JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 6 (2018), Issue 9 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Review

Original Papers

Review

Text Messaging Interventions for Improvement in Physical Activity and Sedentary Behavior in Youth: Systematic Review

Kim Ludwig[1], MSc; Rosie Arthur[1], PhD; Nicholas Sculthorpe[1], PhD; Hollie Fountain[2], PhD; Duncan S Buchan[1], PhD

[1]Institute of Clinical Exercise and Health Science, School of Health and Life Sciences, University of the West of Scotland, Blantyre, United Kingdom
[2]School of Applied Sciences, Edinburgh Napier University, Edinburgh, United Kingdom

Corresponding Author:
Duncan S Buchan, PhD
Institute of Clinical Exercise and Health Science
School of Health and Life Sciences
University of the West of Scotland
Stephenson Place
Hamilton International Technology Park
Blantyre, G72 0LH
United Kingdom
Phone: 44 1698 283100 ext 8508
Email: duncan.buchan@uws.ac.uk

Abstract

Background: The use of text messages (short message service, SMS) to change physical activity and sedentary behavior in youth is of interest due to the need for novel, more effective intervention approaches. Previous reviews have examined a variety of technology-based interventions and their impact on different health behaviors, but evidence regarding the impact of just SMS on physical activity and sedentary behavior is lacking.

Objective: The aim of this study was to assess the effectiveness and use of theory of SMS interventions for improving physical activity and sedentary behavior in youth.

Methods: Authors systematically searched electronic databases from March to November 2017. Citations were sifted using additional reviewers, and a qualitative synthesis of eligible studies was conducted using piloted data extraction forms. To be eligible for inclusion, studies had to be of a randomized controlled or quasi-experimental design, incorporate SMS, involve adolescents between the ages of 10 and 19 years, and assess at least one physical activity or sedentary behavior outcome. Risk of bias was assessed using the Cochrane Collaboration's Risk of Bias tool.

Results: A total of 13 studies reporting 11 interventions were included in the qualitative analysis. Studies included interventions that were conducted in schools, online, or face-to-face. Studies were of high heterogeneity with regard to study duration, participant characteristics, intervention content, and outcome measures. Findings were equivocal with regard to intervention effectiveness for physical activity and sedentary behavior. Overall, 7 interventions resulted in an improvement for physical activity and 6 for sedentary behavior. All studies were judged to be of high risk of bias for at least 1 item.

Conclusions: Some studies in this review showed promising results for using SMS to improve physical activity and sedentary behavior in youth. High heterogeneity of design and outcome measures precluded data pooling and conclusions as to which specific intervention elements are linked to increased effectiveness cannot be drawn. The authors propose incorporating the following elements in future studies: specific focus on desired health behavior; mixed-methods design; include long-term follow-up; include self-monitoring, goal setting, and feedback; combine SMS with a mobile app; and send 3 or more SMS text messages per week. More rigorous studies are needed to explore the relationship between intervention effectiveness and specific intervention components such as content and delivery.

(JMIR Mhealth Uhealth 2018;6(9):e10799) doi:10.2196/10799

KEYWORDS

review; exercise; sedentary lifestyle; text messaging; cell phone; telemedicine; adolescent

