebCite Cache ID 78QAtQOeF]

69. Injection Drug Use and HIV Risk. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/hiv/risk/idu.html [accessed 2019-05-16] [WebCite Cache ID 78QAw4hlM]

70. Laurence C, Wispelwey E, Flickinger TE, Grabowski M, Waldman AL, Plews-Ogan E, et al. Development of PositiveLinks: a mobile phone app to promote linkage and retention in care for people with HIV. JMIR Form Res 2019 Mar 20;3(1):e11578 [FREE Full text] [doi: 10.2196/11578] [Medline: 30892269]

71. Senn TE, Braksmajer A, Coury-Doniger P, Urban MA, Carey MP. Mobile technology use and desired technology-based intervention characteristics among HIV+ Black men who have sex with men. AIDS Care 2017 Apr;29(4):423-427 [FREE Full text] [doi: 10.1080/09540121.2016.1220479] [Medline: 27535069]

72. LeGrand S, Muessig KE, Platt A, Soni K, Egger JR, Nwoko N, et al. Epic allies, a gamified mobile phone app to improve engagement in care, antiretroviral uptake, and adherence among young men who have sex with men and young transgender women who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2018 Apr 5;7(4):e94 [FREE Full text] [doi: 10.2196/resprot.8811] [Medline: 29622527]

73. 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]

XSL•FO

RenderX

74. Dworkin M, Chakraborty A, Lee S, Monahan C, Hightow-Weidman L, Garofalo R, et al. A realistic talking human embodied agent mobile phone intervention to promote HIV medication adherence and retention in care in young HIV-positive African American men who have sex with men: qualitative study. JMIR Mhealth Uhealth 2018 Jul 31;6(7):e10211 [FREE Full text] [doi: 10.2196/10211] [Medline: 30064971]

75. Castel AD, Qasmieh S, Greenberg D, Ellenberger N, Howell TH, Griffith C, et al. Digital gaming to improve adherence among adolescents and young adults living with HIV: mixed-methods study to test feasibility and acceptability. JMIR Serious Games 2018 Oct 15;6(4):e10213 [FREE Full text] [doi: 10.2196/10213] [Medline: 30322838]

76. Care4Today. URL: https://www.care4today.com/ [accessed 2019-12-17]

77. Morano JP, Clauson K, Zhou Z, Escobar-Viera CG, Lieb S, Chen IK, et al. Attitudes, beliefs, and willingness toward the use of mhealth tools for medication adherence in the Florida mhealth adherence project for people living with HIV (FL-MAPP): pilot questionnaire study. JMIR Mhealth Uhealth 2019 Jul 3;7(7):e12900 [FREE Full text] [doi: 10.2196/12900] [Medline: 31271150]

78. Cook PF, Carrington JM, Schmiege SJ, Starr W, Reeder B. A counselor in your pocket: feasibility of mobile health tailored messages to support HIV medication adherence. Patient Prefer Adherence 2015;9:1353-1366 [FREE Full text] [doi: 10.2147/PPA.S88222] [Medline: 26491263]

79. Olalla J, de Lomas JM, Márquez E, González FJ, del Arco A, de la Torre J, et al. Experience of using an app in HIV patients older than 60 years: pilot program. JMIR Mhealth Uhealth 2019 Mar 6;7(3):e9904 [FREE Full text] [doi: 10.2196/mhealth.9904] [Medline: 30839281]

80. Westergaard RP, Genz A, Panico K, Surkan PJ, Keruly J, Hutton HE, et al. Acceptability of a mobile health intervention to enhance HIV care coordination for patients with substance use disorders. Addict Sci Clin Pract 2017 Apr 26;12(1):11 [FREE Full text] [doi: 10.1186/s13722-017-0076-y] [Medline: 28441962]

81. Przybyla SM, Eliseo-Arras RK, Krawiec G, Gower E, Dermen K. Feasibility and acceptability of a smartphone app for daily reports of substance use and antiretroviral therapy adherence among HIV-infected adults. AIDS Res Treat 2016;2016:9510172 [FREE Full text] [doi: 10.1155/2016/9510172] [Medline: 27610243]

82. Horvath KJ, Alemu D, Danh T, Baker JV, Carrico AW. Creating effective mobile phone apps to optimize antiretroviral therapy adherence: perspectives from stimulant-using HIV-positive men who have sex with men. JMIR Mhealth Uhealth 2016 Apr 15;4(2):e48 [FREE Full text] [doi: 10.2196/mhealth.5287] [Medline: 27084049]

83. Rosen RK, Ranney ML, Boyer EW. Formative research for mhealth HIV adherence: the iHAART app. Proc Annu Hawaii Int Conf Syst Sci 2015 Jan;2015:2778-2785 [FREE Full text] [doi: 10.1109/HICSS.2015.336] [Medline: 26644783]

84. Smillie K, van Borek N, Abaki J, Pick N, Maan EJ, Friesen K, et al. A qualitative study investigating the use of a mobile phone short message service designed to improve HIV adherence and retention in care in Canada (WelTel BC1). J Assoc Nurses AIDS Care 2014;25(6):614-625 [FREE Full text] [doi: 10.1016/j.jana.2014.01.008] [Medline: 24768442]

85. Campbell AR, Kinvig K, Côté HC, Lester RT, Qiu AQ, Maan EJ, et al. Health care provider utilization and cost of an mhealth intervention in vulnerable people living with HIV in Vancouver, Canada: prospective study. JMIR Mhealth Uhealth 2018 Jul 9;6(7):e152 [FREE Full text] [doi: 10.2196/mhealth.9493] [Medline: 29986845]

86. Harris LT, Lehavot K, Huh D, Yard S, Andrasik MP, Dunbar PJ, et al. Two-way text messaging for health behavior change among human immunodeficiency virus-positive individuals. Telemed J E Health 2010 Dec;16(10):1024-1029 [FREE Full text] [doi: 10.1089/tmj.2010.0050] [Medline: 21087122]

87. Krishnan A, Ferro EG, Weikum D, Vagenas P, Lama JR, Sanchez J, et al. Communication technology use and mHealth acceptance among HIV-infected men who have sex with men in Peru: implications for HIV prevention and treatment. AIDS Care 2015;27(3):273-282 [FREE Full text] [doi: 10.1080/09540121.2014.963014] [Medline: 25285464]

88. Anand T, Nitpolprasert C, Kerr SJ, Muessig KE, Promthong S, Chomchey N, et al. A qualitative study of Thai HIV-positive young men who have sex with men and transgender women demonstrates the need for eHealth interventions to optimize the HIV care continuum. AIDS Care 2017 Jul;29(7):870-875. [doi: 10.1080/09540121.2017.1286288] [Medline: 28158952]

89. Skrajner MJ, Camp CJ, Heckman TG, Haberman JL, Kochman A. A fluctuating pattern of over- and under-adherence for HAART: a case study from a videophone intervention project. Clin Gerontol 2011 Feb 23;34(2):144-153 [FREE Full text] [doi: 10.1080/07317115.2011.539520] [Medline: 21607118]

90. Tanner AE, Mann L, Song E, Alonzo J, Schafer K, Arellano E, et al. weCARE: a social media-based intervention designed to increase HIV care linkage, retention, and health outcomes for racially and ethnically diverse young MSM. AIDS Educ Prev 2016 Jun;28(3):216-230 [FREE Full text] [doi: 10.1521/aeap.2016.28.3.216] [Medline: 27244190]

91. Cho H, Flynn G, Saylor M, Gradilla M, Schnall R. Use of the FITT framework to understand patients' experiences using a real-time medication monitoring pill bottle linked to a mobile-based HIV self-management app: a qualitative study. Int J Med Inform 2019 Nov;131:103949 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.08.009] [Medline: 31561192]

92. Schnall R, Cho H, Mangone A, Pichon A, Jia H. Mobile health technology for improving symptom management in low income persons living with HIV. AIDS Behav 2018 Oct;22(10):3373-3383 [FREE Full text] [doi: 10.1007/s10461-017-2014-0] [Medline: 29299790]

93. Hinkin CH, Barclay TR, Castellon SA, Levine AJ, Durvasula RS, Marion SD, et al. Drug use and medication adherence among HIV-1 infected individuals. AIDS Behav 2007 Mar;11(2):185-194 [FREE Full text] [doi: 10.1007/s10461-006-9152-0] [Medline: 16897351]

94.   Becker ML, Kasper K, Pindera C, Cheang M, Rodger D, Sanche S, et al. Characterizing the HIV epidemic in the prairie provinces. Can J Infect Dis Med Microbiol 2012;23(1):19-22 [FREE Full text] [doi: 10.1155/2012/408561] [Medline: 23449046]

95.   McInnes CW, Druyts E, Harvard SS, Gilbert M, Tyndall MW, Lima VD, et al. HIV/AIDS in Vancouver, British Columbia: a growing epidemic. Harm Reduct J 2009 Mar 5;6:5 [FREE Full text] [doi: 10.1186/1477-7517-6-5] [Medline: 19265531]

96.   HIV in the United States and Dependent Areas. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/hiv/statistics/overview/ataglance.html [accessed 2019-05-16] [WebCite Cache ID 78QDvJ3y5]

97.   HIV and AIDS in the United Kingdom (UK). Avert. 2019. URL: https://www.avert.org/professionals/hiv-around-world/western-central-europe-north-america/uk [accessed 2019-05-16] [WebCite Cache ID 78QDXSj8e]

98.   Challacombe L. The Epidemiology of HIV in Canada. Catie. 2019. URL: https://www.catie.ca/en/fact-sheets/epidemiology/epidemiology-hiv-canada [accessed 2019-05-16] [WebCite Cache ID 78QDKMFvT]

99.   Shiau S, Arpadi SM, Yin MT, Martins SS. Patterns of drug use and HIV infection among adults in a nationally representative sample. Addict Behav 2017 May;68:39-44 [FREE Full text] [doi: 10.1016/j.addbeh.2017.01.015] [Medline: 28088742]

100.   Courtney KE, Ray LA. Methamphetamine: an update on epidemiology, pharmacology, clinical phenomenology, and treatment literature. Drug Alcohol Depend 2014 Oct 1;143:11-21 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.08.003] [Medline: 25176528]

101.   Zweben JE, Cohen JB, Christian D, Galloway GP, Salinardi M, Parent D, Methamphetamine Treatment Project. Psychiatric symptoms in methamphetamine users. Am J Addict 2004;13(2):181-190. [doi: 10.1080/10550490490436055] [Medline: 15204668]

102.   Bach P, Wood E, Dong H, Guillemi S, Kerr T, Montaner J, et al. Association of patterns of methadone use with antiretroviral therapy discontinuation: a prospective cohort study. BMC Infect Dis 2015 Nov 19;15:537 [FREE Full text] [doi: 10.1186/s12879-015-1255-7] [Medline: 26586238]

103.   Galea S, Vlahov D. Social determinants and the health of drug users: socioeconomic status, homelessness, and incarceration. Public Health Rep 2002;117(Suppl 1):S135-S145 [FREE Full text] [Medline: 12435837]

104.   Schwartz J. The Most Popular Messaging App in Every Country. SimilarWeb: Website Traffic Statistics & Analytics. 2019. URL: https://www.similarweb.com/blog/worldwide-messaging-apps [accessed 2019-05-16] [WebCite Cache ID 78QCS7IYd]

105.   A Habit-formation and Gamification Intervention to Improve ART Adherence Among MSM HIV Patients in Mexico (FUERTES). Clinical Trials. URL: https://clinicaltrials.gov/ct2/show/NCT03410680 [accessed 2019-03-10]

106.   Epic Allies HIV ART Adherence Intervention. Clinical Trials. URL: https://clinicaltrials.gov/ct2/show/NCT02782130 [accessed 2019-03-10]

Abbreviations

ARV: antiretroviral
IVDU: intravenous drug user
mHealth: mobile health
MSM: men who have sex with men
RCT: randomized controlled trial

Original Paper

Mobile Health Usage, Preferences, Barriers, and eHealth Literacy in Rheumatology: Patient Survey Study

Johannes Knitza[1], MD; David Simon[1], MD; Antonia Lambrecht[1], MD; Christina Raab[1], MD; Koray Tascilar[1], MD; Melanie Hagen[1], MD; Arnd Kleyer[1], MD; Sara Bayat[1], MD; Adrian Derungs[2], MSc; Oliver Amft[2], MSc; Georg Schett[1], MD; Axel J Hueber[1,3], MD

[1]Department of Internal Medicine 3 – Rheumatology and Immunology, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

[2]Chair of Digital Health, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

[3]Section Rheumatology, Sozialstiftung Bamberg, Bamberg, Germany

Corresponding Author:
Johannes Knitza, MD
Department of Internal Medicine 3 – Rheumatology and Immunology
Friedrich-Alexander University Erlangen-Nürnberg
Ulmenweg 18
Erlangen, 91054
Germany
Phone: 49 913185 ext 35000
Email: johannes.knitza@uk-erlangen.de

Abstract

Background: Mobile health (mHealth) defines the support and practice of health care using mobile devices and promises to improve the current treatment situation of patients with chronic diseases. Little is known about mHealth usage and digital preferences of patients with chronic rheumatic diseases.

Objective: The aim of the study was to explore mHealth usage, preferences, barriers, and eHealth literacy reported by German patients with rheumatic diseases.

Methods: Between December 2018 and January 2019, patients (recruited consecutively) with rheumatoid arthritis, psoriatic arthritis, and axial spondyloarthritis were asked to complete a paper-based survey. The survey included questions on sociodemographics, health characteristics, mHealth usage, eHealth literacy using eHealth Literacy Scale (eHEALS), and communication and information preferences.

Results: Of the patients (N=193) who completed the survey, 176 patients (91.2%) regularly used a smartphone, and 89 patients (46.1%) regularly used social media. Patients (132/193, 68.4%) believed that using medical apps could be beneficial for their own health. Out of 193 patients, only 8 (4.1%) were currently using medical apps, and only 22 patients (11.4%) stated that they knew useful rheumatology websites/mobile apps. Nearly all patients (188/193, 97.4%) would agree to share their mobile app data for research purposes. Out of 193 patients, 129 (66.8%) would regularly enter data using an app, and 146 patients (75.6%) would welcome official mobile app recommendations from the national rheumatology society. The preferred duration for data entry was not more than 15 minutes (110/193, 57.0%), and the preferred frequency was weekly (59/193, 30.6%). Medication information was the most desired app feature (150/193, 77.7%). Internet was the most frequently utilized source of information (144/193, 74.6%). The mean eHealth literacy was low (26.3/40) and was positively correlated with younger age, app use, belief in benefit of using medical apps, and current internet use to obtain health information.

Conclusions: Patients with rheumatic diseases are very eager to use mHealth technologies to better understand their chronic diseases. This open-mindedness is counterbalanced by low mHealth usage and competency. Personalized mHealth solutions and clear implementation recommendations are needed to realize the full potential of mHealth in rheumatology.

KEYWORDS

Introduction

Rheumatoid arthritis, psoriatic arthritis, and axial spondyloarthritis are complex pathogenetic chronic diseases. These disease entities also require a complex treatment structure with interdisciplinary care by various specialists (rheumatologists, dermatologists, gastroenterologists, ophthalmologists, physiotherapists, etc), and intensive and regular disease monitoring is essential. With this complexity, affected patients often report a lack of understanding of their disease [1].

Mobile health (mHealth) holds promise to improve health care delivery and outcomes for people with chronic diseases [2]. Ideally, mHealth solutions, such as mobile apps and wearable sensors could empower patients, provide individual support, and lead to better outcomes than those available through standard care. Patients with chronic rheumatic diseases already have access to a broad range of mHealth solutions, starting from symptom checkers [3] or referral tools [4]. Once a diagnosis is established, mHealth tools enable patients to better monitor their symptoms passively through sensors [5] and actively by entering data [6,7]. Furthermore, electronic medication reminders can increase medication adherence [8,9], and supporting digital therapy can reduce pain [9] and improve important comorbidities (eg, depression) [10].

Recently the European League Against Rheumatism and Working Group Young Rheumatology of the German Society for Rheumatology (*Arbeitsgemeinschaft Junge Rheumatologie der Deutschen Gesellschaft für Rheumatologie*) published recommendations for the development of mobile apps in rheumatology [11,12]. The early integration of patients in the app development process was stressed in both papers. However, little is known about the patient perspective on mHealth solutions, as current literature focuses on rheumatologists [11,13,14]. To successfully integrate the various mHealth solutions into clinical routine, it is essential to identify barriers and the needs of patients.

The aim of this study was to explore mHealth usage, preferences, barriers, and eHealth literacy reported by German patients living with rheumatic diseases.

Methods

Between December 2018 and January 2019, consecutive patients seen at one rheumatology outpatient clinic of the University Hospital Erlangen were asked to complete a paper-based survey. This study was approved by the Ethics committee (No. 418-18B) and conducted referring to good clinical practice. All patients provided informed consent.

To create the survey (Multimedia Appendix 1), a broad literature review was carried out. Previous mHealth patient surveys in oncology [15] and rheumatology [16] served as a starting basis. The survey comprised four main parts: (1) sociodemographics and health characteristics, (2) mHealth preferences and usage, (3) eHealth literacy, and (4) communication and information preferences. The survey was pilot-tested with 10 patients to detect necessary formatting and wording changes. Minor

revisions were made accordingly. Foreign words and technical term explanations were provided in a footnote. Inclusion criteria were patients (1) aged ≥18 years, (2) who were literate in German, (3) who had the physical and mental ability to fill out a structured questionnaire, and (4) who fulfilled classification criteria for rheumatoid arthritis [17], axial spondyloarthritis [18], or psoriatic arthritis [19].

The sociodemographic and health characteristics included age; gender; residence; diagnosis; disease duration; patient global assessment of disease activity; and current usage of smartphones, tablets, activity trackers, and social media.

The mHealth preferences and usage section included questions about the preferred time and frequency for using a rheumatology app. Patients were asked to rate their preference of app features (5-point Likert scale) and rate the importance of app characteristics (10-point Likert scale). Inquiries were made on internet usage and perceived usefulness, willingness to share recorded app data, general perception about the utility of medical apps, web-based services for improving patient's health, and telemedicine.

Patients' eHealth literacy was measured using the validated German version [20] of the eHealth Literacy Scale (eHEALS) [21]; eHEALS has been translated and validated in multiple languages [20-22]. It is based on a 5-point Likert scale and includes 8 statements concerning self-perceived eHealth literacy.

Patients were asked to state their preferences concerning medication reminders, medical information format, digitally provided information structure, patient diary type, and physician communication type. Rankings did not have to be unique. Characteristics were summarized using means, standard deviations, counts, and percentages as appropriate. We used Pearson correlation to explore relationships between continuous variables. The relationship between the eHEALS score and internet use frequency was examined using a linear regression model with the eHEALS score as the dependent variable and internet use as an ordinal predictor encoded using orthogonal polynomials to characterize nonlinear effects. Relationships between eHEALS score and binary preferences were examined using logistic regression. All models included age and gender adjustments. Two-sided *P* values less than .05 were considered significant. We used Excel (Microsoft Corp) and R (version 3.5.3; R Foundation for Statistical Computing) for data manipulation and analyses.

Results

Patient Characteristics

In total, 224 patients were recruited. Only complete surveys (N=193) were considered for the final analysis. The number of patients rejecting participation was not measured. The study sample's demographics are shown in Table 1. Mean age was 52.1 (SD 13.7) years, with 34.7% (67/193) being at least 60 years old; 59.6% (115/193) were female, and 53.9% (104/193) had been diagnosed with rheumatoid arthritis. The mean disease duration was 8.3 years (SD 8.0). Nearly all patients regularly used a smartphone (176/193, 91.2%), and nearly half of the patients regularly used a tablet (86/193, 44.6%) and social media

(89/193, 46.1%). Only a minority used activity trackers (20/193, 10.4%).

Table 1. Demographic and health characteristics.

Characteristic	Values (N=193)
Age (years), mean (SD)	52.1 (13.7)
Age (years), n (%)	
18-39	46 (23.8)
40-59	80 (41.5)
≥60	67 (34.7)
Gender n (%)	
Female	115 (59.6)
Male	78 (40.4)
Diagnosis, n (%)	
Rheumatoid arthritis	104 (53.9)
Axial spondyloarthritis	37 (19.2)
Psoriatic arthritis	52 (26.9)
Patient global assessment of disease activity (0-10), mean (SD)	3.8 (2.4)
Disease duration (years), mean (SD)	8.3 (8.0)
Disease duration (years), n (%)	
≤1	42 (21.8)
2-5	44 (22.8)
>5	107 (55.4)
Residence, n (%)	
Village	74 (38.3)
Small city	48 (24.9)
Midsized city	35 (18.1)
Big city	36 (18.7)
Regular usage, n (%)	
Smartphone	176 (91.2)
Tablet	86 (44.6)
Activity tracker	20 (10.4)
Social media	89 (46.1)
Medical apps	8 (4.1)

Medical App Acceptance and Willingness to Provide mHealth Data for Research Purposes

Preferences and attitudes regarding potential mHealth apps and data flow were addressed (Table 2). More than two-thirds of the patients (132/193, 68.4%) believed that medical apps are helpful for their health; however, only 4.1% (8/193) patients currently used medical apps, of which none were rheumatology specific apps. Increasing eHEALS scores were associated with a higher probability of expressing belief that apps were helpful after adjusting for age and gender (odds ratio [OR] 1.13, 95% CI 1.07 to 1.19, P<.001). Nearly all patients (188/193, 97.4%) were willing to transfer app data for research purposes, if data security would be ensured. The main barrier for sharing app data with the physician was that personal contact was considered as sufficient (53/193, 63.9%). Second, patients (42/193, 50.6%) were concerned about data transfer. The majority of patients (174/193, 90.2%) wanted to be contacted in case an app detected an abnormality concerning their health, whereas 57.0% (110/193) were also willing to transfer app data to the treating physician. Concerns were data usage, storage, and transfer. Only 28.0% (54/193) were interested in comparing their medication adherence to the medication adherence of other patients. Regarding official app recommendations, 75.6% (146/193) of interviewed patients wanted advice from the national society of rheumatology. Weekly data entry was preferred by 30.6% (59/193) with durations of 5-15 minutes (65/193, 33.7%). Measured on a scale of 10, the most important app characteristics were security (mean 8.9, SD 2.5) and usability (mean 8.5, SD 2.5) (Table 2). Regarding preferred app functions,

patients were most interested in information about medications and diseases and were least interested in direct exchange such as chats with peers with the same disease (Figure 1).

Table 2. Patient attitudes towards medical apps.

Characteristic	Values (N=193)
Patients believing medical apps are helpful for their health, n (%)	132 (68.4)
Patients willing to transfer app data for research purposes, n (%)	188 (97.4)
Patients willing to transfer data to physician with app, n (%)	110 (57.0)
Reason for not willing to transfer data to physician with app, n (%)[a]	
I don't have a suitable device	13 (15.7)
I don't have the technical skills	23 (27.7)
I have concerns about data usage	33 (39.8)
I have concerns about data storage	31 (37.3)
I have concerns about data transfer	42 (50.6)
I have concerns about data protection	24 (28.9)
I only want personal contact with physician	53 (63.9)
I don't find this useful	11 (13.3)
Patients who want to be contacted in case of app-monitoring abnormalities	174 (90.2)
Patients interested to compare medication adherence to other patients, n (%)	54 (28.0)
Patients who want official app recommendations from national society of rheumatology, n (%)	146 (75.6)
Preferred frequency of app usage, n (%)	
Not at all	64 (33.2)
Daily	11 (5.7)
Weekly	59 (30.6)
Monthly	37 (19.2)
Each 3 months	18 (9.3)
Less frequently	4 (2.1)
Preferred time of app usage (minutes), n (%)	
Not at all	64 (33.2)
0-5	45 (23.3)
5-15	65 (33.7)
15-30	18 (9.3)
>30	1 (0.5)
Importance of app characteristics (0-10), mean (SD)	
Interactivity	4.5 (2.9)
Design	4.7 (3.2)
Usability	8.5 (2.5)
Data security	8.9 (2.5)

[a]Multiple answers were possible.

Figure 1. App function preferences (responses to "What app functions would you like?").
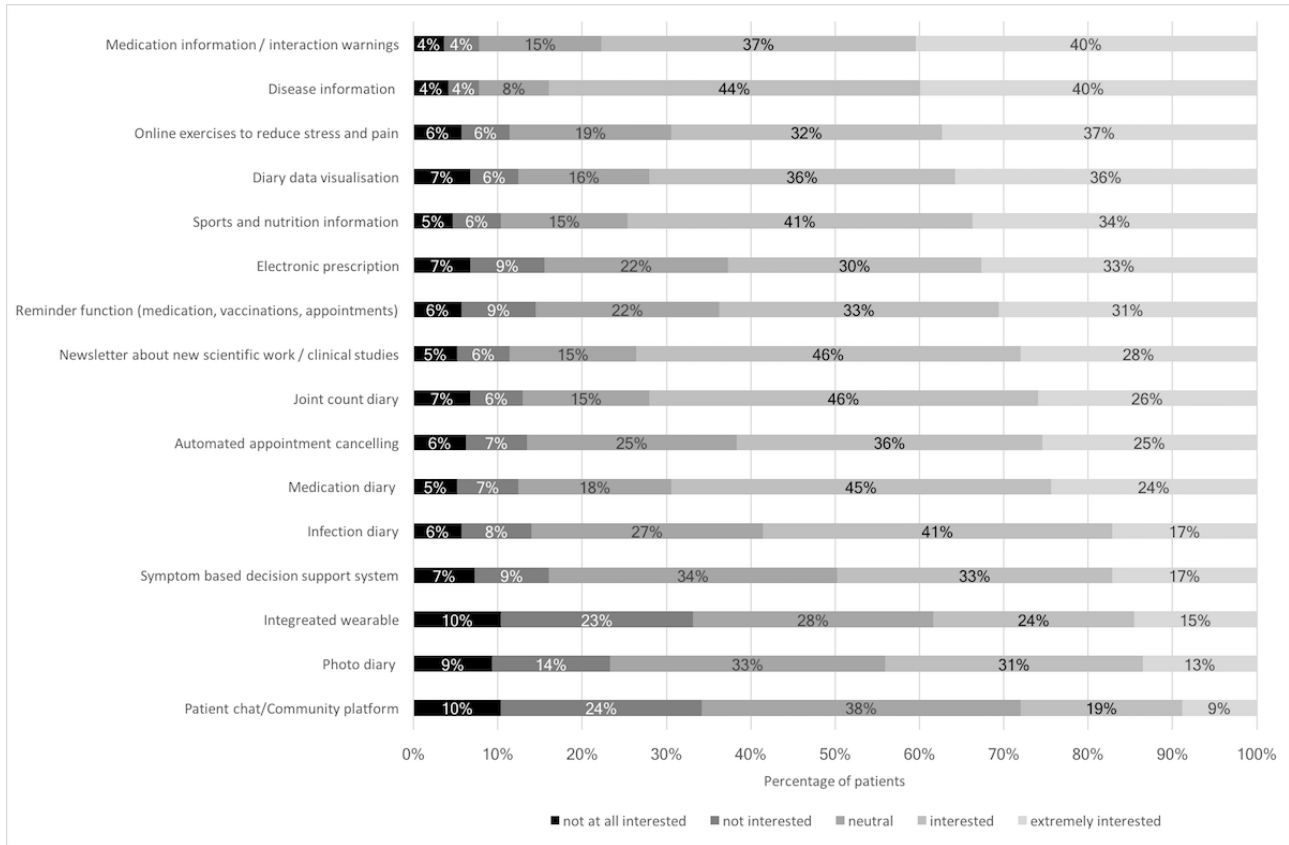
Internet Usage and Perceived Usefulness

To address habits regarding information search, patients were questioned concerning internet usage. The majority of patients (168/193, 87.0%) had previously used the internet to obtain health information (Table 3), and the remaining patients (25/193, 13.0%) lacked the skills or motivation to do a search; 7/25 (28.0%) had no internet access, 7/25 (28.0%) did not think it would be helpful, 6/25 (24.0%) found the information from their physician sufficient, 2/25 (8.0%) did not know how to do a search, and 3/25 (12.0%) stated another reason. Some patients had previously communicated with a physician by email (56/193, 29.0%). Participation in an online health program was rare (3/193, 1.6%). Online support groups were used by patients to post information (14/193, 7.3%), chat with other patients (19/193, 9.8%), or read information (85/193, 44.0%); 19.7% (38/193) were aware of the medication website of the German Society of Rheumatology (*Deutsche Gesellschaft für Rheumatologie*).

Figure 2 presents the type of health information searched on the internet. Most patients looked for medication information (134/168, 79.8%) whereas information on support groups was the least commonly searched (58/168, 34.5%); 124/193 (64.2%) patients preferred filling out medical questionnaires electronically before their clinical visits, 102/193 (52.8%) patients preferred receiving a doctor's letter in an electronic format instead of paper, and 98/193 (50.2%) patients preferred communicating with their rheumatologist by a video call.

Table 3. Internet usage, perceived usefulness, and eHealth literacy.

Characteristic	Values (N=193)
Did you previously look for health information on the internet? n (%)	
Yes	168 (87.0)
No	25 (13.0)
Did you previously communicate with a physician by email? n (%)	
Yes	56 (29.0)
No	137 (71.0)
Did you previously participate in an online health program? n (%)	
Yes	3 (1.6)
No	190 (98.4)
Did you use an online supporting group before to...?[a] n (%)	
Post information	14 (7.3)
Chat with patients	19 (9.8)
Read information	85 (44.0)
I never used one before	103 (53.4)
Do you know useful rheumatology websites or apps? n (%)	
Yes	22 (11.4)
No	171 (88.6)
Do you know the DGRh[b] medication information website? n (%)	
Yes	38 (19.7)
No	155 (80.3)
How useful do you find the internet to make health-related decisions? n (%)	
Not useful at all	17 (8.8)
Not useful	19 (9.8)
Unsure	74 (38.3)
Useful	72 (37.3)
Very useful	11 (5.7)
eHealth literacy, mean (SD)	26.3 (7.1)

[a]Multiple answers were possible.

[b]DGRh: Deutsche Gesellschaft für Rheumatologie (German Society of Rheumatology).

XSL•FO
RenderX

Figure 2. Information previously searched on the internet (responses to "What health information did you look for on the internet?).
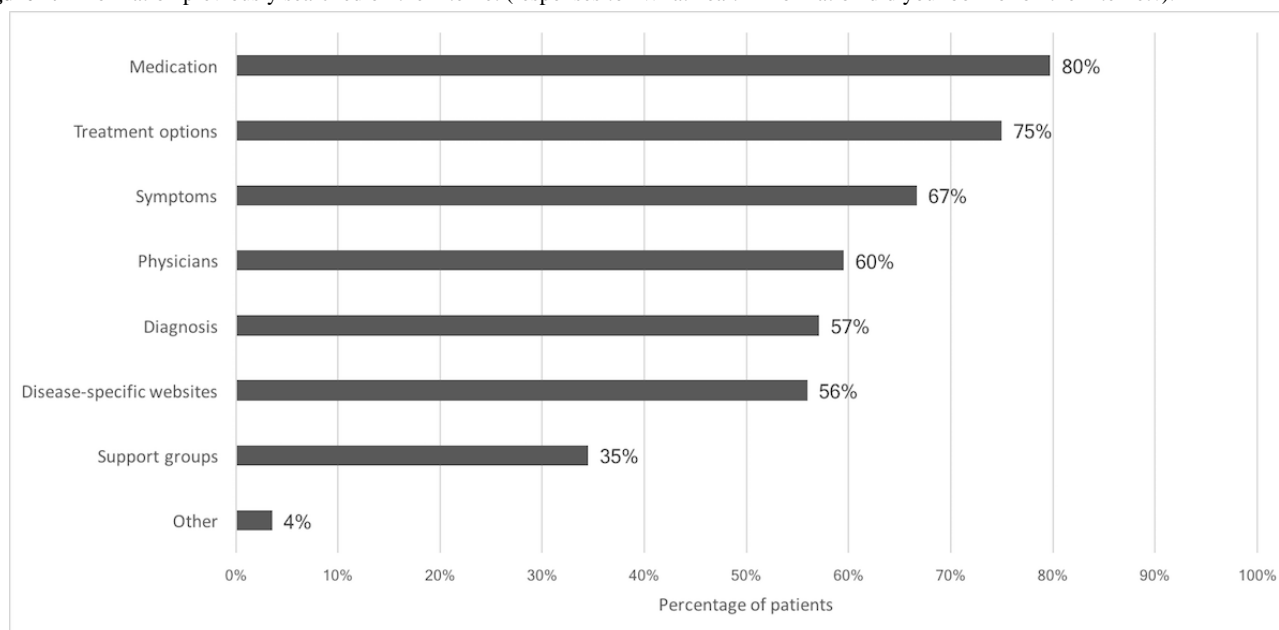
eHealth Literacy

Mean eHealth literacy using eHEALS was 26.3 (SD 7.1) out of 40. Mean scores in women and men were 25.8 and 27.0, respectively, with a mean difference of –1.15 (95% CI –3.14 to 0.84) showing no important effect of gender. Age showed a negative correlation (r=–0.38, 95% CI –0.5 to –0.26) with the eHEALS score (Figure 3). Table 4 shows the distribution of responses to the 8 eHEALS items. The majority of the patients agreed that they know how to use the internet to answer their questions about health (82/193, 71.5%), and a considerable proportion of patients (82/193, 42.5%) felt uncomfortable using information from the internet to make health decisions. A lower eHEALS score was associated with a decreasing frequency of internet usage; the regression analysis shows that, after adjustment for age and gender, the negative association with decreasing frequency of internet use and eHEALS score followed a second-order polynomial (linear: –8.87, 95% CI –11.97 to –5.76; quadratic:–5.64, 95% CI –8.37 to –2.92) meaning that each step of decrease in the use-frequency categories was associated with, not an equal, but a progressively greater decrease in the eHEALS score (Figure 4). Figure 5 shows the usage frequency of different health information sources during the last 3 months prior to the clinical visit. The internet was the most frequently used information source (144/193, 74.6%), with 9.3% (18/193) using it daily, 14.5% (28/193) weekly, and 19.7% (38/193) monthly.

Figure 3. Negative relationship between eHEALS score and age.
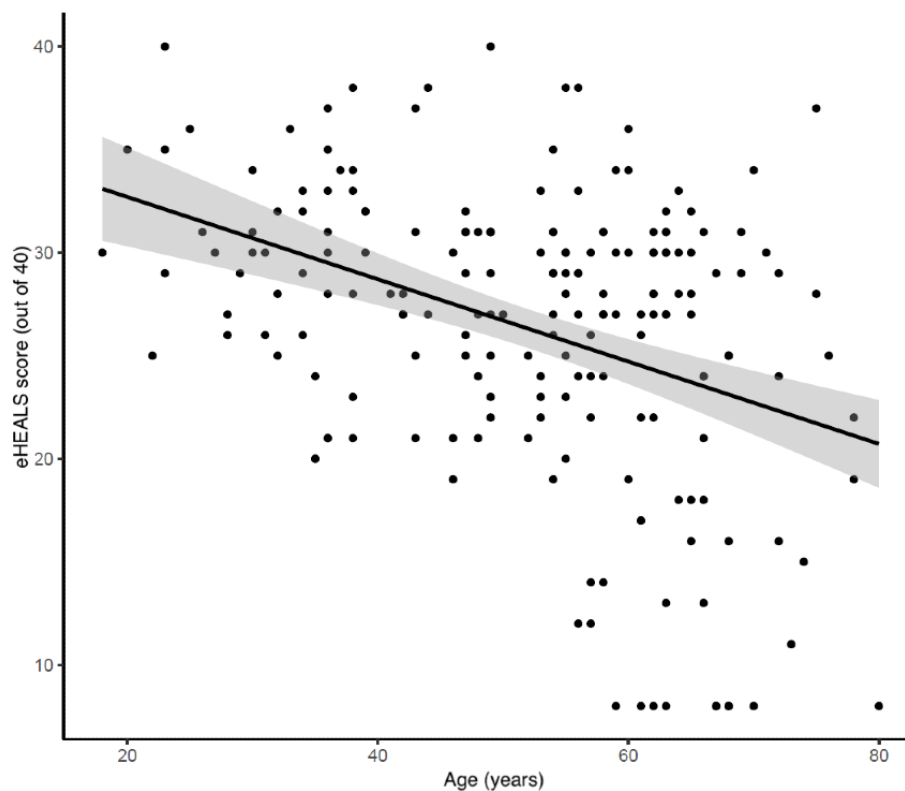
Table 4. eHealth literacy.

eHEALS item	Participants (N=193), n (%)			Score
	Strongly disagree or disagree	Neutral	Agree or strongly agree	Mean (SD)
I know how to find helpful health resources on the Internet	30 (15.5)	50 (25.9)	113 (58.5)	3.5 (1.1)
I know how to use the Internet to answer my questions about health	30 (15.5)	25 (13.0)	138 (71.5)	3.6 (1.1)
I know what health resources are available on the Internet	51 (26.4)	68 (35.2)	74 (38.3)	3.1 (1.1)
I know where to find helpful health resources on the Internet	47 (24.4)	57 (29.5)	89 (46.1)	3.2 (1.1)
I know how to use the health information I find on the Internet to help me	48 (24.9)	49 (25.4)	96 (49.7)	3.2 (1.1)
I have the skills I need to evaluate the health resources I find on the Internet	26 (13.5)	30 (15.5)	137 (71.0)	3.8 (1.1)
I can tell high quality health resources from low quality health resources on the internet	37 (19.2)	52 (26.9)	104 (53.9)	3.4 (1.1)
I feel confident in using information from the Internet to make health decisions	82 (42.5)	82 (42.5)	29 (15.0)	2.6 (1.0)

Figure 4. Association of eHEALS score and frequency of internet use.
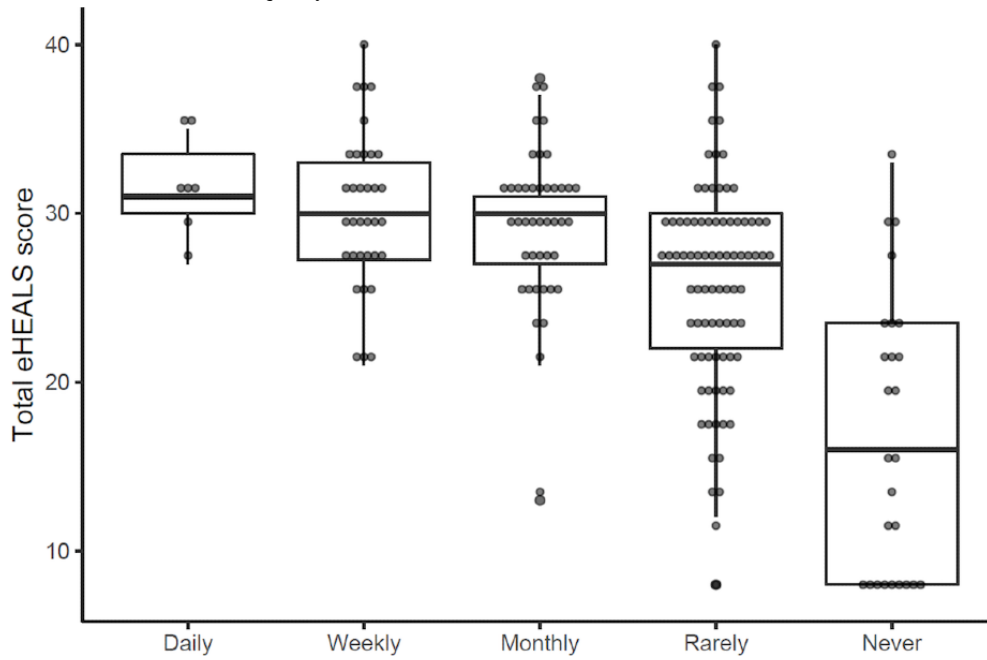
Figure 5. Health information sources used prior to clinical visits (responses to "How often did you use one of the following options to receive health information in the last three months?").
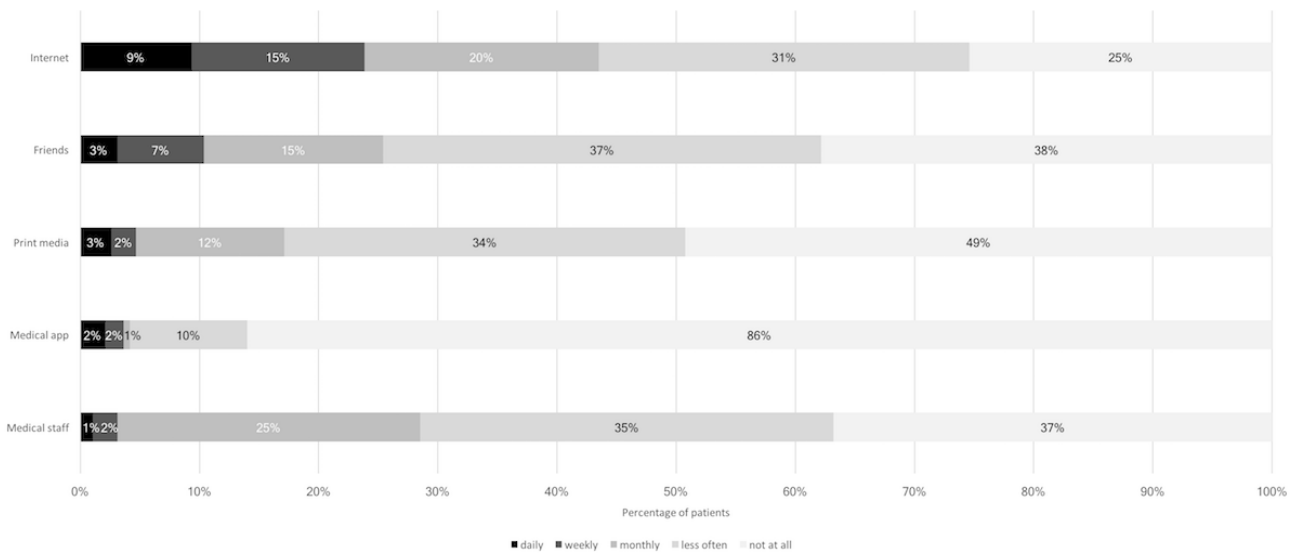
Conservative Communication Preferences

The preferred way to contact the treating rheumatologist was via telephone (136/196, 69.4%), email (42/196, 21.4%), and chat (8/196, 4.1%) (Table 5, last panel). Most patients wanted to be reminded of regular medication intake (198/260, 76.2%). The preferred method for reminders was via an app (118/260, 45.4%) (Table 5). The large majority of patients (173/188, 92.0%) wanted medical information and the preferred media were paper (104/188, 55.3%), app (40/188, 21.3%), and website (29/188, 15.4%). Concerning digitally provided information, patients preferred plain text and images (160/196, 81.6%) over exchange functions (29/196, 14.8%) and game-based learning (7/196, 3.6%). The majority of patients preferred using a patient diary (167/190, 87.9%) (preference for paper was greater than that for an app, which was greater than that for a website). App usage preference increased over the last years (5.4% to 18.1% from 2014 to 2018, see Multimedia Appendix 2).

Table 5. Preferences for medication reminders, medical information format, digitally provided information structure, patient diary type, and physician communication type.

Characteristic[a]	Preference (N=193), n (%)		Score
	First	Last	Mean (SD)
I would like to be reminded of the regular medication intake by...[b]			
Short message (SMS)	39 (15.0)	13 (6.0)	2.8 (1.4)
Mobile app	118 (45.4)	17 (7.9)	3.1 (1.7)
Email	18 (6.9)	10 (4.6)	3.2 (1.3)
Telephone call	20 (7.7)	28 (13.0)	3.5 (1.6)
Not wanted	62 (23.8)	78 (36.1)	3.8 (2.2)
Postcard	3 (1.2)	70 (32.4)	4.9 (1.3)
I prefer medical information ...[c]			
On paper	104 (55.3)	12 (6.2)	1.8 (1.0)
On a website	29 (15.4)	16 (8.2)	2.3 (0.8)
In an app	40 (21.3)	15 (7.7)	2.4 (0.9)
Not wanted	15 (8.0)	151 (77.8)	3.6 (0.9)
For digitally provided information (website/app) this would be important to me:[d]			
Text- and image-based information	160 (81.6)	16 (8.1)	1.3 (0.6)
Exchange with others	29 (14.8)	61 (31.0)	2.2 (0.7)
Game-based learning	7 (3.6)	120 (60.9)	2.6 (0.6)
I would document my state of health and tablet intake in a "patient diary"...[c]			
On paper	92 (48.4)	14 (7.2)	1.9 (1.0)
In an app	53 (27.9)	19 (9.7)	2.2 (0.9)
On a website	22 (11.6)	31 (15.9)	2.6 (0.9)
Not wanted	23 (12.1)	131 (67.2)	3.3 (1.1)
My preferred way to contact my rheumatologist is via...[c]			
Telephone call	136 (69.4)	8 (4.0)	1.5 (0.8)
Email	42 (21.4)	5 (2.5)	2.0 (0.7)
Website/chat	8 (4.1)	32 (16.2)	2.9 (0.7)
Not wanted	10 (5.1)	153 (77.3)	3.6 (0.8)

[a]Answer options could be ranked with the same preference level.

[b]1=preferred option, 6=least preferred option.

[c]1=preferred option, 4=least preferred option.

[d]1=preferred option, 3=least preferred option.

Discussion

Main Findings

Patients with rheumatic and musculoskeletal diseases are ready and willing to use mHealth technologies. Patients believe in the potential of mHealth and are open to participating actively by sharing health data with their physicians and researchers. Structured electronic data acquisition holds great promise to increase data quality, increase quantity, and reduce missing data and bureaucracy. Patients preferred filling out questionnaires before clinical visits, using video calls, and receiving electronic doctor's letters. These measures could drastically improve the clinical experience for patients, cutting down waiting time and long drives to the hospital. For clinicians, automatic data import could reduce administrative efforts. Furthermore, continuously obtained mHealth patient data increases the basis for shared personalized clinical decision making.

Some patients, however, prefer personal contact with their physician and were concerned about data storage and transfer. Also, at the time of the study, no single patient was using a rheumatology specific app, only 4.1% were using medical apps at all, and 11.4% were aware of useful rheumatology websites or apps. The currently scarce mHealth usage and low eHealth literacy highlight an important need for structured mHealth

guidance and patient-adapted information and education. The majority of patients clearly stated this by calling for official app recommendations from the national society of rheumatology. The most popular app features were information about medication and rheumatic diseases. The fact that 87.0% previously searched for various health information on the internet, and that the internet is the most frequently used source for health information, further supports the need for more, better, and personalized medical information. This information should be accurate and reliable, requiring rheumatologists and societies to actively lead and supervise mHealth in rheumatology. The opportunity to support therapeutic online health programs is currently not being fully utilized, despite promising study results [23]. Furthermore, our study shows an enthusiasm on the patients' side to improve medication adherence via mobile apps and diaries. Sharing of this mHealth data with a large research registry would represent an enormous potential to improve treatments and rheumatology research. These results may help patients, developers, clinicians, and researchers to use the full potential of mHealth in rheumatology.

Limitations

The cross-sectional design, self-reported data, and sampling method with relatively small sample size were the most important limitations of this study. Therefore, the results might not be generalizable, and actual mHealth usage might, for example, differ compared to self-reported usage.

Comparison With Prior Work

To our knowledge, this is the first work depicting a German mHealth patient perspective in rheumatology. This work may inform mHealth policy recommendations and adds to the growing body of eHealth rheumatology knowledge [11-14,24,25] by providing detailed patient preferences, needs, and barriers. Hence, we believe that the results of this study could help in devising mHealth solutions that can be integrated into the clinical routine of patients with rheumatic diseases. The importance of including patients in the app development process is stressed in various recommendations [11,12]. *Rheuma-Auszeit* was reported as the only app developed with major patient involvement and scored as the highest quality app [11]. It was also shown that the participation of patients in app development leads to high usage; for example, the ArthritisPower app was used by >18,000 patients with rheumatic diseases in 2018 [26].

Health literacy is critical for patient empowerment [27]. Our work confirms the lack of adequate eHealth skills in rheumatology patients [28], reflected by the borderline mean eHEALS score of 26.3 out of 40. Cutoffs for the eHEALS score show considerable variation in the literature [29-31]; however, a value of 26 was set in a previous study as the cutoff for low eHealth literacy [31]. eHealth literacy was associated with a greater belief in the usefulness of medical apps, and usage of the internet to obtain health information, similar to findings reported by Noblin et al [32]. A previous study [33] showed that younger age and higher eHealth literacy correlate with perceived effectiveness of medical apps. eHealth literacy can change over time [34], and supporting programs should be implemented to increase eHealth literacy [35]. Many mHealth

challenges could be overcome if more support was provided by health providers [36].

Patients demanded clear recommendations from the national society for rheumatology and expressed the highest confidence in an app developed by a rheumatic disease scientific society [25]. Our results suggest that patient app entries should take no longer than 15 minutes and should not be requested more often than weekly. Data transfer should be clearly explained to patients to eliminate this barrier. Customizable app features are needed, as preferences differ, and a one-size-fits-all approach seems to be less effective. App building blocks (videos, information, tools) should be provided so that redundant work is reduced, and a large variety of features can be provided. Ideally, the app content should be created in a joint effort by all stakeholders.

The discrepancy between app use and general belief in usefulness and interest reflect the currently unmet need of effective mHealth solutions, guidance, and education. This discrepancy is also highlighted by the dominance of conservative communication preferences. These results should carefully be interpreted, as most patients currently do not have experience with app-based communication.

In an effort to address this need, the French society for rheumatology recently developed the free patient app Hiboot, which provides patients with trustworthy information about medication and answers to frequently asked questions [37]. App usage was very low among patients (4.1%) compared to medical app usage among German rheumatologists (49%) [13] and an international group of rheumatic patients (79/394, 20%). In contrast to rheumatologists [13], and an international group of rheumatic patients (188/394, 47%) [25]; patients were not aware of any rheumatology specific apps. These results could partly be explained by the paper-based nature of the survey that enabled us to include elderly patients that might have been reluctant to join the study otherwise.

The lack of patient interest was previously identified as a major barrier to app prescription among general practitioners [38]. Besides education and careful evaluation of current mHealth solutions [11], solid evidence for the effectiveness and usability is needed to overcome current barriers and increase app prescription rate.

Concerning app feature preferences, our results were very much in line with previous research [16,25]. Patients do not prefer patient-to-patient communication features; however, patients are very interested in other app features, particularly those providing information.

As reported in previous studies [25,28], our work clearly shows that most patients regularly use the internet to retrieve medical information. In a previous study [3] published in 2016, 47% of the patients consulted the internet to investigate their symptoms. This proportion seems to be increasing, as in our study 67% and in another recent study [25] 95% of patients stated they did so. Compared to German oncology patients interviewed in 2016 [15], more rheumatic patients seem to use mobile devices regularly (69.6% compared to 91.2%); however, the information quality is very heterogeneous which could be misleading.

The obvious enthusiasm of patients to share app data for research purposes and their wish for such data to be made available to the treating rheumatologist for routine care underlines that an improvement to the science and practice of rheumatology could be realized using mHealth. A digital approach allows the inclusion of well-structured patient-generated data to improve clinical shared decision making and clinical research [39]. The interoperability of systems is crucial for the success of such mHealth tools. A part of patients want rheumatologists to review their app input. This could easily cause information overload and an additional workload to physicians; however, rationing appointments, earlier detection, and response to flares would also be possible [40]. In combination with wearable sensors and algorithms, patient monitoring could further be improved.

The medical community, public health system, and private sector need to increase their efforts to improve eHealth competencies and to provide safe and effective digital tools to leverage the way of mHealth into routine rheumatology care.

Conclusion

To our knowledge, this is the first study capturing a detailed mHealth perspective of patients with rheumatic diseases, which could guide rheumatology app development and implementation. Most patients included in this study possessed smartphones and believed that using medical apps could be beneficial for their health. A substantial majority was also willing to share app data for research purposes. The current usage of mHealth among rheumatic patients is, however, very limited and eHealth literacy was rather poor. We could successfully identify unmet needs and patient priorities, which can be used to accelerate and guide the way of mHealth into routine rheumatology care.

Authors' Contributions

This work was performed to fulfill the requirements for obtaining the degree Dr med for AL and CR. JK, DS, AL, CR, KT, and AJH drafted the manuscript. KT and JK performed the statistical analysis. All authors reviewed the draft and provided comments for change. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original German survey.
[PDF File (Adobe PDF File), 347 KB - mhealth_v8i8e19661_app1.pdf]

Multimedia Appendix 2
2014 and 2018/2019 Preferences for medication reminders, medical information format, digitally provided information structure, patient diary type, and physician communication type of patients with rheumatoid arthritis.
[PNG File , 281 KB - mhealth_v8i8e19661_app2.png]

References

1. Kool MB, van Middendorp H, Lumley MA, Schenk Y, Jacobs JWG, Bijlsma JWJ, et al. Lack of understanding in fibromyalgia and rheumatoid arthritis: the Illness Invalidation Inventory (3*I). Ann Rheum Dis 2010 May 24;69(11):1990-1995. [doi: 10.1136/ard.2009.123224]
2. 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]
3. Powley L, McIlroy G, Simons G, Raza K. Are online symptoms checkers useful for patients with inflammatory arthritis? BMC Musculoskelet Disord 2016 Dec 24;17(1):362 [FREE Full text] [doi: 10.1186/s12891-016-1189-2] [Medline: 27553253]
4. Urruticoechea-Arana A, León-Vázquez F, Giner-Ruiz V, Andréu-Sánchez JL, Olivé-Marqués A, Freire-González M, et al. Development of an application for mobile phones (App) based on the collaboration between the Spanish Society of Rheumatology and Spanish Society of Family Medicine for the referral of systemic autoimmune diseases from primary care to rheumatology. Reumatol Clin 2019 Nov 10 [FREE Full text] [doi: 10.1016/j.reuma.2019.09.001] [Medline: 31722849]
5. 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) 2019 Oct;71(10):1336-1343. [doi: 10.1002/acr.23768] [Medline: 30242992]

6. 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]

7. 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]

8. Kang H, Park H. A Mobile App for Hypertension Management Based on Clinical Practice Guidelines: Development and Deployment. JMIR Mhealth Uhealth 2016 Feb;4(1):e12 [FREE Full text] [doi: 10.2196/mhealth.4966] [Medline: 26839283]

9. Yang J, Weng L, Chen Z, Cai H, Lin X, Hu Z, et al. Development and Testing of a Mobile App for Pain Management Among Cancer Patients Discharged From Hospital Treatment: Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 May 29;7(5):e12542. [doi: 10.2196/12542]

10. Economides M, Ranta K, Nazander A, Hilgert O, Goldin PR, Raevuori A, et al. 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 Aug 26;7(8):e14284 [FREE Full text] [doi: 10.2196/14284] [Medline: 31452521]

11. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German Mobile Apps in Rheumatology: Review and Analysis Using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]

12. Najm A, Nikiphorou E, Kostine M, Richez C, Pauling JD, Finckh A, et al. EULAR points to consider for the development, evaluation and implementation of mobile health applications aiding self-management in people living with rheumatic and musculoskeletal diseases. RMD Open 2019 Sep 13;5(2):e001014. [doi: 10.1136/rmdopen-2019-001014]

13. Knitza J, Vossen D, Geffken I, Krusche M, Meyer M, Sewerin P, Arbeitskreis Junge Rheumatologen. [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]

14. Grainger R, Townsley H, White B, Langlotz T, Taylor WJ. Apps for People With Rheumatoid Arthritis to Monitor Their Disease Activity: A Review of Apps for Best Practice and Quality. JMIR Mhealth Uhealth 2017 Feb 21;5(2):e7 [FREE Full text] [doi: 10.2196/mhealth.6956] [Medline: 28223263]

15. Kessel KA, Vogel MM, Kessel C, Bier H, Biedermann T, Friess H, et al. Mobile Health in Oncology: A Patient Survey About App-Assisted Cancer Care. JMIR Mhealth Uhealth 2017 Jun 14;5(6):e81 [FREE Full text] [doi: 10.2196/mhealth.7689] [Medline: 28615159]

16. van DVR, Drossaert CHC, Taal E, van DLMAFJ. 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]

17. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010 Sep;62(9):2569-2581 [FREE Full text] [doi: 10.1002/art.27584] [Medline: 20872595]

18. Rudwaleit M, Landewe R, van der Heijde D, Listing J, Brandt J, Braun J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. Annals of the Rheumatic Diseases 2009 Mar 17;68(6):770-776. [doi: 10.1136/ard.2009.108217]

19. Taylor W, Gladman D, Helliwell P, Marchesoni A, Mease P, Mielants H. Classification criteria for psoriatic arthritis: Development of new criteria from a large international study. Arthritis Rheum 2006 Aug;54(8):2665-2673. [doi: 10.1002/art.21972]

20. Soellner R, Huber S, Reder M. The Concept of eHealth Literacy and Its Measurement. Journal of Media Psychology 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]

21. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]

22. Chung S, Park BK, Nahm E. The Korean eHealth Literacy Scale (K-eHEALS): Reliability and Validity Testing in Younger Adults Recruited Online. J Med Internet Res 2018 Apr 20;20(4):e138. [doi: 10.2196/jmir.8759] [Medline: 29678800]

23. Worm-Smeitink M, Janse A, van Dam A, Evers A, van der Vaart R, Wensing M, et al. Internet-Based Cognitive Behavioral Therapy in Stepped Care for Chronic Fatigue Syndrome: Randomized Noninferiority Trial. J Med Internet Res 2019 Mar 14;21(3):e11276 [FREE Full text] [doi: 10.2196/11276] [Medline: 30869642]

24. Najm A, Gossec L, Weill C, Benoist D, Berenbaum F, Nikiphorou E. Mobile Health Apps for Self-Management of Rheumatic and Musculoskeletal Diseases: Systematic Literature Review. JMIR Mhealth Uhealth 2019 Nov 26;7(11):e14730. [doi: 10.2196/14730]

25. Najm A, Lempp H, Gossec L, Berenbaum F, Nikiphorou E. Needs, Experiences, and Views of People With Rheumatic and Musculoskeletal Diseases on Self-Management Mobile Health Apps: Mixed Methods Study. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e14351. [doi: 10.2196/14351]

26. Nowell W, Curtis D, Curtis J. Arthritis Power Reflects Optimal Characteristics of Apps for Rheumatoid Arthritis. J Clin Rheumatol 2019 Dec 09. [doi: 10.1097/RHU.0000000000001223] [Medline: 31834100]

27. Schulz P, Nakamoto K. Health literacy and patient empowerment in health communication: the importance of separating conjoined twins. Patient Educ Couns 2013 Jan;90(1):4-11 [FREE Full text] [doi: 10.1016/j.pec.2012.09.006] [Medline: 23063359]

28. van DVR, Drossaert CHC, de HM, Taal E, van DLMAFJ. Measuring actual eHealth literacy among patients with rheumatic diseases: a qualitative analysis of problems encountered using Health 1.0 and Health 2.0 applications. J Med Internet Res 2013 Feb;15(2):e27 [FREE Full text] [doi: 10.2196/jmir.2428] [Medline: 23399720]

29. Milne RA, Puts MTE, Papadakos J, Le LW, Milne VC, Hope AJ, et al. Predictors of High eHealth Literacy in Primary Lung Cancer Survivors. J Cancer Educ 2015 Dec;30(4):685-692. [doi: 10.1007/s13187-014-0744-5] [Medline: 25355524]

30. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]

31. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth Literacy: Predictors in a Population With Moderate-to-High Cardiovascular Risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]

32. Noblin AM, Wan TTH, Fottler M. The impact of health literacy on a patient's decision to adopt a personal health record. Perspect Health Inf Manag 2012;9:1-13 [FREE Full text] [Medline: 23209454]

33. Ernsting C, Stühmann LM, Dombrowski SU, Voigt-Antons J, Kuhlmey A, Gellert P. Associations of Health App Use and Perceived Effectiveness in People With Cardiovascular Diseases and Diabetes: Population-Based Survey. JMIR Mhealth Uhealth 2019 Mar 28;7(3):e12179. [doi: 10.2196/12179]

34. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]

35. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011 Nov;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]

36. Vo V, Auroy L, Sarradon-Eck A. Patients' Perceptions of mHealth Apps: Meta-Ethnographic Review of Qualitative Studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817. [doi: 10.2196/13817]

37. Hiboot. Société Française de Rhumatologie. 2019. URL: https://hiboot.larhumatologie.fr/ [accessed 2019-12-30]

38. Byambasuren O, Beller E, Glasziou P. Current Knowledge and Adoption of Mobile Health Apps Among Australian General Practitioners: Survey Study. JMIR Mhealth Uhealth 2019 Jun 3;7(6):e13199. [doi: 10.2196/13199]

39. Ibfelt EH, Jensen D, Hetland ML. The Danish nationwide clinical register for patients with rheumatoid arthritis: DANBIO. CLEP 2016 Oct;Volume 8:737-742. [doi: 10.2147/clep.s99490]

40. Wang P, Luo D, Lu F, Elias JS, Landman AB, Michaud KD, et al. A Novel Mobile App and Population Management System to Manage Rheumatoid Arthritis Flares: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2018 Apr 11;7(4):e84. [doi: 10.2196/resprot.8771]

Abbreviations

eHEALS: eHealth Literacy Scale
mHealth: mobile health
OR: odds ratio

XSL•FO
RenderX

information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

Original Paper

User Experiences With and Recommendations for Mobile Health Technology for Hypertensive Disorders of Pregnancy: Mixed Methods Study

Karin Rolanda Jongsma[1*], PhD; Josephus F M van den Heuvel[2*], MD; Jasmijn Rake[2], MD; Annelien L Bredenoord[1], Prof Dr; Mireille N Bekker[2], MD, PhD

[1]Department of Medical Humanities, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

[2]Obstetrics and Gynaecology, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

[*]these authors contributed equally

Corresponding Author:
Karin Rolanda Jongsma, PhD
Department of Medical Humanities
University Medical Center Utrecht
Utrecht University
PO Box 85500
Utrecht, 3508 GA
Netherlands
Phone: 31 88 75 51351
Email: k.r.jongsma@umcutrecht.nl

Abstract

Background: Hypertensive disorders of pregnancy (HDP) are a primary cause of adverse maternal and neonatal outcomes worldwide. For women at risk of hypertensive complications, guidelines recommend frequent surveillance of blood pressure and signs of preeclampsia. Clinic visits range from every 2 weeks to several times a week. Given the wide ubiquity of smartphones and computers in most countries and a growing attention for self-management, digital technologies, including mobile health (mHealth), constitute a promising component of monitoring (self-measured) blood pressure during pregnancy. Currently, little is known about the experiences of women using such platforms and how mHealth can be aligned with their needs and preferences.

Objective: The objectives were twofold: (1) to explore the experiences of Dutch women who had an increased risk of HDP with a blended care approach (mHealth combined with face-to-face care) for remote self-monitoring of blood pressure and preeclampsia symptoms and (2) to formulate recommendations for the use and integration of mHealth in clinical care.

Methods: Alongside a prospective blended care study (SAFE@home study) that monitors pregnant women at increased risk of HPD with mHealth technology, a mixed methods study was conducted, including questionnaires (n=52) and interviews (n=11). Results were analyzed thematically.

Results: Of the 4 themes, 2 themes were related to the technologies themselves (expectations, usability), and 2 themes were related to the interaction and use of mHealth (autonomy and responsibilities of patients, responsibilities of health care professionals). First, the digital platform met the expectations of patients, which contributed to user satisfaction. Second, the platform was considered user-friendly, and patients favored different moments and frequencies for measuring their blood pressure. Third, patient autonomy was mentioned in terms of increased insight about their own condition and being able to influence clinical decision making. Fourth, clinical expertise of health care professionals was considered essential to interpret the data, which translates to subsequent responsibilities for clinical management. Data from the questionnaires and interviews corresponded.

Conclusions: Blended care using an mHealth tool to monitor blood pressure in pregnancy was positively evaluated by its users. Insights from participants led to 7 recommendations for designing and implementing similar interventions and to enhance future, morally sound use of digital technologies in clinical care.

XSL•FO

RenderX

Introduction

Mobile health (mHealth) refers to the use of mobile devices, mobile phones, and wireless technologies to support the achievement of health objectives [1]. mHealth is expected to improve access to care, enhance patient satisfaction, and reduce clinic visits and admissions without compromising safety of care and is argued to improve interaction with and participation of better-informed and more active patients [2-4]. To date, mHealth has mostly focused on patients with chronic conditions or healthy individuals to improve healthy lifestyle habits [5-7]. As in other domains of health care, including pregnancy care, a shift is currently occurring from hospital-based to home-based services [8]. In search of improved care for pregnant women, tailored care with the integration of mHealth has been suggested as an addition to or partial replacement of frequent prenatal visits [9]. This approach is called blended care, where digital technologies are combined and integrated with face-to-face care. While many of these technologies are being developed and implemented, little is still known about clinical outcomes including safety, effectiveness, patient satisfaction, and ethical considerations.

Hypertensive disorders of pregnancy (HDP) are a primary cause of adverse maternal and neonatal outcomes worldwide and occur in 10% of pregnancies [10]. Risk groups for hypertensive complications include women with chronic hypertension, diabetes, obesity, renal disease, cardiac disease, and preeclampsia in a prior pregnancy. The proportion of women with these risk factors has been steadily rising over recent years [10]. For women considered to be at risk, guidelines recommend frequent observation of the fetal condition and the pregnant woman's blood pressure and signs of preeclampsia [11]. Planned and unplanned visits can range from every 2 weeks to 4 times a week or even daily. The burden of these recurrent visits is significant, for both patients and their spouses and family, as well as for health care services. However, the incidence of preeclampsia with severe features is approximately 3% [12], meaning that a substantial number of monitored women, while at risk, do not develop this condition.

Given the wide ubiquity of smartphones and tablets in most countries, mHealth is a promising alternative for monitoring hypertension during pregnancy. The latest research has shown that pregnant women are willing to undertake repeated self-measurements and a majority of women would like to be involved in their blood pressure management [13,14] and regard remote monitoring important for their pregnancy follow-up [15]. Little is known about the experiences of women using such platforms and how these digital tools can be aligned with their needs and preferences.

This study aimed, firstly, to explore how pregnant women, who have used mHealth as part of a blended care approach for repeated blood pressure measurements and preeclampsia symptom reporting, experience the use of such technology. Second, the study aimed to formulate recommendations based on these user experiences. Based on the insights originating from the users' experiences, we identified several recommendations to design and implement similar interventions and to enhance future use of digital technology in clinical care.

Methods

A mixed methods study, alongside a prospective blended care study (SAFE@home study) [16,17], was performed to explore the understanding of patients' experiences with mHealth [18,19]. Data were collected by means of validated questionnaires and semistructured in-depth interviews with patients that had experience with mHealth for remote monitoring of HDP, to explore their experiences and motivations. The research ethics committee of the University Medical Center Utrecht determined that this study was exempt from the Medical Research Involving Humans Act (reference number 18-898-C).

Context of the Blended Care Approach in Prenatal Care

The overarching prospective study, named Safe@Home, evaluates the use of mHealth technology to remotely monitor blood pressure and preeclampsia symptoms. The data collected within this study were sent by the patient to the digital monitoring team, who reviewed the data each day except for the weekend days. The mHealth technology consisted of an automated blood pressure monitor with Bluetooth connection to a smartphone app for iOS users and a web-based portal for Android users (Figure 1) [16,17]. Digital monitoring started from 16 weeks gestational age and was continued until delivery, with interruption in case of hospital admission. Participation in the blended care approach was offered to pregnant women whom, at intake, presented with one of the following risk factors for hypertensive complications: chronic hypertension, history of prior preeclampsia, or maternal cardiac or renal disease requiring prenatal care in our clinic in the University Medical Centre Utrecht (university hospital) or Diakonessenhuis Utrecht (general teaching hospital). Access to a smartphone or tablet with internet connection and good understanding of either the Dutch or English language were required. More information about the overarching study can be found in [16,17].

Figure 1. Example of the blood pressure trend graph in the Luscii app.
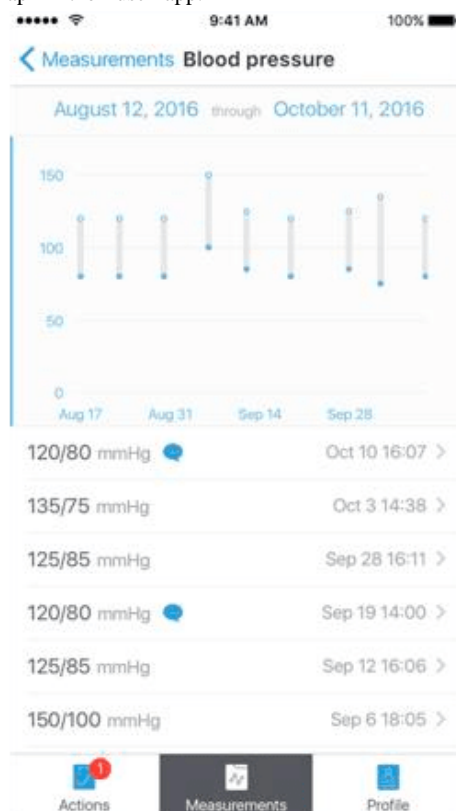
A prenatal visit schedule was predefined for this group of patients, with a reduced number of visits while continuing remote monitoring. Participants were asked to measure their blood pressure every weekday before 10 am and at least 1 hour after waking up. A 9-question symptom score list could be answered in case of hypertension. Predetermined thresholds (systolic blood pressure >140 mm Hg, diastolic blood pressure >90 mm Hg) or self-reported symptoms of preeclampsia in the questionnaire resulted in automatically generated alarm signals on a monitoring dashboard in the hospital. For health care providers, a web portal provided online access to patient-reported questionnaires and blood pressure data. Members of the digital monitoring team (midwives or obstetric nurses) reviewed the data every morning from the outpatient department. The combination of blood pressure measurement and the presence of symptoms was reviewed and if needed, the digital monitoring team could consult the obstetrician for advice. Subsequently, participants were contacted to advise about management or follow-up. The platform was embedded into prenatal care with the use of a reduced predefined prenatal visit schedule, with regular appointments in the outpatient department carried out by hospital midwives and gynecologists (in training).

Data Collection

Questionnaires

At 36 weeks of gestation, two questionnaires were sent by email to all participants of the prospective study. One questionnaire assessed the usability of the mHealth technology (via an app or web portal) and the connected devices, focusing on the ease of use and given instructions. This usability questionnaire consisted of 9 propositions rated on a 5-point Likert scale (strongly agree to strongly disagree) to obtain quantifiable scores (see Textbox 1). Furthermore, the use of the blood pressure monitor, usability of the smartphone app, and content of the app could be rated on a scale from 1 to 10. The usability questionnaire was generated by the study team and not validated before the start of the study. The second questionnaire was the validated Client-Centered Care Questionnaire (CCCQ). The CCCQ was developed as an instrument to measure client-centeredness as experienced by clients of care organizations and to evaluate the effects of interventions aimed at improving the client-centeredness of care services [20]. Themes of the CCCQ include recognition, respect, autonomy, and partnership as perceived by the participants. It consists of 15 questions rated on a 5-point Likert scale, ranging from "totally disagree" to "totally agree" (Multimedia Appendix 1). Results of the CCCQ can be interpreted using a unidimensional application. This is done by aggregating all information in one measure and calculating a total test score. This total score expresses care receivers' perception of client-centered care, with higher scores representing higher perceived client-centered care. Separate questions are discussed thematically, in line with the qualitative results of this study.

Textbox 1. Items on the usability questionnaire.

- The system and its use were easy to understand.

- I felt at ease using the system during pregnancy.

- While using the system, I was able to continue my daily activities.

- I am satisfied with the ease of use of this system.

- I would recommend this system to other pregnant women.

- The instructions for use of the blood pressure monitor were clear.

- The instructions for use of the app were clear.

- It was clear when to provide my measurements.

- It was clear when to contact my health care provider.

Interviews

Semistructured in-depth interviews were conducted with participants of the overarching study after an email invitation. Patients who were willing to be interviewed and were able to speak either Dutch or English were included in the interview study through purposive sampling. The topic list was designed to include the motivations, experiences, and perspectives of patients using the platform. The semistructured format provided participants with the opportunity to discuss matters they believed needed emphasis, while offering guidance throughout the interview. Questions for the interview guide were based on preliminary quantitative results from our questionnaires, which suggested the importance of technical functioning, communication with care providers, and implications for autonomy. The topic list was expanded based on the literature on ethical aspects of digital health, mHealth, and digital monitoring. Interviews were conducted by KJ (assistant professor) and MD (research assistant), both female researchers with experience in qualitative studies using interviews. No relationship with participants was established prior to the interviews. The interviews were conducted until saturation was reached, meaning that no new perspectives or themes were found in consecutive interviews and no new themes emerged from the data. Verbatim transcriptions of interviews and interviewers' notes were compared with audio recordings to check for accuracy. Transcripts were imported into NVivo12 and analyzed thematically, combining inductive and deductive analyses. KJ and MD started with an a priori coding scheme to allow for deductive coding based on topics described in the literature (responsibilities, shared decision making, patient empowerment, motivations). Codes and their meanings were discussed among the research team prior to coding to guarantee intercoder reliability. The inductive part of the thematic analysis combined methods of close reading and constant comparison; codes emerging from the transcripts were clinical expertise, reassurance, burden and stress, and understanding one's own condition. Codes were examined and systematically reviewed for supporting or conflicting evidence concerning emerging themes and codes. We also explored whether there were any differences between nulliparous and parous women and between women with a history of HDP and those without. Where relevant, we explicitly address these differences in the results. Results are reported using the Consolidated Criteria for Reporting Qualitative research checklist.

Results

Patient Characteristics

Of 103 invited participants, a total of 51 participants completed both questionnaires, and one participant only completed the CCCQ (total n=52). The interviews were conducted with 11 women (8 after delivery, 3 during pregnancy) and comprised the qualitative part of this study. All interviews (n=11) took place by phone, as preferred by the participants, and lasted between 35 minutes and 58 minutes. The majority of the interviewed women (8/11) also completed the questionnaires.

The demographic data of both groups are shown in Table 1. Obstetric characteristics of relevance to this topic are indicated for the interviewees and questionnaire participants (see Table 1).

After analysis of the data from the interviews and questionnaires, we identified 4 themes. Of these, 2 themes were related to the mHealth technology itself (themes 1 and 2), and 2 themes were related to the interaction with and use of the mHealth technology (themes 3 and 4).

Table 1. Patient characteristics.

Parameters	Questionnaires (n=52)	Interviews (n=11)
Maternal age (years), mean (SD)	34.40 (4.127)	34.18 (2.529)
BMI (kg/m^2), mean (SD)	24.94 (4.62)	23.88 (2.496)
Ethnicity, n (%)		
Caucasian	47 (90.4)	11 (100)
Afro-Caribbean	0 (0)	0 (0)
Mediterranean	3 (5.8)	0 (0)
Other	2 (3.8)	0 (0)
Level of education, n (%)		
Primary school	1 (1.9)	0 (0)
Secondary school	4 (7.7)	0 (0)
Middle-level applied education	14 (26.9)	3 (27.3)
Higher-level applied education	17 (32.7)	6 (54.5)
Scientific education (university)	13 (25.0)	2 (18.2)
Unknown	3 (5.8)	0 (0)
Nulliparous, n (%)	19 (36.5)	2 (18.2)
HDP[a] prior pregnancy, n (%)		
None	14 (26.9)	3 (27.3)
Chronic hypertension	1 (1.9)	1 (9.1)
Gestational hypertension	5 (9.6)	2 (18.2)
Preeclampsia/HELLP[b]	13 (25.0)	3 (27.3)
Not applicable (nulliparous)	19 (36.5)	2 (18.2)
Initial diagnosis at start of SAFE@home study, n (%)		
Preeclampsia in prior pregnancy	10 (19.2)	2 (18.2)
Chronic hypertension	17 (32.7)	5 (45.5)
Cardiac disease	17 (32.7)	3 (27.3)
Renal disease	8 (15.4)	1 (9.1)
HDP current pregnancy, n (%)		
None	23 (44.2)	2 (18.2)
Chronic hypertension	14 (26.9)	3 (27.3)
Gestational hypertension	6 (11.5)	2 (18.2)
Preeclampsia	9 (17.3)	4 (36.4)

[a]HDP: hypertensive disorders of pregnancy.

[b]HELLP: hemolysis, elevated liver enzymes, low platelet count.

Theme 1: Expectations of and Satisfaction With the mHealth Technology

Quantitative Analysis

Analysis of the usability questionnaire showed that almost all participants (49/51, 96%) felt comfortable using mHealth. The vast majority (45/51, 88%) would recommend it to their friends and family, especially participants who had been pregnant before (97% of multiparous vs. 74% of nulliparous women). Overall, client-centeredness of the blended care approach, based on the CCCQ, was rated at an average 57.5 of 75 points (range 36-75 points), which translates to a score of 77 from a possible total score of 100. This total CCCQ score was comparable between nulliparous women (score of 76, n=19) and parous women (score of 77, n=33). Of all parous women, women with prior HDP (19/33) scored the CCCQ slightly higher (score of 79/100) than those without experience with HDP (14/51; score of 75/100).

Qualitative Analysis

In order to understand what is important for mHealth users, we asked the interview participants what their expectations were before they started using the digital technology in the Safe@Home study and whether their expectations were met. The most often mentioned motivations to start using the technology were the expected reassurance of being closely monitored by a health care professional (9/11; 2 nulliparous and 7 parous), better pregnancy outcomes (6/11; 3 with a history of hypertension and 3 with a history of HPD), and the prospect of fewer hospital visits (5/11; none with a history of HPD). This aligned well with their experiences; most interview participants (8/11) reported they felt reassured and safe because of the close monitoring by their obstetric care professional. The use of mHealth reduced the frequency of visits, which contributed to the users' wellbeing and a more relaxed pregnancy experience (9/11; 2 nulliparous and 7 parous). The blended care approach also enabled timely preventative measures or interventions, which resulted in early detection of abnormalities or risks (2/11). All interview participants considered it a benefit to be able to measure their own blood pressure, especially when they experienced symptoms associated with preeclampsia. Also, when their measurement indicated normal blood pressure, the digital monitoring was considered useful and reassuring, because it would indicate that the symptoms were not caused by hypertension. Comparable to the results of the questionnaire, all interviewed women would recommend the system to other pregnant women.

Some reflections of the interview participants indicated that their expectations did not always match their experiences. A few women were surprised by health care professionals calling when they did not expect it, while at other times, they were not called by the health care professional when they expected it based on their uploaded blood pressure data (2/11; both with a history of hypertension). Participants who needed reassurance that their blood pressure or symptoms were nothing to worry about sometimes called the hospital themselves. Furthermore, one interview participant needed several extra hospital visits because of hard-to-control hypertension, eventually leading to hospital admission. As a result, she was somewhat disappointed that the digital monitoring platform did not live up to her expectations (P5, Table 2).

Table 2. Quotes illustrating interviewees' expectations and satisfaction.

| Topic	Quotes
Reassurance	More relaxed, I'd say. I haven't worried at all about my blood pressure. I considered it under control […] Because you do it continuously [the measurements], it reassures you. (P1)
	It is very pleasant and extremely easy. It's reassuring that you are being monitored [by health care professionals]. (P11)
Frequency of visits	It has given me peace of mind over all those months, primarily because of the significantly reduced number of clinical visits. (P1)
	It is ironic; we expected it because it was announced like that, that we would have to visit the hospital less often, because we would be monitored via the app, but it resulted in more frequent contact. (P5)

Theme 2: Usability of the mHealth Tool

Quantitative Analysis

Analysis of the questionnaires showed that nearly all participants considered the user instructions of the blood pressure monitor (49/51, 96%) and smartphone app or website (48/51, 94%) to be clear and understandable. Similarly, almost everyone (49/51, 96%) found it easy to learn how to use the mHealth technology (Figure 2).

Furthermore, the vast majority of participants (47/51, 92%) was satisfied with the usability of the mHealth technology; 81% (41/51) of the participants said the daily measurements took ≤5 minutes a day (average 4.57 minutes, range 3-15 minutes), and women could easily continue their daily routine while using the technology (50/51, 98%). Some found it difficult to combine digital monitoring with their daily routine (5/51, 10%).

On a scale from 1 to 10, the blood pressure monitor was rated at 8.5 (range 6-10), usage of the smartphone app at 7.6 (range 1-10), and content of the smartphone app at 7.8 (range 1-10).

Figure 2. Results from the user satisfaction questionnaire.
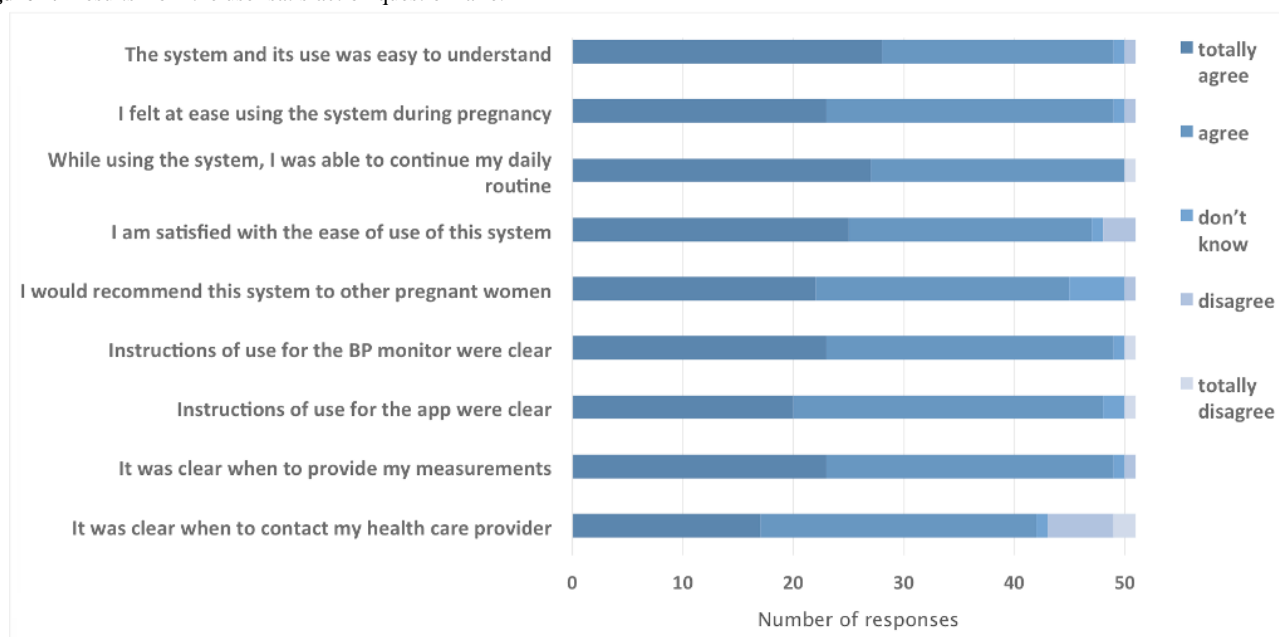
Qualitative Analysis

Similar to the questionnaire, interview participants (6/11) considered the app to be "modern" and easy to use; all users of the web portal (4/11) suggested that an Android app would be more user-friendly. Moreover, the iOS app was regarded to be comprehensive; the symptom survey was considered short but clear and easy to complete — it did not take them much effort and time (4/11; 2 with a history of HPD and 2 without). Other technical aspects that contributed to the ease of use were the reminder function, automatic Bluetooth synchronization, and perceived high accuracy of the measurement. A couple of participants (2/11) noted that technical understanding of the functioning of the app was irrelevant for their user experience.

A few users mentioned that measuring early in the morning was not always easy to combine with either commuting to work or "family rush hour" in the morning (3/11) or not representative, as their morning blood pressure was naturally low (1/11). These users preferred to have the option to measure in the evening instead of the morning. Most considered measuring 5 times a week sufficient; a couple of interview participants measured every day, even during weekends, either because of worries

about her medical condition (1/11; with a history of HPD) or to allow it to become a habit in their daily routine (2/11). Multiple mHealth users (4/11) measured several times a day when they experienced symptoms of preeclampsia or hypertension. At the same time, others (6/11; 4 with a history of HDP) mentioned that daily measurements were too burdensome or medicalizing, especially when they perceived their symptoms or blood pressure to be stable. A couple of interviewees (2/11) mentioned they missed the mHealth tool after giving birth and would have wanted to continue to measure during their postpartum period.

A couple of women (2/11) mentioned technical errors in the synchronization of their measurements with the system used in the hospital. Furthermore, a couple of others (2/11; both parous and with a history of hypertension) felt that the symptom score list to monitor preeclampsia signs was at times confusing because some questions did not match the specific pregnancy term. In particular, the question "Can you feel the baby move?" was considered to be upsetting in the first trimester. Also, one interview participants considered the orange or red lights stressful, as she never saw a green light because of her high values (P9, with a history of HPD; Table 3).

XSL•FO
RenderX

Table 3. Quotes illustrating interviewee perspectives on the usability of the mHealth tool.

Topic	Quote
App vs web portal	At first, I used the web portal, but when I had a closer look, I realized that the app is much easier, because it automatically synchronizes. It is so easy! (P1)
Frequency of measurement	Before I started, I thought it would be burdensome to measure my blood pressure every day and was not convinced that it would be necessary. [...] But eventually, it was very easy. It became part of my routine to measure my blood pressure in the morning before going to work or before bringing the children to school. (P8)
	The app was meant to be used in the morning, which was somewhat a downside, because my blood pressure is fine in the morning. (P3)
	While I was using it, no [I did not experience anything unexpected], but after giving birth and being back home, I continued measuring with my own device, because I missed that sort of information about my body. (P7)
Questions suitable to term	Those questions did not really match with being in the first trimester. Because it asked for example "do you feel contractions," "do you still feel the baby move," But [at that time], I hardly had a belly, and I couldn't even feel the baby yet. [..] I found it difficult and puzzling. (P10)
Alarms	"Those lights [on the blood pressure device], they should get rid of in favor of people who are easily stressed out. They should rather show you a green light when you're fine, orange when there are problems, and red when things are bad. [..] It showed orange so often. Since my blood pressure has been high my whole life, you feel like there is a continuous alarm, while yeah, that was not really the case." (P9)

Theme 3: Autonomy and Responsibility of Patients

Quantitative Analysis

Respondents of the questionnaires were positive about their role within the blended care approach. The majority of the participants felt they were given sufficient opportunity to draw on their own knowledge and experience (40/51, 77%) and to decide about the kind of care they receive (43/52, 83%). Furthermore, they felt they were given enough opportunity to do what they were capable of doing themselves (47/52, 90%). However, only half of the participants (26/52, 50%) felt like they were given enough opportunity to arrange and organize prenatal care themselves. Some (30/52, 57%) would like even more influence in clinical decision making and felt that health care professionals are sometimes too quick to deny a possibility. A minority of the participants (21/52, 40%) felt like they had a say in deciding when the care was provided.

Qualitative Analysis

Interview participants noted two dimensions related to patient autonomy. First, all interview participants (11/11) mentioned that mHealth helped to be informed about HDP. Insights on blood pressure over time, as displayed in a trend line in their app, was especially considered to be informative (6/11; all parous). Such information raised awareness about the symptoms of HDP and when to report to health care professionals (4/11). Some interview participants (2/11) argued that these insights are paired with responsibilities to carefully measure blood pressure and to contact the health care professional when symptoms increase.

A second aspect related to patient autonomy mentioned by participants was that the use of mHealth contributed to them being in control of their own health and to bring their own perspectives to the fore in consultations with health care professionals (7/11). mHealth allowed them to monitor their own symptoms and, when necessary, adapt their behavior (4/11), for example with regard to activities or medication (P8, Table 4).

Table 4. Quotes illustrating interviewee perspectives on autonomy and responsibilities of patients.

Topic	Quote
Being informed	I experienced that I thought I was going to measure hypertension because I felt a headache, but then I didn't measure anything abnormal. That is odd. But exactly because of such experiences, I consider it beneficial to be able to measure, because it provided objective information to really judge it. Because I find it difficult to determine what is the matter, simply by how I feel. (P8)
	I then understood, you know, why they [health care professionals] ask you all these questions and that these are relevant. Because of the symptom score list or due to hypertension and related symptoms of preeclampsia, that I became aware that once you experience such symptoms, you shouldn't think it's normal, but that you have to inform health care professionals. (P2)
Information for lifestyle	It is information, you know. I had to take care of my activity level; when I did not measure hypertension, I would go outside for example. [...]. And in case I would measure hypertension, I would take it easy and ask my husband to take care of the children. It provides information relevant for such activities. That was very helpful. (P8)
Responsibilities	[..] but it is also your own responsibility, the responsibility of the mother or the pregnant woman. Not only because you know your body best, but also because you become aware of aspects because of this research study. And then it's my responsibility to discuss it with the health care professional. (P2)
Control	You both have access to the information. What I see in my overview, the physician can also see, so you can also look at it together. I got the impression that more deliberation is possible, that you do it together like how should I interpret this and the physician can explain it for example. (P8)

Theme 4: Health Care Professionals' Expertise and Responsibilities

Quantitative Analysis

After starting the digital monitoring, it was clear for the majority of the participants when the digital monitoring team needed to receive their measurements (49/51, 96%) and when to contact the physician regardless of their data (42/51, 82%). They felt that their personal wishes were sufficiently considered by the health care professionals (46/52, 89%). Most of the survey respondents said they could tell that their obstetric care professional really listened to them (50/52, 96%) and that they were given enough opportunity to say what kind of care they needed (47/52, 90%).

Qualitative Analysis

In addition to the findings of the quantitative research, interviewees showed that they consider the expertise of the health care professionals important in monitoring HDP. All interview participants (11/11) said that health care professionals have invaluable clinical expertise to oversee the implications of the measurements, as well as to decide the need for additional tests, the interval of clinical visits, and medication or hospitalization. The follow-up initiated by health care professionals — either by phone or via clinical visits — contributed to the feeling of being well taken care of and met the interviewees' expectations regarding responsibilities (5/11). Patients also mentioned that it should remain the health care professionals' responsibility to undertake action when the measurements deviate from the norm (5/11). They felt relieved that monitoring and resulting action are not solely the patient's responsibility (11/11).

Moreover, patients were appreciative that health care professionals acted if a patient would underestimate the severity of their situation (3/11). Patients argued that important decisions about their condition cannot exclusively be based on the information from the tool, but an expert's clinical view is required for interpretation and to make personalized treatment decisions (5/11; Table 5).

Table 5. Quotes illustrating interviewees' perspectives on expertise and responsibilities of health care professionals.

Topic	Quote
Medical expertise	Well, for me, those data, I'm not trained as a health care professional, to interpret my data. I, myself, had the [possibility] to see how my blood pressure developed over time. But the idea that health care professionals see my data and can interpret it and can ask you to come to the hospital when necessary, that is comforting. (P2)
	[..] and that they can interpret it. Like for me, it was the case that it [blood pressure] was higher than 90, even if 90 is the threshold value for me, but [they explained] that for me, you see sometimes other things happening. Then I know that, you know, it's very helpful when a physician helps me and interprets the data. I mean, that they don't simply tell and stick to the threshold values, but also interpret it in your specific situation. I believe the shared effort lies in me conducting the measurements and supplying that information. (P8)
	[..] it's a shared effort. But I think the physician is leading because they know best. I mean, I know more or less which medication I take now and what sort of effect it has on me. But I don't know, for example, whether or when I can stop taking medications and what consequences it would have; I haven't studied for that. (P4)
Active monitoring	I really like having been called after [by a nurse], because then you have confirmation that they will undertake action when it is necessary. [..] I think it was great, that in that way also really something is done with the data you collect every day. (P7)

Discussion

This study analyzed user experiences with a blended care approach for the monitoring of HDP (Safe@Home study). Overall, the results of the questionnaires and interviews corresponded and were supplementary. The effects of using mHealth met the expectations of the participants, who were overall very satisfied with the easy-to-use technology. mHealth was considered to support patient autonomy by providing information and ways to be in control, but the interpretation of the measurements requires the involvement of health care professionals. Participants also noted a few possibilities for improvement. With the focus on future development and implementation of mHealth in care, we extracted multiple recommendations from our results (see Textbox 2).

Textbox 2. Recommendations for the future use and implementation of mHealth technologies in clinical care.

1. Be modest in the communication regarding expected group benefits of the digital health technology to prevent disappointing individual patients who do not experience these specific benefits.

2. Provide the user insight into the data; in particular, a graphic representation over time is a helpful method to foster patient knowledge and can support patients to participate in clinical decision making.

3. The mHealth data should be integrated in (electronic) health records and should be accessible to all health care professionals that are engaged in care.

4. The health care professionals should remain responsible for the interpretation of data obtained via digital monitoring, as the clinical expertise of health care professionals is necessary for the early detection of abnormalities and clinical decision making.

5. Health care professionals should be aware of (pregnant) patients' willingness and capability to self-measure their blood pressure at home.

6. Symptom score lists and blood pressure thresholds should be personalized, meaning that the questions should be adapted to the pregnancy term and thresholds should be set to fit the user's situation.

7. The moment and frequency of measurement should be communicated clearly but should also be sensitive and adaptable to the daily life of the user.

Our Findings in Context

Currently, several digital technologies are being developed that moderate or replace traditional clinical care. The study described here is an excellent example of such digital health technology in the clinical context that replaces some of the care traditionally provided in the clinical setting with digital monitoring at home. Our study confirmed several findings described by other digital monitoring studies. Some comparable studies have reported on remote blood pressure monitoring in pregnancy, without in-clinic monitoring by care professionals [21,22]. For a comparable intervention with clinical monitoring, only survey data were reported [15]. Our study confirmed that pregnant women at risk of HDP are willing to participate in self-monitoring services and are capable of bearing the responsibilities of measuring their own blood pressure [15,21,22]. Our study confirmed that women who experienced HDP in a prior pregnancy, in particular, were strongly in favor of blended care approaches in prenatal care [22]. A comparable intervention for pregnant women with hypertension that included remote monitoring of blood pressure and monitoring by health care professionals reported that 83% of the participants experienced a feeling of safety and that 68% preferred to be contacted within 12 hours after the measurement in case of abnormal measurements, preferably by their midwife or obstetrician [15]. Our study found comparable feelings of appreciation and safety among the users, partly because of the follow-up by health care professionals by phone or via clinical visits. Self-measuring was found to be reassuring; when abnormal values were detected as women took and interpreted their own measurement, it was clear for the participants when to contact the clinic [21]. Other studies have also found that women prefer that blood pressure monitoring should not stop at the delivery date, but should be available postnatally, which was also expressed by our interview participants [21,22].