Introduction

Physical Activity and Sedentary Behavior

Participating in sufficient levels of physical activity (PA) is essential to reduce the risk of all-cause mortality and cardiovascular disease [1,2]. For adolescents, it is recommended that they undertake at least 60 min of moderate to vigorous PA (MVPA) per day [3]. Unfortunately, few adhere to these current activity recommendations with adolescence characterized by declining PA levels in conjunction with increased sedentary time, despite calls for sedentary time to be minimized [4]. For instance, findings from Europe suggest that 83.2% of the adolescents aged 11 to 17 years do not achieve a minimum of 60 min of MVPA per day, whereas globally, it has been estimated that 80.3% of adolescents are insufficiently active [5]. Moreover, global data suggest that adolescents spend 57% of their time in sedentary activities, with 40% of adolescents spending 3 or more hours watching television on weekdays, increasing up to 50% on weekends [6,7]. These findings are particularly concerning as sedentary behavior (SB) is associated with various aspects of poor psychological and physiological health and all-cause and cardiovascular disease-related mortality [8-11]. Conversely, increased PA improves adiposity, blood lipid profile, blood pressure, insulin resistance, aerobic fitness, and bone health [12] while also reducing premature all-cause mortality [13]. Given these relationships, both SB and PA are important therapeutic targets to reduce lifestyle-induced noncommunicable diseases and especially during adolescence, as behaviors developed in younger ages are likely to continue into later life [14,15]. Given the inconsistent success of traditional intervention approaches, there is a need for research to generate new strategies to modify physical inactivity and SB [16].

Mobile Health

Mobile health (mHealth) which draws upon mobile devices for health-related apps has emerged as a promising tool for health-related behavioral interventions [17]. Mobile phones are used by all age groups, with more than 90% of UK children aged 12 to 15 years currently using them [18]. Such high usage suggests that these mobile devices may offer a cost-effective and acceptable means for delivering health behavior change interventions that can fit within people's everyday lives and have population-wide reach. Unsurprisingly, mHealth approaches are also being used to provide health care services worldwide, including Africa, Asia, and South America [19]. In the United Kingdom, the National Health Service is employing the SMS (short message service) text messaging system Florence to support patients in monitoring, managing, and improving their health [20]. mHealth systems can also be used to send appointment or medication reminders to support health care workers by providing training, decision making, and communication tools as well as to implement health promotion and educational interventions [19,21]. However, there is a lack of evidence regarding the effectiveness of mHealth interventions on behavior changes and health outcomes [19,22,23]. Unfortunately, research that has examined the effects of SMS interventions on PA and SB in youth is also scant.

Previous systematic reviews and meta-analyses involving adolescents have included a variety of technologies, such as apps, email, video games, and websites when reviewing the evidence on the most effective means of improving PA and SB [24-32]. However, none of these reviews have assessed the effectiveness of SMS in isolation. Moreover, reviews have included a number of outcomes such as disease state or medication adherence [25,33-36] and have focused on several different health behaviors, such as smoking and diet [25,27,29-32,34]. As such, evidence that has examined the efficacy of mobile devices to influence PA and SB is lacking. Furthermore, and to the best of our knowledge, existing systematic reviews and meta-analyses involving adolescents and SMS as a means for improving PA and SB have not explored the use of theoretical frameworks [24,30-32,34-37].

Theoretical Frameworks

As evidence has shown the increased effectiveness of health interventions using a behavioral theory framework [38,39], it is surprising that many interventions have been developed without a proper underpinning theory. Even in those studies that suggest their intervention was informed by appropriate theory, the specific application of theory often remains unclear [40,41]. In addition to evaluating the evidence of the effectiveness of interventions using mobile phones for improving PA and SB, it is important to evaluate the theory and behavior change techniques (BCTs) that have been used to develop these interventions. Providing this information is essential for health care practitioners to ensure that future mHealth interventions are effectively implemented.

Aims

To provide this evidence, this review aimed to systematically identify mHealth studies that have been developed to increase PA levels and to reduce SB in adolescents. A subsequent aim was to identify the theory and BCTs used in these studies. Findings from this review are expected to provide an insight into the development of future mHealth interventions to maximize their effectiveness.

Methods

Data Reporting

All data are reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement guidelines [42].

Eligibility Criteria

Experimental (randomized controlled trial or quasi-experimental design) studies were included if they involved or reported data separately for participants between the ages of 10 and 19 years with or without known morbidities; used SMS via a mobile phone within the intervention, both in addition to other intervention components or on its own; employed usual care, another intervention, or no intervention as comparator; and assessed at least one outcome related to PA or SB. All outcomes related to PA and SB, such as step count, moderate PA (MPA), and screen time, as well as all subjective and objective outcome measures were eligible for inclusion.