Opportunities for and Challenges With Blended Care Approaches in Clinical Care

With the rapid development and implementation of digital technologies in health care settings, the need for ethical guidance and practical recommendations for the implementation of such technologies, including mHealth, is widely acknowledged by patients, health care professionals, and influential advisory councils [23-25]. With the implementation of these technologies, it becomes possible to move beyond mere speculative debates about the opportunities and challenges of mHealth and to investigate how the practice is developing. Our study explored both user experiences and the expectations of users prior to using mHealth tools for digital monitoring. User experiences depend not only on the quality of the technology but also on the expectations one has before using it. Investigating both expectations and experiences is helpful, not only to understand what may motivate pregnant women to use such technologies but also to assess whether these tools live up to users' expectations. Our study provides several insights in that respect: less frequent hospital visits and better-informed patients were often mentioned as factors contributing to the satisfaction with this technology. This shows that some of the widely discussed promises of mHealth were met in our study. Other claims about mHealth, such as increased accessibility, cost-effectiveness, and more empowered patients [1,2], were not (fully) substantiated by our study.

Furthermore, our study indicates that ethical guidance for the use of digital technologies in health care settings differs in significant ways from concerns about digital health consumers. Using digital technologies, including mHealth, in health care

settings raises a wider range of ethical challenges than have been described in the consumer context [26,27]. Aside from concerns about effectiveness, privacy, and safety, the health care context requires us to carefully assess the delegation of responsibilities to patients, influence on patient autonomy, and proportionality of burden and benefits. Regarding the delegation of responsibilities, our study showed that users are able to bear the responsibility for measuring their own blood pressure, but they did not feel able to bear the responsibility of interpreting their own data. Clinicians play an important role in the responsible use and implementation of these technologies. This indicates that careful consideration is required regarding which tasks and responsibilities can be delegated to technology (instead of face-to-face care) and which can be delegated to patients (instead of the health care professionals) without compromising safety or quality of care. Digital technologies, including mHealth, are not a stand-alone solution in the clinical context and need to be supplemented with clinical expertise. With regard to the influence on patient autonomy, our study has supported evidence that patients can become more familiar with their own body and disease symptoms and are able to use this information in adjusting their behavior or to deliberate with physicians. It is important to recognize that supporting and respecting patient autonomy are not completely in their own hands. Health care professionals involved in blended care play a crucial role. Not only will health care professionals have to recognize and respect wishes of autonomous persons but will also have to navigate between the standardized way of measuring, supported by digital technology, while still being able to personalize the analysis and interpretations to the interests and needs of a specific patient. Lastly, while mHealth technologies have several benefits, such as accessibility of information for both patients and health care professionals, less frequent hospital visits, and better understanding of one's own conditions, these benefits need to outweigh the burden of using these technologies (eg, time investment, user friendliness). Overall, our participants were very positive and satisfied with the mHealth technology, but the interview participants who felt their blood pressure was stable because of prescribed medication argued that the burden of measuring every day became somewhat disproportionate. Less frequent measurements may be a way to balance the burden and benefits for these groups. It also indicates that high levels of satisfaction with this blended care approach might be specific to the high-risk population that was selected for this approach. For the high-risk population, there is much to gain in terms of both health outcomes and time investment, but the balance may tip differently for medium-risk to low-risk groups.

Strengths and Limitations

This is a mixed methods study that benefits from reducing weaknesses inherent to both methods; it expands understanding, while also being comprehensive. Approximately half of the total users of the mHealth technology filled out the validated questionnaires. The sample of interviewees was representative of the participants of the questionnaires in terms of age, BMI, education level, and underlying conditions (Table 1). The findings of the survey and interviews were supplementary and helped to better understand what and for which reasons the mHealth tool was appreciated, which can inform future mHealth health interventions.

Our results must be interpreted in the context of the following limitations. Selection bias (self-selection) might have influenced the results, as participants of the prospective study agreed to take part in this innovative strategy with digital monitoring and may thus have a positive attitude in general to mHealth. Furthermore, the women willing to participate in this study may have had a relatively positive experience with this specific technology. Also, the experience of participants could have been biased by the outcome of their pregnancy. However, as the findings of the questionnaires (collected during pregnancy) and the interview data (8 postpartum, 3 during pregnancy) correspond, the influence may be marginal. The interviewed patients were fairly highly educated and may therefore not be representative of pregnant women in other socioeconomic situations. This explorative study has a relatively small sample size; in both the quantitative and qualitative aspects of the study, the provided ratios and percentages were not statistically powered and therefore cannot be fully generalized to other populations or care settings. Although saturation was reached on the identified codes and themes, further research could investigate these topics in more depth.

Conclusions and Recommendations

Our study explored the perspectives of pregnant women regarding the use of mHealth in a blended care approach to remotely monitor blood pressure in pregnancy. Based on the experiences of the users, several recommendations have been formulated. These recommendations draw on the needs, experiences, and views of the patients, meaning that following these recommendations will contribute to better-aligned and patient-centered care. These recommendations can help other scholars or physicians to guide the process of implementation and design of similar mHealth technologies.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Client-centered-care questionnaire (CCCQ).
[PDF File (Adobe PDF File), 393 KB - mhealth_v8i8e17271_app1.pdf]

References

1. mHealth: New horizons for health through mobile technologies: second global survey on eHealth. Geneva: World Health Organization; 2014:156-192.
2. Topol E. The Patient Will See You Now: The Future of Medicine is in Your Hands. Jackson, TN: Ingram Publisher Services; 2015.
3. World Health Organization. WHO Guideline: Recommendations on digital interventions for health system strengthening. 2019. URL: https://www.who.int/reproductivehealth/publications/digital-interventions-health-system-strengthening/en/ [accessed 2020-07-16]
4. European Commission. Green Paper: mobile Health ("mHealth"). Shaping Europe's digital future. 2014 Apr 10. URL: https://ec.europa.eu/digital-single-market/en/news/green-paper-mobile-health-mhealth [accessed 2020-07-16]
5. Mol A. What Diagnostic Devices Do: The Case of Blood Sugar Measurement. Theoretical Medicine and Bioethics 2000 Jan;21(1):9. [doi: 10.1023/a:1009999119586]
6. Lupton D. The digitally engaged patient: Self-monitoring and self-care in the digital health era. Soc Theory Health 2013 Jun 19;11(3):256-270. [doi: 10.1057/sth.2013.10]
7. Pinnock H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicentre, randomised controlled trial. BMJ 2013 Oct 17;347(3):f6070-f6070 [FREE Full text] [doi: 10.1136/bmj.f6070] [Medline: 24136634]
8. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the Next-Generation Perinatal Care: An Overview of the Literature. J Med Internet Res 2018 Jun 05;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
9. Woo VG, Lundeen T, Matula S, Milstein A. Achieving higher-value obstetrical care. Am J Obstet Gynecol 2017 Mar;216(3):250.e1-250.e14. [doi: 10.1016/j.ajog.2016.12.033] [Medline: 28041927]
10. Hutcheon JA, Lisonkova S, Joseph K. Epidemiology of pre-eclampsia and the other hypertensive disorders of pregnancy. Best Pract Res Clin Obstet Gynaecol 2011 Aug;25(4):391-403. [doi: 10.1016/j.bpobgyn.2011.01.006] [Medline: 21333604]
11. Hypertension in pregnancy: diagnosis and management: NICE guideline [NG133]. National Institute for Health and Care Excellence. 2019 Jun 25. URL: https://www.nice.org.uk/guidance/ng133 [accessed 2020-07-16]
12. Lisonkova S, Sabr Y, Mayer C, Young C, Skoll A, Joseph K. Maternal Morbidity Associated With Early-Onset and Late-Onset Preeclampsia. Obstetrics & Gynecology 2014;124(4):771-781. [doi: 10.1097/aog.0000000000000472]
13. Hodgkinson JA, Tucker KL, Crawford C, Greenfield SM, Heneghan C, Hinton L, et al. Is self monitoring of blood pressure in pregnancy safe and effective? BMJ 2014 Nov 18;349(1):g6616-g6616. [doi: 10.1136/bmj.g6616] [Medline: 25406132]
14. Magee LA, von Dadelszen P, Chan S, Gafni A, Gruslin A, Helewa M, Group For The CHIPS Pilot Trial Collaborative. Women's views of their experiences in the CHIPS (Control of Hypertension in Pregnancy Study) Pilot Trial. Hypertens Pregnancy 2007 Jul 07;26(4):371-387. [doi: 10.1080/10641950701547549] [Medline: 18066957]
15. Lanssens D, Vandenberk T, Lodewijckx J, Peeters T, Storms V, Thijs IM, et al. Midwives', Obstetricians', and Recently Delivered Mothers' Perceptions of Remote Monitoring for Prenatal Care: Retrospective Survey. J Med Internet Res 2019 Apr 15;21(4):e10887 [FREE Full text] [doi: 10.2196/10887] [Medline: 30985286]
16. van den Heuvel JF, Kariman SS, van Solinge WW, Franx A, Lely AT, Bekker MN. SAFE@HOME - Feasibility study of a telemonitoring platform combining blood pressure and preeclampsia symptoms in pregnancy care. Eur J Obstet Gynecol Reprod Biol 2019 Sep;240:226-231. [doi: 10.1016/j.ejogrb.2019.07.012] [Medline: 31330428]
17. van den Heuvel JFM, Lely AT, Huisman JJ, Trappenburg JCA, Franx A, Bekker MN. SAFE@HOME: Digital health platform facilitating a new care path for women at increased risk of preeclampsia - a case-control study. Pregnancy Hypertens 2020 Jul 18;22:30-36 [published online ahead of print, 2020 Jul 18]. [doi: 10.1016/j.preghy.2020.07.006] [Medline: 32717653]
18. Lilford RJ, Foster J, Pringle M. Evaluating eHealth: how to make evaluation more methodologically robust. PLoS Med 2009 Nov 24;6(11):e1000186 [FREE Full text] [doi: 10.1371/journal.pmed.1000186] [Medline: 19956674]
19. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
20. de Witte L, Schoot T, Proot I. Development of the client-centred care questionnaire. J Adv Nurs 2006 Oct;56(1):62-68. [doi: 10.1111/j.1365-2648.2006.03980.x] [Medline: 16972919]
21. Hinton L, Tucker KL, Greenfield SM, Hodgkinson JA, Mackillop L, McCourt C, et al. Blood pressure self-monitoring in pregnancy (BuMP) feasibility study; a qualitative analysis of women's experiences of self-monitoring. BMC Pregnancy Childbirth 2017 Dec 19;17(1):427. [doi: 10.1186/s12884-017-1592-1] [Medline: 29258469]
22. Sheehan E, Khalil A, Kay L. Using a smartphone app to identify signs of pre-eclampsia and/or worsening blood pressure. British Journal of Midwifery 2019 Feb 02;27(2):92-99. [doi: 10.12968/bjom.2019.27.2.92]

XSL•FO
RenderX

23. Lucivero F, Jongsma KR. A mobile revolution for healthcare? Setting the agenda for bioethics. J Med Ethics 2018 Oct 15;44(10):685-689 [FREE Full text] [doi: 10.1136/medethics-2017-104741] [Medline: 29907579]

24. Ruckenstein M, Schüll ND. The Datafication of Health. Annu. Rev. Anthropol 2017 Oct 23;46(1):261-278. [doi: 10.1146/annurev-anthro-102116-041244]

25. The Nuffield Council on Bioethics. The collection, linking and use of data in biomedical research and health care: ethical issues. 2015 Mar 02. URL: https://www.nuffieldbioethics.org/publications/biological-and-health-data [accessed 2020-07-26]

26. Lanzing M. "Strongly Recommended" Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies. Philos. Technol 2018 Jun 6;32(3):549-568 [FREE Full text] [doi: 10.1007/s13347-018-0316-4]

27. Gostin LO, Halabi SF, Wilson K. Health Data and Privacy in the Digital Era. JAMA 2018 Jul 17;320(3):233-234. [doi: 10.1001/jama.2018.8374] [Medline: 29926092]

Abbreviations

CCCQ: Client-Centered Care Questionnaire
HDP: hypertensive disorders of pregnancy
mHealth: mobile health

XSL•FO

RenderX

Original Paper

Use of Tablets and Smartphones to Support Medical Decision Making in US Adults: Cross-Sectional Study

Aisha Langford[1], PhD, MPH; Kerli Orellana[1], BA; Jolaade Kalinowski[1], EdD, MA; Carolyn Aird[1], MPH; Nancy Buderer[2], MS

[1]Department of Population Health, NYU Langone Health, New York, NY, United States
[2]Nancy Buderer Consulting, LLC, Oak Harbor, OH, United States

Corresponding Author:
Aisha Langford, PhD, MPH
Department of Population Health, NYU Langone Health
30 E 30th Street, Room 611
New York, NY, 10016
United States
Phone: 1 646 501 2914
Email: aisha.langford@nyulangone.org

Abstract

Background: Tablet and smartphone ownership have increased among US adults over the past decade. However, the degree to which people use mobile devices to help them make medical decisions remains unclear.

Objective: The objective of this study is to explore factors associated with self-reported use of tablets or smartphones to support medical decision making in a nationally representative sample of US adults.

Methods: Cross-sectional data from participants in the 2018 Health Information National Trends Survey (HINTS 5, Cycle 2) were evaluated. There were 3504 responses in the full HINTS 5 Cycle 2 data set; 2321 remained after eliminating respondents who did not have complete data for all the variables of interest. The primary outcome was use of a tablet or smartphone to help make a decision about how to treat an illness or condition. Sociodemographic factors including gender, race/ethnicity, and education were evaluated. Additionally, mobile health (mHealth)- and electronic health (eHealth)-related factors were evaluated including (1) the presence of health and wellness apps on a tablet or smartphone, (2) use of electronic devices other than tablets and smartphones to monitor health (eg, Fitbit, blood glucose monitor, and blood pressure monitor), and (3) whether people shared health information from an electronic monitoring device or smartphone with a health professional within the last 12 months. Descriptive and inferential statistics were conducted using SAS version 9.4. Weighted population estimates and standard errors, univariate odds ratios, and 95% CIs were calculated, comparing respondents who used tablets or smartphones to help make medical decisions (n=944) with those who did not (n=1377), separately for each factor. Factors of interest with a *P* value of <.10 were included in a subsequent multivariable logistic regression model.

Results: Compared with women, men had lower odds of reporting that a tablet or smartphone helped them make a medical decision. Respondents aged 75 and older also had lower odds of using a tablet or smartphone compared with younger respondents aged 18-34. By contrast, those who had health and wellness apps on tablets or smartphones, used other electronic devices to monitor health, and shared information from devices or smartphones with health care professionals had higher odds of reporting that tablets or smartphones helped them make a medical decision, compared with those who did not.

Conclusions: A limitation of this research is that information was not available regarding the specific health condition for which a tablet or smartphone helped people make a decision or the type of decision made (eg, surgery, medication changes). In US adults, mHealth and eHealth use, and also certain sociodemographic factors are associated with using tablets or smartphones to support medical decision making. Findings from this study may inform future mHealth and other digital health interventions designed to support medical decision making.

XSL·FO
RenderX

Introduction

As of June 2019, approximately 81% of US adults owned smartphones and nearly 52% of US adults owned tablets according to the Pew Research Center [1]. By race/ethnicity, the proportion of smartphone ownership for White, Black, and Hispanic people was 82%, 80%, and 79%, respectively [1]. By gender, 79% of women and 84% of men owned a smartphone in 2019 [1]. Although smartphone and tablet ownership have increased over the past decade, the degree to which people use mobile devices to help them make medical decisions remains unclear.

A broad goal of informed decision making is to provide individuals (and family members when appropriate and desired) with the amount of understandable, accurate, and balanced information needed to make a high-quality decision about a screening, treatment, or other health-related option (eg, end-of-life issues, circumcision, vaccines, genetic tests) [2,3]. Shared decision making is a related concept to informed decision making and is broadly conceptualized as a collaborative process that allows patients and their providers to make health care decisions together, taking into account the best scientific evidence available as well as patients' values and preferences [4]. Some studies have shown that mobile health (mHealth) and electronic health (eHealth) can improve shared decision-making opportunities and encourage greater patient participation in medical decision making [5-7]. Other studies have shown differences in who uses mHealth and eHealth by age, gender, race/ethnicity, education, and history of health conditions [8-10].

Little is known about how mHealth and eHealth tools may impact medical decision making in nonclinical, population-based samples. A better understanding of this gap in knowledge is important given that strategies to support informed decision making and shared decision making are increasingly being advocated by decision science experts and health care organizations [11-14]. The purpose of this study is to explore factors associated with self-reported use of tablets or smartphones to support medical decision making.

Methods

Brief Overview of the Health Information National Trends Survey

The Health Information National Trends Survey (HINTS) [15] is a probability-based, nationally representative cross-sectional survey of noninstitutionalized US adults aged 18 and over. It is the only national survey exclusively devoted to monitoring trends in health communication and the health information environment. HINTS was developed by the National Cancer Institute's Health Communication and Informatics Research Branch and has been administered approximately every 2 years since 2003. Details about HINTS methodology are reported elsewhere [16,17] and can be seen on the HINTS website [15].

This study evaluated cross-sectional participant data from HINTS 5, Cycle 2. The sample design consisted of a single-mode postal mail survey, using the Next Birthday Method for respondent selection. Data were collected between January and May 2018. There were 3504 responses in the full HINTS 5, Cycle 2 data set; 2321 responses remained after eliminating those with missing data for any factor of interest. A total of 1183 participants were not included in this analysis: 752 respondents were removed due to unusable responses on the main outcome variable including not having a smartphone or device, 102 were removed because they replied "not applicable" to the question regarding sharing health information from an electronic monitoring device or smartphone with a health professional, and 329 were removed due to missing information on at least one of the other factors of interest (eg, demographics or medical information).

Measures

Use of Tablets or Smartphones for Medical Decision Making

The primary outcome was worded as, "Has your tablet or smartphone helped you make a decision about how to treat an illness or condition?" Response options were yes/no.

Use of Electronic Devices Other Than Tablets and Smartphones to Monitor Health

Participants were asked, "In the past 12 months, have you used an electronic wearable device to monitor or track your health or activity? For example, a Fitbit, Apple Watch, or Garmin Vivofit." Response options were yes/no.

Sharing Health Information From an Electronic Device

Participants were asked, "Have you shared health information from either an electronic monitoring device or smartphone with a health professional within the last 12 months?" Response options were yes, no, and not applicable.

Health and Wellness Apps

Participants were asked, "On your tablet or smartphone, do you have any 'apps' related to health and wellness?" Response options were yes, no, and don't know.

History of Medical Conditions

Participants were asked if a doctor or other health professional ever told them that they had any of the following medical conditions (yes/no): (1) diabetes or high blood sugar; (2) high blood pressure or hypertension; (3) heart condition such as heart attack, angina, or congestive heart failure; (4) chronic lung disease, asthma, emphysema, or chronic bronchitis; (5) arthritis or rheumatism; and (6) depression or anxiety disorder. Participants were also asked if they had ever been diagnosed as having cancer (yes/no)? Each medical condition was evaluated individually, as well as "at least one medical condition" compared with none.

Demographics

Demographics such as age in years (median and categorical), gender (male and female), race/ethnicity (White, Black or African American, Hispanic, Asian, and Other), and education (<high school, high-school graduate, some college or trade school, and college graduate) were also assessed.

Data Analysis

Weighted population estimates and standard errors were calculated to describe the sample. Odds ratios (ORs) and 95% CI were calculated, comparing respondents who used tablets or smartphones to help make medical decisions (n=944) with those who did not (n=1377), separately for each factor using univariate logistic regression models. A multivariate logistic regression model for the odds of using tablets or smartphones to make medical decisions was attempted for factors that were univariately significant with $P<.10$. CIs for ORs that do not contain 1 are considered significant for the purpose of this study. All analyses were conducted using SAS version 9.4 (SAS Institute).

Results

Sociodemographic characteristics and univariate ORs are presented in Table 1. Briefly, categorical age, gender, mHealth, and eHealth factors were significantly associated with use of tablets or smartphones for helping participants make medical decisions. However, race/ethnicity, educational attainment, and history of various medical conditions were not significant.

In the multivariate model (Table 2), we found that men had lower odds of reporting that a tablet or smartphone helped them make a medical decision compared with women (OR 0.59, 95% CI 0.42-0.81; $P=.002$). By contrast, participants who had health and wellness apps on tablets or smartphones (OR 1.54, 95% CI 1.06-2.23; $P=.02$), used other electronic devices to monitor health (OR 1.46, 95% CI 1.08-1.97; $P=.01$), and shared information from devices or smartphones with health care professionals (OR 1.87, 95% CI 1.37-2.56, $P<.001$) had higher odds of reporting that their tablet or smartphone helped them make a medical decision, compared with those who did not. Respondents aged 75 and older had lower odds of using a tablet or smartphone compared with younger respondents aged 18-34 (OR 0.38, 95% CI 0.20-0.72).

Table 1. Description of population estimates and univariate odds ratios for using tablet or smartphone to help make decisions about how to treat an illness or condition.[a]

Characteristics	Use tablet/smartphone to help with medical decisions (n=944), weighted percentage (standard error)	Do not use tablet/smartphone (n=1377), weighted percentage (standard error)	Odds ratio (CI)	*P* value
Demographics				
Age (years)				Overall .003
18-34	28.4 (3.4)	26.8 (2.1)	Reference	
35-49	31.2 (2.8)	30.1 (1.8)	0.98 (0.61-1.57)	.93
50-64	31.5 (2.3)	28.8 (1.6)	1.03 (0.67-1.60)	.89
65-74	7.0 (0.7)	9.4 (0.7)	0.70 (0.44-1.11)	.13
75+	2.0 (0.3)	4.8 (0.5)	0.38 (0.22-0.67)	.001
Median age (years)	45.1 (0.96)	46.2 (0.70)	0.99 (0.99-1.00)	.27
Gender				
Female	58.6 (2.4)	44.2 (1.8)	Reference	
Male	41.4 (2.4)	55.8 (1.8)	0.56 (0.41-0.76)	<.001
Race ethnicity				Overall .16
NH[b] White	62.7 (2.2)	68.1 (1.6)	Reference	
NH African American	11.4 (1.2)	8.9 (0.8)	1.40 (0.98-1.99)	.06
Hispanic	17.0 (1.7)	14.7 (1.1)	1.26 (0.87-1.80)	.21
NH Asian	6.4 (1.5)	4.6 (0.8)	1.51 (0.65-3.48)	.33
Other	2.5 (0.7)	3.8 (0.5)	0.73 (0.31-1.68)	.45
Highest level of school				Overall .67
College graduate+	35.0 (1.9)	31.8 (1.3)	Reference	
Some college	41.2 (2.5)	42.6 (1.7)	0.88 (0.66-1.17)	.37
High-school graduate	18.9 (2.3)	19.4 (1.7)	0.88 (0.57-1.36)	.57
Less than high school	4.9 (1.1)	6.1 (1.1)	0.73 (0.38-1.41)	.35
Use of electronic devices other than tablets and smartphones to monitor health				
Yes	47.9 (2.7)	31.3 (2.0)	2.02 (1.48-2.75)	<.001
No	52.1 (2.7)	68.7 (2.0)	Reference	
Share health information from an electronic monitoring device or smartphone with a health professional				
Yes	26.8 (1.9)	13.5 (1.4)	2.36 (1.75-3.18)	<.001
No	73.2 (1.9)	86.5 (1.4)	Reference	
Presence of health and wellness apps on a tablet or smartphone				
Yes	61.6 (2.6)	44.3 (3.0)	2.02 (1.41-2.91)	<.001
No or Don't know	38.4 (2.6)	55.7 (3.0)	Reference	
Medical conditions				
Diabetes				
Yes	14.3 (1.2)	13.5 (1.2)	1.07 (0.81, 1.41)	.63
No	85.7 (1.2)	86.5 (1.2)	Reference	
High blood pressure				

Characteristics	Use tablet/smartphone to help with medical decisions (n=944), weighted percentage (standard error)	Do not use tablet/smartphone (n=1377), weighted percentage (standard error)	Odds ratio (CI)	P value
Yes	30.4 (1.7)	34.4 (2.4)	0.84 (0.63-1.11)	.22
No	69.6 (1.7)	65.6 (2.4)	Reference	
Heart disease				
Yes	4.8 (0.8)	4.5 (0.8)	0.94 (0.53-1.66)	.82
No	95.2 (0.8)	95.5 (0.8)	Reference	
Lung disease				
Yes	10.2 (1.3)	11.6 (1.3)	0.87 (0.58-1.30)	.48
No	89.8 (1.3)	88.4 (1.3)	Reference	
Depression or anxiety				
Yes	26.3 (2.0)	22.0 (2.1)	1.26 (0.92-1.73)	.15
No	73.7 (2.0)	78.0 (2.1)	Reference	
Ever had cancer				
Yes	7.2 (0.9)	8.2 (0.7)	0.87 (0.58-1.31)	.49
No	92.8 (0.9)	91.8 (0.7)	Reference	
Arthritis				
Yes	19.6 (1.8)	16.9 (1.4)	1.20 (0.86-1.68)	.28
No	80.4 (1.8)	83.1 (1.4)	Reference	
One or more of the above medical conditions				
Yes	60.2 (2.7)	60.9 (2.7)	0.97 (0.70-1.35)	.87
No	39.8 (2.7)	39.1 (2.7)	Reference	

[a]Total number of respondents (unweighted n).

[b]Native Hawaiian.

Table 2. Multivariable logistic regression model for using a tablet or smartphone to help make decisions about how to treat an illness or condition.

Variable	Odds ratio (CI)	P value
Age (years)		Overall 0.007
18-34	Reference	
35-49	0.95 (0.58-1.55)	.83
50-64	0.97 (0.60-1.56)	.90
65-74	0.65 (0.39-1.08)	.09
75+	0.38 (0.20-0.72)	.004
Male (vs female)	0.59 (0.42-0.81)	.002
Use of electronic devices other than tablets and smartphones to monitor health (vs no)	1.46 (1.08-1.97)	.01
Share health information from an electronic monitoring device or smartphone with a health professional (vs no)	1.87 (1.37-2.56)	<.001
Presence of health and wellness apps on a tablet or smartphone (vs no or don't know)	1.54 (1.06-2.23)	.02

Discussion

Principal Findings

The goal of this study was to explore factors associated with self-reported use of tablets or smartphones to support medical decision making in a large, nationally representative sample of US adults. In summary, we found that some sociodemographic factors including categorial age and gender, presence of health and wellness apps, and use of mHealth or eHealth tools to monitor health were associated with using a tablet or smartphone to support medical decision making. Although sociodemographic

differences in tablet and smartphone use are well documented [8,18,19], this study is the first-known investigation on the use of tablets or smartphones in the specific context of medical decision making.

Over the last decade, there has been growing use of mHealth and eHealth interventions. Notably, the more contemporary term *digital health* includes categories such as mHealth, health information technology, wearable devices, telehealth and telemedicine, and personalized medicine [20]. Our findings complement prior studies exploring various aspects of mHealth and eHealth to support patients' behavior change and self-management across health conditions—often done via telemedicine, SMS text messaging, or smartphone apps [21-26]—as well as studies showing that tablets and smartphones can help support chronic disease management [18]. Novel uses of mobile devices are being explored to support shared decision via *digital phenotyping* [27] or "moment-by-moment quantification of the individual-level human phenotype in-situ using data from smartphones and other personal digital devices" [28]. While it has been established that people with more than one health condition are more likely to use digital health tools [29,30], we did not observe an association between the number and type of health conditions with the use of tablets or smartphones to support medical decision making. Moreover, we did not observe an effect of race/ethnicity or educational attainment with regard to using tablets or smartphones to support medical decision making.

A key finding from this study was that HINTS participants aged 75 and older had lower odds of using tablets or smartphones compared with participants aged 18-34. While this finding is not surprising given that older US adults generally have lower rates of tablets and smartphones ownership compared with their younger counterparts, it should not be interpreted as unwillingness or inability of people aged 75 or older to use mobile devices. For example, Parker et al. [31] found that 85% of adults in their study aged 60 years or older were willing to try mHealth devices. In that same study, barriers to mHealth use included concerns about cost and unfamiliarity with technology, while facilitators included prior training on how to use mHealth devices and devices tailored to the functional needs of older adults [31]. In a different study, Seifert et al. [32] explored the willingness of adults aged 50 years or older to share mobile health data with researchers and found that approximately 57% were willing to do so.

Moving forward, increased use of mobile devices to support decision making may be realized as more trainings on how to use these tools and resources to lessen the financial burden of getting these devices become available in community and health system settings [33]. Additionally, as future research studies about digital health include more older adults, researchers will have a better understanding of the information needs and digital design preferences of this group. For example, user-centered design processes can help inform how design choices made in the development stages may positively or negatively affect medical decision making in older adults.

With regard to gender, we found that men had lower odds of using tablets or smartphones to support medical decision making compared with women. While somewhat speculative, it is possible that women use tablets and smartphones more often than men to support medical decision making for at least three reasons. First, women may be more likely than men to manage health-related decisions for family members including children, parents, and partners. Second, women may be more likely than men to make and keep routine health care appointments, which in turn, may present more opportunities for medical decision making. Third, women may have more health-related decisions to make due to physiology. For example, women often make choices related to birth control, childbirth, breastfeeding, breast and cervical cancer screening and related treatment, breast reconstruction after a mastectomy, menopause, and uterine fibroids. Not surprisingly, we also found that the use of other electronic devices beyond tablets and smartphones to monitor health (eg, Fitbit, blood glucose meters, and blood pressure monitors), the presence of health and wellness apps on a tablet or smartphone, and sharing information from an electronic monitoring device or smartphone with a health professional were associated with the use of tablets or smartphones to support medical decision making. These findings may be partially explained by the notion that people who are comfortable using mobile devices may have higher levels of eHealth and digital health literacy compared with those who do not use these tools, and are therefore more likely to use mHealth tools to support decision making.

There are several plausible pathways by which tablets or smartphones can help support medical decision making. As shown in Figure 1, we offer 4 potential explanations. It should be noted that our starting point is access to a tablet or smartphone. We specifically use the term *access* to a tablet or smartphone instead of *ownership* because some clinical research projects provide participants with mobile devices as part of a study. Additionally, some people may use another person's mobile device, particularly if they are in the same household.

Figure 1. Conceptualization of How Mobile Devices May Help Support Medical Decision Makingd.
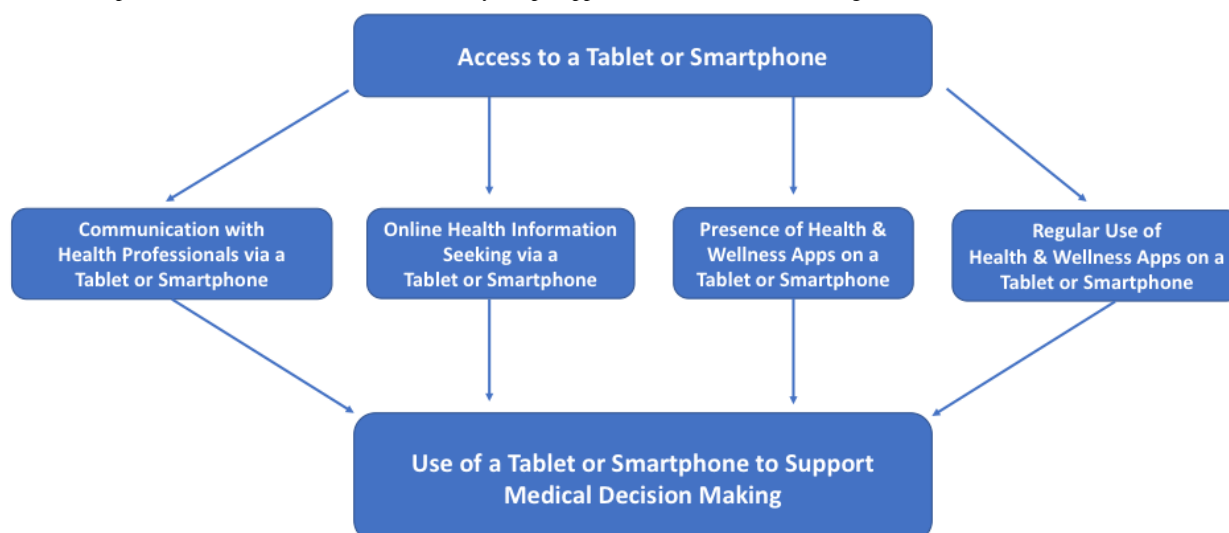
First, it is possible that people use their tablets or smartphones to communicate with health professionals. For example, people may call, email, send SMS text messages, and use apps or other platforms to have virtual face-to-face visits with health professionals. People may also share health information from an electronic monitoring device or smartphone with health professionals. These communication opportunities may help patients get their health-related questions and concerns addressed, which in turn can help them make a medical decision. Second, people may use tablets or smartphones for online health information-seeking purposes [34-38]. In 2016, approximately 36.6 million online users in the US reported that they accessed the internet exclusively via mobile devices according to Statista [39]. Furthermore, in 2019, approximately 74% of US adults reported that they accessed the internet daily via a mobile device according to HINTS data [40]. It is possible that people are using their tablets or smartphones to search for health information online to help them make a medical decision. On the one hand, these online searches may take people to reputable sources of health information such as the Centers for Disease Control and Prevention (CDC), National Institutes of Health (NIH), MedlinePlus [41], disease-specific organizations (eg, American Heart Association), or the Ottawa Patient Decision Aid Inventory [42]. On the other hand, online health information seeking may also expose people to misinformation and biased sources of health information [36]. Nevertheless, regardless of the actual or perceived quality of health information found online, it may impact how individuals make medical decisions for themselves or a loved one. Third, it is possible that the presence of health and wellness apps on a tablet or smartphone may impact how a person makes a decision about how to treat a health condition or illness, even if the app is used infrequently or one time. For example, some patient portals are now available as mobile apps (eg, MyChart in Epic). Through these apps, patients can access their medical records online and see laboratory results, clinical notes, and medication lists. It is possible that seeing a one-time laboratory result is enough to initiate medical decision making. For example, if a person gets a laboratory report showing high triglycerides, they may decide to make lifestyle changes or consider taking prescription medications. Fourth, regular use of health and wellness apps may also support medical decision making. For example, some patients use health and wellness apps regularly to (1) document home readings for conditions such as stage 1 hypertension and type 2 diabetes; (2) track symptoms such as pain, cognitive problems, or disrupted sleep; and (3) monitor health behaviors such as physical activity and diet. The data gathered from regular use of health and wellness apps may inform future decisions about how to treat a health condition or an illness.

Strengths, Limitations, and Future Directions

Strengths of this study include the use of HINTS, a nationally representative survey designed to track health communication and health information technology, and the study's focus on the role of tablets and smartphones for medical decision making. Despite these strengths, our results should be interpreted in the context of both known and potential limitations. For example, we do not know the specific medical decisions that participants were referring to when completing the survey (eg, to pursue surgery, start or change medications, get genetic testing). Based on the wording of the HINTS question about health and wellness apps on a tablet or smartphone, we are not able to distinguish between the simple *presence* of apps and whether or not participants *used* these apps on a regular basis. Finally, while we offered some explanations in Figure 1 regarding plausible ways by which mobile devices may support decision making, we do not know all of the potential mechanisms by which tablets or smartphones may help support medical decision making.

Future research should explore how, if at all, frameworks such as the Ottawa Decision Support Framework [43] and models such as Technology Acceptance Model may help explain our findings [44]. Related, more research is needed to understand the various mechanisms by which electronic wearable devices to monitor health (ie, activity trackers or blood pressure monitors) and the sharing of information from these devices with health professionals may have an impact on subsequent medical decision making and behavioral outcomes such as cancer screening, weight management, and dietary patterns. Future work should also evaluate which attributes of health and wellness apps have the biggest impact on reported use of tablets

and smartphones to support medical decision making. For example, the use of icon arrays and pictographs to convey risk, sixth grade or less level readability of text, use of values clarification methods, and a balanced presentation of health care options are commonly used principles in print and web-based decision aids [45,46]. However, the literature on how to incorporate these principles into health and wellness apps is less developed. Finally, future studies should evaluate the role of digital health literacy and general health literacy in the context of using mobile devices to support medical decision making [47].

This manuscript is especially timely given that it was written in April 2020 as the coronavirus pandemic was spreading throughout the US. Due to social distancing recommendations and widespread cancelations of nonurgent, in-person medical appointments [48,49], there is now even greater attention given to the importance of mHealth and eHealth with regard to (1) conducting telehealth visits and ensuring equitable access to racial/ethnic and other sociodemographic groups, (2) supporting informed decision making about participation in clinical trial for COVID-19 (coronavirus 2019), and (3) shared decision making between patients and clinicians about how to manage chronic conditions for patients who are concerned about being seen in-person as clinics begin to open.

Conclusion

Tablets and smartphones hold promise for supporting medical decision making and may provide new opportunities for facilitating prevention, early diagnosis of diseases, and management of chronic conditions outside of traditional health care settings [20]. Moreover, mobile technologies are prevalent in minority communities and may serve as an important bridge to inclusivity and access to resources, thereby enhancing opportunities for informed and shared decision making [50,51]. Given the growing use of digital health tools available to the general public (eg, health and wellness apps), digital health technology in clinical practice, and focus on supporting informed and shared decision for US adults, data on the role of tablets and smartphones in supporting medical decision making are important to track over time.

Conflicts of Interest

None declared.

References

1. The Pew Research Center. Mobile Fact Sheet. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-08-01]
2. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, International Patient Decision Aids Standards (IPDAS) Collaboration. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. BMJ 2006 Aug 26;333(7565):417 [FREE Full text] [doi: 10.1136/bmj.38926.629329.AE] [Medline: 16908462]
3. Volk RJ, Llewellyn-Thomas H, Stacey D, Elwyn G. Ten years of the International Patient Decision Aid Standards Collaboration: evolution of the core dimensions for assessing the quality of patient decision aids. BMC Med Inform Decis Mak 2013 Nov 29;13(S2). [doi: 10.1186/1472-6947-13-s2-s1]
4. Barry MJ, Edgman-Levitan S. Shared Decision Making — The Pinnacle of Patient-Centered Care. N Engl J Med 2012 Mar;366(9):780-781. [doi: 10.1056/nejmp1109283]
5. Omboni S, Caserini M, Coronetti C. Telemedicine and M-Health in Hypertension Management: Technologies, Applications and Clinical Evidence. High Blood Press Cardiovasc Prev 2016 Apr 12;23(3):187-196. [doi: 10.1007/s40292-016-0143-6]
6. Rhoads SJ, Serrano CI, Lynch CE, Ounpraseuth ST, Gauss CH, Payakachat N, et al. Exploring Implementation of m-Health Monitoring in Postpartum Women with Hypertension. Telemedicine and e-Health 2017 Oct;23(10):833-841. [doi: 10.1089/tmj.2016.0272]
7. Hsu WC, Lau KHK, Huang R, Ghiloni S, Le H, Gilroy S, et al. Utilization of a Cloud-Based Diabetes Management Program for Insulin Initiation and Titration Enables Collaborative Decision Making Between Healthcare Providers and Patients. Diabetes Technology & Therapeutics 2016 Feb;18(2):59-67. [doi: 10.1089/dia.2015.0160]
8. Langford AT, Solid CA, Scott E, Lad M, Maayan E, Williams SK, et al. Mobile Phone Ownership, Health Apps, and Tablet Use in US Adults With a Self-Reported History of Hypertension: Cross-Sectional Study. JMIR Mhealth Uhealth 2019 Jan 14;7(1):e12228. [doi: 10.2196/12228] [Medline: 31344667]
9. Senft N, Everson J. eHealth Engagement as a Response to Negative Healthcare Experiences: Cross-Sectional Survey Analysis. J Med Internet Res 2018 Dec 05;20(12):e11034. [doi: 10.2196/11034] [Medline: 30518513]
10. Reiners F, Sturm J, Bouw LJ, Wouters EJ. Sociodemographic Factors Influencing the Use of eHealth in People with Chronic Diseases. IJERPH 2019 Feb 21;16(4):645. [doi: 10.3390/ijerph16040645]
11. Leng G, Clark CI, Brian K, Partridge G. National commitment to shared decision making. BMJ 2017 Oct 20:j4746. [doi: 10.1136/bmj.j4746]
12. Hargraves I, LeBlanc A, Shah ND, Montori VM. Shared Decision Making: The Need For Patient-Clinician Conversation, Not Just Information. Health Affairs 2016 Apr;35(4):627-629. [doi: 10.1377/hlthaff.2015.1354]
13. Légaré F, Thompson-Leduc P. Twelve myths about shared decision making. Patient Education and Counseling 2014 Sep;96(3):281-286. [doi: 10.1016/j.pec.2014.06.014]

14. Sepucha KR, Simmons LH, Barry MJ, Edgman-Levitan S, Licurse AM, Chaguturu SK. Ten Years, Forty Decision Aids, And Thousands Of Patient Uses: Shared Decision Making At Massachusetts General Hospital. Health Affairs 2016 Apr;35(4):630-636. [doi: 10.1377/hlthaff.2015.1376]

15. National Cancer Institute. Health Information National Trends Survey. 2020. URL: hints.cancer.gov [accessed 2020-08-05]

16. Finney Rutten LJ, Davis T, Beckjord EB, Blake K, Moser RP, Hesse BW. Picking Up the Pace: Changes in Method and Frame for the Health Information National Trends Survey (2011–2014). Journal of Health Communication 2012 Sep;17(8):979-989. [doi: 10.1080/10810730.2012.700998]

17. Nelson D, Kreps G, Hesse B, Croyle R, Willis G, Arora N, et al. The Health Information National Trends Survey (HINTS): Development, Design, and Dissemination. Journal of Health Communication 2004 Sep;9(5):443-460. [doi: 10.1080/10810730490504233]

18. Kim BY, Lee J. Smart Devices for Older Adults Managing Chronic Disease: A Scoping Review. JMIR Mhealth Uhealth 2017 May 23;5(5):e69. [doi: 10.2196/mhealth.7141] [Medline: 28536089]

19. 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. [doi: 10.2196/jmir.5604] [Medline: 28428170]

20. U.S. FoodDrug Administration. Digital Health. 2020. URL: https://www.fda.gov/medical-devices/digital-health [accessed 2020-08-03]

21. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone Apps to Support Self-Management of Hypertension: Review and Content Analysis. JMIR Mhealth Uhealth 2019 May 28;7(5):e13645. [doi: 10.2196/13645] [Medline: 31140434]

22. Lancaster K, Abuzour A, Khaira M, Mathers A, Chan A, Bui V, et al. The Use and Effects of Electronic Health Tools for Patient Self-Monitoring and Reporting of Outcomes Following Medication Use: Systematic Review. J Med Internet Res 2018 Dec 18;20(12):e294. [doi: 10.2196/jmir.9284] [Medline: 30563822]

23. Morita PP, Yeung MS, Ferrone M, Taite AK, Madeley C, Stevens Lavigne A, et al. A Patient-Centered Mobile Health System That Supports Asthma Self-Management (breathe): Design, Development, and Utilization. JMIR Mhealth Uhealth 2019 Jan 28;7(1):e10956. [doi: 10.2196/10956] [Medline: 30688654]

24. Rogers E, Aidasani SR, Friedes R, Hu L, Langford AT, Moloney DN, et al. Barriers and Facilitators to the Implementation of a Mobile Insulin Titration Intervention for Patients With Uncontrolled Diabetes: A Qualitative Analysis. JMIR Mhealth Uhealth 2019 Jul 31;7(7):e13906. [doi: 10.2196/13906] [Medline: 31368439]

25. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015 Sep 07;24(9):CD002098-CD002041 [FREE Full text] [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]

26. Mahmood A, Kedia S, Wyant DK, Ahn S, Bhuyan SS. Use of mobile health applications for health-promoting behavior among individuals with chronic medical conditions. DIGITAL HEALTH 2019 Oct 10;5:205552076198218. [doi: 10.1177/2055207619882181]

27. Panda N, Solsky I, Haynes AB. Redefining shared decision-making in the digital era. European Journal of Surgical Oncology 2019 Dec;45(12):2287-2288. [doi: 10.1016/j.ejso.2019.07.025]

28. Torous J, Kiang MV, Lorme J, Onnela J. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. JMIR Mental Health 2016 May 05;3(2):e16. [doi: 10.2196/mental.5165] [Medline: 27150677]

29. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth Interventions to Support Self-Management of Long-Term Conditions: A Systematic Metareview of Diabetes, Heart Failure, Asthma, Chronic Obstructive Pulmonary Disease, and Cancer. J Med Internet Res 2017 May 17;19(5):e172. [doi: 10.2196/jmir.6688] [Medline: 28526671]

30. 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. [doi: 10.2196/jmir.4883] [Medline: 27185295]

31. Parker SJ, Jessel S, Richardson JE, Reid MC. Older adults are mobile too!Identifying the barriers and facilitators to older adults' use of mHealth for pain management. BMC Geriatr 2013 May 6;13(1). [doi: 10.1186/1471-2318-13-43]

32. Seifert A, Christen M, Martin M. Willingness of Older Adults to Share Mobile Health Data with Researchers. GeroPsych 2018 Mar;31(1):41-49. [doi: 10.1024/1662-9647/a000181]

33. Gordon NP, Crouch E. Digital Information Technology Use and Patient Preferences for Internet-Based Health Education Modalities: Cross-Sectional Survey Study of Middle-Aged and Older Adults With Chronic Health Conditions. JMIR Aging 2019 Apr 04;2(1):e12243. [doi: 10.2196/12243] [Medline: 31518291]

34. Langford A, Loeb S. Perceived Patient-Provider Communication Quality and Sociodemographic Factors Associated With Watching Health-Related Videos on YouTube: A Cross-Sectional Analysis. J Med Internet Res 2019 May 17;21(5):e13512. [doi: 10.2196/13512] [Medline: 31102372]

35. Langford AT, Roberts T, Gupta J, Orellana KT, Loeb S. Impact of the Internet on Patient-Physician Communication. European Urology Focus 2020 May;6(3):440-444. [doi: 10.1016/j.euf.2019.09.012]

XSL•FO
RenderX

36. Loeb S, Sengupta S, Butaney M, Macaluso JN, Czarniecki SW, Robbins R, et al. Dissemination of Misinformative and Biased Information about Prostate Cancer on YouTube. European Urology 2019 Apr;75(4):564-567. [doi: 10.1016/j.eururo.2018.10.056]

37. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: Systematic review and meta-analysis. Journal of Affective Disorders 2019 Feb;245:270-278. [doi: 10.1016/j.jad.2018.11.037]

38. Tan SS, Goonawardene N. Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review. J Med Internet Res 2017 Jan 19;19(1):e9. [doi: 10.2196/jmir.5729] [Medline: 28104579]

39. Statista. Number of mobile-only internet users in the United States from 2015 to 2021. Statista. 2019. URL: https://www.statista.com/statistics/297504/number-of-us-mobile-only-internet-users/ [accessed 2020-08-03]

40. National Cancer Institute. How often do you access the Internet through a mobile device? 2019;. Health Information National Trends Survey (HINTS). URL: https://hints.cancer.gov/view-questions-topics/question-details.aspx?PK_Cycle=11&qid=1493 [accessed 2020-08-03]

41. U.S. National Library of Medicine. MedlinePlus. URL: MedlinePlus.gov [accessed 2020-08-05]

42. Ottawa Hospital Research Institute. Ottawa Patient Decision Aid Inventory. URL: decisionaid.ohri.ca [accessed 2020-08-05]

43. O'Connor AM. Ottawa Decision Support Framework to Address Decisional Conflict. 2006. URL: decisionaid.ohri.ca/docs/develop/ODSF.pdf [accessed 2020-08-05]

44. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the Technology Acceptance Model to Explore User Experience, Intent to Use, and Use Behavior of a Patient Portal Among Older Adults With Multiple Chronic Conditions: Descriptive Qualitative Study. J Med Internet Res 2019 Apr 08;21(4):e11604. [doi: 10.2196/11604] [Medline: 30958272]

45. Stacey D, Légaré F, Lewis K. Decision aids for people facing health treatment or screening decisions. Cochrane Database of Systematic Reviews(4) 2017 Apr 12. [doi: 10.1002/14651858.cd001431.pub5]

46. Witteman HO, Scherer LD, Gavaruzzi T, Pieterse AH, Fuhrel-Forbis A, Chipenda Dansokho S, et al. Design Features of Explicit Values Clarification Methods. Med Decis Making 2016 Jan 29;36(4):453-471. [doi: 10.1177/0272989x15626397]

47. van der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: Measuring a Broad Spectrum of Health 1.0 and Health 2.0 Skills. J Med Internet Res 2017 Jan 24;19(1):e27. [doi: 10.2196/jmir.6709] [Medline: 28119275]

48. Healthcare Facilities: Managing Operations During the COVID-19 Pandemic. Centers for Disease Prevention and Control (CDC). 2020 Jun 28. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/guidance-hcf.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fhealthcare-facilities%2Fguidance-hcf.html

49. Jennifer D. Regional Strategies for Academic Health Centers to Support Primary Care During the COVID-19 Pandemic: A Plea From the Front Lines. Jama Health Forum. 2020 Apr 8. URL: https://jamanetwork.com/channels/health-forum/fullarticle/2764405?resultClick=1 [accessed 2020-08-03]

50. Kathryn Z. Digital differences. The Pew Research Center?s Internet and American Life Project. URL: https://www.pewresearch.org/internet/2012/04/13/digital-differences/ [accessed 2012-04-13]

51. James DCS, Harville C. Smartphone Usage, Social Media Engagement, and Willingness to Participate in mHealth Weight Management Research Among African American Women. Health Educ Behav 2017 Jun 13;45(3):315-322. [doi: 10.1177/1090198117714020]

Abbreviations

CDC: Centers for Disease Control and Prevention
HINTS: Health Information National Trends Survey
NIH: National Institutes of Health
OR: odds ratio

Original Paper

Wearable Device Heart Rate and Activity Data in an Unsupervised Approach to Personalized Sleep Monitoring: Algorithm Validation

Jiaxing Liu[1], MSc; Yang Zhao[1,2], PhD; Boya Lai[1], MSc, PhD; Hailiang Wang[1,2], MPhil, MSc, PhD; Kwok Leung Tsui[1,2], MPhil, PhD

[1]School of Data Science, City University of Hong Kong, Kowloon, China (Hong Kong)

[2]Centre for Systems Informatics Engineering, City University of Hong Kong, Kowloon, China (Hong Kong)

Corresponding Author:
Yang Zhao, PhD
Centre for Systems Informatics Engineering
City University of Hong Kong
Tat Chee Avenue
Kowloon, 999077
China (Hong Kong)
Phone: 852 34425792
Email: yang.zhao@my.cityu.edu.hk

Abstract

Background: The proliferation of wearable devices that collect activity and heart rate data has facilitated new ways to measure sleeping and waking durations unobtrusively and longitudinally. Most existing sleep/wake identification algorithms are based on activity only and are trained on expensive and laboriously annotated polysomnography (PSG). Heart rate can also be reflective of sleep/wake transitions, which has motivated its investigation herein in an unsupervised algorithm. Moreover, it is necessary to develop a personalized approach to deal with interindividual variance in sleep/wake patterns.