Furthermore, only studies that were written in the English language and where full text was available were included. Studies were excluded if they solely used other technologies such as apps, websites, or email.

Information Sources

A systematic search of the following electronic databases was conducted in March 2017 and updated in November 2017: Web of Science (coverage 1864-2017), PubMed (1809-2017), MEDLINE (1946-2017), Cumulative Index to Nursing and Allied Health Literature Complete (1937-2017), PsycINFO (1800s-2017; not available for search update and replaced by PsycARTICLES 1894-2017), and SPORTDiscus (1930-2017). All databases except PubMed (November 7, 2017) were last searched on November 8, 2017. During the initial search, KL searched bibliographies and contacted corresponding authors of eligible studies. Bibliographies of existing systematic reviews and meta-analyses identified during the initial search process were also screened for eligible studies [24-37,43,44].

Search

Search terms and combinations of the electronic database search are shown in Table 1.

Study Selection

Study citations from the electronic search were imported into the reference manager software Zotero (Version 5.0, online and standalone). KL manually removed duplicates. For the initial search, KL and HF independently screened titles and abstracts of all remaining studies. Following the search update, KL and DSB independently reviewed new titles and abstracts with the full texts of relevant titles obtained to confirm eligibility. KL and HF (DSB for search update) discussed discrepancies until consensus was reached. KL hand-searched bibliographies of eligible studies and contacted corresponding authors for additional manuscripts. All eligible studies were then included in the qualitative analysis.

Data Collection Process

Data extraction was conducted based on the Cochrane Collaboration's Data Extraction Template for Included Studies (Version 1.8) [45]. Items of interest for this review such as the content of SMS and interactivity were added to the Cochrane Data Extraction Template. KL piloted the updated template on 2 randomly chosen studies eligible for this review. Subsequently, the piloted form was revised where necessary. Thereafter, KL and HF (DSB after search update) independently extracted required data using the revised form. Extractions were compared and discussed until consensus was reached for all items. Content was then synthesized for analysis.

Data Items

Data extracted included (1) general study information (such as country, aims, and target health behavior); (2) methods (such as study design and duration of intervention); (3) participants (such as population description, number recruited, age, sex, and health status); (4) intervention and control groups (such as name of group, number of participants randomized, intervention mode, content, use of theory, message content, frequency, device, interaction, and adherence); (5) outcomes (assessed PA and SB outcomes, method of PA/SB outcome assessment, timing of PA/SB outcome assessment); (6) results and conclusion (including additional results information and relevant conclusions); (7) other information (including funding source and conflicts of interest). Where data were missing or clarification was sought, study authors were contacted. Where multiple studies reported on multiple follow-up periods or outcomes of the same intervention, outcomes from the longest follow-up time point available for each outcome were extracted.

Risk of Bias in Individual Studies

Assessment of risk of bias was conducted at study level. KL and HF (DSB after search update) reviewed all included manuscripts using the Cochrane Collaboration's risk of bias assessment tool [46]. KL employed this assessment tool using RevMan (software, version 5.3). Due to the nature of behavioral interventions, blinding of participants and personnel is challenging and rarely incorporated [47]. This item was therefore not included in the assessment. The following remaining domains were judged: selection bias (random sequence allocation and allocation concealment), detection bias (blinding of outcome assessment), attrition bias (incomplete outcome data), reporting bias (selective reporting), and other bias. KL and HF (DSB after search update) ranked each item as high, low, or unclear risk for each study and discussed discrepancies until a consensus was reached.

Table 1. Electronic database search terms and combinations. Asterisks were used to search for words beginning with these letters.