Objective: We aimed to develop an unsupervised personalized sleep/wake identification algorithm using multifaceted data to explore the benefits of incorporating both heart rate and activity level in these types of algorithms and to compare this approach's output with that of an existing commercial wearable device's algorithms.

Methods: In this study, a total of 14 community-dwelling older adults wore wearable devices (Fitbit Alta; Fitbit Inc) 24 hours a day and 7 days a week over period of 3 months during which their heart rate and activity data were collected. After preprocessing the data, a model was developed to distinguish sleep/wake states based on each individual's data. We proposed the use of hidden Markov models and compared different modeling schemes. With the best model selected, sleep/wake patterns were characterized by estimated parameters in hidden Markov models, and sleep/wake states were identified.

Results: When applying our proposed algorithm on a daily basis, we found there were significant differences in estimated parameters between weekday models and weekend models for some participants.

Conclusions: Our unsupervised approach can be effectively implemented based on an individual's multifaceted sleep-related data from a commercial wearable device. A personalized model is shown to be necessary given the interindividual variability in estimated parameters.

Introduction

Background

Sleep plays a vital role in maintaining health [1,2]. Adequate sleep can help to maintain a high quality of life [3]. In contrast, short sleep duration may lead to adverse health outcomes, such as obesity, insulin resistance, depression, hypertension, and cardiovascular disease [4-8]. Changes in sleep duration have been associated with declined cognitive function and increased mortality rate in middle-aged population [9,10]. Thus, to detect changes in sleep patterns early on or for sleep disorder diagnosis,

XSL•FO
RenderX

it is essential to measure sleep duration accurately and longitudinally.

Polysomnography is the gold standard for sleep duration and sleep quality assessment; various devices are used to record multiple body functions such as brain activity, eye movements, skeletal muscle movement, and heart rhythm during sleep, and it is typically done in a sleep lab or clinic [11]. Health care professionals use these physiological measures to assess sleep states, however, the cost of overnight PSG may range from US $600 to $5000 each night [12]. Such assessment is expensive and intrusive for consumers, and labor-intensive and resource-demanding for health care professionals to achieve [13-15], making it hard to use for long-term sleep monitoring at home.

An emerging trend has been to adopt sensor-based wearable devices to assess sleep duration and achieve long-term sleep monitoring. According to the reports from the International Data Corporation, 83.8 million wearable devices were shipped during the first two quarters of 2019 [16,17]. One of the most common technologies for consumer sleep-monitoring wearable devices is accelerometer-based actigraphy [18], which tracks physical movement and determines when a person is asleep or awake based on whether a low or high amount of activity is recorded [19,20].

Related Work

A recent paper [21] reviewed and validated existing supervised sleep-scoring algorithms using actigraphy in a large cohort. Many types of supervised algorithms have been used, such as linear discriminant analysis [22,23], decision trees [24], artificial neural networks [24], long short-term memory [21], and convolutional neural networks [21]; however, all require training with annotated PSG, collected and labeled at great expense. Moreover, the labeled daily-living sleep/wake PSG data are challenging to collect during daytime making 24-hour accuracy hard to evaluate.

Because annotated PSG is hard to acquire for model building, it is intuitive to use an unsupervised method. Commonly used techniques are based on rules and thresholds. For instance, the Actiwatch (Mini Mitter Co Inc) software determined the start of sleep when there were 10 consecutive minutes below a certain mobility threshold and determined the end of sleep with 10 consecutive minutes above the threshold [25]. This approach was commonly adopted in commercial wearable devices because of its simplicity; however, the choice of the mobility threshold was relatively arbitrary. Few studies have used other unsupervised machine learning approaches such as clustering [15] and hidden Markov model (HMM) [26].

Hidden Markov models are well suited to modeling time series data in a probabilistic way by using latent states [27]. Temporal dependency can be learned, and the parameters of the model are interpretable. In the context of bioinformatics, HMMs have been applied to monitor circadian rhythmicity using physical activity data to characterize interindividual variability [28]. In other high-frequency physiological data collected during PSG—such as electroencephalography, electrooculography, and electromyography—HMMs have been used to classify sleep stages [29,30]. These papers showed that the sleep/wake cycle could be inferred from physiological data using HMMs; therefore, we wanted to extend this approach and its strengths to activity and heart rate data from wearable devices.

In addition to modeling sleep/wake transitions via activity data, heart rate is also reflective of the circadian cycle, and it can be recorded accurately using photoplethysmography in wearable devices [31]. In a 24-hour assessment study in ambulatory patients, heart rate varied significantly in sleep and wake periods [32]. Willemen et al [33] showed that heart rate, along with activity, can predict sleep and wake well with the use of support vector machine algorithms on healthy participant data; however, this approach also required supervised training (ie, labeled PSG).

Furthermore, the generalizability of existing algorithms is of concern especially when different sleeping patterns and habits can be observed in different people [15]. Person-to-person differences in demography and physiology can have significant effects on sleep/wake patterns and characteristics [34,35]. Personalized sleep-scoring algorithms are needed to avoid interindividual variance problems, since algorithms learn from individual lifestyle and physiological patterns using long-term contextual history. Existing studies have shown that the sleep/wake state can be better inferred using a personalized approach from actigraphy [36,37].

Objectives

In this paper, we aimed to explore the feasibility of using HMM to analyze heart rate and activity data collected by a wearable device and to develop a personalized and unsupervised sleep/wake identification approach. To our knowledge, there is little research focused on personalized and unsupervised sleep/wake identification algorithms using a wearable device. Also, the approach enables advantageous complementary use of both heart rate and activity data. The algorithm is demonstrated using a real-world data set collected with commercial wearable devices in the older adult population and its performance is illustrated with case studies.

We describe recruitment and data collection, data preprocessing steps, HMM, rescoring rules, and comparison scoring results. We also demonstrate the approach with a case study, investigate the fusion effect of heart rate and activity data in modeling, compare our scoring results with Fitbit's results individually, and investigate pattern changes using daily models.

Methods

Participant Recruitment and Data Collection

We collaborated with an older adult care center in Hong Kong to recruit participants who met the following criteria: aged 60 years or above, community-dwelling Hong Kong residents, and willing to take participate in a 3-month longitudinal observational study from December 18, 2017 to February 28, 2018. The Research Ethics Committee of the City University of Hong Kong approved this pilot study (reference number 3-2-201803_02). All participants provided written consent.

Heart rate data and activity data were collected using Fitbit Alta (Fitbit Inc). Participants were asked to wear the device on their

nondominant hand for the full 24 hours each day for 3 consecutive months. For activity data, the most common choice for use in sleep/wake classification algorithms is activity count. However, since activity count was unavailable in Fitbit, we used step count instead, which was also a reflection of activity intensity in an epoch. Fitbit Alta reported heart rate every 1 minute and reported step count every 15 minutes.

Data Preprocessing

Since the study was conducted in a free-living home environment, some participants removed the devices when showering or at night. When the participant removed the device, the device reported heart rate and step count as zero, and the nonwear time could be inferred. The recordings were examined, and nonwear days were identified and removed before analysis if (1) more than 30 minutes of heart rate data were in that day or (2) there was a step count of zero on that day. We excluded participants who had more than 50% of the days in the observation period identified as nonwear from analysis.

After the elimination of nonwear days, any remaining data with missing values could be kept and dealt with in HMM. Next, we further preprocessed the step count data. In order to facilitate the fusion of step count and heart rate data in the models, downscaling was used to deal with the multigranularity data [38]. It was achieved by disaggregating the 15-minute step count data and simulating the of 1-minute step count time series. We assumed that the 15-minute step count U^{STEP} was evenly distributed to every minute. Thus, 1-minute step counts were generated by

(**1**)

The total step count in 15 minutes was closely preserved.

Hidden Markov Models

Definitions

Hidden Markov models are composed of paired stochastic variables: hidden states and observed variables. The model assumes that an observed sequence has been generated by distributions, which conditionally depend on the hidden states in an underlying and unobserved Markov process. In our sleep/wake identification problem, we considered two-state hidden Markov models. The hidden states were $S=\{s_1, s_2\}$. Each observed bivariate time series (of heart rate and activity data) was denoted as

(**2**)

where $t \in \{1,...,T\}$ and T was the total length. The two-state chain was initialized by the initial state distribution, $\pi=\{\pi_1, \pi_2\}$ where $\Sigma\pi_k=1$. The sequence of hidden states was $\mathbf{Z_T}=\{z_1,z_2,...,z_T\}$, where $z_t \in S$ for any t. The structure of a standard multivariate HMM is shown in Figure 1.

The unobserved process was assumed to satisfy the Markov property. The transition probability matrix was denoted by Γ as

(**3**)

whose (i,j) entry represented the probability of state s_i transitioning to state s_j:

$$P(z_{t+1}=s_j \mid z_t=s_i) = \gamma_{ij}, \; s_i, s_j \in S \; (4)$$

The emission density function

$$p(x_t \mid Z_t=s_i) \; (5)$$

was associated with hidden states, which denoted the density of the observation $\mathbf{X_T}$ if the hidden state was s_i at time t. For multivariate time series, there were 2 schemes to develop HMMs: (1) Specify the state-dependent joint distributions of the observed variables for different states or (2) assume contemporaneous conditional independence.

Figure 1. An illustration graph of the structure of multivariate hidden Markov model.

Model Scheme M1: Specification of the State-Dependent Joint Distributions

In the multivariate case, it can be straightforward to specify the joint distribution in the context of our application, since heart rate is highly correlated with activity intensity. The bivariate normal distribution was considered because of its practical uniqueness. Thus, the emission density function could be written as follows if we assumed a bivariate joint distribution for $\mathbf{x}_t \mid s_i$:

(**6**)

with

(**7**)

The correlation between heart rate and activity level can be directly characterized by ρ_i.

Model Scheme M2: Contemporaneous Conditional Independence

Specifying suitable joint distributions can be sometimes difficult, and for simplicity, contemporaneous conditional independence can be assumed. This means that the state-dependent joint distribution is the product of the corresponding marginal distributions:

(**8**)

Note, contemporaneous conditional independence does not mean the two observed time series are mutually independent since the Markov chain can induce dependent pairs [27], and the marginal distributions need not necessarily belong to the same family of distribution. Thus, we can assume the univariate distributions in different states according to prior information. The choices of distributions are discussed in our real-world case study.

Model Fitting and Decoding

After the model was fully specified. The likelihood could be obtained by summing the values assumed by $z_1, z_2, ..., z_T$:

(**9**)

The likelihood function was evaluated by the forward algorithm. An advantage of the HMM is that missing data can be dealt with simply by adjusting the likelihood computation. The corresponding state-dependent probabilities were replaced by 1 for all states. Parameter estimation was achieved by numerical maximization or Baum-Welch algorithm [27]. Next, we decoded the series of the hidden states globally by maximizing the conditional probability of the whole sequence $P(\mathbf{Z}_T \mid \mathbf{X}_T)$. The optimal path was found using the Viterbi algorithm with estimated parameters. The hidden states were matched with

sleep and wake states. The state with higher estimated mean heart rate, μ_{HR}, and mean activity level, μ_{ACT}, represented the active status of the participants, and related to a waking state. On the other hand, the state with lower estimated mean heart rate and mean activity level represented the resting status, which was related to a sleeping state (heart rate variance, σ^2_{HR}, and activity level variance, σ^2_{ACT}).

Implementation of the Hidden Markov Models

After preprocessing, the data were ready for implementation of the HMMs. For each participant, we plotted the kernel density estimates of heart rate and of log-transformed step count (Figure 2) to explore suitable prior joint emission distributions. From the kernel density plot of heart rate, the overdispersed and nonsymmetric observations suggested a bimodal distribution, which may be modeled by a mixture model, and very likely a two-component Gaussian mixture; however, mixture models do not take temporal dependency into account, which prompted us to adopt the hidden Markov model. A Poisson distribution is a natural choice for modeling count data; however, the step count ranged from 0 to 160. Since it would have been computationally expensive to estimate, especially for our long sequence, we took the log transformation of step count and found that they were also distributed marginally as a two-component Gaussian mixture.

Based on the marginal density plots, we proposed two schemes for fitting multivariate HMM. For model scheme M1, we specified bivariate Gaussian distribution for heart rate and $\log(X^{STEP}+1)$ for both states; for model scheme M2, we assumed univariate Gaussian distribution for heart rate for both states and univariate Gaussian distribution for log-transformed step count for both states.

Two HMMs were fitted for each participant, one HMM modeled using model scheme M1 and another using model scheme M2. The best model was chosen based on Akaike information criterion (AIC) and Bayesian information criterion (BIC). Moreover, the goodness of fit of the best model was further assessed by ordinary normal pseudo-residuals [27]:

(**10**)

If the observations $\mathbf{x}_1, ..., \mathbf{x}_T$ were indeed generated by the model $\mathbf{X}_t \sim \mathbf{F}_t$, the ordinary normal pseudo-residuals would be distributed standard normal.

With the estimated parameters for the best model, we found a globally optimal path for the observations. The decoding of the hidden states was passed to rescoring rules, and the final scoring results of sleep and wake states were decided. The implementation of HMM was built based on the dependent mixture models package (depmixS4; version 1.4-2) in R software (version 3.5.1) [39]

To study the fusion effect of heart rate and activity in HMM, two HMM models observing single source data was considered for comparison. We fitted a two-state hidden Markov model observing activity level X^{ACT}, which was $\log(X^{STEP}+1)$, denoted

as the activity HMM, and a model observing heart rate denoted as the heart rate HMM. In both the activity HMM and the heart rate HMM, the emission distributions for both states were assumed to be normal distributions according to the empirical data analysis.

**Figure 2.** Kernel density plots for heart rate and log(X_STEP+1) for all participants.
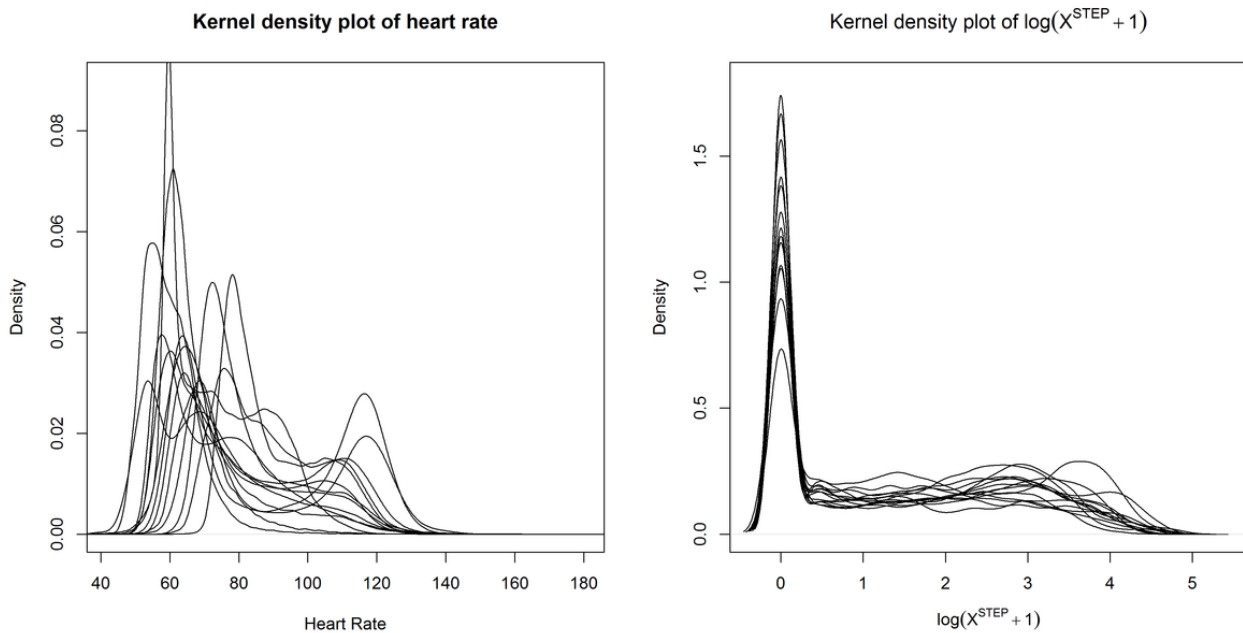
Rescoring Rules

In actigraphy algorithms, Webster et al [40] reported that the most common error was scoring wake as sleep. In order to correct this systematic error, they developed these rescoring rules, which were further validated by different researchers [21,22]. Webster rescoring rules can be described as (1) after at least 4 minutes scored as wake, the first minute scored sleep will be rescored wake; (2) after at least 10 minutes scored as wake, the first 3 minutes scored sleep will be rescored wake;

(3) after at least 15 minutes scored as wake, the first 4 minutes scored sleep will be rescored as wake; (4) 6 minutes or less scored sleep surrounded by at least 10 minutes (before or after) scored as wake are rescored wake; and (5) 10 minutes or less scored as sleep surrounded by at least 20 minutes (before or after) scored as wake are rescored wake. These rules were applied to 1-minute decoding results from HMM sequentially.

The workflow of our sleep/wake identification approach can be summarized in Figure 3.

Figure 3. The workflow of our hidden Markov model–based sleep/wake identification approach. S/W: sleep/wake.
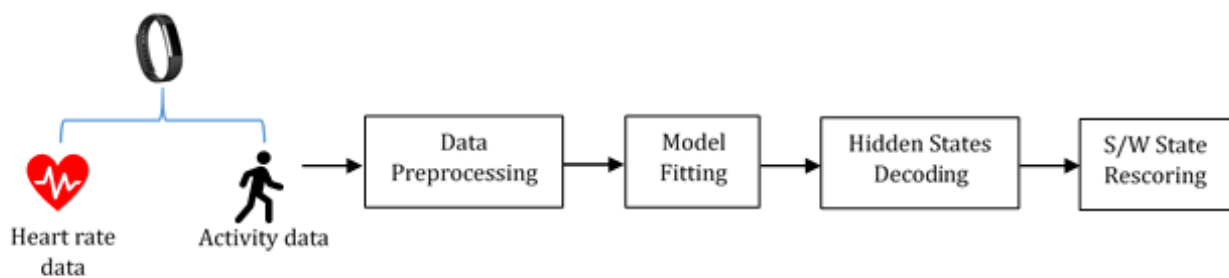
Comparison of Scoring Algorithms

Fitbit Alta automatically detects sleep based on activity—"When your body is completely at rest and you haven't moved for about an hour, your Fitbit device records that you're asleep" [41];—however, the exact scoring algorithm is proprietary. Only the sleep/wake scoring results can be compared with Fitbit's output (7 sleep-related states). For comparison, *asleep*, *deep*, *light*, and *REM* were reclassified as *sleep*, while *restless*, *awake*, and *wake* were reclassified as *wake*.

Daily Basis Model

In addition to applying the proposed model to 3-month time periods for each participant, as a pilot experiment, we applied the personalized algorithms on a daily basis for each person. To investigate whether there was any difference in estimated parameters between weekdays (Monday to Friday) and weekends (Saturday and Sunday), we used two-tailed independent t tests to compare the parameters of the two for each participant ($P<.05$ was deemed significant).

Results

Overview

After nonwear exclusion, there were 14 participants whose data qualified for analysis (aged from 61 to 91 years old; 12 women and 2 men); 6 participants had hypertension, 5 had high cholesterol, 2 had diabetes mellitus, 3 had cancer, and 1 had a stroke. The percentage of missing heart rate data ranged from 0.31% to 0.96% (mean 0.64%). Examples of 24 hours of step count and heart rate data are shown in Figure 4 in which the circadian cycle was quite clearly evident, and the segmentation of activity and heart rate over the 24 hours was visible.

Figure 4. A 24-hour example plot of step count every 15 minutes and heart rate every 1 minute for participant EL01 from 8 AM to 7:59 AM the following morning. Note: Times are in 24-hour format.
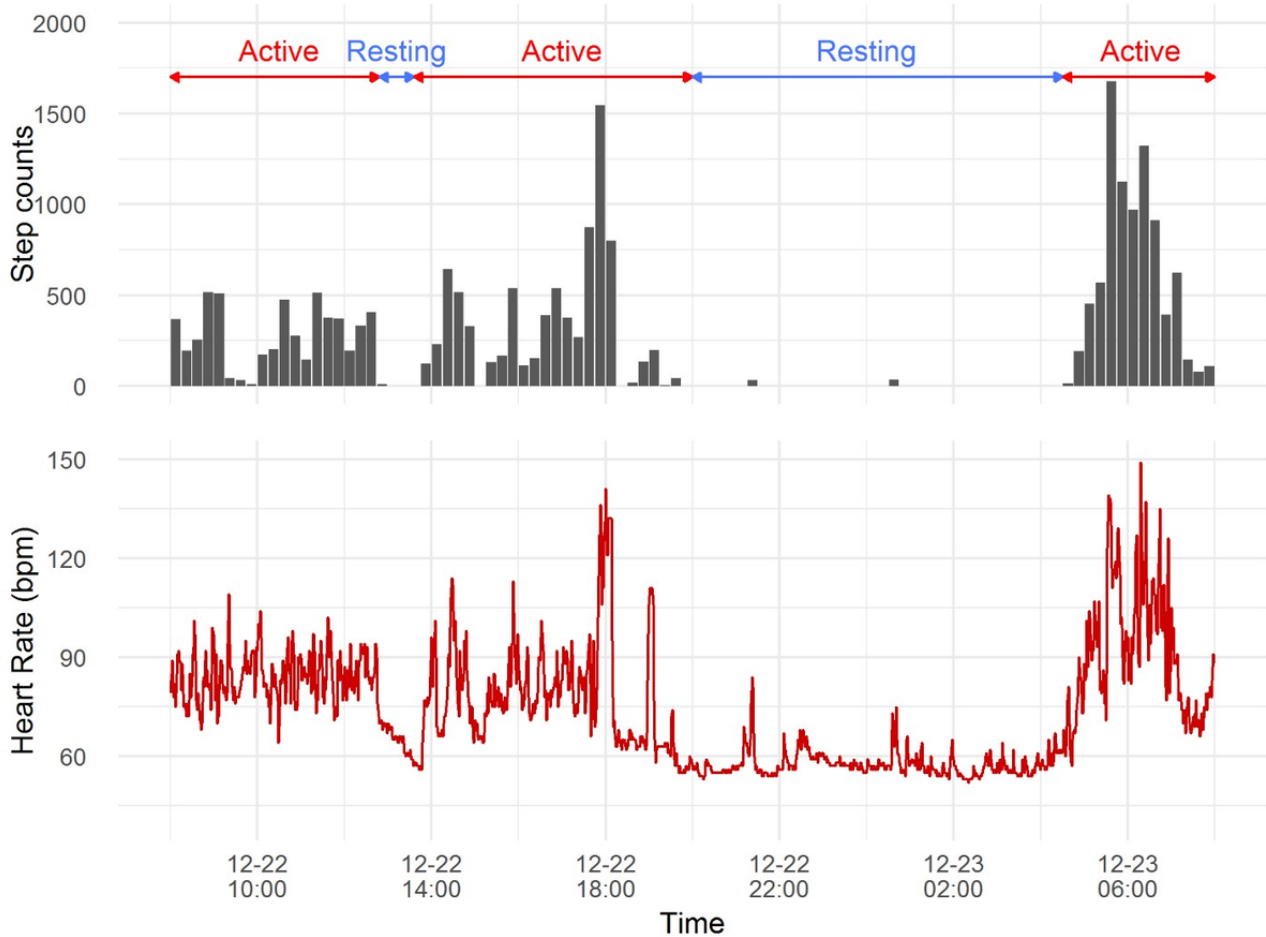
Model Selection and Parameter Estimation Results

In order to illustrate our proposed approach, we used the sleep and wake identification of two typical examples, participants EL02 and EL21, for demonstration. We first examined the recordings to eliminate nonwear days. There were 16 days were eliminated from analysis for EL02 and 9 days were eliminated from analysis for EL21.

To compare two schemes of modeling multivariate time series, we present model selection results, including log likelihoods, AIC, and BIC in Table 1. According to both AIC and BIC, the model with bivariate normal joint distribution specified for both states was more appropriate for EL02. For the other 13 participants, both AIC and BIC also tended to favor model scheme M1, which assumed bivariate normal emission distributions for both states.

Table 1. Comparison of models fitted to heart rate and log(XSTEP+1) for EL02.

Model scheme	Emission distribution	df	Log likelihood	AIC	BIC
M1	Bivariate normal	13	–419428.3	838882.7	839004.3
M2	Conditional independence	11	–424798.5	849721.9	849721.9

We checked the general goodness of fit of the best model with ordinary normal pseudo-residuals. If the fitted model was valid, the pseudo-residuals would be distributed normally. By visual inspection of the quantile-quantile (QQ) plot of the pseudo-residuals for each participant's best model, we found they fit well. We present QQ-plots for EL02 and EL21 in Figure 5; we could observe that for EL02 and EL21 the pseudo-residuals were, in general, distributed normally, though

XSL•FO
RenderX

the distribution had heavier tails for EL21. Overall, our 2-state HMM model assuming bivariate Gaussian was adequate and valid for the participants in our study.

Figure 5. Quantile-quantile plots of ordinary normal pseudo-residuals in model scheme M2 for EL02 and EL21.
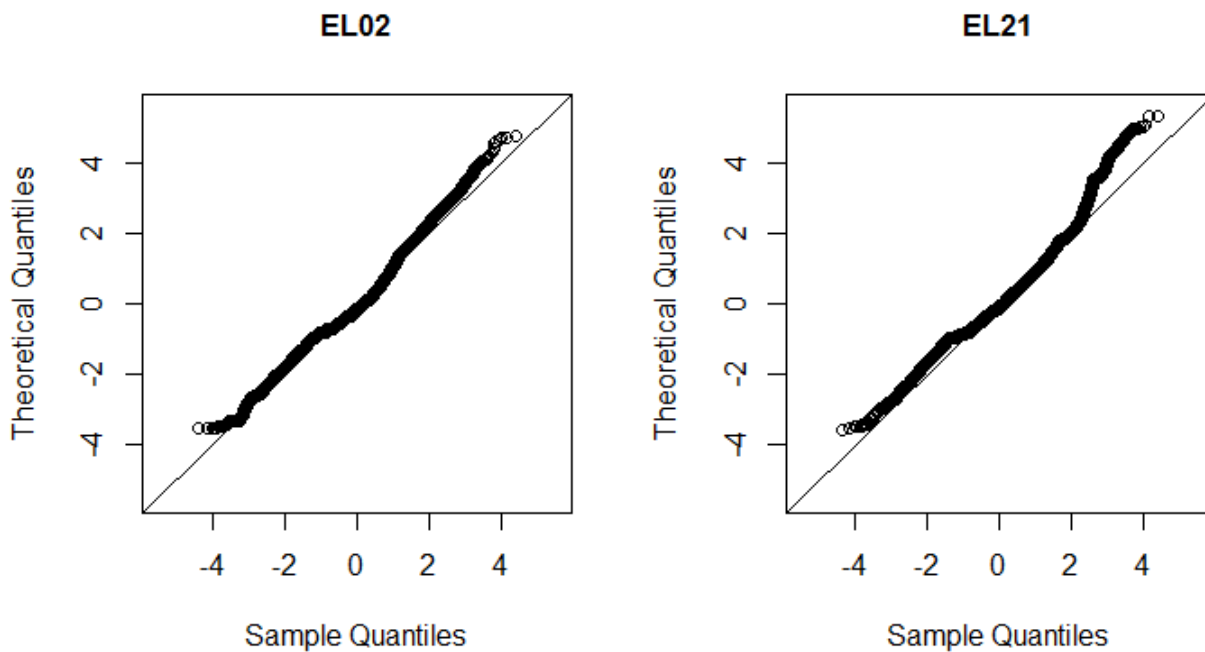
The estimated parameters in emission distributions for EL02 are shown in Table 2 for the best model chosen by AIC and BIC. According to the estimated parameters for emission distribution in different states, we can generally classify the 2 hidden states as sleep and wake. Wake was the state with higher mean heart rate and mean activity level in the emission density distribution. In the wake state, the estimated variance for heart rate was 213.68 and for activity level was 0.83, which were much larger than those of 157.37 for heart rate and 0.15 for activity level during the sleep state. This reflected the variability of activities during the participant's waking period. The estimated transition probability for EL02 was as follows:

From the estimated transition probability, it was more likely that, in a given minute, the participant would stay in the same state as the state in the previous minute. It also supported the necessity of the use of HMM to deal with the time dependence in observations.

For participant EL21, the estimated parameters in emission density distributions are also shown in Table 2. The transition probability for EL21 was estimated as

The general pattern of estimated parameters for EL21 was similar to EL02. While the difference between estimated mean values in the two states were much lower than that of EL02. For the sleep state, the activity level was roughly zero (μ_{ACT} <0.01). For the wake state, the heart rate (μ_{HR}=80.17) and the activity level (μ_{ACT}=2.22) were lower than those for EL02, showing that the mean intensity of activity of EL21 was lower than EL02. In terms of transition probability, was 0.0082 for EL21 and 0.0097 for EL02, which might suggest that for participant EL21, it was slightly harder to fall asleep when awake than for EL02.

Since the models were fitted individually, we report the estimated parameters for 14 individuals in Table 3. For the wake state, mean was 87.18 (SD 12.52), while mean for the sleep state was 66.37 (SD 7.82). The difference in variance in the estimated means in two states among individuals potentially reflected the person-to-person diversity in lifestyle and heart rhythm. Moreover, the individual-specific parameters helped characterize the sleep/wake cycle.

Table 2. Estimated parameters in model M2 for EL02 and EL21.

Participant and emission parameters	Wake state	Sleep state
EL02		
Activity		
μ_{ACT}	2.98	0.27
σ^2_{ACT}	0.83	0.15
Heart rate		
μ_{HR}	110.21	74.39
σ^2_{HR}	213.68	157.37
Correlation, ρ	0.41	0.04
EL21		
Activity		
μ_{ACT}	2.22	<0.01
σ^2_{ACT}	1.08	<0.01
Heart rate		
μ_{HR}	80.17	56.46
σ^2_{HR}	276.27	58.56
Correlation, ρ	0.56	0.01

Table 3. Mean estimated HMM parameters for the sample of participants.

Parameters	Wake state, mean (SD)	Sleep state, mean (SD)
Emission		
Activity		
μ_{ACT}	2.24 (0.29)	0.02 (0.07)
σ^2_{ACT}	1.16 (0.19)	0.01 (0.04)
Heart rate		
μ_{HR}	87.18 (12.52)	66.37 (7.82)
σ^2_{HR}	241.16 (110.75)	47.18 (36.91)
Correlation, ρ	0.54 (0.07)	0.01 (0.01)
Transition		
Wake state	0.989 (0.003)	0.011 (0.003)
Sleep state	0.017 (0.002)	0.983 (0.002)

Investigating the Effect of the Fusion of Heart Rates and Activity in the Model

We compared the final scoring results from approaches based on the heart rate HMM, the activity HMM and our fusion approach minute-by-minute for each participant. Table 4 presents duration, heart rate, and activity level in different combinations of possible results for EL02 as an example.

For EL02, 49.30% (42,599/86,400 minutes) of the recordings were scored as wake, and 33.87% (29,264/86,400 minutes) were scored as sleep by all three methods, which indicated the monotonic relationship between heart rate and activity level for most of the time, whether sleep or wake. There were 13.42% (11,593/86,400) that changed states by our fusion approach compared to that indicated using only one data-source type. Our approach rarely (2/86,400, <0.001%) scored one minute as sleep state if either the heart rate HMM or the activity HMM had classified the minute as wake state. The determination of sleep state in our approach was a combination of activity and heart rate for EL02.

Table 4. Comparison between the activity HMM, the heart rate HMM, and our fusion approach for participant EL02.

HMM			Comparison		
Activity only	Heart rate only	Fusion	Duration (total minutes=86,400), n (%)	Heart rate, mean (SD)	Activity level, mean (SD)
wake	wake	wake	42,599 (49.30)	113.37 (11.28)	3.13 (0.91)
wake	wake	sleep	0 (0)	—	—
wake	sleep	wake	2825 (3.27)	76.12 (8.94)	2.14 (0.96)
wake	sleep	sleep	1 (0.00)	74 (—)	0.76 (—)
sleep	wake	wake	8768 (10.15)	101.21 (13.74)	0.82 (0.67)
sleep	wake	sleep	1 (0.00)	96 (—)	0.38 (—)
sleep	sleep	wake	2942 (3.41)	80.72 (8.37)	0.97 (0.68)
sleep	sleep	sleep	29,264 (33.87)	69.62 (5.10)	0.28 (4.24)

For recordings that the activity HMM scored as wake, the heart rate HMM scored as sleep, and our approach scored as wake, the mean heart rate was 76.12, and the mean activity level was 2.14 (equivalent to 7.5 steps per minute). The nontrivial activity level led our approach to score that minute as wake. For recordings that the activity HMM scored sleep, the heart rate HMM scored wake, and our approach scored as wake, the mean heart rate was 101.21, and the mean activity level was 0.82.

Furthermore, we present an example of 24 hours of scoring results from the three model types along with observational data for EL02 in Figure 6. The bars below the observations indicated the scored sleep or wake states in three models. From the highlighted period in Figure 6, we can see that if we used activity data alone, it was very likely to be classified as sleep due to the extremely low activity level. However, the high and fluctuating heart rate might suggest the person was awake.

Figure 6. An example plot of observations and scoring results from Heart Rate HMM, Activity HMM, and our approach for EL02. HMM: hidden Markov model.
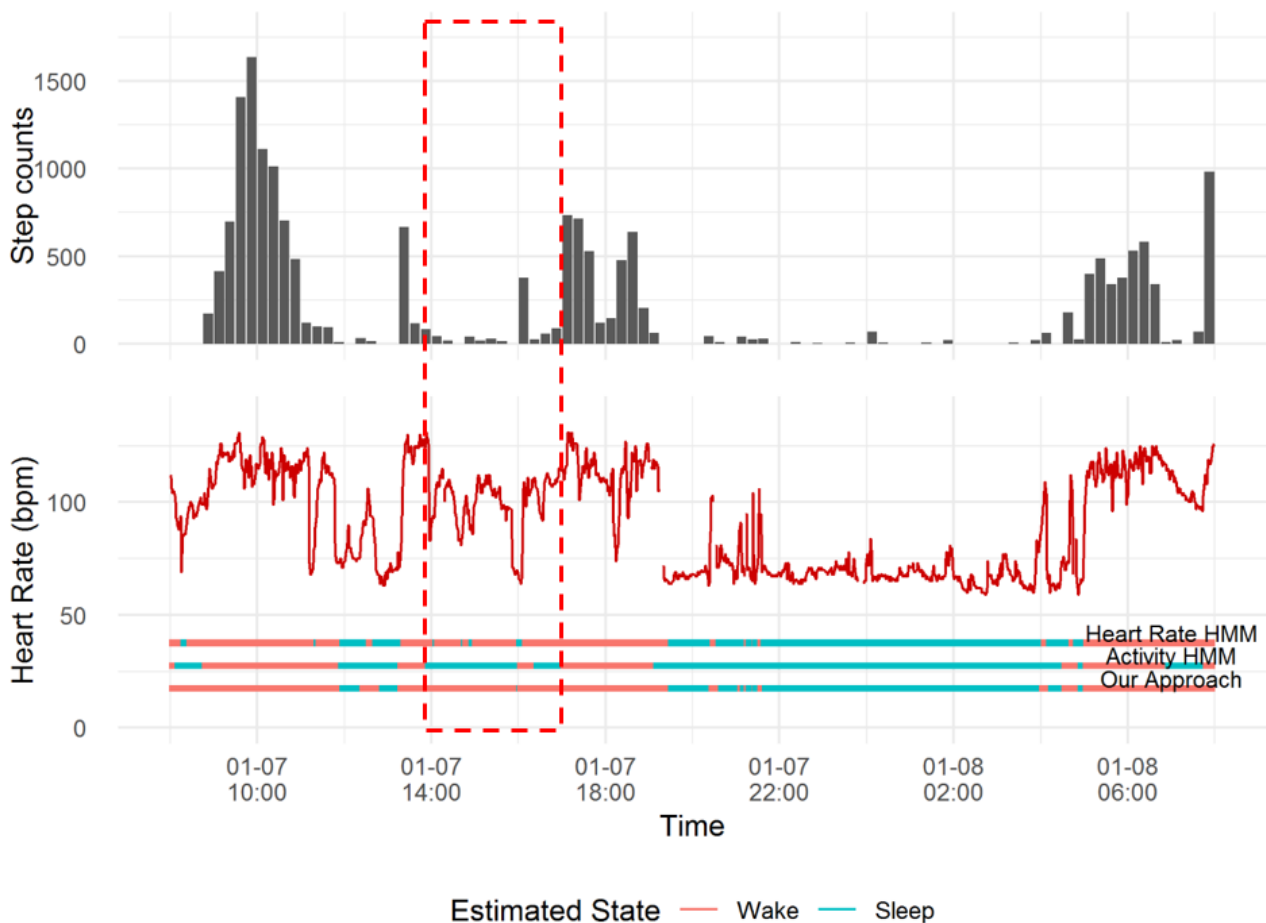
Except for epoch-by-epoch comparison, we also evaluated the performance on a crucial sleep metric, the total amount of sleeping time in the recording period [42]. For all participants, we calculated daily total sleep time at night for all available days. The nighttime sleep period of each participant was collected with the Pittsburgh Sleep Quality Index indicating when they usually went to bed at night and got up in the morning [43]. Figure 7 displays the boxplot of estimated total sleep time

during their bedtime for all participants. We could see that the median estimated total sleep time at bedtime varied a lot from person-to-person. For EL01, EL02, EL06, EL24, and EL25, the median total sleep time estimated by our approach was less than those from the heart rate HMM and the activity HMM. For the other participant, they were nearly the same as the activity HMM.

Figure 7. Boxplot of estimated daily total sleep time during bedtime at night from heart rate HMM, activity HMM, and fusion HMM (our approach) for all participants. HMM: hidden Markov model; TST: total sleep time.
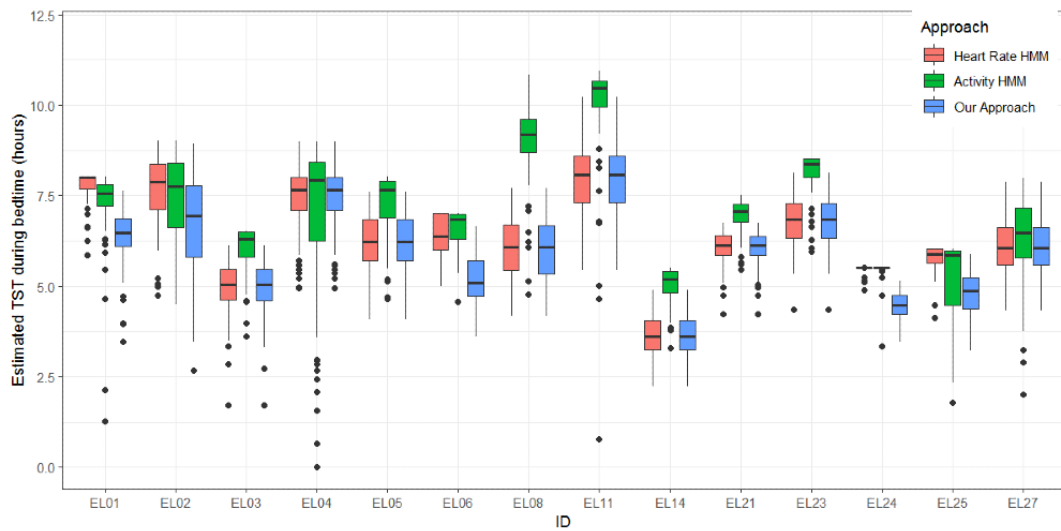
Comparison With Fitbit's Sleep-Wake Scoring Results

We compared results from our approach with those from Fitbit's scoring algorithm minute-by-minute. We treated Fitbit's sleep/wake scoring algorithm as representative of existing methods. The mean agreement between our approach and Fitbit's scoring was 87.31%, (range 82.90% to 91.04%). As for total sleep time, the boxplot of the estimated total sleep time at night using our approach and Fitbit's algorithm are displayed

in Figure 8. There were several days where Fitbit's had no sleep/wake records but had continuous regular heart rate recordings. The median estimates of total sleep time from our approach were lower than those from Fitbit in 12 of 14 participants, which indicated our approach tended to score more wake epochs. The dispersion of estimated sleep duration using our approach was relatively narrower than that of Fitbit's algorithm, which may suggest a stable indicator of participants' habitual nighttime sleep duration.

Figure 8. Boxplot of estimated daily total sleep time during bedtime at night from our approach and Fitbit's approach for all participants. HMM: hidden Markov model; TST: total sleep time.
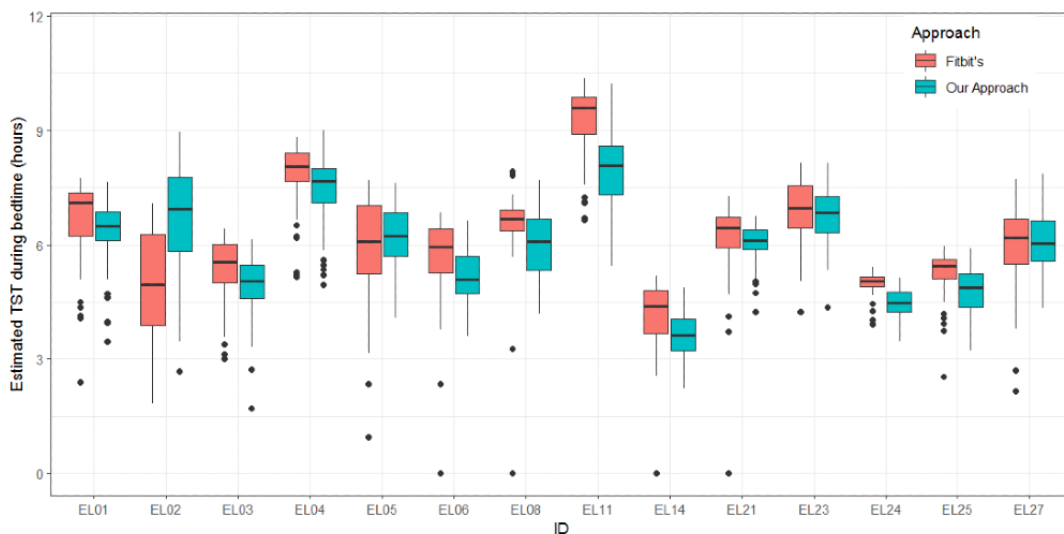
Investigating Pattern Changes Using Daily Basis Model

The P values for independent two-tailed t tests comparing estimated parameter values between weekdays and weekends for each person are shown in Table 5. There was no significant difference for γ_{21}, σ^2_{HR} for wake state, and μ_{ACT} for sleep state between weekends and weekdays for any participant. As shown in Table 5, 5 of 14 participants had no significant difference in

any of the estimated parameters between weekday models and weekend models, while 9 participants had differences in at least one parameter. For example, EL04 had a significantly lower transition probability from wake to sleep (0.015 for weekdays; 0.018 for weekends). EL11 had a higher mean estimated variance of activity level in wake state on weekdays (1.49 for weekdays; 1.29 for weekends).