Category	Search term
Intervention mode	
1	“mobile phone”
2	smartphone
3	“cell phone”
4	“handheld device”
5	text messag*
6	SMS[a]
7	“messag* service”
8	“messaging system”
9	mHealth
10	telehealth
11	“online health”
12	e-Health
13	eHealth
14	“mobile health”
15	“digital media”
16	ICT[b]
17	(1-16) combined with OR
Study design	
18	“randomised controlled”
19	“randomized controlled”
20	RCT[d]
21	“controlled trial”
22	quasi-experimental
23	(18-22) combined with OR
Participants	
24	adolescen*
25	youth
26	“young people”
27	“young adult*”
28	child*
29	paediatric
30	pediatric
31	teen*
32	“school age”
33	“school-aged”
34	highschool
35	“secondary school”
36	(24-35) combined with OR
Behavior	
37	activity

Category	Search term
38	sport
39	exercise
40	health*
41	"behaviour change"
42	lifestyle
43	sedentary
44	sitting
45	(37-44) combined with OR
46	(17,23,36,45) combined with AND

[a]SMS: short message service.

[b]ICT: information and communication technology.

[c]RCT: randomized controlled trial.

Results

Study Selection

The electronic database and hand search produced 5565 and 266 studies, respectively. After removal of duplicates, 2365 studies were screened. A total of 2295 records were excluded, and 70 full-text articles were assessed. Moreover, 13 eligible full-text articles assessing 11 different interventions remained and were included in the qualitative analysis. A flowchart of the systematic literature search is displayed in Figure 1.

Study Characteristics

Study characteristics of included studies are shown in Tables 2 and 3. A total of 12 studies targeted PA [48-59] and 7 targeted SB [48-51,54,59,60]. Additionally, most studies also focused on dietary behaviors [49-52,54,57,59,60].

Some studies focused on participants with specific characteristics, including those not meeting current PA guidelines [48,53], not participating in physical education lessons or organized sports [54], having type 1 diabetes [56], being at high risk for diabetes [57], having a body mass index ≥ the eighty-fifth percentile [49,59], and being ≥1 year post cancer therapy [55]. When including overweight or obese participants, rates ranged between 23.7% (62/262) [52] and 55% (22/40) [49] for overweight and between 6.7% (15/225) [52] and 45% (18/40) [49] for obesity. The mean age of participants ranged between 12.5 [52] and 17.3 years [58]. One intervention only included female participants [50,51,54]. A total of 12 studies consisted of ≥50% female participants [48,50-60].