Table 5. P values for t tests on estimated parameter values for weekday and weekend for each participant.

ID	P values											
	Transition[a]		Wake state					Sleep state				
			Activity		Heart rate		Corr[b]	Activity		Heart rate		Corr
	γ_{12}	γ_{21}	μ_{ACT}	σ^2_{ACT}	μ_{HR}	σ^2_{HR}	ρ	μ_{ACT}	σ^2_{ACT}	μ_{HR}	σ^2_{HR}	ρ
EL01	.28	.89	.30	.66	.22	.72	.04[c]	.94	.33	.34	.26	.83
EL02	.64	.94	.90	.85	.44	.99	.74	.90	.54	.63	.55	.86
EL03	.38	.55	.14	.96	.18	.25	.17	.16	.09	.10	.08	.02[c]
EL04	.047[c]	.70	.53	.31	.51	.63	.68	.78	.50	.82	.53	.08
EL05	.45	.58	.64	.48	.91	.39	.69	.66	.74	.70	.03[c]	.68
EL06	.20	.95	.59	.05	.96	.14	.92	.33	.59	.92	.39	.09
EL08	.39	.46	.12	.23	.11	.73	.93	.96	.06	.12	.61	.53
EL11	.70	.70	.63	.002[c]	.60	.16	.17	.18	.42	.73	.43	.63
EL14	.58	.11	.84	.04[c]	.34	.50	.90	.56	.36	.58	.78	.75
EL21	.008[c]	.07	.86	.79	.006[c]	.94	.67	.14	.07	.008[c]	.04[c]	.006[c]
EL23	.97	.85	.52	.56	.56	.62	.54	.28	.43	.32	.62	.63
EL24	.97	.98	.08	.55	.002[c]	.38	.54	.07	.048[c]	.06	.33	.02[c]
EL25	.19	.76	.55	.47	.10	.21	.54	.28	.50	.69	.89	.38
EL27	.66	.54	.009[c]	.08	.30	.44	.01[c]	>.999	.54	.73	.046[c]	.32

[a]γ_{11} and γ_{22} are not reported because they have the same results as γ_{21} and γ_{21}.

[b]Corr denotes correlation.

[c]Value is significant $P<.05$.

Discussion

Principal Findings

Longitudinal monitoring of sleep duration can objectively help detect sleep disorders and reduce the risk of related diseases. In order to facilitate personalized home-based monitoring, it is essential to record sleep and wake states using wearable devices efficiently and nonintrusively. In this study, we proposed a novel personalized and unsupervised approach for sleep/wake identification using both heart rate and activity data from a commercial wearable device. The approach was successfully implemented in case studies of community-dwelling older adults.

Our proposed approach is the first unsupervised and personalized sleep/wake classification approach, to our knowledge. It does not require any time-consuming and costly PSG annotation, which is hard to obtain simultaneously with

wearable device data [15,36]. Furthermore, our approach was efficient enough to be adaptive to a different participant without requiring PSG annotations. In our case study, the variance in estimated parameters in HMM between participants also proved the necessity of a personalized model.

The data-level fusion of activity and heart rate data for sleep/wake scoring in wearable device was explored. Based on comparison among scoring results from HMMs using heart rate only, activity data only, and both data sources, we concluded that our approach could potentially help identify more wake epochs for people who have distinguishable heart rate patterns between sleep and wake. This coincided with the findings of significant different heart rate during sleep and wake states [32] and its classification power in sleep/wake identification algorithms [33,44].

Our approach had results that were mostly consistent with those of Fitbit, a commonly used commercial device. For most of the

participants, our approach tended to score more wake epochs during bedtime, which may be potentially useful as it has been shown by Montgomery-Downs et al [45] that many commercial wearable devices, when compared to PSG, tend to overestimate sleep epochs. This should be further investigated with PSG annotations.

Strengths

In addition to sleep and wake identification and total sleep time estimation results, this paper proposed an approach that provides a new probabilistic way to characterize and quantify activity patterns and cardiac patterns during sleep and wake for each participant with estimated HMM parameters. A low estimated mean activity in wake states suggest a sedentary behavior style when the participant is wake, which should be of concern to the participant or their health care provider [46]. Abnormally high estimated mean heart rate in sleep states can be an indicator of autonomic nervous system dysfunction or of the development of chronic fatigue syndrome [47]. A high estimated probability of transitioning from sleep to wake might suggest disturbed sleep. Monitoring these parameters for clinical use is promising and remains to be explored in specific tasks.

In addition, we demonstrated how to characterize cardiac and behavior patterns on a daily basis. During our 3-month study, some participants exhibited significantly different patterns between weekdays and weekends. On weekdays, participants had regular visits to older adult centers in their community, which provided various activities. This could explain why some participants tended to have higher estimated mean and variance of activity level for wake state and were less likely to transition from wake to sleep on weekdays. There exists not only interindividual variability in sleep patterns but also intraindividual variability [34]. This analysis can also provide a reference to improve the accuracy of the personalized model to infer sleep/wake states by including covariates (eg, weekday or weekend) in the transition matrix or emission distributions in HMM.

Limitations

We acknowledge that there are also some limitations to our work. First, direct conclusions about our approach's accuracy cannot be drawn because of the lack of PSG recordings in this study and the small sample size. The estimated sleep/wake states might also be better interpreted as resting-active states at this stage. Second, the comparison with Fitbit's scoring may not be very fair since the exact algorithm is not publicly known and the data type used for sleep/wake scoring was unknown as well. We also cannot reasonably compare the proposed approach with existing methods on the same data set because of the lack of personalized and unsupervised approaches in sleep/wake scoring (to our knowledge). Third, we collected heart rate and step count data using Fitbit, which were reported in different granularities. In order to achieve data-level fusion and prepare it for modeling, we simulated 1-minute step data from 15-minute step data which might have resulted in some imprecision. We strongly call for different reporting granularity options in commercial wearable devices to further facilitate research and their use in health care monitoring systems.

Future Work

In the future, we plan to compare the proposed approach with PSG to further validate the accuracy of the scoring. Some future research directions include (1) exploring the relationship between the length of data and accuracy of sleep/wake classification to yield a reliable algorithm, (2) exploring the existence of not only interindividual variability in sleep patterns but also intraindividual variability [34] (such as developing an incremental approach for daily sleep/wake duration reporting where the sleep/wake pattern can be re-estimated as new data are captured), (3) with the strength of HMM, it would be interesting to see whether sleep characteristics shown in periodically estimated HMM parameters can be correlated with health condition or circadian rhythm changes, and (4) exploring the association between subjective Pittsburgh Sleep Quality Index and estimated HMM parameters in different populations.

Conflicts of Interest

None declared.

References

1. Consensus Conference Panel, Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, et al. Joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society on the recommended amount of sleep for a healthy adult: methodology and discussion. Sleep 2015 Aug 01;38(8):1161-1183 [FREE Full text] [doi: 10.5665/sleep.4886] [Medline: 26194576]
2. Irwin MR. Why sleep is important for health: a psychoneuroimmunology perspective. Annu Rev Psychol 2015 Jan 03;66:143-172 [FREE Full text] [doi: 10.1146/annurev-psych-010213-115205] [Medline: 25061767]
3. Faubel R, Lopez-Garcia E, Guallar-Castillón P, Balboa-Castillo T, Gutiérrez-Fisac JL, Banegas JR, et al. Sleep duration and health-related quality of life among older adults: a population-based cohort in Spain. Sleep 2009 Aug;32(8):1059-1068 [FREE Full text] [Medline: 19725257]

XSL•FO

RenderX

4. Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. Eur Heart J 2011 Jun;32(12):1484-1492. [doi: 10.1093/eurheartj/ehr007] [Medline: 21300732]

5. Chaput J, McNeil J, Després JP, Bouchard C, Tremblay A. Short sleep duration as a risk factor for the development of the metabolic syndrome in adults. Prev Med 2013 Dec;57(6):872-877. [doi: 10.1016/j.ypmed.2013.09.022] [Medline: 24099879]

6. St-Onge M, Grandner MA, Brown D, Conroy MB, Jean-Louis G, Coons M, American Heart Association Obesity, Behavior Change, Diabetes,Nutrition Committees of the Council on LifestyleCardiometabolic Health; Council on Cardiovascular Disease in the Young; Council on Clinical Cardiology;Stroke Council. Sleep duration and quality: impact on lifestyle behaviors and cardiometabolic health: a scientific statement from the American Heart Association. Circulation 2016 Nov 01;134(18):e367-e386 [FREE Full text] [doi: 10.1161/CIR.0000000000000444] [Medline: 27647451]

7. Chaput J, Després J, Bouchard C, Tremblay A. Short sleep duration is associated with reduced leptin levels and increased adiposity: results from the Quebec family study. Obesity (Silver Spring) 2007 Jan;15(1):253-261 [FREE Full text] [doi: 10.1038/oby.2007.512] [Medline: 17228054]

8. Itani O, Jike M, Watanabe N, Kaneita Y. Short sleep duration and health outcomes: a systematic review, meta-analysis, and meta-regression. Sleep Med 2017 Apr;32:246-256. [doi: 10.1016/j.sleep.2016.08.006] [Medline: 27743803]

9. Ferrie JE, Shipley MJ, Cappuccio FP, Brunner E, Miller MA, Kumari M, et al. A prospective study of change in sleep duration: associations with mortality in the Whitehall II cohort. Sleep 2007 Dec;30(12):1659-1666 [FREE Full text] [doi: 10.1093/sleep/30.12.1659] [Medline: 18246975]

10. Ferrie JE, Shipley MJ, Akbaraly TN, Marmot MG, Kivimäki M, Singh-Manoux A. Change in sleep duration and cognitive function: findings from the Whitehall II Study. Sleep 2011 May 01;34(5):565-573 [FREE Full text] [doi: 10.1093/sleep/34.5.565] [Medline: 21532949]

11. Berry RB, Brooks R, Gamaldo CE, Harding SM, Lloyd RM, Marcus CL, et al. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Darien, Illinois: American Academy of Sleep Medicine; 2017.

12. Peters B. What is an overnight sleep study (polysomnogram)? Verywell Health. URL: https://www.verywellhealth.com/what-to-expect-in-a-sleep-study-3015121 [accessed 2020-07-16]

13. Blackwell T, Redline S, Ancoli-Israel S, Schneider JL, Surovec S, Johnson NL, Study of Osteoporotic Fractures Research Group. Comparison of sleep parameters from actigraphy and polysomnography in older women: the SOF study. Sleep 2008 Feb;31(2):283-291 [FREE Full text] [doi: 10.1093/sleep/31.2.283] [Medline: 18274276]

14. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. Chronobiol Int 2018 Apr;35(4):465-476. [doi: 10.1080/07420528.2017.1413578] [Medline: 29235907]

15. El-Manzalawy Y, Buxton O, Honavar V. Sleep/wake state prediction and sleep parameter estimation using unsupervised classification via clustering. 2017 Presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 13-16 Nov, 2017; Kansas City, MO, USA p. 718-723. [doi: 10.1109/BIBM.2017.8217742]

16. Ongoing demand fuels a strong growth trajectory for wearable devices in q1 2019 with wrist-worn and ear-worn leading the market, according to IDC. International Data Corporation. 2019 May 30. URL: https://www.idc.com/getdoc.jsp?containerId=prUS45115019 [accessed 2020-07-16]

17. Wrist worn wearables maintain a strong growth trajectory in q2 2019, according to IDC. International Data Corporation. 2019 Sep 12. URL: https://www.idc.com/getdoc.jsp?containerId=prUS45521319 [accessed 2020-07-16]

18. Bianchi MT. Sleep devices: wearables and nearables, informational and interventional, consumer and clinical. Metabolism 2018 Dec;84:99-108. [doi: 10.1016/j.metabol.2017.10.008] [Medline: 29080814]

19. Sadeh A, Hauri PJ, Kripke DF, Lavie P. The role of actigraphy in the evaluation of sleep disorders. Sleep 1995 May;18(4):288-302. [doi: 10.1093/sleep/18.4.288] [Medline: 7618029]

20. Karlen W, Mattiussi C, Floreano D. Improving actigraph sleep-wake classification with cardio-respiratory signals. 2008 Presented at: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 20-25 Aug, 2008; Vancouver, BC, Canada p. 5262-5265. [doi: 10.1109/iembs.2008.4650401]

21. Palotti J, Mall R, Aupetit M, Rueschman M, Singh M, Sathyanarayana A, et al. Benchmark on a large cohort for sleep-wake classification with machine learning techniques. NPJ Digit Med 2019;2:50 [FREE Full text] [doi: 10.1038/s41746-019-0126-9] [Medline: 31304396]

22. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992 Oct;15(5):461-469. [Medline: 1455130]

23. Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. Sleep 1994;17(3):201-207. [Medline: 7939118]

24. Tilmanne J, Urbain J, Kothare MV, Wouwer AV, Kothare SV. Algorithms for sleep-wake identification using actigraphy: a comparative study and new results. J Sleep Res 2009 Mar;18(1):85-98 [FREE Full text] [doi: 10.1111/j.1365-2869.2008.00706.x] [Medline: 19250177]

25. CamNtech. The Actiwatch user manual. Salusa. 2008 Jul 10. URL: https://www.salusa.se/Filer/Produktinfo/Aktivitet/TheActiwatchUserManualV7.2.pdf [accessed 2020-07-16]

26. Li X, Zhang Y, Jiang F, Zhao H. A novel machine learning unsupervised algorithm for sleep/wake identification using actigraphy. Chronobiol Int 2020 Apr 28:1-14. [doi: 10.1080/07420528.2020.1754848] [Medline: 32342702]

27. Zucchini W, MacDonald IL, Langrock R. Hidden Markov Models for Time Series: An Introduction Using R, Second Edition. Boca Raton, FL, USA: CRC Press; 2016.

28. Huang Q, Cohen D, Komarzynski S, Li X, Innominato P, Lévi F, et al. Hidden Markov models for monitoring circadian rhythmicity in telemetric activity data. J R Soc Interface 2018 Feb;15(139) [FREE Full text] [doi: 10.1098/rsif.2017.0885] [Medline: 29436510]

29. Pan S, Kuo C, Zeng J, Liang S. A transition-constrained discrete hidden Markov model for automatic sleep staging. Biomed Eng Online 2012 Aug 21;11:52 [FREE Full text] [doi: 10.1186/1475-925X-11-52] [Medline: 22908930]

30. Langrock R, Swihart BJ, Caffo BS, Punjabi NM, Crainiceanu CM. Combining hidden Markov models for comparing the dynamics of multiple sleep electroencephalograms. Stat Med 2013 Aug 30;32(19):3342-3356 [FREE Full text] [doi: 10.1002/sim.5747] [Medline: 23348835]

31. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. JMIR Mhealth Uhealth 2019 Mar 11;7(3):e10828 [FREE Full text] [doi: 10.2196/10828] [Medline: 30855232]

32. Furlan R, Guzzetti S, Crivellaro W, Dassi S, Tinelli M, Baselli G, et al. Continuous 24-hour assessment of the neural regulation of systemic arterial pressure and RR variabilities in ambulant subjects. Circulation 1990 Feb;81(2):537-547. [doi: 10.1161/01.cir.81.2.537] [Medline: 2297860]

33. Willemen T, Van Deun D, Verhaert V, Vandekerckhove M, Exadaktylos V, Verbraecken J, et al. An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification. IEEE J Biomed Health Inform 2014 Mar;18(2):661-669. [doi: 10.1109/JBHI.2013.2276083] [Medline: 24058031]

34. Long X, Haakma R, Leufkens TRM, Fonseca P, Aarts RM. Effects of between- and within-subject variability on autonomic cardiorespiratory activity during sleep and their limitations on sleep staging: a multilevel analysis. Comput Intell Neurosci 2015;2015:583620 [FREE Full text] [doi: 10.1155/2015/583620] [Medline: 26366167]

35. Rusterholz T, Tarokh L, Van Dongen HPA, Achermann P. Interindividual differences in the dynamics of the homeostatic process are trait-like and distinct for sleep versus wakefulness. J Sleep Res 2017 Apr;26(2):171-178 [FREE Full text] [doi: 10.1111/jsr.12483] [Medline: 28019041]

36. Khademi A, El-Manzalawy Y, Master L, Buxton OM, Honavar VG. Personalized sleep parameters estimation from actigraphy: a machine learning approach. Nat Sci Sleep 2019;11:387-399 [FREE Full text] [doi: 10.2147/NSS.S220716] [Medline: 31849551]

37. Karlen W, Floreano D. Adaptive sleep-wake discrimination for wearable devices. IEEE Trans Biomed Eng 2011 Apr;58(4):920-926. [doi: 10.1109/TBME.2010.2097261] [Medline: 21172750]

38. Deliang C, Inger H. Empirical-statistical downscaling. Danvers, MA, USA: World Scientific Publishing Company; 2008.

39. Visser I, Speekenbrink M. depmixS4: an R Package for Hidden Markov Models. J Stat Soft 2010;36(7). [doi: 10.18637/jss.v036.i07]

40. Webster JB, Kripke DF, Messin S, Mullaney DJ, Wyborney G. An activity-based sleep monitor system for ambulatory use. Sleep 1982;5(4):389-399. [doi: 10.1093/sleep/5.4.389] [Medline: 7163726]

41. How do I track my sleep with my Fitbit device? Fitbit. URL: https://help.fitbit.com/articles/en_US/Help_article/1314.htm [accessed 2020-07-16]

42. Shrivastava D, Jung S, Saadat M, Sirohi R, Crewson K. How to interpret the results of a sleep study. J Community Hosp Intern Med Perspect 2014;4(5):24983 [FREE Full text] [doi: 10.3402/jchimp.v4.24983] [Medline: 25432643]

43. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [Medline: 2748771]

44. Ye Y, Yang K, Jiang J, Ge B. Automatic sleep and wake classifier with heart rate and pulse oximetry: derived dynamic time warping features and logistic model. 2016 Presented at: Annual IEEE Systems Conference (SysCon); 2016; Orlando, FL, USA p. 1-6. [doi: 10.1109/syscon.2016.7490623]

45. Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. Sleep Breath 2012 Sep;16(3):913-917. [doi: 10.1007/s11325-011-0585-y] [Medline: 21971963]

46. Stamatakis E, Davis M, Stathi A, Hamer M. Associations between multiple indicators of objectively-measured and self-reported sedentary behaviour and cardiometabolic risk in older adults. Prev Med 2012 Jan;54(1):82-87. [doi: 10.1016/j.ypmed.2011.10.009] [Medline: 22057055]

47. Boneva RS, Decker MJ, Maloney EM, Lin J, Jones JF, Helgason HG, et al. Higher heart rate and reduced heart rate variability persist during sleep in chronic fatigue syndrome: a population-based study. Auton Neurosci 2007 Dec 30;137(1-2):94-101. [doi: 10.1016/j.autneu.2007.08.002] [Medline: 17851136]

Abbreviations

AIC: Akaike information criterion
BIC: Bayesian information criterion

XSL•FO
RenderX

HMM: hidden Markov model
PSG: polysomnography
QQ: quantile-quantile

Original Paper

Using Smart Bracelets to Assess Heart Rate Among Students During Physical Education Lessons: Feasibility, Reliability, and Validity Study

Jiangang Sun[1], MEd; Yang Liu[1,2], PhD

[1]School of Physical Education and Sport Training, Shanghai University of Sport, Shanghai, China

[2]Shanghai Research Center for Physical Fitness and Health of Children and Adolescents, Shanghai University of Sport, Shanghai, China

Corresponding Author:
Yang Liu, PhD
School of Physical Education and Sport Training
Shanghai University of Sport
650 Qingyuanhuan Rd
Shanghai, 200438
China
Phone: 86 21 6550 7989
Fax: 86 21 6550 7125
Email: docliuyang@hotmail.com

Abstract

Background: An increasing number of wrist-worn wearables are being examined in the context of health care. However, studies of their use during physical education (PE) lessons remain scarce.

Objective: We aim to examine the reliability and validity of the Fizzo Smart Bracelet (Fizzo) in measuring heart rate (HR) in the laboratory and during PE lessons.

Methods: In Study 1, 11 healthy subjects (median age 22.0 years, IQR 3.75 years) twice completed a test that involved running on a treadmill at 6 km/h for 12 minutes and 12 km/h for 5 minutes. During the test, participants wore two Fizzo devices, one each on their left and right wrists, to measure their HR. At the same time, the Polar Team2 Pro (Polar), which is worn on the chest, was used as the standard. In Study 2, we went to 10 schools and measured the HR of 24 students (median age 14.0 years, IQR 2.0 years) during PE lessons. During the PE lessons, each student wore a Polar device on their chest and a Fizzo on their right wrist to measure HR data. At the end of the PE lessons, the students and their teachers completed a questionnaire where they assessed the feasibility of Fizzo. The measurements taken by the left wrist Fizzo and the right wrist Fizzo were compared to estimate reliability, while the Fizzo measurements were compared to the Polar measurements to estimate validity. To measure reliability, intraclass correlation coefficients (ICC), mean difference (MD), standard error of measurement (SEM), and mean absolute percentage errors (MAPE) were used. To measure validity, ICC, limits of agreement (LOA), and MAPE were calculated and Bland-Altman plots were constructed. Percentage values were used to estimate the feasibility of Fizzo.

Results: The Fizzo showed excellent reliability and validity in the laboratory and moderate validity in a PE lesson setting. In Study 1, reliability was excellent (ICC>0.97; MD<0.7; SEM<0.56; MAPE<1.45%). The validity as determined by comparing the left wrist Fizzo and right wrist Fizzo was excellent (ICC>0.98; MAPE<1.85%). Bland-Altman plots showed a strong correlation between left wrist Fizzo measurements (bias=0.48, LOA=–3.94 to 4.89 beats per minute) and right wrist Fizzo measurements (bias=0.56, LOA=–4.60 to 5.72 beats per minute). In Study 2, the validity of the Fizzo was lower compared to that found in Study 1 but still moderate (ICC>0.70; MAPE<9.0%). The Fizzo showed broader LOA in the Bland-Altman plots during the PE lessons (bias=–2.60, LOA=–38.89 to 33.69 beats per minute). Most participants considered the Fizzo very comfortable and easy to put on. All teachers thought the Fizzo was helpful.

Conclusions: When participants ran on a treadmill in the laboratory, both left and right wrist Fizzo measurements were accurate. The validity of the Fizzo was lower in PE lessons but still reached a moderate level. The Fizzo is feasible for use during PE lessons.

XSL•FO
RenderX

KEYWORDS

Introduction

The health benefits of moderate to vigorous intensity physical activity (MVPA) have been documented [1,2]. In China and some European countries, only 20% of children and adolescents achieve the recommended 60 minutes of MVPA daily [3-7]. Physical education (PE) lessons are an important way to promote MVPA in children and adolescents. Such lessons not only provide a chance for students to be active, directly accumulating MVPA over time, but also provide opportunities for students to learn different types of motor or sport skills that may increase their MVPA as well [8,9]. Recognizing the importance of PE lessons for the physical fitness and health of children and adolescents [10], the Centers for Disease Control and Prevention of the United States [11] and Association for Physical Education of the United Kingdom [12] have recommended that children and adolescents engage in MVPA for at least 50% of the total PE lesson time [7]. However, many studies have found that students often do not meet this recommendation [13,14]. In addition, the fact that students cannot reach 50% MVPA in PE lessons can greatly reduce their chance of reaching 60 minutes MVPA daily [15]. Thus, it is important to estimate students' physical activity during PE lessons and implement interventions to increase students' participation in MVPA [16].

It is common to evaluate physical activity via HR monitoring, and HR measurements can provide a lot of information about the intensity of physical activity [17]. Monitoring HR can help teachers determine students' levels of physical activity and improve PE lessons. An increasing number of studies have used objective measurements, such as HR-based devices, to estimate physical activity during PE lessons [18]. The accuracy of these devices has been confirmed [19], but most of them lack feasibility and are too expensive [15]. Therefore, there is a need to identify accurate, inexpensive, convenient, and feasible devices. Increasingly more researchers and consumers are paying attention to wrist-worn wearables [20-22], some of which can be used to evaluate physical activity [23]. However, few studies have examined the use of wrist-worn wearables during PE lessons.

The Fizzo Smart Bracelet (Fizzo) is a wrist-worn wearable made in China, based on photoplethysmography and dependent on optical HR measurement. The Fizzo is not specifically designed to measure physical activity in young children but to measure HR in children, adolescents, and adults. A customized algorithm was designed for the device to measure physical activity based on HR in children and adolescents during PE lessons in school settings. The aims of this paper are the following: (1) to examine the reliability and validity of Fizzo in measuring HR in a laboratory setting; and (2) to examine the feasibility and validity of Fizzo for children and adolescents during PE lessons.

Methods

Ethics

The study was approved by the Institutional Review Board of the Shanghai University of Sport (102772019RT034). Before enrollment in this study, written informed consent was provided by every participant.

Participants

This study was divided into two substudies: Study 1 and Study 2. In Study 1, 11 students from the Shanghai University of Sport, School of Physical Education and Sport Training (Shanghai, China) volunteered to participate. Recruitment was conducted by word of mouth around the university. The participants did not have any musculoskeletal injuries or illnesses.

In Study 2, we randomly selected 10 schools (2 elementary schools, 7 junior high schools, and 1 high school) from Xuhui District in Shanghai, each of which enrolled 2-3 students in PE lessons, for a total of 28 students. At the end of the PE lessons, the students and their teachers (1 teacher/lesson) completed short questionnaires to assess the feasibility of using the Fizzo. Data obtained from the students (Study 1 and Study 2) included age, gender, height, weight, and body mass index (Table 1).

Data were collected in October 2018 (Study 1) and May 2019 (Study 2). The number of participants (n=28) was in line with Wallen et al [24], considering a power of 0.5 and a probability of type I error of 50%. The sample sizes of similar studies ranged from 20 to 60 [21].

Devices

This study evaluated the Fizzo (The Fifth Zone Fitness Laboratory Company), which is a wrist-worn wearable (cost of ~300¥; US $2.85) that has a battery life of 3 to 8 days. It uses optical sensors to deduce relative volumetric changes in blood perfusion and calculate HR. The Fizzo has a triaxial accelerometer that can measure steps. The data are uploaded to a website [25], where information is stored for up to 90 days, and can be downloaded through an app.

We chose the 2008 Polar Team2 Pro (Polar Electro Oy) as the reference tool in this study. The accuracy of this chest strap has been examined and it is considered to be a standard for the assessment of HR during exercise and training [26,27].

Both devices have a feedback function, but the presentation of data requires a computer or iPad (Apple Inc). In this study, we showed the Fizzo-collected data to the teachers using an iPad.

Study Procedures and Data Collection

The protocol of Study 1 and 2 are shown in Figure 1. In Study 1, we instructed the students to wear comfortable sportswear and shoes before coming to the laboratory. Before the test, each student put on two Fizzos, one each on their left and right wrists. At the same time, a Polar Team2 Pro (Polar) was placed correctly on the participant's chest with the help of the

researchers and served as the standard. All three devices were tightly secured to ensure skin contact [28]. We chose running speeds of 6 km/h and 12 km/h to correspond with moderate and vigorous intensity movement, respectively. The left Fizzo was compared to the right Fizzo to examine the reliability of the device. Additionally, the left and right Fizzo measurements were compared with the Polar measurements to examine validity. Before the test, participants ran on a treadmill at a self-selected speed for 3 minutes to warm up and to adapt to the environment. Participants ran at speeds of 6 km/h (moderate intensity) and 12 km/h (vigorous intensity) for 12 and 5 minutes, respectively. Between stages, the participants stood for 2 minutes to rest [29]. All participants ran at an incline of zero degrees and all tests took about 22 minutes to complete. Participants could stop at any time if they felt uncomfortable, and every participant successfully finished the test. HR was recorded every second during the test.

In Study 2, we went to 10 schools and measured the HR of 28 students during PE lessons. Before the PE lessons, the students placed a Fizzo on one wrist (according to their own preference; all students chose the right hand) and a Polar band on their chest, which were both secured tightly to ensure skin contact. The start and end time of the lessons depended on the teacher's

instructions. When the teachers started the class, we began to measure HR; we stopped when the teachers ended the class. All PE lessons were completed outdoors.

According to studies by Lee et al [17], McNamara et al [30], and Cruz [31], the feasibility of a wrist-worn wearable relies heavily on its accuracy, acceptability, applicability, and usefulness. At the end of the PE lessons, the students and their teachers completed a questionnaire with the help of trained staff. At the same time, we recorded whether or not students removed the Fizzo during the PE lesson. The questionnaire consisted of three questions about the Fizzo that addressed its comfort, application, and helpfulness.

Using a 5-point Likert-type scale, comfort indicators ranged from "very comfortable" to "very uncomfortable," and the level of difficulty in putting on and removing the Fizzo ranged from "very easy" to "very hard." The last question asked the PE teachers whether the Fizzo was helpful, with yes/no response options. A total of 28 students and 10 teachers finished the questionnaires. In addition, we invited the teachers to indicate whether or not the Fizzo was helpful for PE lessons. We also asked them to give suggestions, although this was not mandatory. All teachers accepted the invitation.

Figure 1. Study 1 and 2 protocol. PE: physical education.
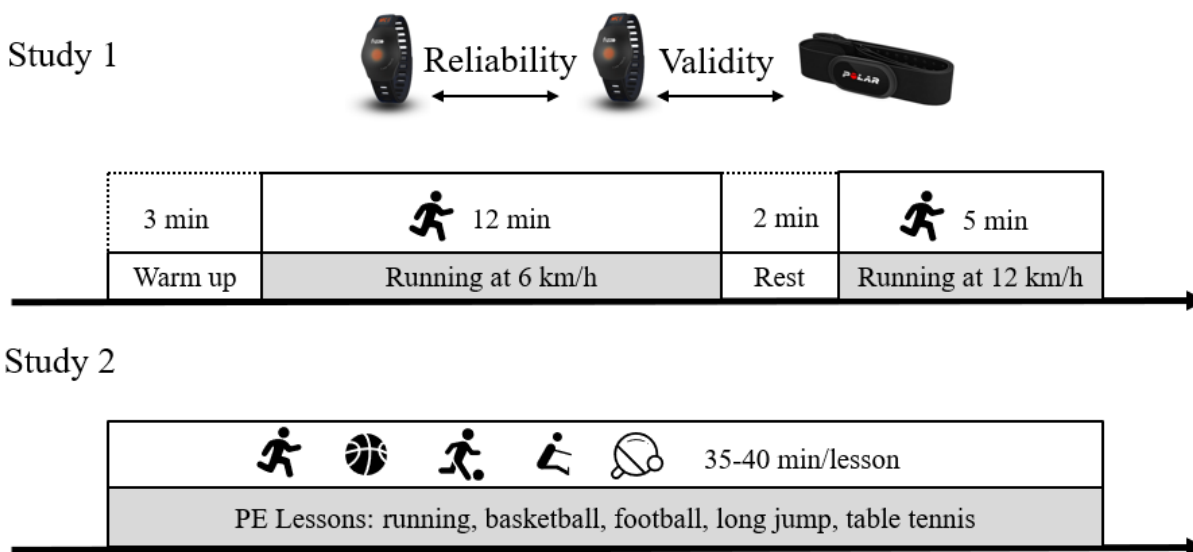
Statistical Analysis

The data acquired from the Fizzo and Polar were analyzed using SPSS Statistics (Version 24; IBM Corp). Statistical significance and assumptions for linear statistics were set at $P \le .05$ [32-34]. Before the statistical analyses, all data had been tested for missing values. The data were retained if <10% of the data included missing data points and outliers. In Study 1, there was no missing data, so all data were retained. During PE lessons, 4 students took off their Polar device and the missing data was more than 10% of the total, so their data were deleted. To maintain authenticity, we did not remove outliers, as the Fizzo's instructions state that HR is accurately measured when placed on either the left or right wrist. Reliability was determined by

comparing the left Fizzo data to the right Fizzo data using intraclass correlation coefficients (ICC) for single measurements and absolute agreement from a two-way mixed effect model [35]. ICC shows a measure of relative agreement.

Reliability was determined by calculating the ICC between the two devices with a 95% CI. ICC≥0.9 was considered excellent, 0.90>ICC≥0.75 was considered good, 0.75>ICC≥0.60 was considered moderate, and ICC<0.60 was considered poor [21]. Theoretically, the values of ICC were all positive [36], while the other values were set to zero [37]. Additionally, the reliability between the left Fizzo and right Fizzo was calculated by the mean differences (MDs) and standard error of the mean (SEM) [37,38]. The degree of error can be shown by the MD (SEM) value, with high values indicating high error [34].

To determine validity, ICC with 95% CI for single measurements and absolute agreement from a two-way mixed effect model were used to calculate the relative agreement between the Polar and Fizzo devices in Study 1 and Study 2 [38,39]. In addition, mean absolute percent error (MAPE) was used to assess the degree of error between the standard and the Fizzo, with the equation MAPE = [(Fizzo – Polar)/Polar] × 100%. According to some previous studies [20,40], MAPE≤10% can be considered good, whereas MAPE>10% is considered poor. Finally, the level of agreement was examined using a Bland-Altman analysis and 95% limits of agreement (LOA) between the Polar and Fizzo across the range of HR data (a narrower range is better); this method is recommended to estimate the agreement of medical devices [41].

To determine the feasibility of Fizzo, the number and percentage of "very comfortable" and "comfortable," "very easy" and "easy," and "yes" responses were used; a percentage ≥90% was considered good. The number and percentage of students who removed the Fizzo were recorded; a percentage ≤10% was considered good. In addition, the teachers' views about the Fizzo were analyzed.

Results

Overview

The physical characteristics of participants (for both Study 1 and Study 2) are presented in Table 1. There were 11 participants (median age 22.0 years, IQR 3.75 years) in Study 1 (October 2018). Every participant came to the laboratory and was tested twice (on different days, with an interval of 1 week). Finally, we obtained 22 sets of HR data (10 males and 12 females). All participants were right hand dominant.

Data from 4 students were excluded because they removed the Polar chest band during the PE lesson; ultimately, we collected data from 24 students (median age 14.0 years, IQR 2.0 years) in Study 2 (May 2019).

Table 1. Participants characteristics for Study 1 and Study 2.

Study and characteristics	Values
Study 1 (n=11)	
Age (years), median (IQR)	22.0 (3.75)
Males, n (%)	5.0 (45.5)
Weight (kg), median (IQR)	61.0 (21.9)
Height (cm), median (IQR)	170.0 (10.8)
BMI (kg/m^2), median (IQR)	21.5 (3.8)
Study 2 (n=24)	
Age (years), median (IQR)	14.0 (2.0)
Males, n (%)	11.0 (45.9)
Weight (kg), median (IQR)	52.9 (22.5)
Height (cm), median (IQR)	162.3 (15.6)
BMI (kg/m^2), median (IQR)	22.9 (4.3)

Reliability in Study 1

Table 2 provides the ICC, MD (SEM), and MAPE (SD) for interdevice reliability when running on a treadmill in the laboratory. HR data from the right Fizzo were similar to those from the left Fizzo. The total values for interdevice reliability were evaluated at two running speeds and the devices demonstrated good reliability (ICC=0.99 [95% CI 0.99-0.99]; MD=0.05 [SEM 0.03]; MAPE=1.43% [SD 1.67]). The reliability between the right and left Fizzo at a running speed of 6 km/h was ICC=0.98 (95% CI 0.98-0.98); MD=0.42 (SEM 0.03); MAPE=1.42% (SD 1.64). At a running speed of 12 km/h, reliability values were ICC=0.99 (95% CI 0.99-0.99); MD=–0.66 (SEM 0.056); MAPE=1.44% (SD 1.72). The MD (SEM) at 6 km/h was slightly better than that at 12 km/h; the HR data of the right Fizzo were slightly higher than those of the left Fizzo at 6 km/h and the result was the opposite at 12 km/h. Overall, the reliability between the left and right Fizzo was high (ICC=0.99; MD<0.7; MAPE<2%).

Table 2. Interdevice reliability measures of the left versus right Fizzo for heart rate data captured during treadmill running in Study 1.

Characteristics	6 km/h	12 km/h	Total
Intraclass correlation coefficient (95% CI)	0.978 (0.977-0.979)	0.988 (0.988-0.990)	0.990 (0.990-0.991)
Mean difference, right-left (standard error of measurements)	0.42 (0.03)	–0.66 (0.06)	0.05 (0.03)
Mean absolute percentage error, % (SD)	1.42 (1.64)	1.44 (1.72)	1.43 (1.67)

Validity in Study 1

The outcomes showed that the mean (SD) HR from the left and right Fizzo were similar to the Polar (Table 3). The left Fizzo (ICC=0.99, 95% CI 0.99-0.99) and right Fizzo (ICC=0.99, 95% CI 0.99-0.99) showed a strong relationship with the Polar device. Additionally, the MAPE (SD) were all small: left Fizzo=1.62% (1.65); right Fizzo=1.82% (2.02). The total validity of both the right and left Fizzo were excellent in the laboratory. The mean difference and agreement between the Fizzo and Polar HR measurements were also shown by the Bland-Altman (Figure 2). Overall, at all running speeds, the left Fizzo had a mean error of 0.64 beats per minute (bpm; lower LOA to upper LOA=–5.18 to 6.45 bpm) and an MAPE (SD) of 1.62% (1.65). The right Fizzo had a mean error of 0.69 bpm (lower LOA to upper LOA=–5.96 to 7.24 bpm) and an MAPE (SD) of 1.82 % (2.02). As the speed increased, mean error changed slightly but the 95% LOA range was larger: left Fizzo 95% LOA=0.78 bpm (lower LOA to upper LOA=–3.95 to 5.52 bpm) and right Fizzo 95% LOA=1.20 bpm (lower LOA to upper LOA=–4.36 to –6.76 bpm) at a speed of 6 km/h; left Fizzo 95% LOA=0.34 bpm (lower LOA to upper LOA=–7.15 to 7.83 bpm) and right Fizzo 95% LOA=–0.32 bpm (lower LOA to upper LOA=–8.13 to 7.48 bpm) at a speed of 12 km/h. The range of HR measurements at 12 km/h was larger than at 6 km/h and the magnitude of the change was small.

At a running speed of 12 km/h, the right Fizzo underestimated HR. Additionally, the left and right Fizzo tended to overestimate HR, and the mean differences were all small.

Table 3. Validity of Fizzo versus Polar in Study 1.

Characteristics	6 km/h	12 km/h	Total
Polar, mean (SD)	125.9 (13.8)	158.5 (26.1)	136.8 (24.3)
Left Fizzo, mean (SD)	126.7 (13.3)	158.8 (27.0)	137.4 (24.3)
Right Fizzo, mean (SD)	127.1 (13.3)	158.1 (27.1)	137.4 (24.1)
Left Fizzo, intraclass correlation coefficient (95% CI)	0.984 (0.984 to 0.985)	0.990 (0.989 to 0.990)	0.993 (0.992 to 0.993)
Right Fizzo, intraclass correlation coefficient (95% CI)	0.978 (0.977 to 0.979)	0.989 (0.988 to 0.989)	0.990 (0.990 to 0.991)
Left Fizzo, limits of agreement (lower, upper)	0.78 (–3.95 to 5.52)	0.34 (–7.15 to 7.83)	0.64 (–5.18 to 6.45)
Right Fizzo, limits of agreement (lower, upper)	1.20 (–4.36 to 6.67)	–0.32 (–8.13 to 7.48)	0.69 (–5.96 to 7.24)
Left Fizzo, MAPE[a] (SD) (%)	1.56 (1.52)	1.74 (1.88)	1.62 (1.65)
Right Fizzo, MAPE (SD) (%)	1.80 (2.05)	1.85 (1.95)	1.82 (2.02)

[a]MAPE: mean absolute percentage error.

XSL·FO

RenderX

Figure 2. Bland-Altman plot of Fizzo and Polar values for Study 1. A, C, and E are for the left Fizzo and Polar at a running speed of 6 km/h, 12 km/h, and total, respectively. B, D, and F are for the right Fizzo and Polar at a running speed of 6 km/h, 12 km/h, and total, respectively.
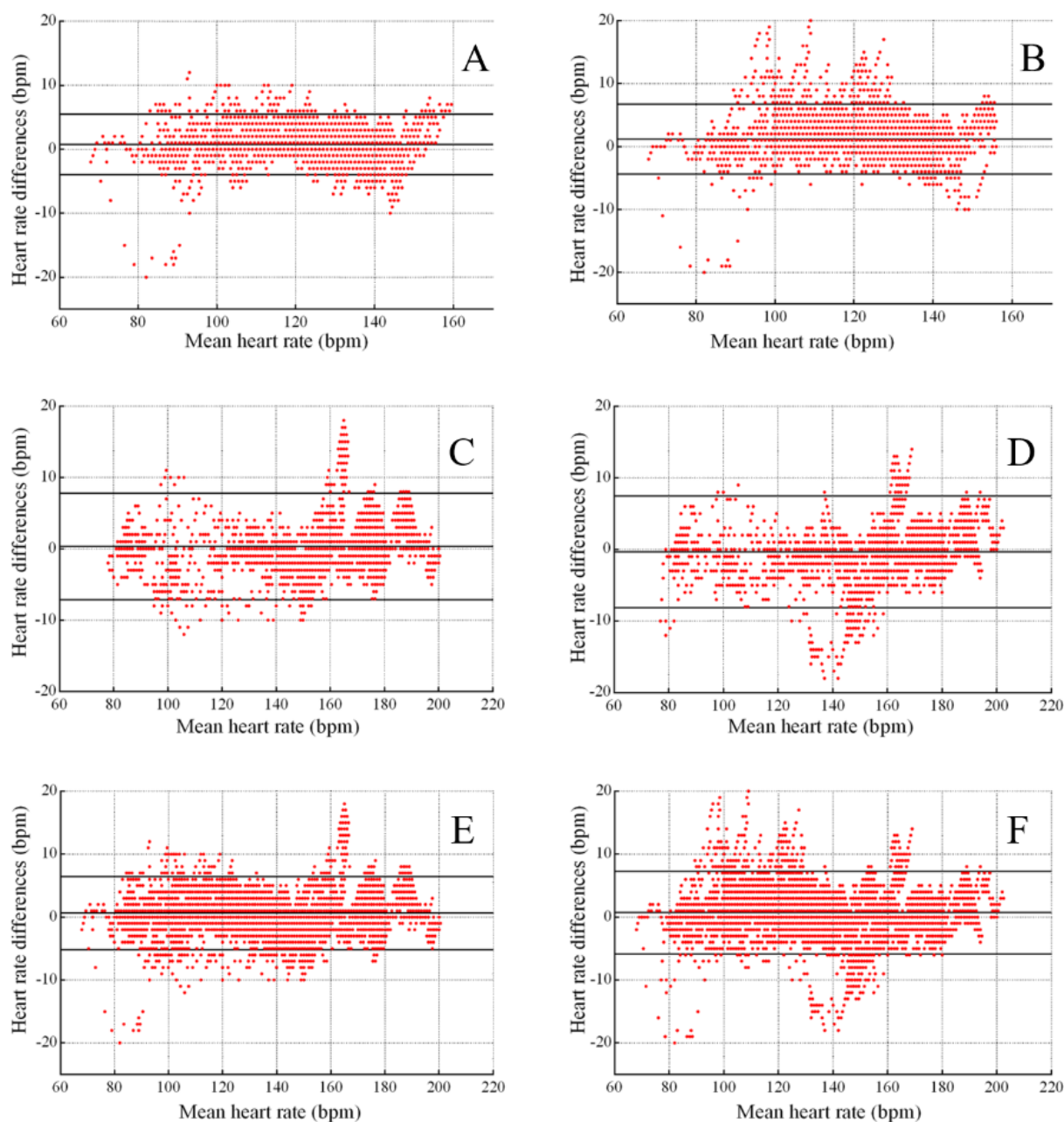
Validity in Study 2

The contents of the PE lessons included running, basketball, football, long jump, and table tennis. Across the PE lessons, the mean (SD) HR between the Fizzo and Polar devices were similar: 137.6 bpm (26.8 bpm) and 140.2 bpm (24.7 bpm), respectively (Table 4). Bland-Altman analysis showed that the Fizzo had a mean error of –2.60 bpm, while the 95% LOA between the two devices ranged from –38.89 to –33.69 bpm. The ICC between the Fizzo and Polar devices was 0.742 lower than in the laboratory. The MAPE was 8.89% for the PE lessons, higher than in the laboratory (1.82%). The Fizzo slightly underestimated HR compared to the Polar during PE lessons.

Table 4. Validity of Fizzo versus Polar in Study 2.

Characteristics	Value
Polar, mean (SD)	140.2 (24.7)
Fizzo, mean (SD)	137.6 (26.8)
Fizzo versus Polar, intraclass correlation coefficient (95% CI)	0.742 (0.739 to 0.746)
Fizzo versus Polar, limits of agreement (lower, upper)	–2.60 (–38.89 to 33.69)
Fizzo versus Polar, mean average percentage error, % (SD)	8.89 (11.04)

The range of the LOA was greater than in the laboratory. The HR data had a mean error of –2.60 bpm (lower LOA to upper LOA=–38.89 to 33.69 bpm). The LOA of the Bland-Altman plots are presented in Figure 3. The Fizzo had the narrowest LOA in the laboratory condition, and broader LOA in the PE lesson condition. The ICC was lower during the PE lessons (0.748) than in the laboratory (>0.99). The MAPE was larger for the PE lessons (8.89%) than in the laboratory (1.82%). The validity of Fizzo in the laboratory is better than in the PE lessons.

Figure 3. Bland-Altman plot of Fizzo and Polar values in Study 2.
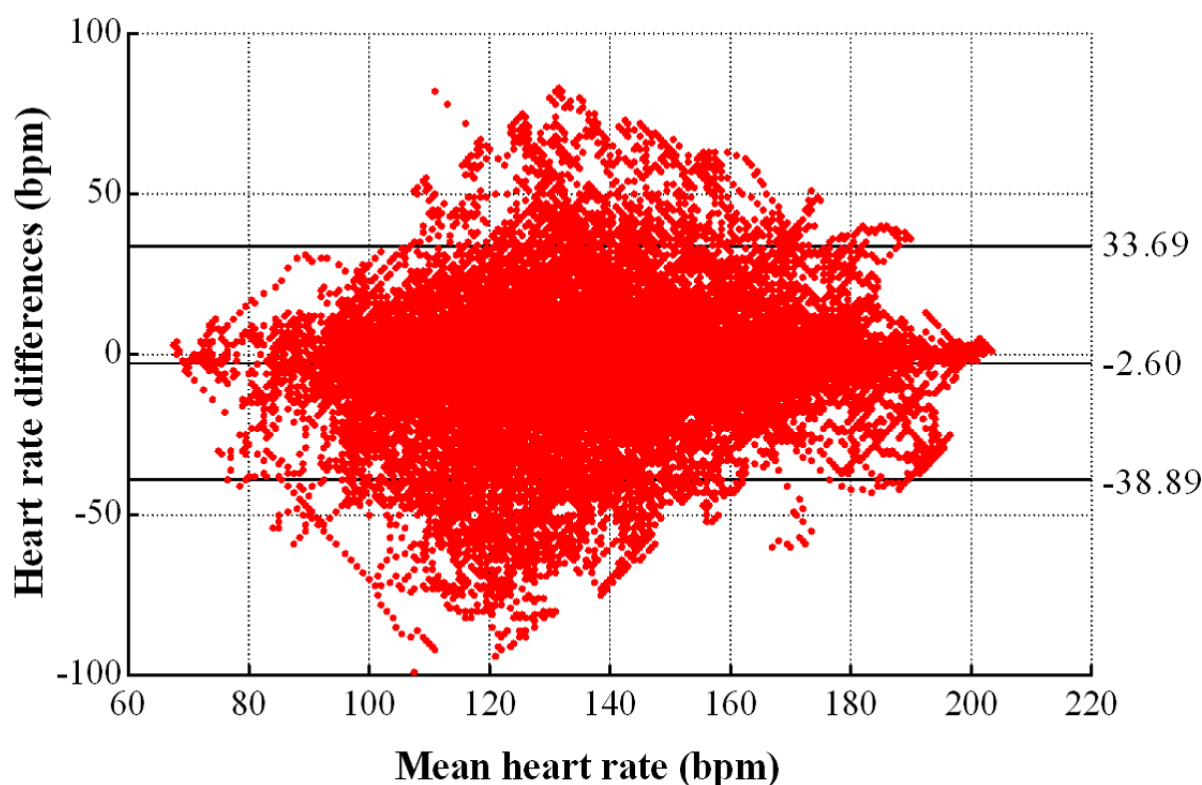
Feasibility in Study 2

The feasibility responses from students and teachers were highly consistent (Table 5). All students chose "very easy," and 71% of students chose "very comfortable" while 29% of students chose "comfortable"; no one chose the "neutral," "uncomfortable," or "very uncomfortable" options. The comfort level of Fizzo is high; none of the students felt uncomfortable during the PE lessons. Regarding application, all students chose "very easy," which means that Fizzo is convenient to wear and remove, and it is easy to put on for PE lessons. All teachers chose "yes" for "helpful"; they believed that the Fizzo has practical value and it is useful for PE lessons. They think that the Fizzo helps them track the students' physical activity levels more conveniently and helps them create more reasonable course content. Furthermore, they think that the Fizzo is very suitable for PE lessons. All the teachers were interested in using the Fizzo in the future. No students tried to remove the Fizzo during the PE lessons. The results showed that it is feasible to use the Fizzo during PE lessons.

XSL•FO
RenderX

Table 5. Questionnaire results for feasibility.

Questionnaire responses	Students (n=28), n (%)	Teachers (n=10), n (%)
Comfort, n (%)		
Very comfortable	20 (71)	—[a]
Comfortable	8 (29)	—
Neutral	0 (0)	—
Uncomfortable	0 (0)	—
Very uncomfortable	0 (0)	—
Ease of application and removal, n (%)		
Very easy	28 (100)	—
Easy	0 (0)	—
Neutral	0 (0)	—
Hard	0 (0)	—
Very hard	0 (0)	—
Helpful, n (%)		
Yes	—	10 (100)
No	—	0 (0)

[a]Not applicable.

Discussion

Principal Findings

The results indicated that the validity and reliability of the Fizzo were excellent in the laboratory. The validity of the Fizzo in the PE lessons was lower than in the laboratory setting, but still moderate. The Fizzo was feasible for PE lessons.

The intensity, duration, and frequency of physical activity can be inferred through HR. Measuring the HR of students accurately and conveniently has great benefit when aiming to increase MVPA to achieve better health preservation effects. The emergence of wrist-worn wearables provides a realistic basis for the achievement of these goals [42]. However, there are still few studies assessing the validity and feasibility of wrist-worn wearables during PE lessons. The purpose of this study was to examine the reliability and validity of a new wrist-worn wearable, Fizzo, in the laboratory and during PE lessons.

Some studies have stated that trackers' reliability and validity may be affected by which wrist the device is worn on [43,44], as there may be some differences in reliability or validity between the right and left wrist (dominant and nondominant). However, in Study 1, we found that validity and reliability were not affected by which wrist the Fizzo was worn on; the HR values as measured by the left and right Fizzo were similar to the HR values obtained from the Polar device. The validity and reliability of Fizzo were almost unaffected by running speed, and Fizzo maintained its high performance when the running speed was increased. This finding is different from previous studies, which found that accuracy became worse when the running or jogging intensity was increased [16]. The stronger relationship between the Polar and Fizzo devices during the

laboratory study can be attributed to more stable experimental controls. Furthermore, when running on a treadmill, the movements of the left and right arms were similar and were not affected by external forces, which may be another reason for this differing result [45].

In Study 2, the validity of the Fizzo was lower during the PE lessons compared to the laboratory, but still showed moderate accuracy (ICC=0.742; mean error=–2.60 bpm, LOA=–38.89 to 33.69; MAPE=8.89%). ICC was very close to the "good" level (0.75); increasing the sample size may impact this result. The MAPE and the range of LOA during PE lessons were larger than in the laboratory and the Fizzo tends to underestimate HR during PE lessons. The Fizzo was easy to use and comfortable for students. Furthermore, it was helpful for teachers during PE lessons and the teachers were interested in using it in the future, which demonstrated the feasibility of the use of the Fizzo. As indicated by the different performance levels of the Fizzo between Study 1 and 2, the validity of the Fizzo may be not influenced by physical activity intensity (running speed) but may be affected by sex, environment, and the type of physical activity [45]. In PE lessons, the students' arms, where the Fizzo was worn, were often subjected to external forces, such as slapping a basketball. These conditions may cause the device to lose skin contact and leak light, resulting in measurement errors. It is also possible that sweat between the device and the skin may cause errors. Compared to Study 1, it was not only the type of exercise that changed in Study 2, but also the environment; these changes may have caused the reduced validity. In the laboratory, the environmental light was relatively stable, but in the PE lessons, it was always changing. The Fizzo uses an optical sensor to measure HR, so it is very sensitive to light. If the Fizzo is not worn tight enough, changes in light may affect its accuracy.

With the development of new technologies, wrist-worn wearables for measuring HR will be an alternative to the chest strap [46]. Wearing the Fizzo on the wrist was more convenient and easier to use than the chest strap; all students considered the Fizzo as comfortable, as well as easy to wear and take off. A chest-worn device is more troublesome, and clothes must be taken off. In addition, it needs to be wiped with alcohol, so it is inconvenient, particularly when the weather is cold. The break between different classes is about 10 minutes in primary and secondary school in China; before the PE lessons started, there were only one or two students who had enough time to apply the Polar bands with the help of assistants. It is very difficult to apply the chest band by oneself. The content of the PE lessons is heterogeneous and complicated and could cause chest-worn devices to fall off, which was why we went to 10 schools but only got 24 samples. Conversely, the Fizzo is worn on the wrist and was less restrictive during physical activity compared with the Polar; the wrist-worn wearable is more comfortable and more acceptable. The teachers considered it helpful for monitoring the physical activity level of their students. According to the HR values of the students, the teachers could choose more reasonable physical activity intensity levels. All teachers believed the Fizzo has practical value and the device showed high feasibility.

When conducting a long-term physical activity surveillance study, a valid survey is usually defined as >10 hours of wear time every day for at least 4 days [47]. The wrist-worn wearables may be more convenient than the chest bands and devices in other positions [48]. Therefore, we need to pay attention not only to accuracy but also to the adherence of participants. Although the Fizzo performed well in PE lessons, further studies are required to prove it can be used in daily life.

To avoid the social desirability response bias with objective measures of physical activity [48], both the Polar and Fizzo do not have a display. However, sometimes feedback is needed; the Fizzo can show almost all students' data (60 students) at the same time on a smartphone or an iPad, while the Polar can show data from 15 students on a computer. Some devices can receive and show the data up to a distance of 200 meters, but sometimes the scope of activities of students in PE lessons far exceeds this distance. The Fizzo can cover the entire playground if small antennas are put near the perimeter of the venue. The Fizzo is more feasible for the monitoring of PE lessons.

Like Polar, most devices for the collection of physical activity data are too expensive; the price of the Polar device is quite high (about 100,000¥ [US $948.7] for 15 Polar bands and a computer). The price of 15 Fizzo is significantly lower (about 4500¥ [US $42.75]), and the computer or iPad can be purchased by oneself, so it is more affordable for researchers. When research funding is limited, researchers can choose cheaper equipment. The Fizzo may be a good choice as it has good validity, reliability, and feasibility during PE lessons. Although many types of PE classes are not included in this study, the available results indicate that the Fizzo has a relatively large application potential in a PE class setting.

Limitations

This study has limitations. The participants were healthy students, while subgroups with known arrhythmia and many types of PE lessons were not included in this study; thus, we cannot be sure of the accuracy of the Fizzo in other populations, including people with heart disease. The small sample size may have affected the results of Study 2, which may limit its generalizability.

To match the end of the Fizzo measurements with the end of the Polar measurements, we only measured HR while participants were running; we did not measure HR after the participant stopped running.

Students were of different ages and included elementary, secondary, and high school students; there may be some reliability and validity differences among these three groups during PE lessons, but that was not distinguished in this study. Physical activity types may have an impact on the reliability and validity of the device, but our study did not classify different types of physical activity; therefore, we cannot be sure which types of physical activity were causing the changes in reliability and validity.

Conclusions

This study shows that the Fizzo has good reliability and validity during moderate and vigorous intensity running on a treadmill in the laboratory. Compared to the laboratory results, the validity of the Fizzo was decreased in PE lessons but still reached a moderate level. The main factor affecting device reliability and validity may not be the intensity of physical activity but the type of physical activity. More research is needed to determine which types of physical activity affected the reliability and validity. Ultimately, the Fizzo is accurate, comfortable to wear, easy to apply and remove, and has a high application value in a PE lesson setting.

XSL•FO
RenderX

Authors' Contributions

JS and YL conceptualized and designed this study and analyzed and interpreted the data. JS drafted the manuscript. YL contributed intellectually to improving the manuscript. JS and YL assisted in revising the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Global Recommendations on Physical Activity for Health. Geneva: WHO Press: World Health Organization; 2010. URL: https://www.who.int/dietphysicalactivity/factsheet_young_people/en/

2. Kang M, Mahar MT, Morrow JR. Issues in the Assessment of Physical Activity in Children. Journal of Physical Education, Recreation & Dance 2016 Jul 29;87(6):35-43. [doi: 10.1080/07303084.2016.1192943]

3. Bucksch J, Inchley J, Hamrik Z, Finne E, Kolip P, HBSC Study Group Germany. Trends in television time, non-gaming PC use and moderate-to-vigorous physical activity among German adolescents 2002-2010. BMC Public Health 2014 Apr 12;14:351 [FREE Full text] [doi: 10.1186/1471-2458-14-351] [Medline: 24725269]

4. Manz K, Schlack R, Poethko-Müller C, Mensink G, Finger J, Lampert T, KiGGS Study Group. [Physical activity and electronic media use in children and adolescents: results of the KiGGS study: first follow-up (KiGGS wave 1)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2014 Jul;57(7):840-848. [doi: 10.1007/s00103-014-1986-4] [Medline: 24950833]

5. Verloigne M, Van Lippevelde W, Maes L, Yıldırım M, Chinapaw M, Manios Y, et al. Levels of physical activity and sedentary time among 10- to 12-year-old boys and girls across 5 European countries using accelerometers: an observational study within the ENERGY-project. Int J Behav Nutr Phys Act 2012 Mar 31;9:34 [FREE Full text] [doi: 10.1186/1479-5868-9-34] [Medline: 22462550]

6. Zhu Z, Tang Y, Zhuang J, Liu Y, Wu X, Cai Y, et al. Physical activity, screen viewing time, and overweight/obesity among Chinese children and adolescents: an update from the 2017 physical activity and fitness in China-the youth study. BMC Public Health 2019 Feb 15;19(1):197 [FREE Full text] [doi: 10.1186/s12889-019-6515-9] [Medline: 30767780]

7. Hollis JL, Williams AJ, Sutherland R, Campbell E, Nathan N, Wolfenden L, et al. A systematic review and meta-analysis of moderate-to-vigorous physical activity levels in elementary school physical education lessons. Prev Med 2016 May;86:34-54. [doi: 10.1016/j.ypmed.2015.11.018] [Medline: 26592691]

8. Zhou Y, Wang L. Correlates of Physical Activity of Students in Secondary School Physical Education: A Systematic Review of Literature. Biomed Res Int 2019;2019:4563484 [FREE Full text] [doi: 10.1155/2019/4563484] [Medline: 30911543]

9. Hills AP, Dengel DR, Lubans DR. Supporting public health priorities: recommendations for physical education and physical activity promotion in schools. Prog Cardiovasc Dis 2015;57(4):368-374. [doi: 10.1016/j.pcad.2014.09.010] [Medline: 25269062]

10. Lonsdale C, Rosenkranz RR, Peralta LR, Bennie A, Fahey P, Lubans DR. A systematic review and meta-analysis of interventions designed to increase moderate-to-vigorous physical activity in school physical education lessons. Prev Med 2013 Feb;56(2):152-161. [doi: 10.1016/j.ypmed.2012.12.004] [Medline: 23246641]

11. Centers for Disease Control and Prevention. National Center on Chronic Disease Prevention and Health Promotion. 2012. URL: https://www.cdc.gov/chronicdisease/

12. Josephine PH. Association for Physical Education Health Position Paper. 2017 Jun 16. URL: https://www.afpe.org.uk [accessed 2020-07-28]

13. Hollis JL, Sutherland R, Williams AJ, Campbell E, Nathan N, Wolfenden L, et al. A systematic review and meta-analysis of moderate-to-vigorous physical activity levels in secondary school physical education lessons. Int J Behav Nutr Phys Act 2017 Apr 24;14(1):52 [FREE Full text] [doi: 10.1186/s12966-017-0504-0] [Medline: 28438171]

14. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. JMIR mHealth uHealth 2018 Apr 12;6(4):e94 [FREE Full text] [doi: 10.2196/mhealth.9754] [Medline: 29650506]

15. Hartwig TB, Del Pozo-Cruz B, White RL, Sanders T, Kirwan M, Parker PD, et al. A monitoring system to provide feedback on student physical activity during physical education lessons. Scand J Med Sci Sports 2019 Sep;29(9):1305-1312. [doi: 10.1111/sms.13438] [Medline: 31033042]

16. Reddy RK, Pooni R, Zaharieva DP, Senf B, El Youssef J, Dassau E, et al. Accuracy of Wrist-Worn Activity Monitors During Common Daily Physical Activities and Types of Structured Exercise: Evaluation Study. JMIR mHealth uHealth 2018 Dec 10;6(12):e10338 [FREE Full text] [doi: 10.2196/10338] [Medline: 30530451]

17. Lee J, Lee DH, Oh TJ, Kim KM, Choi SH, Lim S, et al. Clinical Feasibility of Monitoring Resting Heart Rate Using a Wearable Activity Tracker in Patients With Thyrotoxicosis: Prospective Longitudinal Observational Study. JMIR mHealth uHealth 2018 Jul 13;6(7):e159 [FREE Full text] [doi: 10.2196/mhealth.9884] [Medline: 30006328]

18. Bassett DR. Device-based monitoring in physical activity and public health research. Physiol Meas 2012 Nov;33(11):1769-1783. [doi: 10.1088/0967-3334/33/11/1769] [Medline: 23110847]

19. Terbizan DJ, Dolezal BA, Albano C. Validity of Seven Commercially Available Heart Rate Monitors. Measurement in Physical Education and Exercise Science 2002 Dec;6(4):243-247. [doi: 10.1207/s15327841mpee0604_3]

20. Nelson BW, Allen NB. Accuracy of Consumer Wearable Heart Rate Measurement During an Ecologically Valid 24-Hour Period: Intraindividual Validation Study. JMIR mHealth uHealth 2019 Mar 11;7(3):e10828 [FREE Full text] [doi: 10.2196/10828] [Medline: 30855232]

21. Falter M, Budts W, Goetschalckx K, Cornelissen V, Buys R. Accuracy of Apple Watch Measurements for Heart Rate and Energy Expenditure in Patients With Cardiovascular Disease: Cross-Sectional Study. JMIR mHealth uHealth 2019 Mar 19;7(3):e11889 [FREE Full text] [doi: 10.2196/11889] [Medline: 30888332]

22. Herkert C, Kraal JJ, van Loon EMA, van Hooff M, Kemps HMC. Usefulness of Modern Activity Trackers for Monitoring Exercise Behavior in Chronic Cardiac Patients: Validation Study. JMIR mHealth uHealth 2019 Dec 19;7(12):e15045 [FREE Full text] [doi: 10.2196/15045] [Medline: 31855191]

23. Murakami H, Kawakami R, Nakae S, Yamada Y, Nakata Y, Ohkawara K, et al. 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 Aug 02;7(8):e13938 [FREE Full text] [doi: 10.2196/13938] [Medline: 31376273]

24. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of Heart Rate Watches: Implications for Weight Management. PLoS One 2016;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]

25. Fizzo. URL: http://Fizzo.cn/kanban [accessed 2020-07-27]

26. 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]

27. Thivel D, Maso F, Aouiche S, Coignet B, Doré E, Duché P. Nutritional responses to acute training sessions in young elite rugby players. Appetite 2015 Jan;84:316-321. [doi: 10.1016/j.appet.2014.10.030] [Medline: 25450892]

28. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of Wrist-Worn Heart Rate Monitors. JAMA Cardiol 2016 Oct 12. [doi: 10.1001/jamacardio.2016.3340] [Medline: 27732703]

29. Pina IL, Balady GJ, Hanson P, Labovitz AJ, Madonna DW, Myers J. Guidelines for clinical exercise testing laboratories. A statement for healthcare professionals from the Committee on Exercise and Cardiac Rehabilitation, American Heart Association. Circulation 1995 Feb 01;91(3):912-921. [doi: 10.1161/01.cir.91.3.912] [Medline: 7828326]

30. McNamara RJ, Tsai LLY, Wootton SL, Ng LWC, Dale MT, McKeough ZJ, et al. Measurement of daily physical activity using the SenseWear Armband: Compliance, comfort, adverse side effects and usability. Chron Respir Dis 2016 May;13(2):144-154 [FREE Full text] [doi: 10.1177/1479972316631138] [Medline: 26879695]

31. Cruz J, Brooks D, Marques A. Impact of feedback on physical activity levels of individuals with chronic obstructive pulmonary disease during pulmonary rehabilitation: A feasibility study. Chron Respir Dis 2014 Nov;11(4):191-198. [doi: 10.1177/1479972314552280] [Medline: 25278009]

32. Schaffer SD, Holzapfel SD, Fulk G, Bosch PR. Step count accuracy and reliability of two activity tracking devices in people after stroke. Physiother Theory Pract 2017 Oct;33(10):788-796. [doi: 10.1080/09593985.2017.1354412] [Medline: 28777710]

33. Weatherall J, Paprocki Y, Kudel I, Meyer T, Witt EA. Concurrent Validity of the Armour39 Heart Rate Monitor Strap: Erratum. J Strength Cond Res 2017 Apr;31(4):e70. [doi: 10.1519/JSC.0000000000000557] [Medline: 28328722]

34. Mazoteras Pardo V, Losa Iglesias ME, López Chicharro J, Becerro de Bengoa Vallejo R. The QardioArm App in the Assessment of Blood Pressure and Heart Rate: Reliability and Validity Study. JMIR mHealth uHealth 2017 Dec 15;5(12):e198 [FREE Full text] [doi: 10.2196/mhealth.8458] [Medline: 29246880]

35. McGraw KO, Wong SP. Forming Inferences about Some Intraclass Correlation Coefficients. Psychological Methods 1996;1(4):390-390. [doi: 10.1037/1082-989x.1.4.390]

36. Giraudeau B. Negative values of the intraclass correlation coefficient are not theoretically possible. J Clin Epidemiol 1996 Oct;49(10):1205-1206. [doi: 10.1016/0895-4356(96)00053-4] [Medline: 8827004]

37. Fokkema T, Kooiman TJM, Krijnen WP. Reliability and Validity of Ten Consumer Activity Trackers Depend on Walking Speed. Med Sci Sports Exerc 2017 Apr;49(4):793-800. [doi: 10.1249/MSS.0000000000001146] [Medline: 28319983]

38. Gatti AA, Stratford PW, Brenneman EC, Maly MR. GT3X+ accelerometer placement affects the reliability of step-counts measured during running and pedal-revolution counts measured during bicycling. J Sports Sci 2016 Jun;34(12):1168-1175. [doi: 10.1080/02640414.2015.1096018] [Medline: 26487374]

39. Gorny AW, Liew SJ, Tan CS, Müller-Riemenschneider F. Fitbit Charge HR Wireless Heart Rate Monitor: Validation Study Conducted Under Free-Living Conditions. JMIR mHealth uHealth 2017 Oct 20;5(10):e157 [FREE Full text] [doi: 10.2196/mhealth.8233] [Medline: 29055881]

40. Feito Y, Garner HR, Bassett DR. Evaluation of ActiGraph's low-frequency filter in laboratory and free-living environments. Med Sci Sports Exerc 2015 Jan;47(1):211-217. [doi: 10.1249/MSS.0000000000000395] [Medline: 24870583]

41.   Zaki R, Bulgiba A, Ismail R, Ismail NA. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review. PLoS One 2012;7(5):e37908 [FREE Full text] [doi: 10.1371/journal.pone.0037908] [Medline: 22662248]

42.   Müller AM, Wang NX, Yao J, Tan CS, Low ICC, Lim N, et al. Heart Rate Measures From Wrist-Worn Activity Trackers in a Laboratory and Free-Living Setting: Validation Study. JMIR mHealth uHealth 2019 Oct 02;7(10):e14120 [FREE Full text] [doi: 10.2196/14120] [Medline: 31579026]

43.   Gjoreski M, Gjoreski H, Luštrek M, Gams M. How Accurately Can Your Wrist Device Recognize Daily Activities and Detect Falls? Sensors (Basel) 2016 Jun 01;16(6) [FREE Full text] [doi: 10.3390/s16060800] [Medline: 27258282]

44.   Dieu O, Mikulovic J, Fardy PS, Bui-Xuan G, Béghin L, Vanhelst J. Physical activity using wrist-worn accelerometers: comparison of dominant and non-dominant wrist. Clin Physiol Funct Imaging 2017 Sep;37(5):525-529. [doi: 10.1111/cpf.12337] [Medline: 26749436]

45.   Gilgen-Ammann R, Schweizer T, Wyss T. Accuracy of the Multisensory Wristwatch Polar Vantage's Estimation of Energy Expenditure in Various Activities: Instrument Validation Study. JMIR mHealth uHealth 2019 Oct 02;7(10):e14534 [FREE Full text] [doi: 10.2196/14534] [Medline: 31579020]

46.   Horton JF, Stergiou P, Fung TS, Katz L. Comparison of Polar M600 Optical Heart Rate and ECG Heart Rate during Exercise. Med Sci Sports Exerc 2017 Dec;49(12):2600-2607. [doi: 10.1249/MSS.0000000000001388] [Medline: 29135785]

47.   Pedišić Z, Bauman A. Accelerometer-based measures in physical activity surveillance: current practices and issues. Br J Sports Med 2015 Feb;49(4):219-223. [doi: 10.1136/bjsports-2013-093407] [Medline: 25370153]

48.   Ayabe M, Kumahara H, Morimura K, Tanaka H. Epoch length and the physical activity bout analysis: an accelerometry research issue. BMC Res Notes 2013 Jan 18;6:20 [FREE Full text] [doi: 10.1186/1756-0500-6-20] [Medline: 23331772]

Abbreviations

bpm: beats per minute
Fizzo: Fizzo Smart Bracelet
HR: heart rate
ICC: intraclass correlation coefficient
LOA: limits of agreement
MAPE: mean absolute percentage error
MD: mean difference
MVPA: moderate to vigorous intensity physical activity
PE: physical education
Polar: Polar Team2 Pro

XSL•FO
RenderX

Original Paper

Respiration Rate Estimation Based on Independent Component Analysis of Accelerometer Data: Pilot Single-Arm Intervention Study

JeeEun Lee[1], MS; Sun K Yoo[2], PhD

[1]Graduate Program of Biomedical Engineering, Yonsei University, Seoul, Republic of Korea

[2]Department of Medical Engineering, Yonsei University College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Sun K Yoo, PhD
Department of Medical Engineering
Yonsei University College of Medicine
Yonsei-ro, Seodaemun-gu
Seoul,
Republic of Korea
Phone: 82 10 3458 2435
Email: sunkyoo@yuhs.ac

Abstract

Background: As the mobile environment has developed recently, there have been studies on continuous respiration monitoring. However, it is not easy for general users to access the sensors typically used to measure respiration. There is also random noise caused by various environmental variables when respiration is measured using noncontact methods in a mobile environment.

Objective: In this study, we aimed to estimate the respiration rate using an accelerometer sensor in a smartphone.

Methods: First, data were acquired from an accelerometer sensor by a smartphone, which can easily be accessed by the general public. Second, an independent component was extracted to calibrate the three-axis accelerometer. Lastly, the respiration rate was estimated using quefrency selection reflecting the harmonic component because respiration has regular patterns.

Results: From April 2018, we enrolled 30 male participants. When the independent component and quefrency selection were used to estimate the respiration rate, the correlation with respiration acquired from a chest belt was 0.7. The statistical results of the Wilcoxon signed-rank test were used to determine whether the differences in the respiration counts acquired from the chest belt and from the accelerometer sensor were significant. The *P* value of the difference in the respiration counts acquired from the two sensors was .27, which was not significant. This indicates that the number of respiration counts measured using the accelerometer sensor was not different from that measured using the chest belt. The Bland-Altman results indicated that the mean difference was 0.43, with less than one breath per minute, and that the respiration rate was at the 95% limits of agreement.

Conclusions: There was no relevant difference in the respiration rate measured using a chest belt and that measured using an accelerometer sensor. The accelerometer sensor approach could solve the problems related to the inconvenience of chest belt attachment and the settings. It could be used to detect sleep apnea through constant respiration rate estimation in an internet-of-things environment.

Introduction

Background

Owing to the rapid recent development of mobile medical monitoring, an increasing number of people desire to manage their health through health care services in daily life. To provide information for health management, it is important to monitor biosignals constantly [1,2]. Among biosignals, respiration is the easiest to measure and provides various types of information about health management based on a number of respiration parameters and respiration patterns [2]. Respiration data collected in daily life can be used as a health care index involving factors such as feedback for sleep apnea syndrome

and other issues [3]. In addition, patients could discover abnormalities early through constant monitoring outside of medical facilities and thereby prevent dangerous escalation of their issues [4].

Currently, the sensors used to measure respiration are for medical and experimental use, so they are not easily accessible for general users and their prices are high [5]. In addition, they are of the contact type, requiring users to attach multiple sensors to the body, which makes them difficult to adopt and difficult to use for constant measurement of respiration [6]. To solve these problems, there have been studies on methods to measure respiration using devices such as accelerometers, radar sensors, and thermal cameras [6]. For use with such noncontact monitoring equipment, smartphones allow easy accessibility for general users and integrate easily with mobile monitoring environments [7]. For these reasons, in this study, the number of respirations was calculated using data from an accelerometer sensor passed to a smartphone, which was suitable for monitoring the respiration rate in daily life.

The accelerometer sensor has three axes and has sensitivity related to the degree of inclination and to the direction in which it is resting [8]. Therefore, axial correction is required to use the values from the accelerometer sensor. The typical method for axial correction is to calculate the root sum square of the magnitude [9]. Other methods include the use of the mean along the z-axis, the magnitude of each axis, the calculation of correlation, and the calculation of the average of peak frequency [10]. However, these methods have limitations for removing the uncertainty inherent with accelerometer sensors when they are used in an indoor environment.

There are several methods to estimate respiration. In particular, respiration has been analyzed in the frequency domain using general fast Fourier transform (FFT) and short-time Fourier transform (STFT) [11]. Since respiration is quasi-periodic [12], the dominant component could be found using FFT and STFT. However, when noise is caused from situations, such as power lines and movement, it changes the dominant frequency component [13]. Additionally, after filtration to divide the frequency domain, the number of respirations has been calculated using correlation analysis of filtered respiration signal peak number counters, a Wiener filter, and autocorrelation [11,14-16]. The Wiener filter estimates the signal through assumption of the noise spectrum. When sensors are attached to the body of a subject, the Wiener filter is frequently used to remove body movement [14]. This preprocessing method is difficult to use for measuring respiration accurately owing to its sensitivity to noises. In particular, when respiration is measured with the noncontact method in a mobile environment, various environmental noises occur. Autocorrelation separates the signal and noise. The correlation of a signal can be estimated using a lagged dependent variable [15]. However, it has limitations because respiration is exactly periodic. To solve the

problem, we used quefrency selection. A respiration signal is a harmonic component because of its quasi-periodic characteristic. Additionally, a search range related to respiration is set to minimize any noise component.

Objectives

An accelerometer sensor in a smartphone was used because it is one of the representative internet-of-things (IoT) devices. To reduce the uncertainty caused by external noise, the vector by which the accelerometer sensor independence is maximized was calculated using independent component analysis (ICA). In addition, respiration has regular patterns, so the respiration rate was estimated using quefrency selection reflecting harmonic information. The respiration rate with the suggested method and the respiration rate with the gold standard using a respiration belt were compared to evaluate significance. When the accelerometer sensor in a smartphone is used as an IoT environment, data can be acquired at various locations. Additionally, the specification of the accelerometer sensor is different according to the smartphone device. For this problem, a case study was performed to determine the difference according to location and smartphone. The aim of the study was to estimate the respiration rate based on ICA of an accelerometer sensor.

Methods

Data Acquisition

Before data acquisition, recruitment notices were posted on a notice board. Only if a subject agreed to the research, the subject joined the study. Additionally, subjects who were likely to participate in this study were not excluded based on social and economic conditions. The study did not register vulnerable subjects. Before recruiting the subjects, the subjects provided information about their health status, medications, and diseases. Subjects who did not have mental or physical diseases joined the experiment.

This is a single-arm intervention study without a control group and without randomization. Data acquisition was performed in a controlled environment by the laboratory. Each subject signed a written consent form prior to the experiment. The respiration signal was collected while the subject reclined on a bed.

The device used to measure respiration was the accelerometer sensor of the Samsung Galaxy S8 smartphone, and nonlinear sampling was performed at an average of 500 Hz. Respiration signals from a chest belt were measured at the same time as actual respiration signals for comparison with the respiration signals calculated using the accelerometer sensor. Respiration measured from the chest belt was sampled at 500 Hz using BIOPAC MP 150TM equipment. As shown in Figure 1, the accelerometer sensor was located near the left shoulder of the subject, and the chest belt was secured across the chest.

Figure 1. Experimental environment. Solid-line box: Samsung Galaxy S8 accelerometer sensor; dashed-line box: chest belt.


ICA-Based Accelerometer Calibration

The accelerometer sensor of Samsung Galaxy S8 has a nonlinear sampling rate. To convert this to fixed sampling, time-stamp and accelerometer sensor values were stored simultaneously. Figure 2 shows the process for conversion of the rate of

sampling by the accelerometer sensor. The stored original data were up-sampled at 1000 Hz, and missing values were filled in using the same values measured at the previous time stamp. The filled-in missing data were down-sampled again at 500 Hz, which is the same sampling rate used with the original data.

Figure 2. Processing for fixed sampling.
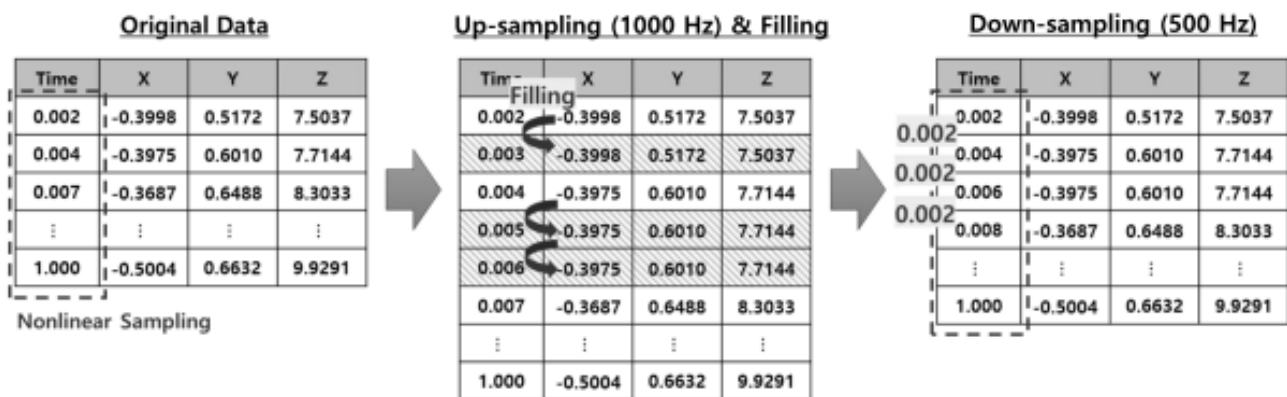
The data measured using the accelerometer sensor contained not only respiration data but also data of various signals, such as motion and external noises. Among these, respiration appeared in the frequency spectrum at less than 0.4 Hz. Thus, respiration signals were preprocessed through a 0.4-Hz low-pass filter [17].

When an accelerometer sensor is used in an IoT environment, a variety of noises exist in the environment. In addition to

respiration, these various noise sources in the environment were also measured (and mixed into the data) by the accelerometer sensor. Therefore, the respiration signals were separated using ICA. Because the original ICA signals were created by other physical processes, it was assumed that they have independent and irregular distributions [18]. Figure 3 presents the ICA model used to distinguish respiration and the signals of the various sources being measured by the accelerometer. It shows the process of separating the sources by estimating U.

Figure 3. Independent component analysis model for respiration separation.
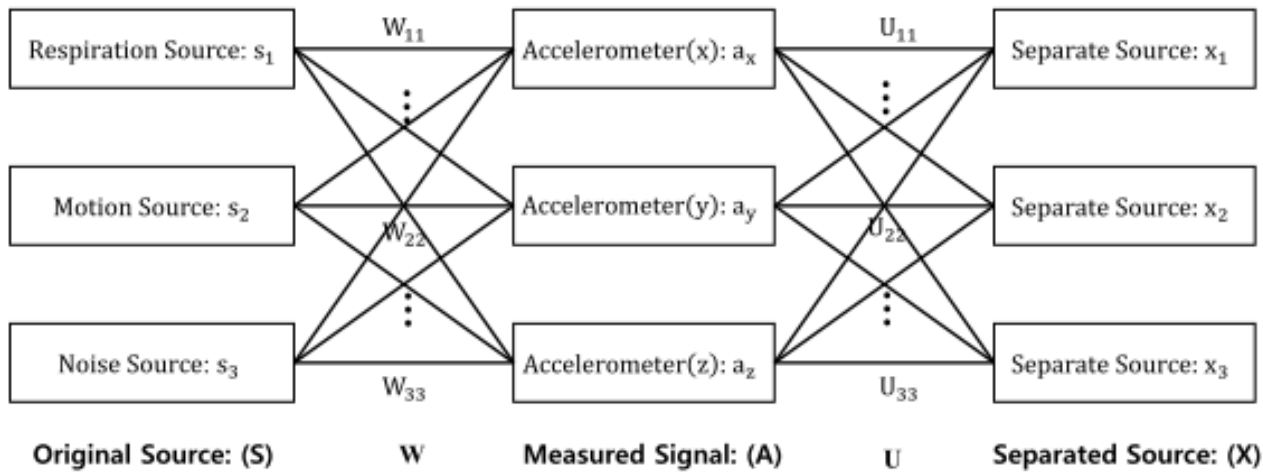
The separation of sources was calculated using equations 1-3. Equation 1 shows the measured signal (*A*). This signal (*A*) shows the vector measured from the three-axis accelerometer sensor according to time (*t*) and has the value $(3 \times [sr \times t])$, where *sr* is the sampling rate. In this study, the sampling rate was 500 Hz. To separate the original signals of the accelerometer sensor, equations 2 and 3 were used. In this case, $U = W^{-1}$ is the mixing matrix, *S* is the original source, and *X* = *S* is the estimated separated source [18].

Quefrency Selection for Respiration Rate Estimation

Respiration is a regular signal, so it is possible to estimate the respiration rate through the estimation of the interval measured by the accelerometer sensor.

For detection of the regularity of *X*, a cepstrum was used. The cepstrum was calculated through inverse Fourier transform. In this way, the harmonic component of signals could be acquired [19,20].

Equation 4 was used for the quefrency selection, which is the harmonic component of *X*, and the term CA_{max_peak} is the point

where the maximum peak of the cepstrum of *X* appears [21]. It is needed to set the search range related to respiration. The respiration-related search range defines searching point (SP). To spot the maximum peak dependent on the respiration signal SP, SP was designated using equation 5. The term RR_{sec} shows the respiration rate per second, and SP was designated as (1251-6000) using equation 5.

In addition, whether the second harmonic exists is detected through the harmonic component adjacent to the search point. The respiration rate was estimated according to the existence of the second harmonic, and RR_{min}, the number of respirations per minute, was calculated using equation 6 [21].

Figure 4 shows a diagram displaying how to estimate the respiration rate. The filtered signals from the accelerometer sensor were separated based on ICA. The independent component analyzed using cepstrum and selected the maximum point at the range related to respiration.

Figure 4. Diagram for respiration rate estimation. ICA: independent component analysis.
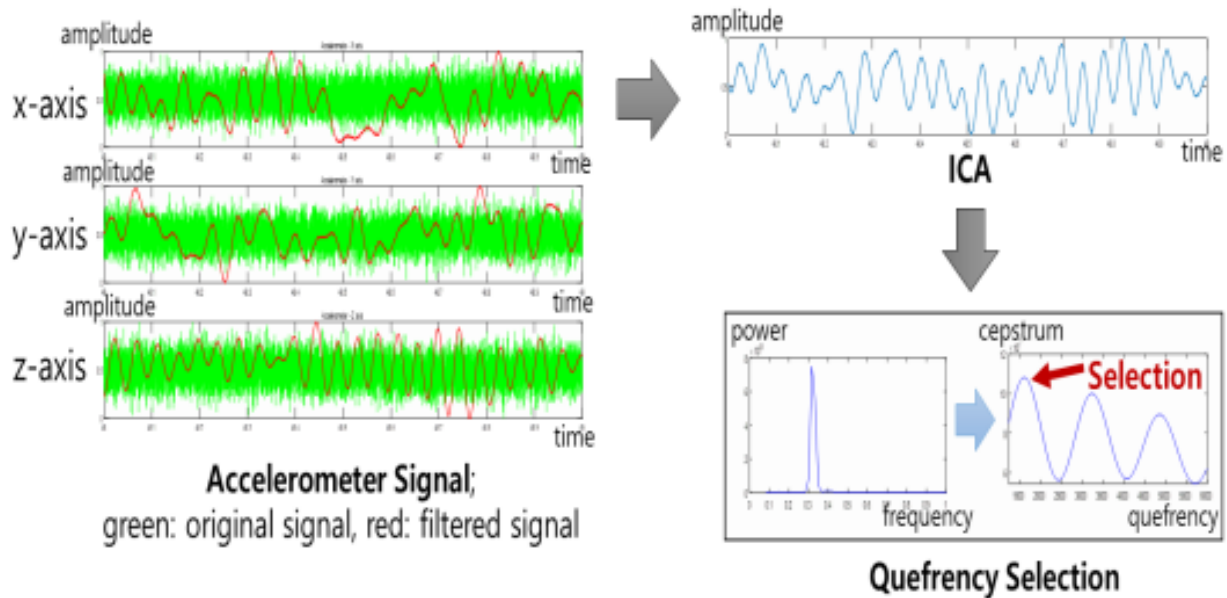
Results

Data Overview

The study was approved by the Institutional Review Board of Yonsei University health system (IRB number: 4-2018-0411). From April 2018, we enrolled 30 male participants. Among the subjects, the mean age was 26.67 (SD 2.41) years, mean height was 173.8 (SD 5.33) cm, and mean weight was 74.43 (SD 9.52) kg.

Comparison With the Accelerometer Calibration Method

For analysis of the method presented in this study, the signals acquired from the chest belt were used as the standard. The chest belt signals were segmented with 1-minute epochs, and the number of respirations per minute was calculated by counting the number of maximum peaks. This study presents an accelerometer calibration method using ICA to extract the original signals from the accelerometer sensor.

For assessment of the method presented here, the Pearson correlation (Pearson r) was used. With Pearson r, the correlation of the number and size of respirations calculated from the actual respiration counts and ICA calibration was determined. Because r is close to −1 or 1, it is very similar to the real respiration rate.

The ICA methods proposed for accelerometer calibration of each axis, the root sum square (RSS) of each axis, and principal component analysis (PCA) methods (which were used previously) were verified using Pearson r. Each axis is affected by tilting and direction [8]. Since the signal is acquired while the subject and smartphone are set on the bed, it is important to confirm which axis is more sensitive after preprocessing such as filtering. If one axis is correlated with respiration, the axis has to be selected to reduce calculation. The RSS is calculated using the square root sum of each axis, and it is frequently used for calibration of the accelerometer sensor [9]. PCA is similar to ICA as one of the dimension reduction methods. It finds new principal axes while preserving the variance and transforms data from a high-dimensional space to a low-dimensional space without linear correlation [22].

Figure 5 shows Pearson r skeletal box-and-whisker plots for respiration rate estimation according to each calibration method. In the skeletal box-and-whisker plots, the red line is the median, and the skeletal box is represented from the first to third quartiles of Pearson r. The minimum and maximum of Pearson r without outliers are shown as whiskers with end caps [23]. When calibrated with ICA, it has the highest correlation (0.7). In addition, ICA has a lower difference in the range between the maximum and minimum values compared with other values, so the result with it is more stable than those with other calibration methods.

Figure 5. Skeletal box-and-whisker plot according to the calibration method. ICA: independent component analysis; PCA: principal component analysis; RSS: root sum square.
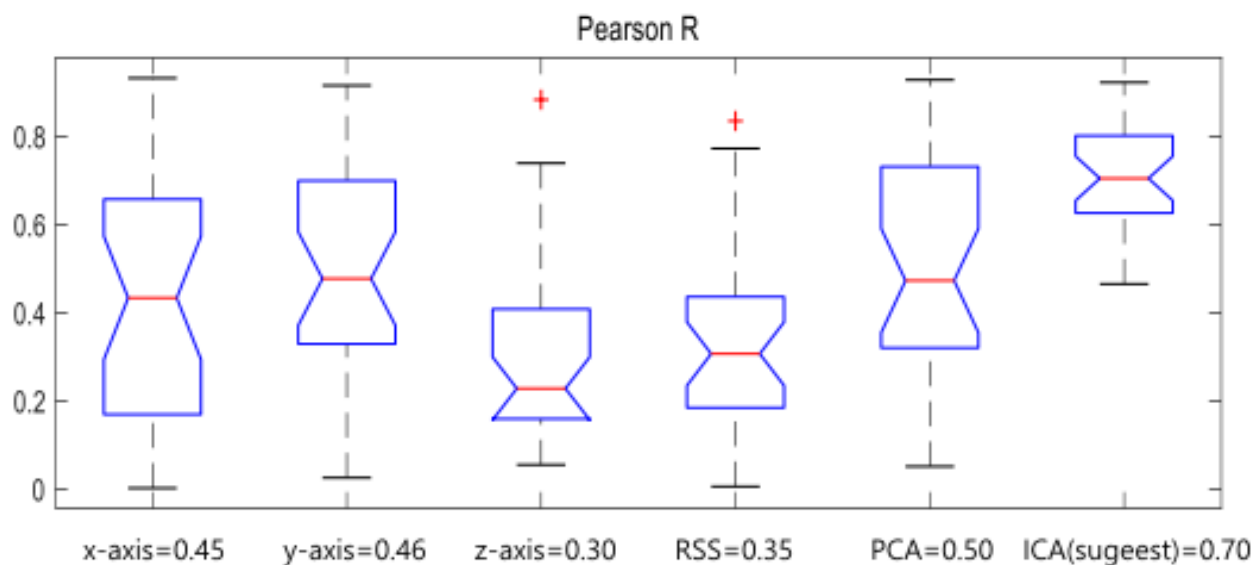
In this study, statistical evaluation was performed on whether the ICA calibration presented with the use of the Pearson r value was significant compared with other methods (x-axis, y-axis, z-axis, RSS, and PCA). Prior to statistical evaluation, visual regularity verification was performed using a quantile-quantile plot. When identified using a quantile-quantile plot, all the data were hard to consider as linear, so a nonparametric test was performed. Because significance was evaluated according to each method using the same subjects, a Wilcoxon signed-rank test was used. Table 1 presents the statistical results of the Wilcoxon signed-rank test. The Pearson r of the respiration rate estimation according to the ICA method was significant ($P<.001$) compared with the other methods.

Table 1. Statistical evaluation of the calibration method using the Wilcoxon rank-sum test.

| Method	W	z	P value
x-axis	647	−3.95	<.001
y-axis	635	−4.13	<.001
z-axis	533	−5.64	<.001
RSS[a]	567	−5.14	<.001
PCA[b]	713	−2.98	.003

[a]RSS: root sum square.

[b]PCA: principal component analysis.

Comparison With Conventional Algorithms

After the accelerometer sensor was calibrated using ICA, conventional methods to extract the respiration rate estimation and the proposed quefrency selection method were verified using Pearson r. The compared conventional methods were processed with filtering of the acquisition data. There were three methods as follows: peak count, spectral peak transition, and autocorrelation. In the three compared methods, the peak count determines the peaks from the ICA calibration data. To estimate the respiration dominant frequency, the highest spectral peak transition is selected using FFT. The selected spectral peak transition shows the signal period. Autocorrelation is calculated using correlation with the lagged values [15].

Figure 6 shows Pearson r skeletal box-and-whisker plots for the respiration rate estimation according to each conventional method. When the respiration rate was estimated by quefrency selection, it was determined to have the highest correlation value (0.7). In addition, there was little difference in the range between the maximum and minimum values compared with other methods, so it was more stable than other conventional algorithms.

Figure 6. Skeletal box-and-whisker plot according to the conventional algorithm. AC: autocorrelation; PC: peak count; SP: spectral peak transition.
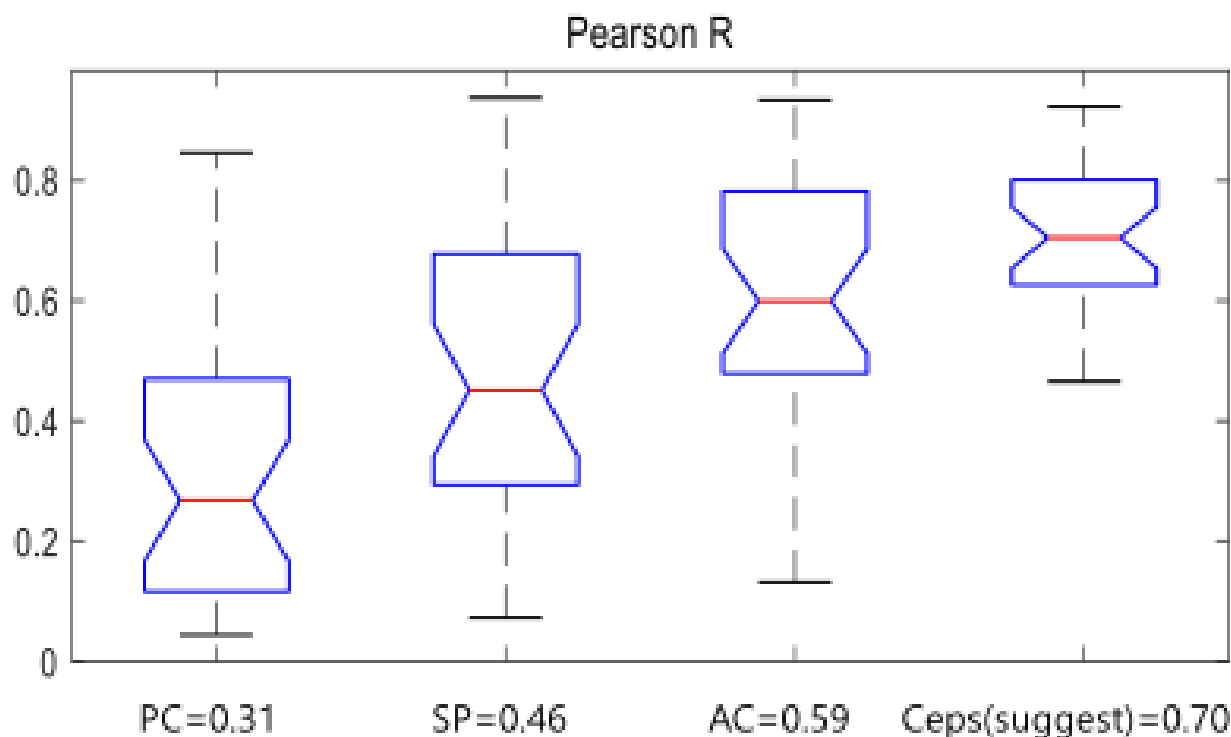
In this study, statistical evaluation was performed to determine whether the quefrency selection presented with use of the Pearson r value was significantly different from the other methods (peak count, spectral peak transition, and autocorrelation). The statistical results of the Wilcoxon signed-rank test in Table 2 show that the respiration rate estimation by the quefrency selection method has more relevant results.

Table 2. Statistical evaluation of conventional algorithms using the Wilcoxon rank-sum test.

Method	W	z	P value
Peak count	521	−5.82	<.001
Spectral peak transition	680	−3.47	<.001
Autocorrelation	774	−2.08	.04

Comparison With the Chest Belt

Table 3 shows the results from comparing the average respiration count per minute acquired using the chest belt ($R.R_{belt}$) with that acquired using the accelerometer sensor ($R.R_{est}$) for each subject. The difference in the respiration rate per minute between the two sensors was at the most two times the respiration count. When the peaks were counted using the chest belt, a peak counting error occurred with the respiratory waveforms of a half cycle at the first and the last respiration. On the other hand, the quefrency selection was independent in respiratory waveforms of a half cycle, so errors were reduced in the detection of the respiration rate. Table 3 shows that there was no significant difference in the performance of respiration rate estimation between the accelerometer sensor and the chest belt (W=842, z=−1.11, P=.27).

Table 3. Results of respiration rate estimation.

Number	$R.R_{belt}$ [a]	$R.R_{est}$ [b]	Difference
1	15	17	2
2	18	18	0
3	16	16	0
4	16	17	1
5	17	17	0
6	18	18	0
7	16	15	1
8	18	18	0
9	17	19	2
10	18	20	2
11	17	18	1
12	15	16	1
13	17	18	1
14	16	17	1
15	18	20	2
16	16	16	0
17	15	15	0
18	15	17	2
19	16	16	0
20	16	17	1
21	17	16	1
22	16	15	1
23	17	16	1
24	16	16	0
25	15	16	1
26	16	17	1
27	15	15	0
28	18	18	0
29	17	16	1
30	17	17	0

[a]$R.R_{belt}$: average respiration count per minute acquired using the chest belt.

[b]$R.R_{est}$: average respiration count per minute acquired using the accelerometer sensor.

In addition, the statistical results of the Wilcoxon signed-rank test were used to determine whether the differences in the respiration count acquired from the chest belt and from the accelerometer sensor were significant. The P value of the difference in the respiration count acquired from the two sensors was .27, which was not significant. This indicates that the number of respiration counts measured using the accelerometer sensor was not different from that measured using the chest belt. Therefore, it is possible to use an accelerometer sensor for estimation of the respiration rate, instead of a chest belt.

A Bland-Altman analysis was performed to evaluate the performance of respiration rate estimation from the accelerometer sensor compared with the respiration belt. The Bland-Altman analysis showed a correlation between the accelerometer sensor and respiration belt (Figure 7A). Figure 7B shows the Bland-Altman results for the mean difference and the bias of the 95% confidence interval [24]. In Figure 7, the correlation coefficient was 0.7. The Bland-Altman results indicate that the mean difference was 0.43, with less than one breath per minute, and that the respiration rate was at the 95% limits of agreement. The accelerometer sensor could produce results ranging from a 2.3 breaths per minute overestimation to 1.5 breaths per minute underestimation.

Figure 7. Bland-Altman results between the respiration belt and accelerometer sensor. $R.R_{est}$: average respiration counts per minute acquired using the accelerometer sensor; $R.R_{belt}$: average respiration counts per minute acquired using the chest belt.
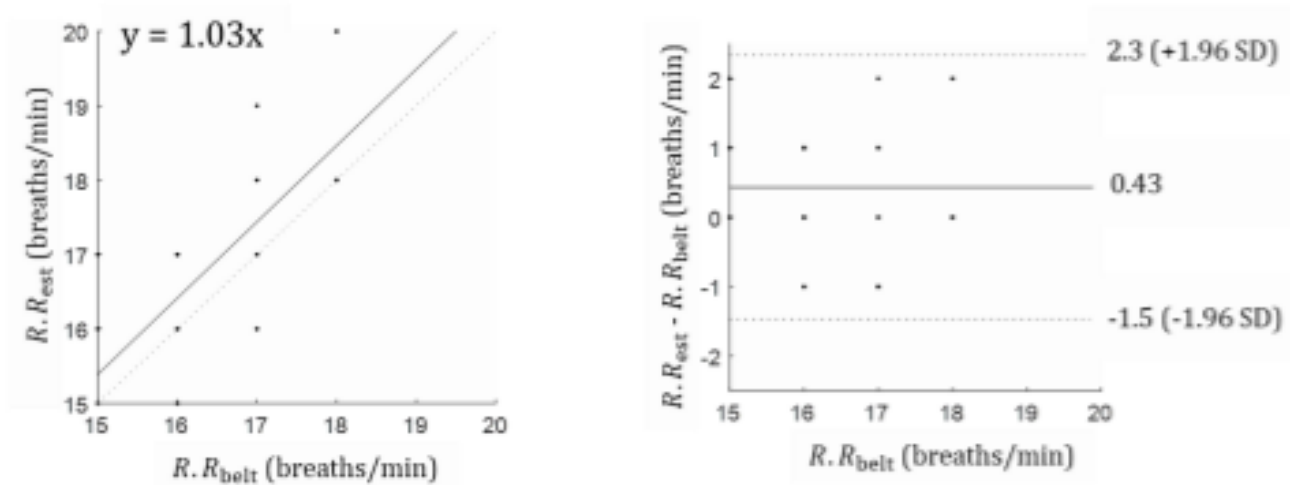
Case Study

Here, Loc_{up} is the case in which the smartphone was located next to the left shoulder of the subject and Loc_{down} is the case in which it was located under the left foot. Table 4 shows the results of respiration estimation according to location. The difference in respiration count per minute by location was at minimum zero times and at maximum one time. In most cases, no difference in the respiration count occurred, which indicates that the quefrency selection method using an accelerometer sensor has low sensitivity to location (W=97, z=−0.65, P=.52).

Table 4. Results of respiration rate estimation by location.

Number	Loc_{up} [a]	Loc_{down} [b]	Difference
1	17	17	0
2	18	19	1
3	15	15	0
4	16	17	1
5	16	15	1
6	16	16	0
7	18	18	0
8	17	17	0
9	18	18	0
10	17	17	0

[a]Loc_{up}: smartphone is located next to the left shoulder.

[b]Loc_{down}: smartphone is located under the left foot.

In Figure 8, the correlation coefficient is 0.93. The Bland-Altman results indicated that the mean difference was 0.1, with less than one breath per minute, and that the respiration rate was at the 95% limits of agreement. The accelerometer sensor could produce results ranging from a 1.2 breaths per minute overestimation to 1.0 breath per minute underestimation.

Figure 8. Bland-Altman results according to sensor location. Loc_{down}: smartphone is located under the left foot; Loc_{up}: smartphone is located next to the left shoulder.
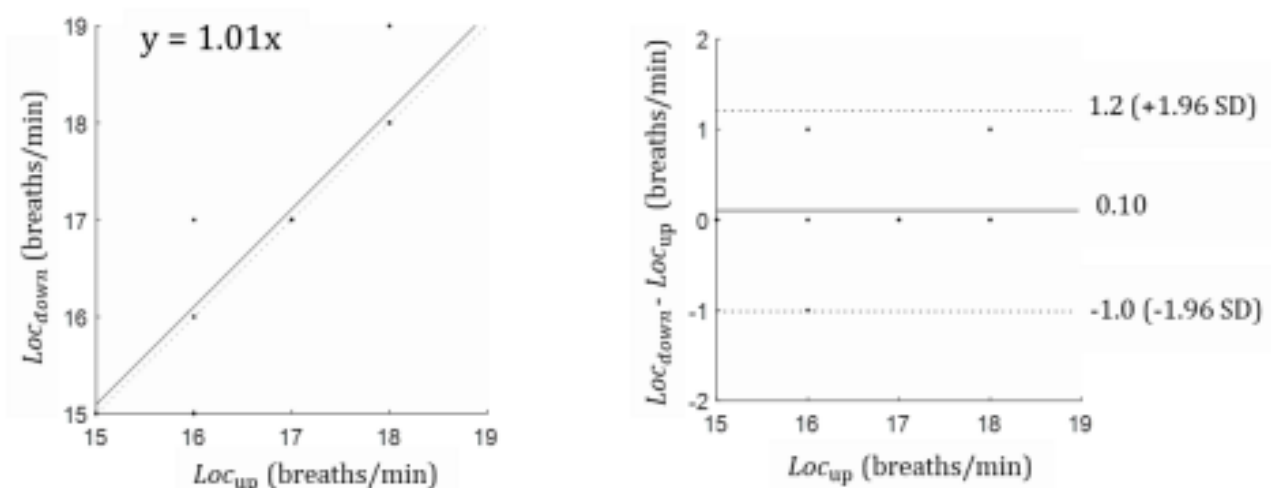
There is a difference in the accelerometer sensor of different types of smartphones. Therefore, to identify the difference according to the type of smartphone, a case study was performed in 10 subjects. The smartphones used for the case study were Samsung Galaxy S8 (smartphone 1) and Samsung Galaxy S7 (smartphone 2). Table 5 shows the results of respiration rate estimation using each smartphone. The difference in the respiration count per minute for each smartphone type was at minimum zero times and at maximum one time, which indicates that there was no relevant difference in the results with these two smartphones.

The Wilcoxon signed-rank test was used to determine statistically whether the estimated respiration count was significantly different according to the type of smartphone and location. The P value of the difference in the estimated respiration count by location was .52, which was not significant. This shows that there was no significant difference in the respiration count estimated at different locations. The P value of the difference in the estimated respiration count by smartphone type was .88, which was also not significant (W=103, z=−0.16). This shows that there was no significant difference in the respiration count estimated using the two different smartphones.

Table 5. Results of respiration rate estimation by smartphone.

| Number	Smartphone 1[a]	Smartphone 2[b]	Difference
1	18	18	0
2	17	18	1
3	18	18	0
4	18	19	1
5	17	17	0
6	17	17	0
7	18	18	0
8	18	18	0
9	18	18	0
10	19	19	0

[a]Smartphone 1: Samsung Galaxy S8.

[b]Smartphone 2: Samsung Galaxy S7.

In Figure 9, the correlation coefficient is 0.79. The Bland-Altman results indicate that the mean difference was 0.2, with less than one breath per minute, and that the respiration rate was at the 95% limits of agreement. The accelerometer sensor could produce results ranging from a 1.0 breath per minute overestimation to 0.63 breaths per minute underestimation.

Figure 9. Bland-Altman results according to smartphone type. Smartphone 1: Samsung Galaxy S8; smartphone 2: Samsung Galaxy S7.
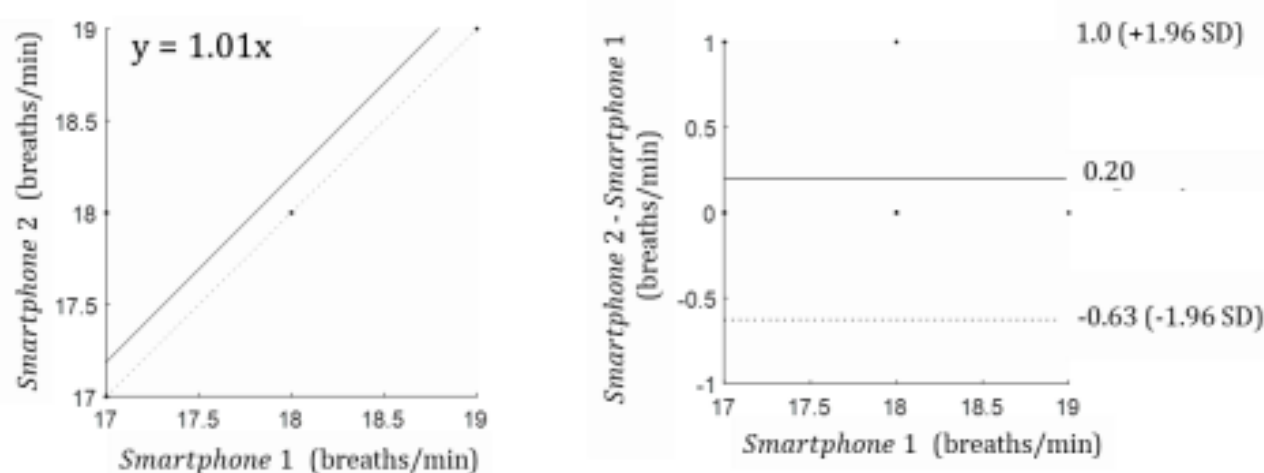
Discussion

Principal Findings

In this study, respiration counts were estimated based on data from an accelerometer sensor in a smartphone. During recent advances in the mobile medical monitoring environment, a variety of smart devices have been developed [2,3]. There is also an increasing desire to manage health by measuring and analyzing data using such smart devices [1,5]. In particular, the development of various wearable devices, such as smartphones and smart bands, provides the potential to acquire various types of health care information such as movement and heartbeat. Respiration is a biosignal directly related to body activity [3,5]. The estimation of respiration can prevent dangerous incidents by predicting diseases and detecting sleep apnea [4]. However, the number of smart devices able to present respiration rates is limited, and the sensors that can accurately detect respiration are not easy to use in daily life. Thus, the purpose of this study was to estimate respiration counts using an accelerometer sensor in a smartphone that is easy to access by normal people.

This study acquired signals using the accelerometer sensor in a smartphone for a long time to identify the feasibility of constantly estimating respiration during natural motion. Because the accelerometer sensor in a smartphone samples nonlinearly, linear sampling was imposed. The accelerometer sensor is a three-axis type, which requires calibration. In this study, components that maximize independence from the signals acquired from the three axes of the accelerometer sensor were distinguished using ICA, and signals showing the range of respiration were extracted. When the statistical significance of the difference in results for the conventional method and Pearson r of the respiration counts estimated from the chest belt and accelerometer sensor were determined, ICA showed a significant result.

Lastly, accelerometer sensors are greatly affected by environmental noises. However, while such noises have random characteristics, respiration has regularity, and this regularity has harmonic components. Thus, respiration rates were estimated using quefrency selection. It was confirmed that there was a relevant difference in the data acquired using Pearson r for the

conventional method and quefrency selection, as well as statistical verification. It was also confirmed that the performance of respiration rate estimation was excellent when quefrency selection was applied to signals acquired from the accelerometer sensor. This was determined through verification of the difference and statistical significance of the number of respiration counts calculated using the chest belt and the proposed (accelerometer) method.

The use of the accelerometer sensor in a smartphone is a noncontact method and has the advantage of constant respiration monitoring. This enables easy measurement of respiration in daily life. However, when it is used in an actual environment, a number of environmental variables exist. Specifically, the same user could put the smartphone at different locations during the measurements, and the performances of embedded sensors could also be different according to smartphone type. Therefore, case studies were performed according to sensor location and smartphone type. The results indicated that respiration rate detection is possible independent of location and smartphone type.

Limitations

The situation in which respiration rate estimation is needed the most is during sleep [3,4]. Diseases can be predicted and emergency situations can be judged by detecting sleep apnea. Therefore, in this study, respiration was estimated while the subjects were lying down. The feasibility for long-term estimation was also confirmed. In the future, signals during sleep could be acquired and analyzed to apply actual respiration rate estimation during sleep. However, the experiment environment was controlled. During the respiration rate estimation, subjects could change position and lay laterally or in the prone position. Therefore, a further study about position change is needed for application of the approach in a real environment.

Conclusions

In this study, respiration rates were estimated using data from the accelerometer sensor of a smartphone as an IoT device. This study showed differentiation of the respiration rate estimation achieved through ICA calibration and quefrency selection. The

respiration estimation sensors that are currently used are not easy to access and not easy to use in daily life owing to the need for multiple sensors with direct contact. However, smartphones are easy to use in daily life and some are equipped with accelerometer sensors, which makes them suitable for respiration rate estimation. There is a difference in performance according to the calibration method used with the accelerometer sensor. The accuracy of the respiration rate estimation can be enhanced when independent components are detected from three-axis ICA signals and quefrency selection is applied. This new approach could solve the problems related to the inconvenience of electrode attachment and equipment settings that affect respiration rate estimation. Furthermore, it could be used to detect sleep apnea through constant respiration rate estimation in an IoT environment.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Kristiani DG, Triwiyanto T, Nugraha PC, Irianto BG, Syaifudin, Titisari D. The Measuring of Vital Signs Using Internet Of Things Technology (Heart Rate And Respiration). 2019 Presented at: 2019 International Seminar on Application for Technology of Information and Communication (iSemantic). IEEE; September 21-22, 2019; Semarang, Jawa Tengah, Indonesia p. 417-422. [doi: 10.1109/isemantic.2019.8884312]

2. Faust-Christmann CA, Taetz B, Zolynski G, Zimmermann T, Bleser G. A Biofeedback App to Instruct Abdominal Breathing (Breathing-Mentor): Pilot Experiment. JMIR Mhealth Uhealth 2019 Sep 30;7(9):e13703 [FREE Full text] [doi: 10.2196/13703] [Medline: 31573919]

3. Schade MM, Bauer CE, Murray BR, Gahan L, Doheny EP, Kilroy H, et al. Sleep Validity of a Non-Contact Bedside Movement and Respiration-Sensing Device. J Clin Sleep Med 2019 Jul 15;15(7):1051-1061. [doi: 10.5664/jcsm.7892] [Medline: 31383243]

4. Nam Y, Kim Y, Lee J. Sleep Monitoring Based on a Tri-Axial Accelerometer and a Pressure Sensor. Sensors (Basel) 2016 May 23;16(5) [FREE Full text] [doi: 10.3390/s16050750] [Medline: 27223290]

5. Bari R, Adams RJ, Rahman MM, Parsons MB, Buder EH, Kumar S. rConverse: Moment by Moment Conversation Detection Using a Mobile Respiration Sensor. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar;2(1) [FREE Full text] [doi: 10.1145/3191734] [Medline: 30417165]

6. Al-Khalidi FQ, Saatchi R, Burke D, Elphick H, Tan S. Respiration rate monitoring methods: a review. Pediatr Pulmonol 2011 Jun;46(6):523-529. [doi: 10.1002/ppul.21416] [Medline: 21560260]

7. Bhattacharya R, Bandyopadhyay N, Kalaivani S. Real time Android app based respiration rate monitor. 2017 Dec 18 Presented at: 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA); April 20-22, 2017; Coimbatore, India. [doi: 10.1109/iceca.2017.8203633]

8. Wannenburg J, Malekian R. Physical Activity Recognition From Smartphone Accelerometer Data for User Context Awareness Sensing. IEEE Trans. Syst. Man Cybern, Syst 2017 Dec;47(12):3142-3149. [doi: 10.1109/tsmc.2016.2562509]

9. Wang F, Chan H, Hsu M, Lin C, Chao P, Chang Y. Threshold-based fall detection using a hybrid of tri-axial accelerometer and gyroscope. Physiol Meas 2018 Oct 11;39(10):105002. [doi: 10.1088/1361-6579/aae0eb] [Medline: 30207983]

10. Liu G, Guo Y, Zhu Q, Huang B, Wang L. Estimation of respiration rate from three-dimensional acceleration data based on body sensor network. Telemed J E Health 2011 Nov;17(9):705-711 [FREE Full text] [doi: 10.1089/tmj.2011.0022] [Medline: 22035321]

11. Novak P, Novak V. Time/frequency mapping of the heart rate, blood pressure and respiratory signals. Med. Biol. Eng. Comput 1993 Mar;31(2):103-110. [doi: 10.1007/bf02446667]

12. Sharma H, Sharma K. ECG-derived respiration using Hermite expansion. Biomedical Signal Processing and Control 2018 Jan;39:312-326. [doi: 10.1016/j.bspc.2017.08.016]

13. Omlin X, Crivelli F, Heinicke L, Zaunseder S, Achermann P, Riener R. Effect of Rocking Movements on Respiration. PLoS One 2016;11(3):e0150581 [FREE Full text] [doi: 10.1371/journal.pone.0150581] [Medline: 26954500]

14. Fedotov AA, Akulov SA, Akulova AS. Motion artifacts reduction in wearable respiratory monitoring device. 2017 Presented at: Joint conference of European Medical and Biological Engineering Conference (EMBEC) and Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC) 2017; June 12-14, 2017; Tampere, Finland p. 1121-1124. [doi: 10.1007/978-981-10-5122-7_280]

15. Kwasniewska A, Ruminski J, Szankin M. Improving Accuracy of Contactless Respiratory RateEstimation by Enhancing Thermal Sequences withDeep Neural Networks. Applied Sciences 2019 Oct 17;9(20):4405. [doi: 10.3390/app9204405]

16.   Huang R. Contact-free breathing rate monitoring with smartphones: a sonar phase approach. Auburn University. 2018. URL: https://etd.auburn.edu/handle/10415/6403 [accessed 2020-06-30]

17.   Norton MP, Nelson FC. Fundamentals of Noise and Vibration Analysis for Engineers. The Journal of the Acoustical Society of America 1990 Oct;88(4):2044-2044. [doi: 10.1121/1.400194]

18.   Kim BS, Yoo SK. Motion Artifact Reduction in Photoplethysmography Using Independent Component Analysis. IEEE Trans. Biomed. Eng 2006 Mar;53(3):566-568. [doi: 10.1109/tbme.2005.869784]

19.   Heo H, Sung D, Lee K. Note onset detection based on harmonic cepstrum regularity. 2013 Presented at: 2013 IEEE International Conference on Multimedia and Expo (ICME); July 15-19, 2013; San Jose, California, USA. [doi: 10.1109/icme.2013.6607461]

20.   Skowronski MD, Shrivastav R, Hunter EJ. Cepstral Peak Sensitivity: A Theoretic Analysis and Comparison of Several Implementations. J Voice 2015 Nov;29(6):670-681 [FREE Full text] [doi: 10.1016/j.jvoice.2014.11.005] [Medline: 25944288]

21.   Lee J, Yoo SK. Radar-Based Detection of Respiration Rate with Adaptive Harmonic Quefrency Selection. Sensors (Basel) 2020 Mar 13;20(6) [FREE Full text] [doi: 10.3390/s20061607] [Medline: 32183139]

22.   Bates A, Ling MJ, Mann J, Arvind DK. Respiratory rate and flow waveform estimation from tri-axial accelerometer data. 2010 Presented at: 2010 International Conference on Body Sensor Networks; June 7-9, 2010; Singapore. [doi: 10.1109/bsn.2010.50]

23.   Spitzer M, Wildenhain J, Rappsilber J, Tyers M. BoxPlotR: a web tool for generation of box plots. Nat Methods 2014 Mar;11(2):121-122 [FREE Full text] [doi: 10.1038/nmeth.2811] [Medline: 24481215]

24.   Martin Bland J, Altman D. Statistical Methods For Assessing Agreement Between Two Methods of Clinical Measurement. The Lancet 1986 Feb;327(8476):307-310. [doi: 10.1016/s0140-6736(86)90837-8]

Abbreviations

FFT: fast Fourier transform
ICA: independent component analysis
IoT: internet-of-things
PCA: principal component analysis
RSS: root sum square
STFT: short-time Fourier transform

XSL•FO
RenderX

<u>Corrigenda and Addenda</u>

Correction: Effectiveness of Mobile App-Assisted Self-Care Interventions for Improving Patient Outcomes in Type 2 Diabetes and/or Hypertension: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Kaifeng Liu[1], PhD; Zhenzhen Xie[1], MSc; Calvin Kalun Or[1], PhD

Department of Industrial and Manufacturing Systems Engineering, University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Calvin Kalun Or, PhD
Department of Industrial and Manufacturing Systems Engineering
University of Hong Kong
Room 8-7, 8/f., Haking Wong Building,
University of Hong Kong
Hong Kong,
China (Hong Kong)
Phone: 852 39172587
Fax: 852 28586535
Email: klor@hku.hk

Related Article:

Correction of: https://mhealth.jmir.org/2020/8/e15779/

In "Effectiveness of Mobile App-Assisted Self-Care Interventions for Improving Patient Outcomes in Type 2 Diabetes and/or Hypertension: Systematic Review and Meta-Analysis of Randomized Controlled Trials" (JMIR mHealth 2020;8(8):e15779) the authors noted an error.

In Table 4, the symbol/indicator for "Logging–Medication" and "HbA$_{1c}$ reduction" should be a solid dot (referring to footnote g) instead of a cross (referring to footnote i).

The correction will appear in the online version of the paper on the JMIR Publications website on August 19, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

XSL•FO
RenderX

XSL•FO

RenderX

<u>Original Paper</u>

Using mHealth to Increase the Reach of Local Guidance to Health Professionals as Part of an Institutional Response Plan to the COVID-19 Outbreak: Usage Analysis Study

Olivier Windisch[1*], MD; Ido Zamberg[2,3,4*], MD; Marie-Céline Zanella[5], MD; Angèle Gayet-Ageron[3,6], MD, PhD; Katherine Blondon[3,7], MD, PhD; Eduardo Schiffer[3,8], MD, PhD; Thomas Agoritsas[2,9], MD, PhD

[1]Division of Urology, Department of Surgery, University Hospitals of Geneva, Geneva, Switzerland

[2]Division of General Internal Medicine, Department of Medicine, University Hospitals of Geneva, Geneva, Switzerland

[3]Faculty of Medicine, University of Geneva, Geneva, Switzerland

[4]School of Education, Johns Hopkins University, Baltimore, MD, United States

[5]Division of Infectious Diseases, Department of Medicine, University Hospitals of Geneva, Geneva, Switzerland

[6]Division of Clinical Epidemiology, Department of Community Health and Medicine, University Hospitals of Geneva, Geneva, Switzerland

[7]Medical Directorate, University Hospitals of Geneva, Geneva, Switzerland

[8]Division of Anesthesiology, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University Hospitals of Geneva, Geneva, Switzerland

[9]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Olivier Windisch, MD
Division of Urology
Department of Surgery
University Hospitals of Geneva
Rue Gabrielle-Perret-Gentil
Geneva
Switzerland
Phone: 41 79 777 78 89
Email: olivier.windisch@hcuge.ch

Abstract

Background: The ongoing coronavirus disease (COVID-19) pandemic forced health jurisdictions worldwide to significantly restructure and reorganize their medical activities. In response to the rapidly evolving body of evidence, a solid communication strategy is needed to increase the reach of and adherence to locally drafted and validated guidance to aide medical staff with COVID-19–related clinical decisions.

Objective: We present a usage analysis of a dedicated mobile health (mHealth) platform as part of an institutional knowledge dissemination strategy of COVID-19–related guidance to all health care workers (HCWs) in a large academic hospital.

Methods: A multidisciplinary team of experts drafted local guidance related to COVID-19. In total, 60 documents and 17 external links were made available through the platform. Documents were disseminated using a recently deployed mHealth platform for HCWs. Targeted dissemination of COVID-19–related content began on March 22, 2020. Using a third-party statistics tool, data concerning user activity and content use was anonymously collected. A quantitative analysis of user activity was performed over a 4-month period, separated into 3 periods: 2 months before (Period A), 2 weeks after (Period B), and 6 weeks following (Period C) targeted dissemination. Regional epidemiological data (daily new COVID-19 cases and total COVID-19–related hospitalizations) was extracted from an official registry.

Results: During the study period, the platform was downloaded by 1233 new users. Consequently, the total number of users increased from 1766 users before Period A to a total of 2999 users at the end of Period C. We observed 27,046 document views, of which 12,728 (47.1%) were COVID-19–related. The highest increase in activity occurred in Period B, rapidly following targeted dissemination, with 7740 COVID-19–related content views, representing 71.2% of total content views within the abovementioned period and 550 daily views of COVID-19–related documents. Total documents consulted per day increased

from 117 (IQR 74-160) to 657 (IQR 481-1051), *P*<.001. This increase in activity followed the epidemiological curbing of newly diagnosed COVID-19 cases, which peaked during Period B. Total active devices doubled from 684 to 1400, daily user activity increased fourfold, and the number of active devices rose from 53 (IQR 40-70) to 210 (IQR 167-297), *P*<.001. In addition, the number of sessions per day rose from 166 (IQR 110-246) to 704 (IQR 517-1028), *P*<.001. A persistent but reduced increase in total documents consulted per day (172 [IQR 131-251] versus 117 [IQR 74-160], *P*<.001) and active devices (71 [IQR 64-89] versus 53 [IQR 40-70]) was observed in Period C compared to Period A, while only 29.8% of the content accessed was COVID-19–related. After targeted dissemination, an immediate increase in activity was observed after push notifications were sent to users.

Conclusions: The use of an mHealth solution to disseminate time-sensitive medical knowledge seemed to be an effective solution to increase the reach of validated content to a targeted audience.

KEYWORDS

Introduction

Since the World Health Organization (WHO) declared the novel coronavirus (severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2]) a pandemic on March 11, 2020, it has spread rapidly worldwide. As of July 26, 2020, there were more than 15.7 million confirmed cases and more than 640,000 COVID-19–related deaths [1,2]. Authorities and hospitals worldwide have braced themselves to face the unprecedented challenge of providing adequate patient care while trying to avoid health care system saturation and guaranteeing maximum protection for health care workers (HCWs) from nosocomial infections and psychological consequences [3,4].

To respond to the present challenge, health institutions worldwide have been forced to deeply reorganize and reprioritize their daily activities. Hospitals have been restructured to better streamline the increasing number of patients and to separate suspected and confirmed COVID-19 cases from patients without COVID-19 to avoid nosocomial infections [5]. In certain countries, field hospitals and drive-through screening facilities have been built to reduce the load on local health institutions and decrease the community transmission rate [6,7]. Medical and surgical units have been converted for the care of patients with COVID-19 and intensive care units as well as operating rooms were reshaped to prepare for a large number of patients in need of acute respiratory support [8]. Medical staff have also been delocalized or newly recruited to help in such units, with very little initial training on the management of COVID-19. In parallel, local, national, and international health institutions have dealt with important scientific knowledge gaps concerning SARS-CoV-2 infection vectors, transmission dynamics, epidemiology, and treatment possibilities [9].

At the University Hospitals of Geneva (HUG), all nonemergency surgeries were cancelled to mobilize anesthesiology staff to the intensive care unit and to spare essential medical equipment. Outpatient clinic activity was reduced significantly and appointments were conducted via telephone and telemedicine consultations when feasible. Non–COVID-19 medical and surgical activities were transferred to other local hospitals, and COVID-19–related activity was centralized at the authors' institution. New dedicated wards were created, and others were

transformed to admit patients with SARS-CoV-2 under the responsibility of the General Internal Medicine Division. More than 120 medical and surgical staff from a large number of departments (including gynecology and pediatrics) were recruited to the abovementioned units, as well as to the Intensive Care and Emergency Departments to respond to the increasing number of COVID-19–related hospitalizations and consultations. In addition, military paramedical personnel and advanced medical students were recruited to support patient care.

To respond to the rapidly evolving body of evidence and staff training needs, a local multidisciplinary team of experts was formed, with the mission to critically appraise and adopt published evidence and produce institutional guidance adapted to local epidemiology and resources. As new and sometimes contradictory COVID-19–related evidence was published on a daily basis, the guidance was updated frequently as new information became available. Thus, there was a crucial need to be able to disseminate the drafted guidance to medical staff and to rapidly communicate any updates.

Communication between health institutions and health professionals is critical, and many steps exist in the transmission of information from health authorities and local medical leadership to health care providers. Previous research suggests there is a discrepancy between the perceived ability of health authorities or medical leadership to disseminate information and the actual delivery of information to the targeted audience [10,11]. A solid communication strategy is therefore crucial, especially in a pandemic context, where information is time-sensitive, and evidence is constantly and rapidly evolving [12]. Recruiting medical staff from different clinical backgrounds and medical students, who are possibly unfamiliar with the equipment and specific procedures used in COVID-19 care, means that more supervision and training is required, further underscoring the need for readily available validated information. Mobile health technology solutions, referred to as mHealth, are increasingly used and gaining relevance among health professionals, and may represent an interesting solution to disseminate information to health care providers [13,14]. Such mHealth solutions are not only used to disseminate information to HCWs, but also to patients or guardians. One example is ICNexchange, which is a Pinterest-based platform that provides medical information to pediatric patients under

treatment for inflammatory bowel disease. ICNexchange's feasibility, acceptability, and utility was shown in a recent quantitative analysis of user activity [15].

We present here an assessment of an mHealth platform used to disseminate institutional COVID-19–related guidance to medical staff in our institution during the pandemic as part of hospital's reorganization and knowledge dissemination strategy. Our hypothesis was that the use of such an mHealth solution would increase the reach of validated knowledge to HCWs. Since adequate reach to HCWs is difficult to assess, we performed a quantitative assessment using real-time and anonymous statistics to assess user activity and content use before, shortly after, and over an extended period following the targeted dissemination of information on the platform.

Methods

mHealth Solution Description

Our team developed a mobile smartphone platform called HeadToToe to help medical staff easily access locally validated and endorsed medical content [16,17]. It provides an institutional knowledge dissemination solution and has been organized by medical specialty to offer internationally and locally validated and up-to-date medical guidance in the form of documents, videos, and clinical scores, previously selected and validated by senior physicians. The platform offers an administration interface allowing easy content updating; updates are promptly and automatically available for users. All content has a planned obsolescence (ie, it will expire after a predefined date) set by the content's owner, guaranteeing that all available content on the platform is up-to-date. In addition, automatic and anonymous real-time statistics are collected with Yahoo Flurry [18], which provides data concerning user activity and content viewing patterns. A detailed description of the instrument's development process and features is described in previous publications [16,17]. The platform is available on iOS and Android and both versions provide the same features and user experience to HCWs.

Guidelines Team Activity

Due to the urgency of the current situation and abovementioned scientific knowledge gaps, and as part of the institutional response plan to COVID-19, a multidisciplinary team of experts was created, which was independent from the mHealth platform team. The institutional team's mission was that of critically appraising the available information on COVID-19 and creating trustworthy, actionable, and evidence-based local guidance for the hospital's HCWs considering local epidemiology and resources. Topics covered were local containment measures, HCW protection procedures, diagnostic criteria, testing strategies, patient orientation and triage, inpatient management and treatment strategies, outpatient follow-up and management, pharmacological considerations and interactions, and specific information for patients. As new studies and information were published on a daily basis, guidance was frequently modified and updated in accordance with recent evidence.

The objective of the team was to disseminate information to all of the hospital's HCWs to ensure, to the best of their abilities,

HCWs were adequately informed and there were no unwarranted differences in patient care.

COVID-19 Guidance Dissemination

COVID-19–related content was progressively made available through a HUG website and the mHealth platform. After a successful 18-day pilot implementation in the Children's Hospital of the HUG [12], institutional leadership decided to deploy the platform to all medical departments on March 22, 2020.

At the time of analysis, a total of 60 local procedures were drafted and made available through the platform with an additional 17 links to governmental websites and other validated resources.

The guidance topics distribution was as follows: 4 procedures concerning general overview and identification of suspected cases, 3 concerning orientation strategies, 13 concerning inpatient management, 21 procedures related to pharmacological considerations, 10 for acute care management, 6 for outpatient management, and 3 procedures related to COVID-19–related death.

Data Analysis and Statistical Methods

User activity and content viewing patterns were compared between three distinct periods. Period A was defined as the 2-month period preceding the targeted dissemination of COVID-19–related content in the institution (from January 21, 2020, to March 21, 2020). Period B was defined as the 14-day period beginning on the day of targeted dissemination, in parallel with the hospital reorganization and reallocation of medical staff (from March 22, 2020, to April 4, 2020). Period C was an extended period of 6 additional weeks after targeted dissemination (from April 5, 2020, to May 21, 2020). Since our mHealth solution was already used during Period A in the Children's Hospital, pediatric documents regarding COVID-19 were already available on the platform at the time of targeted dissemination. Data retrieved from Yahoo Flurry included the number of active devices, new device installations (defined further as new devices), and total sessions, as well as data concerning content use. To assess the use of COVID-19–related documents, we collected and classified all documents used during the three periods, and manually identified which were COVID-19–related. Local epidemiologic data was extracted from an official and public national registry of the number of new daily cases of COVID-19 and total hospitalizations in the canton of Geneva, Switzerland [19]. It should be noted that all regional confirmed cases of COVID-19 requiring hospitalization were centralized in our institution (HUG). Epidemiological data were graphically presented and superimposed on the platform's activity. Data are presented by their mean, median, and interquartile range (IQR) for continuous variables; relative frequencies and percentages are used for categorical variables. We compared data between the time periods using a nonparametric Mann-Whitney U test. Stata (Version 16, StataCorp LLC) was used for all statistical analysis [20]. Significance refers to a P value $<.05$.

Results

By the end of the study, 1236 new users downloaded the app. The total number of users increased from 1766 users before Period A to 2999 users at the end of Period C. In total, documents were consulted 27,046 times during the study period, of which 12,728 (47.1%) were COVID-19–related.

During Period B, shortly following targeted dissemination, the mobile app was downloaded 912 times and the total active devices using the platform doubled. User activity increased significantly, observed by a four-fold increase in the number of active devices per day (53 versus 211, $P<.001$) and the number of sessions per day (166 versus 704, $P<.001$). The average time spent on the app per device per day increased by 24% ($P=.02$). Table 1 provides additional information on app use. Figure 1 shows a significant increase in document views per day (117 versus 657, $P<.001$), of which 70.5% were COVID-19–related.

Table 1. General data concerning app use[a].

App use variables	Period A, median (IQR)	Period B, median (IQR)	Change (B vs A), %	P value (B vs A)	Period C, median (IQR)	Change (C vs A), %	P value (C vs A)
Daily document views	114 (74-160)	657 (481-1051)	+671	<.001	172	+51	<.001
Active devices per day	53 (40-70)	211 (167-297)	+296	<.001	71 (64-89)	+34	<.001
Number of sessions per day	166 (110-246)	704 (517-1028)	+322	<.001	225 (185-282)	+36	<.001
New devices per day	2 (1-5)	26 (13-56)	+1200	<.001	2 (1-4)	0	.75
Time per device per day (minutes)	4.0 (3.1-5.1)	5.0 (4.5-5.6)	+24	.02	3.9 (3.3-4.7)	–3	.88

[a]Period A is defined as the 2-month period preceding targeted dissemination (from January 21, 2020, to March 21, 2020). Period B is defined as a 14-day period beginning on the day of targeted dissemination (from March 22, 2020 to April 4, 2020). Period C is defined as a 6-week period (from April 5, 2020 to May 21, 2020) following targeted dissemination. Where the P value was <.05 in a two-sided test, significance was considered reached.

Figure 1. Total and coronavirus disease–related document views distribution within all three observation periods. An asterisk indicates a significant increase when compared to Period A ($P<.01$). COVID-19: coronavirus disease.
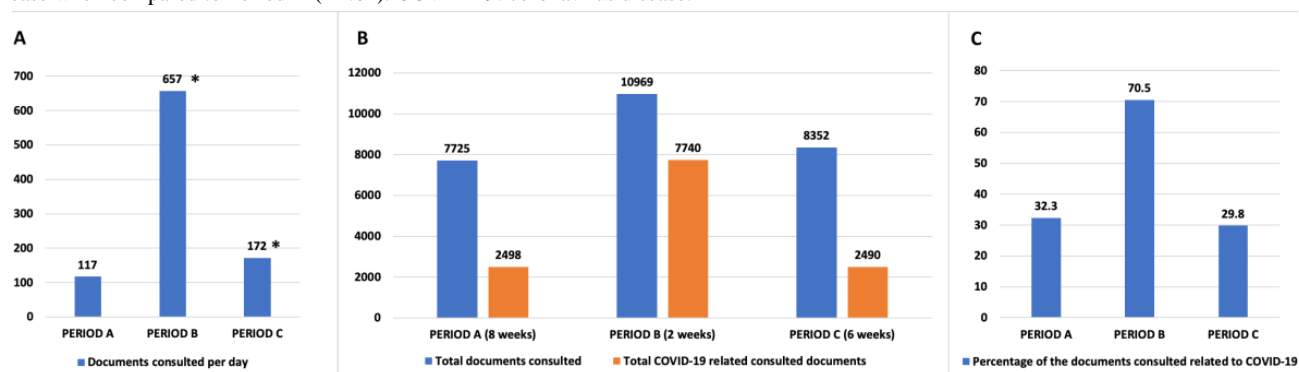
In Period C, there were an additional 136 downloads of the mHealth platform. User activity had a persistent increase in Period C compared to Period A, as measured by total active devices during the period (1029 in Period C versus 684 in Period A), with a 35% increase in daily active devices (71 in Period C versus 53 in Period A, $P<.001$), and number of sessions per day (225 versus 166, $P<.001$). Figure 1 shows a persistent increase in documents consulted per day compared to Period A (172 in Period C versus 117 in Period A, $P<.001$), while only 29.8% of the total documents consulted were COVID-19–related. Global user activity was reduced compared to Period B as shown in Table 1. Table 2 provides additional details concerning new and active devices and reveals that a higher proportion of users used the iOS app throughout all periods.

Table 2. New and active devices and associated operating systems.

Device details	Period A (before dissemination), n (%)	Period B (Weeks 1-2 since dissemination), n (%)	Change, %	Period C (Weeks 3-8 after dissemination), n (%)	Change, %
Active devices					
Total	684 (100)	1400 (100)	+105	1029 (100)	+50
iOS	464 (67.8)	892 (63.7)	+92	705 (68.5)	+52
Android	220 (32.2)	508 (36.3)	+131	324 (31.5)	+47
New devices					
Total	185 (100)	912 (100)	+392	136 (100)	–32
iOS	121 (65.4)	597 (65.5)	+393	90 (66.2)	–26
Android	64 (34.6)	313 (34.5)	+389	46 (33.8)	–28

Figure 2 shows mHealth platform use compared with COVID-19 local epidemiological data. Maximum platform activity was observed the day following targeted dissemination, with an increased use that paralleled the increase in newly diagnosed cases. An increase in user activity, measured by active devices and sessions per day, was observed when push notifications were sent to inform users of new or updated content related to COVID-19.

Figure 2. User activity compared with local coronavirus disease (COVID-19) epidemiological data.
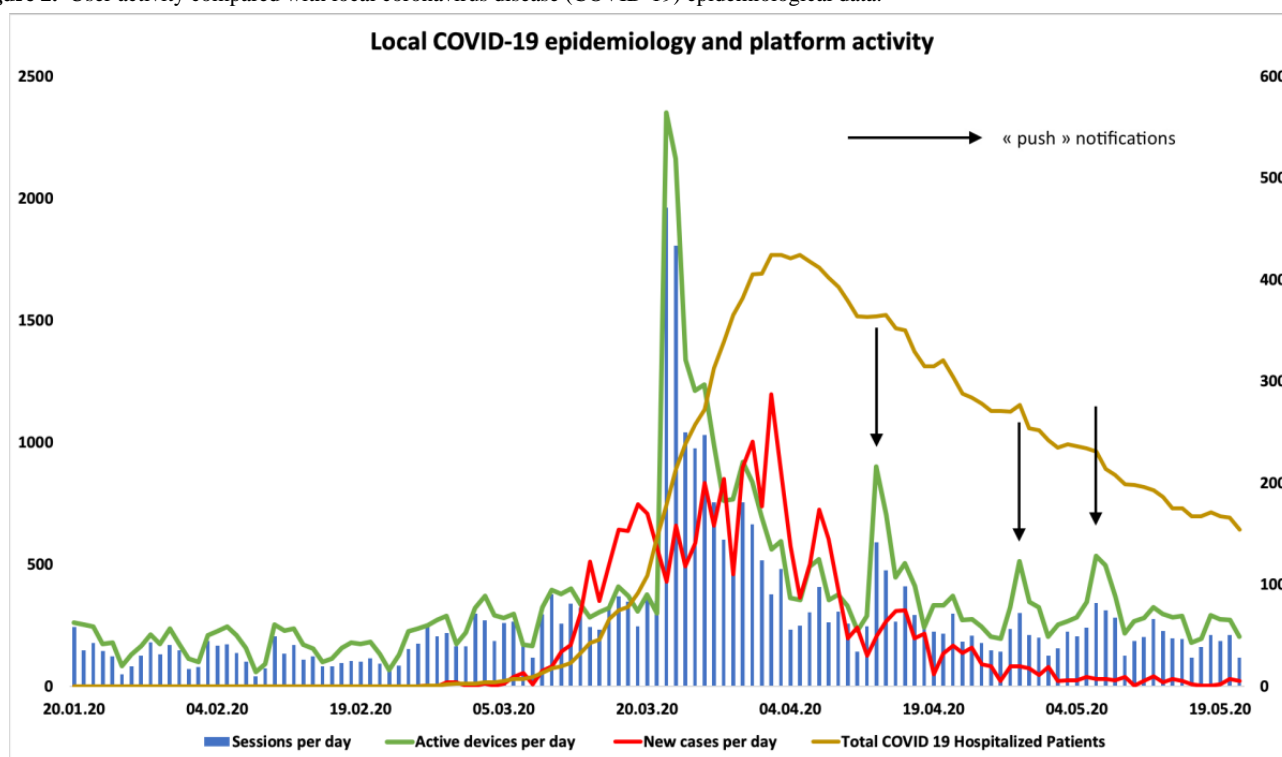
Discussion

Principal Findings

As part of the response plan to the COVID-19 crisis, our study focused on disseminating locally endorsed and validated information for HCWs during the pandemic. A locally developed mHealth solution was adopted by institutional leadership as an information dissemination channel for COVID-19–related content, following a recent successful pilot study in the Children's Hospital of our institution [12]. In this quantitative study, we observed a significant increase in user activity immediately after targeted dissemination. During a period of 2 weeks, 912 people downloaded the mHealth platform and user

activity, measured by the number of users and sessions per day, increased fourfold, showing a significant increase from baseline activity. The increase in user activity was especially notable in the first 2 weeks following targeted dissemination, with HCWs showing a high interest in COVID-19–related content; in total, 71.2% (n=7740) of content consulted during this period was COVID-19–related, an average of 550 documents per day.

Variations in user activity between observation periods could be explained by local epidemiology and circumstances. In fact, our institution centralized all regional COVID-19–related care and paused all nonurgent surgical and outpatient activity. Thus, the main daily clinical activity in our institution during Period B was COVID-19–related. As shown in Figure 2, this period happened early in the local epidemiology, when the number of

new daily cases and new hospitalized patients were highest, and seemed to correspond to the highest user activity on the platform. Therefore, we postulate that increased user activity in this period was related to the acute need of HCWs to assimilate knowledge concerning COVID-19–related care. Subsequently, the standardized aspect of COVID-19–related care combined with the lack of other types of medical activity and the decrease in new cases and hospitalizations, as well as the experience and knowledge gained by HCWs during previous weeks could explain the decrease in user activity in Period C. Interestingly, during Period C, user activity stayed significantly higher than before targeted dissemination of the platform (Period A). This relative increase could be explained by the increased number of total users and might show that HCWs adopted the platform as a knowledge source. The latter might be confirmed by the fact that the percentage of COVID-19–related documents viewed decreased during this period, amounting to only 29.8% of total content viewed, while the number of hospitalized patients remained stable. This might suggest that HCWs had a general interest in the content offered on the platform.

The COVID-19 pandemic presents several unprecedented challenges to international and local governments as well as to health authorities and institutions. Prevention of infection, reduction of the transmission rate, and adequate management of patients with COVID-19 in the context of a constantly evolving body of evidence are currently a priority for jurisdictions worldwide. An infodemic has been occurring in parallel with the pandemic, as unfiltered information is easily accessible, which may allow fake news and false rumors to spread. In its last situation report, the WHO warned about fake products used for diagnosis, prevention, and treatment of COVID-19, showing an urgent need for better awareness among the general public, as well as HCWs [21].

Hospital reorganization and restructuring created a reality where a large number of HCWs from different clinical backgrounds were required to learn a large amount of new clinical procedures and practices in a short period of time. In addition, medical students worldwide have played an active role in the response to this crisis [22], and represent a special population of HCWs that have not completed their training, and thus might require more supervision and could benefit from easily accessible validated information [12].

A solid communication strategy requires an assessment of its efficiency, which can be difficult to achieve since this requires feedback from the sender on how the dissemination process was perceived and from the targeted audience on whether and how the information was received. A high discrepancy rate was reported during the H1N1 crisis, with 81% of local health departments perceiving their ability to disseminate information as very good or excellent, while only 52% of surveyed physicians reported receiving information, and as few as 16% reported using this information [11]. A systematic review showed that same discrepancy and showed that the effectiveness of communication between health authorities or leadership and HCWs is not well documented in the literature and needs to be evaluated in a more rigorous manner [10]. The same team found that email was a preferred communication method compared

to fax, SMS text messaging, or no message at all, but did not investigate an mHealth solution at that point [12].

mHealth solutions are increasingly used, gaining relevance among health professionals, and may represent an interesting solution to disseminate information to and among HCWs [13-15]. Other technological initiatives have been described during the COVID-19 pandemic, such as the elaboration of an electronic health record tool to support clinical care for patients with COVID-19 and to monitor different clinical characteristics (case identification, isolation procedures, adherence to patient screening, real-time data sharing) in their hospital [23]. This type of initiative underlines the role of evolving technology, especially when it concerns time-sensitive information, and emphasizes the need for quick reactions from medical leadership and information technology staff.

The role of our mHealth platform has been previously described for students' clinical training and exam preparation as well as for medical residents' training [16]. In a recent quantitative and qualitative analysis involving HCWs of the Children's Hospital in HUG completed at the beginning of the COVID-19 epidemic in Switzerland, the platform appeared to be of value and it was rapidly implemented and seen as time-effective and informative [12]. In the current study, we confirm that the platform was effectively used during Period A, when 31.2% of documents were COVID-19–related. We also showed that an mHealth solution could offer stakeholders quantitative feedback on user activity, which might be a valid and appropriate method for measuring the reach of information. Indeed, in this study, 2 weeks after the large-scale drafting and dissemination of local guidance, user activity and content use increased significantly. This increased user activity and the high percentage of COVID-19–related content use may be further evidence of the pertinence of mHealth solutions as efficient communication tools to increase the reach of validated knowledge to HCWs.

Push notifications are another interesting feature of mHealth solutions, and could have the potential to increase the reach of information as well as knowledge updates for HCWs. In our institution, as illustrated in Figure 2, user activity increased when push notifications concerning updates and new content were sent. This is a major advantage of mHealth solutions compared to classic communications methods, as it might reduce the need for HCWs to actively seek new and updated information and increase the visibility of knowledge updates.

Another potentially interesting aspect of our platform is that many of our medical students were already in possession of the mobile app and familiar with accessing local recommendations, which might have helped with their integration into clinical practice.

Finally, thanks to increased content visibility, a larger amount of medical content was made available through the platform and more medical and paramedical specialized divisions in our institution requested to share their own COVID-19–related content about specific and specialized clinical situations.

The main limitations of this study are the lack of assessment of adherence to the drafted guidance and the lack of qualitative user feedback, both of which could be considered important

XSL•FO

RenderX

markers of successful reach. A qualitative assessment of the platform was recently performed in the Children's Hospital of the HUG, which showed that HCWs felt reassured by content dissemination through the platform, found it time-efficient, and had less need to seek other information sources [12]; however, an assessment of adherence to the guidance needs to be included in future research. Nevertheless, we believe that sustained user activity, a high percentage of COVID-19–related content use, and increased activity observed shortly after push notifications suggest that disseminated information properly reached medical staff. Moreover, we believe that further validation of knowledge reach could be achieved with short, mandatory, and automatically generated quizzes concerning specific content items, potentially providing stakeholders with a higher degree of feedback and quantitative measurement of content reach.

Conclusions

An mHealth solution seemed to be an effective and quick way to increase the reach of validated information to medical staff, in particular for time-sensitive and rapidly evolving medical guidance. Push notifications are an interesting feature of mHealth solutions for knowledge updates and could further increase the reach of information. In addition, real-time statistics could give quantitative feedback on the efficiency of the information dissemination strategy. Further research should be done to assess the clinical impact of mHealth solutions, such as adherence to validated guidance, quality of clinical practice, and patient outcomes. The use of automatically generated and mandatory quizzes concerning important content could provide stakeholders with additional and valuable feedback on content reach and would be the subject of future research. Finally, a medicoeconomical assessment should be done to fully understand the impact of mHealth solutions on information dissemination within institutions.

Authors' Contributions

OW is a main investigator and cofounder of the project and idea, took part in the analysis of results and literature review, and wrote the manuscript. OW also took part in the implementation at graduate and postgraduate levels. IZ is a main investigator and cofounder of the project and idea, programmed the code for the platform, took part in the analysis of results and literature review, and wrote the manuscript. IZ also took part in the implementation at graduate and postgraduate levels. MCZ is a member of the COVID-19 guidance multidisciplinary team, took an active part in guidance drafting, and critically revised the manuscript. AGA is a member of the COVID-19 guidance multidisciplinary team, took an active part in guidance drafting, critically revised the manuscript, and took part in the analysis of results. KB is a member of the COVID-19 guidance multidisciplinary team, took an active part in guidance drafting, and is a member of the Medical Directorate. KB also had an active role in the medical knowledge dissemination strategy, critically revised the manuscript, and helped implement the platform in the hospital. ES is a team leader and cofounder of the project and idea, supervised the project, critically revised the manuscript, and helped implement the platform in the hospital and in the Faculty of Medicine of Geneva University. TA is a member of the COVID-19 guidance multidisciplinary team, and took an active part in guidance drafting as well as in writing and critically revising the manuscript. TA had an active role in the medical knowledge dissemination strategy and helped in the implementation of the platform in the hospital.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19) situation reports. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-07-26]
2. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Feb 26;76:71-76. [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
3. Wax RS, Christian MD. Practical recommendations for critical care and anesthesiology teams caring for novel coronavirus (2019-nCoV) patients. Can J Anaesth 2020 May;67(5):568-576 [FREE Full text] [doi: 10.1007/s12630-020-01591-x] [Medline: 32052373]
4. The Lancet. COVID-19: protecting health-care workers. Lancet 2020 Mar 21;395(10228):922 [FREE Full text] [doi: 10.1016/S0140-6736(20)30644-9] [Medline: 32199474]
5. Buoro S, Di Marco F, Rizzi M, Fabretti F, Lorini FL, Cesa S, et al. Papa Giovanni XXIII Bergamo Hospital at the time of the COVID-19 outbreak: Letter from the warfront. Int J Lab Hematol 2020 Jun;42 Suppl 1:8-10 [FREE Full text] [doi: 10.1111/ijlh.13207] [Medline: 32222091]
6. Chen Z, He S, Li F, Yin J, Chen X. Mobile field hospitals, an effective way of dealing with COVID-19 in China: sharing our experience. Biosci Trends 2020 Jul 17;14(3):212-214. [doi: 10.5582/bst.2020.01110] [Medline: 32238673]
7. Kwon KT, Ko JH, Shin H, Sung M, Kim JY. Drive-Through Screening Center for COVID-19: a Safe and Efficient Screening System against Massive Community Outbreak. J Korean Med Sci 2020 Mar 23;35(11):e123 [FREE Full text] [doi: 10.3346/jkms.2020.35.e123] [Medline: 32193904]

XSL•FO
RenderX

8.   Zangrillo A, Beretta L, Silvani P, Colombo S, Scandroglio AM, Dell'Acqua A, et al. Fast reshaping of intensive care unit facilities in a large metropolitan hospital in Milan, Italy: facing the COVID-19 pandemic emergency. Crit Care Resusc 2020 Apr 01;22(2):91-94. [Medline: 32227819]

9.   Khot WY, Nadkar MY. The 2019 Novel Coronavirus Outbreak - A Global Threat. J Assoc Physicians India 2020 Mar;68(3):67-71. [Medline: 32138488]

10. Revere D, Nelson K, Thiede H, Duchin J, Stergachis A, Baseman J. Public health emergency preparedness and response communications with health care providers: a literature review. BMC Public Health 2011 May 18;11:337 [FREE Full text] [doi: 10.1186/1471-2458-11-337] [Medline: 21592390]

11. Dearinger AT, Howard A, Ingram R, Wilding S, Scutchfield D, Pearce KA, et al. Communication efforts among local health departments and health care professionals during the 2009 H1N1 outbreak. J Public Health Manag Pract 2011;17(1):45-51. [doi: 10.1097/PHH.0b013e3181f54110] [Medline: 21135660]

12. Zamberg I, Manzano S, Posfay-Barbe K, Windisch O, Agoritsas T, Schiffer E. A Mobile Health Platform to Disseminate Validated Institutional Measurements During the COVID-19 Outbreak: Utilization-Focused Evaluation Study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18668 [FREE Full text] [doi: 10.2196/18668] [Medline: 32250958]

13. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]

14. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]

15. McLinden D, Myers S, Seid M, Busch M, Davis D, Murphy J. The Learning Exchange, a Community Knowledge Commons for Learning Networks: Qualitative Evaluation to Test Acceptability, Feasibility, and Utility. JMIR Form Res 2019 Mar 14;3(1):e9858 [FREE Full text] [doi: 10.2196/formative.9858] [Medline: 30869643]

16. Zamberg I, Windisch O, Agoritsas T, Nendaz M, Savoldelli G, Schiffer E. A Mobile Medical Knowledge Dissemination Platform (HeadToToe): Mixed Methods Study. JMIR Med Educ 2020 May 27;6(1):e17729 [FREE Full text] [doi: 10.2196/17729] [Medline: 32249758]

17. Windisch O, Zamberg I, Iselin C, Schiffer E. [Head To Toe, a medical knowledge distribution platform : a practical example in urology]. Rev Med Suisse 2019 Nov 27;15(673):2205-2208. [Medline: 31778051]

18. Mobile App Analytics Platform for Android & iOS. Flurry. URL: www.flurry.com [accessed 2020-08-11]

19. StataCorp. Stata Statistical Software. 2019. URL: https://www.stata.com/ [accessed 2020-08-11]

20. COVID-19: Situation épidémiologique à Genève. URL: https://www.ge.ch/document/covid-19-situation-epidemiologique-geneve [accessed 2020-07-12]

21. Leung GM, Leung K. Crowdsourcing data to mitigate epidemics. The Lancet Digital Health 2020 Apr 20;2(4):e156-e157 [FREE Full text] [doi: 10.1016/S2589-7500(20)30055-8] [Medline: 32296776]

22. Representatives of the STARSurg Collaborative, EuroSurg Collaborative, TASMAN Collaborative. Medical student involvement in the COVID-19 response. Lancet 2020 Apr 18;395(10232):1254 [FREE Full text] [doi: 10.1016/S0140-6736(20)30795-9] [Medline: 32247322]

23. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid Response to COVID-19: Health Informatics Support for Outbreak Management in an Academic Health System. J Am Med Inform Assoc 2020 Mar 24 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]

Abbreviations

COVID-19: coronavirus disease
HCW: health care worker
HUG: Geneva University Hospitals
mHealth: mobile health
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Original Paper

Acceptability of App-Based Contact Tracing for COVID-19: Cross-Country Survey Study

Samuel Altmann[1*], MPhil; Luke Milsom[1*], MPhil; Hannah Zillessen[1*], MPhil; Raffaele Blasone[1], BSc; Frederic Gerdon[2], MA; Ruben Bach[2], PhD; Frauke Kreuter[2,3,4], PhD; Daniele Nosenzo[5,6], PhD; Séverine Toussaert[1], PhD; Johannes Abeler[1], PhD

[1]University of Oxford, Oxford, United Kingdom
[2]University of Mannheim, Mannheim, Germany
[3]University of Maryland, College Park, MD, United States
[4]Institute for Employment Research, Nürnberg, Germany
[5]Aarhus University, Aarhus, Denmark
[6]University of Nottingham, Nottingham, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Johannes Abeler, PhD
University of Oxford
Manor Road
Oxford, OX1 3UQ
United Kingdom
Phone: 44 1865 281440
Email: johannes.abeler@economics.ox.ac.uk

Abstract

Background: The COVID-19 pandemic is the greatest public health crisis of the last 100 years. Countries have responded with various levels of lockdown to save lives and stop health systems from being overwhelmed. At the same time, lockdowns entail large socioeconomic costs. One exit strategy under consideration is a mobile phone app that traces the close contacts of those infected with COVID-19. Recent research has demonstrated the theoretical effectiveness of this solution in different disease settings. However, concerns have been raised about such apps because of the potential privacy implications. This could limit the acceptability of app-based contact tracing in the general population. As the effectiveness of this approach increases strongly with app uptake, it is crucial to understand public support for this intervention.

Objective: The objective of this study is to investigate the user acceptability of a contact-tracing app in five countries hit by the pandemic.

Methods: We conducted a largescale, multicountry study (N=5995) to measure public support for the digital contact tracing of COVID-19 infections. We ran anonymous online surveys in France, Germany, Italy, the United Kingdom, and the United States. We measured intentions to use a contact-tracing app across different installation regimes (voluntary installation vs automatic installation by mobile phone providers) and studied how these intentions vary across individuals and countries.

Results: We found strong support for the app under both regimes, in all countries, across all subgroups of the population, and irrespective of regional-level COVID-19 mortality rates. We investigated the main factors that may hinder or facilitate uptake and found that concerns about cybersecurity and privacy, together with a lack of trust in the government, are the main barriers to adoption.

Conclusions: Epidemiological evidence shows that app-based contact tracing can suppress the spread of COVID-19 if a high enough proportion of the population uses the app and that it can still reduce the number of infections if uptake is moderate. Our findings show that the willingness to install the app is very high. The available evidence suggests that app-based contact tracing may be a viable approach to control the diffusion of COVID-19.

KEYWORDS

COVID-19; contact tracing; proximity tracing; app; digital; user acceptability; mHealth; epidemiology

Introduction

The COVID-19 pandemic is the greatest public health threat of the last 100 years. In the absence of an effective treatment or vaccination (as of June 2020), the public health response has so far relied on nonpharmaceutical measures to limit the spread of the epidemic, such as physical distancing, case isolation, and manual contact tracing [1]. These measures have not been sufficient to stop the epidemic. Many countries have therefore resorted to partial or full "lockdown" measures to control the epidemic, severely limiting social and economic interactions among their citizens. Although lockdowns may help countries to keep the number of infections under control [2], they come at a great social and economic cost [3-6].

COVID-19 is difficult to trace by traditional methods as COVID-19 cases are infectious 1-2 days before experiencing symptoms and contacts on average become infectious 3-4 days after exposure. The window to achieve containment by manual contact tracing is thus extremely short. Ferretti and colleagues [7] have proposed digital (app-based) contact tracing as an alternative measure to contain the epidemic without the large economic costs of lockdowns. The idea is to use low-energy Bluetooth connections between phones to record the interactions users have with others, particularly those interactions that may pose a higher risk of infection (eg, spending more than 15 minutes within 2 meters of another person). If a user is diagnosed with COVID-19, they can use the app to declare the diagnosis, which then notifies all other users who have come in close contact with the infected person, asking them to isolate at home for 14 days or until they have been tested by the public health authority. The main advantage over traditional (manual) forms of contact tracing is that the app allows instantaneous notification of contacts, which is a key determinant of the effectiveness of case isolation and contact tracing strategies for COVID-19 [7]. Other advantages are that the automatic recording of contacts scales up easily, and avoids the loss of information due to patients' recall bias and/or imperfect knowledge of the people they have been in contact with.

More and more countries are currently developing various types of contact-tracing apps and several countries have already launched one (eg, Singapore [8], Germany [9], and France [10]). The success of app-based contact tracing, however, critically depends on people's willingness to use the app. Hinch and colleagues' [11] epidemic simulation in the United Kingdom shows that the app reduces infections at all levels of uptake but that it is only sufficient to stop the epidemic if approximately 60% of the population use it. It is therefore important to gauge the strength of public support for this approach and to understand the factors that may hinder or facilitate uptake. For

instance, since the app would need to trace individuals' interactions with others, privacy concerns may undermine support and adoption [12]. It is also possible that such a technological solution may not work as well for the less digitally literate share of the population, further increasing the unequal impact of the COVID-19 pandemic within and across countries [13]. In this sense, an "opt-out" installation policy, where mobile phone providers or Apple and Google [14] would automatically install the app on phones, could maximize uptake. It is unclear, however, whether the public would be willing to support this more intrusive solution.

In light of the many open questions surrounding the viability of app-based contact tracing, we designed a survey to measure public support for this approach in five countries that are currently hit by the COVID-19 pandemic: France, Germany, Italy, the United Kingdom, and the United States. The specific objectives of our study are to (a) assess the overall acceptability among the public of app-based contact tracing under different installation policies (eg, voluntary installation or automatic installation by the government); (b) uncover country-level and individual-level variation in support for the app; and (c) understand the main mechanisms that may facilitate or impede app usage across various subgroups of countries and individuals.

Methods

Survey Design

We conducted large online surveys in five countries (France, Germany, Italy, the United Kingdom, and the United States) to measure the acceptability of app-based contact tracing for COVID-19 before apps were introduced in any of the five countries. A complete description of the survey can be found in the Multimedia Appendix 1; here, we provide an overview. At the beginning, after collecting respondents' informed consent, we described the app, explaining how a general version would function as well as the purpose it would serve (Textbox 1). We abstracted from any details about centralized versus decentralized data storage procedures. Respondents had to pass a comprehension check to proceed further. We then asked respondents how likely they would be to install the app on their phone, if it became available to download voluntarily ("opt-in" installation policy). Respondents were then asked about their main reasons for and against installing the app as well as their compliance with self-isolation requests. Next, we assessed to what degree respondents would be open to an "opt-out" policy, where mobile phone providers would automatically install the app on all phones, but users would be able to uninstall the app at any time. We then collected demographic information and concluded the survey with questions about respondents' attitudes toward the government under different installation regimes.

Textbox 1. App description in the UK survey.

"Imagine there was an app that you could install on your mobile phone. This app would automatically alert you if you had been in close contact for at least 15 minutes with someone who was infected with the coronavirus. Such an app does not exist yet in the UK. But we, researchers from the University of Oxford, are interested in understanding what you would think about such an app. [...]

The app would be developed by the NHS. You would need to install the app by simply clicking a link.

Once installed, the app would register which other users are close to you. The app would do this by using Bluetooth and your location.

The app would NOT access your contacts, photos, or other data held on your phone. Only the NHS would have access to the data collected. [...]

If the NHS diagnoses the coronavirus in somebody you have been in close contact with, the app would notify you automatically. The app would give you targeted advice on what to do. It will ask you to self-isolate at home for 14 days or until you have been tested for the virus.

This would be useful since people can infect others even before they have a fever or a cough. Self-isolating would thus protect your family, friends and colleagues from being infected by you. At the same time, only people who were in contact with an infected person would need to self-isolate.

If you had not been in close contact with a confirmed case, then the app would show you an "all clear" message. [...]

If you are diagnosed with coronavirus, the app would notify all people you have been in close contact with, without identifying you to them, and advise them to self-isolate. This would increase the chance of finding all the people you might have infected and help make sure they can keep their loved ones safe as well. If enough people use the app, it will slow down the epidemic and might even stop it entirely."

The versions for the other countries differed with regard to the technology used (GPS and Bluetooth vs only Bluetooth) and lockdown restrictions in place generally and tied to app usage specifically. For details, see the Multimedia Appendix 1.

We kept the survey design as similar as possible across all five countries, with a few exceptions to accommodate country differences with regard to lockdown measures in place at the time of taking the survey. The US survey (deployed last) contained a few additional questions, including robustness checks. See Section A in Multimedia Appendix 1 for more details.

Ethics approval was obtained from the University of Oxford (reference number ECONCIA20-21-06).

No personal information was collected as part of the study.

Target Population, Sample Size, and Attrition

The surveys were administered between March 20 and April 10, 2020. We recruited respondents through Lucid, an online panel provider. We targeted a sample size of 1000 respondents in each of the four European countries, and 2000 in the United States, with quotas set for the samples to be representative of the overall population in terms of gender, age, and region of residence. A total of 10,375 individuals started the survey and 10,308 consented to participate (participation rate=99%). Out of the people who consented to participate, 6166 passed the comprehension check and started the main questionnaire. After removing incomplete responses and duplicates, we had a sample of 6061 complete and unique responses (completion rate=59%). Finally, we removed 66 respondents who either did not own a mobile phone or did not disclose their gender, leaving us with a final sample of 5995 respondents. To control for the potential effect of our recruitment method, we repeated the German survey with a probability-based sample in an offline recruited online panel. See Section B in Multimedia Appendix 1 for further details on recruitment, filtering and attrition, and the final sample.

Statistical Analysis

Our main outcome variables measure respondents' intention to have the app installed on their phone under the two installation regimes (opt-in vs opt-out). The outcomes were measured on a 5-point ordinal scale (opt-in: from *definitely install* to *definitely won't install*; opt-out: from *definitely keep* to *definitely uninstall*). In our regression analysis, we dichotomized these outcome measures (1=definitely or probably install/keep the app; 0=otherwise).

We used multivariate regression analysis (linear probability models; probit and ordered logit in additional analyses presented in Multimedia Appendix 1) to examine the relationship between intention to install and a number of covariates: age, gender, country, presence of comorbidities (diabetes, high blood pressure, heart or breathing problems), usage of mobile phone outside the house, frequency of social interactions, ability to work from home during the lockdown, ability to obtain sick pay while working from home, trust in national government, and incidence of COVID-19 deaths in a respondent's region of residence (see Section C.3 in Multimedia Appendix 1 for more details). Table B.2 in Multimedia Appendix 1 presents a summary of these covariates.

Results

We find broad support for app-based contact tracing. Support is high in all countries, across all subgroups of the population, and under both installation regimes (opt-in and opt-out). Panel A of Figure 1 shows that, under the voluntary (opt-in) installation regime, 4484 out of 5995 respondents (74.8%) across all countries say they would probably or definitely download the contact-tracing app, if it was available. Panel B shows that 4059 out of 5995 respondents (67.7%) say they would probably or definitely keep the app installed on their phone under the automatic (opt-out) installation regime. In both regimes, the share of respondents who say they would not have the app installed on their phone is very small (red portion of the bars in Figure 1).

Figure 1. Likelihood of having the app installed, under opt-in and opt-out regimes and by country. Light/dark red bars correspond to *probably/definitely won't install* in Panel A and *probably/definitely uninstall* in Panel B.
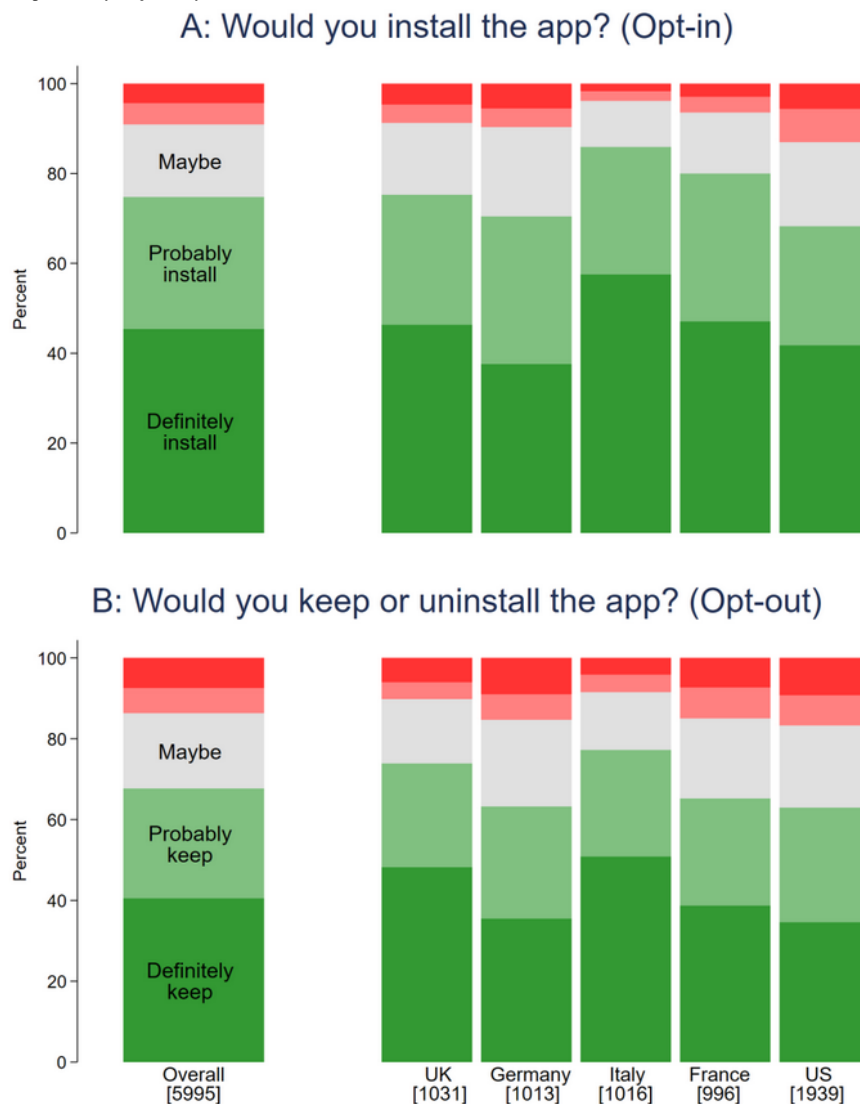

Support is high in all five countries where we implemented the survey: in each country, at least 68% of respondents say that they would install or keep the app. Moreover, Figures 9–11 in Multimedia Appendix 1 show that support for the app is generally high across various subgroups of the population (eg, across men and women, across different age groups, etc), suggesting widespread acceptability of the app-based contact tracing solution to the COVID-19 pandemic.

Despite the broad and widespread acceptability of the app, we find that support varies systematically across countries and individuals. For instance, Figure 1 shows that Germany and the United States are relatively less supportive of the app compared to the other countries. This is the case both under the opt-in and

opt-out regimes. Among individual characteristics, we find that those who have less trust in their national government are more hesitant to have the app installed on their phones (Figure 11 in Multimedia Appendix 1).

We further explore this heterogeneity using multivariate regression analysis, where we examine the relationship between support for the app and a variety of individual- and country-level covariates. Figure 2 shows the impact that these covariates have on the probability of definitely or probably installing the app under the opt-in regime, using a linear probability model (see Section C.1 in Multimedia Appendix 1 for a similar analysis of the opt-out regime).

Figure 2. Determinants of stating *definitely install* or *probably install*. Note: the dependent variable is an indicator variable taking the value 1 if a respondent chose *definitely install* or *probably install* when asked whether they would install the app or not, and 0 otherwise. We use a Linear Probability Model. Lines represent 95% CIs calculated with heteroskedasticity-robust standard errors. All coefficients are the result of a single regression and thus display marginal effects. A coefficient of 0.1 implies a respondent who chose this option is 10 percentage points more likely to state they would definitely or probably install the app relative to the base category.
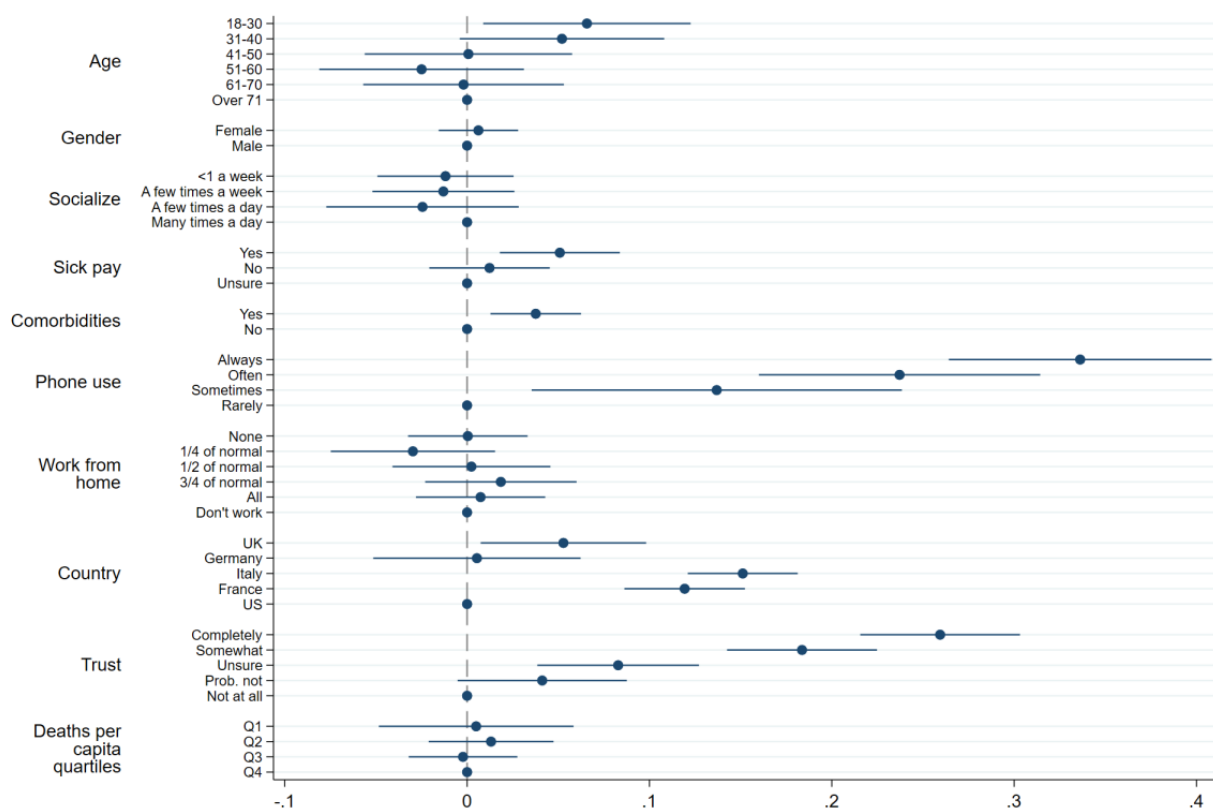
The analysis confirms that Germany and the United States are significantly less supportive of the app, especially compared to France and Italy. Taking the two most extreme cases, respondents in Italy are 15.1 percentage points (95% CI 12.1-18.1) more likely to support the app than respondents in the United States. Surprisingly, Figure 2 shows very little correlation between regional-level COVID-19 mortality rates and support for the app.

Among individual-level characteristics, we find that people who carry their phone with them more often are more likely to install the app. Those who always carry their phone with them are 33.6 percentage points (95% CI 26.4-40.8) more likely to support the app than those who carry their phone only rarely. App support is also 3.7 percentage points (95% CI 1.3-6.2) larger among respondents with one or more comorbidities. Moreover, the probability of installing the app increases with trust in the government. People who completely trust the government are 25.9 percentage points (95% CI 21.6-30.3) more likely to install the app than those who do not have any trust in the government.

We found similar results using an ordered logit model, a linear probability model dichotomizing on just *definitely install*, and when using a probit model (Multimedia Appendix 1). Finally, results are also qualitatively similar when considering installation intentions under opt-out rather than opt-in (Figure 8 in Multimedia Appendix 1). Interestingly, under the opt-out

regime, trust in government displays an even stronger correlation with the intention to keep the app installed on one's phone.

We can use the data on respondents' reasons for or against installing the app to better understand the nature of the observed variation in app support across countries and individuals. A first set of reasons against the app revolved around concerns about government surveillance at the end of the epidemic (mentioned by 2518 out of 5995 respondents, 42%) and cybersecurity (fears that the app could make the phone vulnerable to hackers; 2098/5995, 35%). Respondents also reported that usage of the app may increase feelings of anxiety (1559/5995, 26%), possibly reflecting aversion to feedback about a possible infection. The most frequent reasons in favor of the app were willingness to protect family and friends (4077/5995, 68%), a sense of responsibility toward the community (3177/5995, 53%), and a hope that the app may stop the epidemic (3297/5995, 55%). Figures 16 and 17 in Multimedia Appendix 1 show the relationship between the probability of selecting a particular reason and country- and individual-level characteristics.

Several patterns are of interest. First, we find that, compared to other countries, respondents in Germany and the United States are more likely to mention concerns about government surveillance as one of the reasons against installing the app. In these countries, we also see a larger share of respondents expressing concerns about security of the app, especially compared to Italy and the United Kingdom. Thus, concerns

about privacy and security seem to be an important impediment to the adoption of the app, particularly in Germany and the United States.

Among individual-level characteristics, we find that respondents who have less trust in their national government are also more likely to express concerns about government surveillance. This suggests that privacy concerns play a role in the negative relationship between trust in government and the probability of installing the app found in Figure 2. In contrast, we find that frequent usage of mobile phones is related to a stronger perception of the potential benefits of the app: respondents who more often carry their phone with them are more likely to believe that the app would benefit them, by helping them stay healthy and keeping them informed about the risks of infection.

Discussion

Principal Findings

In our study, we find high support for app-based contact tracing—irrespective of age, gender, region, or even country of residence. Since the effectiveness of app-based contact tracing crucially depends on a sufficient level of uptake, our findings are encouraging for the prospects of this approach. Although support is high in all countries and subgroups of the population, the data reveal that concerns about cybersecurity and privacy, coupled with trust in government, are important determinants of support. Countries with stronger privacy and security concerns (Germany and the United States) are relatively less supportive of app-based contact tracing. Individuals who have less trust in their national government are also less supportive.

Implications

The lack of trust in government can have far-reaching implications. Our analysis shows that this factor has a negative effect on people's intention to install a contact-tracing app on their phones. Furthermore, supplementary analysis (Section C.6 in Multimedia Appendix 1) also shows that people with lower trust in government are more in favor of an opt-in installation policy than an opt-out regime where the government asks mobile phone providers to automatically install the app on all phones. An opt-out regime is likely to translate into higher effective installation rates, for instance, by reducing the negative effects of procrastination or unawareness [15]. However, our data suggests that only governments that enjoy a relatively high level of trust from their citizens may be able to resort to more paternalistic approaches. A policy implication of these findings is that governments should consider delegating the organization of app-based contact tracing to a highly-reputable and transparent public health authority at arm's length from the government. If the mobile phone's operating system (eg, iOS or Android) does the contact tracing directly, trust in, for example, Apple or Google, would become more important.

Our results also point toward the need to address privacy and cybersecurity concerns with an app design that respects user personal data as much as possible. Research on the privacy implications of app-based contact tracing, and the potential solutions to these concerns, is currently underway [12,16,17]. Interestingly, however, when we asked our respondents how the data collected by the app should be treated, we find that nearly 60% would consent to making the deidentified data available to research.

Limitations

Our study has some limitations that we tried to address in different ways. First, respondents recruited online may not be representative of the entire population. In particular, digital literacy and willingness to share data could be higher among such respondents. To ensure that our results do not hinge on our specific sample, we replicated an abridged version of the German survey with a different panel provider that randomly recruits its participants offline. Our results remain almost completely unchanged (Section B.3 in Multimedia Appendix 1).

Second, our survey asked hypothetical questions about future behavior. However, high levels of intended installations may not directly translate into actual installations. While studies often find good correlation between what people declare they would do in surveys and actual behavior [18-22], even in relation to app installations [23-26], many things will have changed between the time of the surveys and when countries eventually introduce the app. For example, at the time of the survey, Italy's epidemic was close to its peak and the urgency of the situation was very salient. As the epidemic recedes, the perceived need to do something about the epidemic will also recede (eg, [22,27]). Moreover, we made participants aware of the app and explained the app and its potential effects. In reality, many potential users will not be aware of the app, might not engage with the concept, or will not be willing or able to spend the time to find and install the app. More generally, a reported willingness to install is only a necessary first stage to adoption, and our findings about heterogeneity in support point toward specific subgroups of the population that may need stronger encouragements to adoption. We show in Section C.4 of Multimedia Appendix 1 that respondents who would install the app mention far more reasons for its adoption than those who would not install it (but a similar number of reasons against). We show in Section C.9 that respondents in our replication study who did not answer the comprehension questions correctly were less willing to install. Stressing the various benefits of the app, to oneself and others, and explaining the function and purpose of the app may be a particularly effective strategy to foster adoption. Further research will be needed to understand how to translate a person's willingness to install into the person actually installing the app.

Third, in our survey, we measured support for the general concept of app-based contact tracing, leaving out specific details regarding the implementation, which were not available to us at the time respondents took the survey. One downside of only surveying about the general idea is that it might be harder for respondents to visualize how such a system could work, which may increase hypothetical bias. However, we find that the details we gave about implementation (eg, whether the app uses Bluetooth or GPS) seem to have very little impact on support. This suggests that our general measure of support for app-based contact tracing may be portable across different implementation settings.

Fourth, our results analyzing heterogeneity by age rely on coarse age binning. Such banding is subject to flaws if the sample is not distributed well across bands, which is not possible to verify in our study. Thus, results by age should be read with this limitation in mind.

Finally, our survey respondents were recruited from a specific subset of industrialized Western democracies. Attitudes towards app-based contact tracing may vary across countries with different levels of development and political regimes. It is nevertheless encouraging, in terms of external validity, that we observe a strong similarity in responses across the five countries we sampled and that analogous findings have been reported in ongoing surveys conducted in Australia and Taiwan [28]. In developing countries and among disadvantaged populations, the more limited access to smartphones raises both efficacy and equity issues; the development of low-cost Bluetooth devices with similar functionalities could improve access to digital contact tracing.

Conclusions

In conclusion, our study shows strong public support for app-based contact tracing to tackle COVID-19. This is an important finding since public support is a necessary condition for the viability of the approach. Further research is needed to gauge the extent to which public support for app-based contact tracing translates into actual app adoption and, more generally, to evaluate its potential for epidemic control.

Authors' Contributors

SA and LM were responsible for figures, study design, data collection, data analysis, and writing; HZ for study design, data collection, data analysis, and writing; RB for literature search and data collection; FG and RB for data collection and data analysis; FK for study design, data collection, and data analysis; DN, ST, and JA for study design, data collection, data analysis, and writing.

Authors' Contributions

SA and LM were responsible for figures, study design, data collection, data analysis, and writing; HZ for study design, data collection, data analysis, and writing; RB for literature search and data collection; FG and RB for data collection and data analysis; FK for study design, data collection, and data analysis; DN, ST, and JA for study design, data collection, data analysis, and writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Online appendix containing additional results, additional information about the samples, and the full questionnaire.
[PDF File (Adobe PDF File), 3209 KB - mhealth_v8i8e19857_app1.pdf]

References

1. Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Imperial College London 2020:E [FREE Full text] [doi: 10.25561/77482]
2. Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. The Lancet Public Health 2020 May 17;5(5):e279-e288 [FREE Full text] [doi: 10.1016/S2468-2667(20)30090-6] [Medline: 32311320]
3. Economic outlook - 26 March 2020. INSEE. 2020. URL: https://www.insee.fr/en/statistiques/4473305?sommaire=4473307 [accessed 2020-04-22]
4. Barrot J, Grassi B, Sauvagnat J. Sectoral Effects of Social Distancing. SSRN Journal 2020. [doi: 10.2139/ssrn.3569446]
5. Bayham J, Fenichel EP. Impact of school closures for COVID-19 on the US health-care workforce and net mortality: a modelling study. The Lancet Public Health 2020 May;5(5):e271-e278. [doi: 10.1016/s2468-2667(20)30082-7]
6. Adams-Prassl A, Boneva T, Golin M, Rauh C. Inequality in the impact of the coronavirus shock: Evidence from real time surveys. Journal of Public Economics 2020 Sep;189:104245 [FREE Full text] [doi: 10.1016/j.jpubeco.2020.104245]
7. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):513-551 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
8. Government of Singapore. TraceTogether. 2020. URL: https://www.tracetogether.gov.sg/ [accessed 2020-04-22]
9. Infektionsketten digital unterbrechen mit der Corona-Warn-App. RKI. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Warn_App.html [accessed 2020-04-22]

10.   Government of France. StopCovid. 2020. URL: https://www.economie.gouv.fr/stopcovid [accessed 2020-04-22]

11.   Hinch R. Effective Configurations of a Digital Contact Tracing App: A report to NHSX. 2020. URL: https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf?1587531217 [accessed 2020-04-25]

12.   Abeler J, Bäcker M, Buermeyer U, Zillessen H. COVID-19 Contact Tracing and Data Protection Can Go Together. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e19359 [FREE Full text] [doi: 10.2196/19359] [Medline: 32294052]

13.   Hargittai E. Second-Level Digital Divide: Differences in People's Online Skills. firstmonday.org 2002;7(4):e [FREE Full text] [doi: 10.5210/fm.v7i4.942]

14.   Google. Apple and Google partner on COVID-19 contact tracing technology. Google Blog. 2020. URL: https://blog.google/inside-google/company-announcements/apple-and-google-partner-covid-19-contact-tracing-technology/ [accessed 2020-04-27]

15.   Johnson EJ, Goldstein D. Medicine. Do defaults save lives? Science 2003 Nov 21;302(5649):1338-1339. [doi: 10.1126/science.1091721] [Medline: 14631022]

16.   Hyunghoon C, Daphne I, Yun WY. Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-offs. arXiv. 2020. URL: https://arxiv.org/abs/2003.11511 [accessed 2020-04-11]

17.   Christoph G, Michael G, Daniel G. Tracing Contacts to Control the COVID-19 Pandemic. arXiv. 2020. URL: https://arxiv.org/abs/2004.00517 [accessed 2020-04-13]

18.   Hainmueller J, Hangartner D, Yamamoto T. Validating vignette and conjoint survey experiments against real-world behavior. Proc Natl Acad Sci U S A 2015 Feb 24;112(8):2395-2400 [FREE Full text] [doi: 10.1073/pnas.1416587112] [Medline: 25646415]

19.   Chandon P, Morwitz VG, Reinartz WJ. Do Intentions Really Predict Behavior? Self-Generated Validity Effects in Survey Research. Journal of Marketing 2018 Oct 16;69(2):1-14. [doi: 10.1509/jmkg.69.2.1.60755]

20.   Waterfield G, Kaplan S, Zilberman D. Willingness to Pay versus Willingness to Vote: Consumer and Voter Avoidance of Genetically Modified Foods. American Journal of Agricultural Economics 2019 Dec 29;102(2):505-524 [FREE Full text] [doi: 10.1002/ajae.12001]

21.   Jordan J, Yoeli E, Rand D. Don't get it or don't spread it? Comparing self-interested versus prosocial motivations for COVID-19 prevention behaviors. PsyArXiv 2020 [FREE Full text] [doi: 10.31234/osf.io/yuq7x]

22.   Zhang B, Kreps S, McMurry N. Americans' perceptions of privacy and surveillance in the COVID-19 Pandemic. OSF Preprints. 2020. URL: https://osf.io/9wz3y/ [accessed 2020-06-29]

23.   Keusch F, Struminskaya B, Antoun C, Couper MP, Kreuter F. Willingness to Participate in Passive Mobile Data Collection. Public Opin Q 2019 Jul 27;83(Suppl 1):210-235 [FREE Full text] [doi: 10.1093/poq/nfz007] [Medline: 31337924]

24.   Kreuter F, Haas G, Keusch F, Bähr S, Trappmann M. Collecting Survey and Smartphone Sensor Data With an App: Opportunities and Challenges Around Privacy and Informed Consent. Social Science Computer Review 2018 Dec 18;38(5):533-549. [doi: 10.1177/0894439318816389]

25.   Wenz A, Jackle A, Couper MP. Willingness to use mobile technologies for data collection in a probability household panel. Survey Research Methods 2019 Apr;13(1):1-22 [FREE Full text] [doi: 10.18148/srm/2019.v1i1.7298]

26.   Jäckle A, Burton J, Couper MP, Lessof C. Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: coverage and participation rates and biases. Survey Research Methods 2019;13(1):23-44 [FREE Full text] [doi: 10.18148/srm/2019.v1i1.7297]

27.   Betsch C, Korn L, Felgendref L, Eitze S, Schmid P, Sprengholz P, et al. German COVID-19 Snapshot Monitoring (COSMO) - Welle 10 (05.05.2020). PsychArchives. 2020 May. URL: https://hdl.handle.net/20.500.12034/2521 [accessed 2020-06-28]

28.   Dennis S, Dennis S, Lorenz-Spreen P, Oberauer K, Okan Y, Goldstone R, et al. Social Licensing of Privacy-Encroaching Policies to Address the COVID-19 Pandemic. 2020 Jul. URL: https://stephanlewandowsky.github.io/UKsocialLicence/index.html [accessed 2020-04-25]

XSL·FO
RenderX

XSL•FO
RenderX