Intervention Design and Content

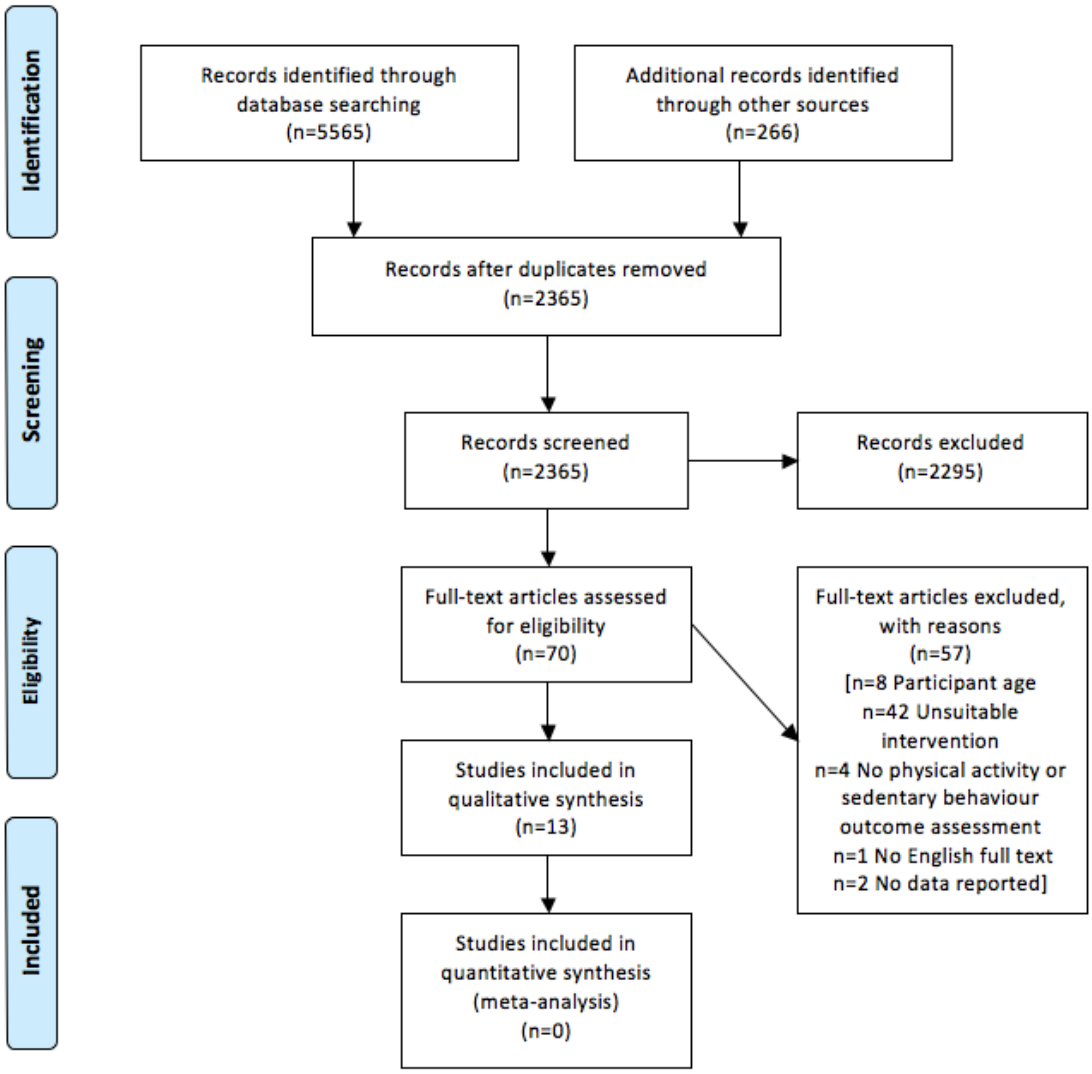
A total of 2 interventions included SMS in addition to a school program [50-52,54]. A total of 5 interventions used SMS text messages as part of an online intervention [49,53,55,57,60] and others used pedometers [56], group sessions and telephone calls [59], apps [48,49,55], and Fitbit trackers (Fitbit, Inc.) [49,55]. Only one intervention consisted solely of SMS [58]. Moreover, 2 interventions consisted of different types of SMS [48,58]. Depending on group allocation, one employed SMS focusing on affective or instrumental beliefs [58], whereas the other involved SMS from different senders, including a parent, peer, or behavioral health specialist [48]. School-based interventions using SMS included elements such as sports and PA opportunities, educational (group) seminars, provision of healthy foods, self-monitoring tools, and printed or email materials promoting healthy lifestyles [50-52,54]. One intervention also used a Facebook group to promote healthy lifestyles and keep participants informed about the intervention [52]. Interventions that included an online component also consisted of a variety of elements, such as forums, diet analysis, videos, educational games, challenges, educational materials, expert advice, behavioral skill training, goal setting, monitoring, feedback, and tutorials on behavioral change strategies [49,53,57,60]. One study included access to a private Facebook group, which provided rewards for achievements, encouragement, and a discussion board, as well as using Fitbit trackers and an app to monitor progress toward individualized goals [55].

In another study, participants wore pedometers that were used to encourage PA and facilitate recording progress [56]. Another study included group sessions that provided education on health behaviors and achieving successful behavior change. In this study, participants also received phone coaching during the 12-month maintenance period post intervention [59]. One study using an app for monitoring and reporting of PA also included autonomous and external goal setting as well as daily feedback [48]. Depending on which condition participants were assigned for that day, SMS text messages were sent by a behavioral health specialist, parents, or a peer [48].

Content of Text Messages

SMS text messages were used to encourage, motivate, reinforce, and prompt participants to be physically active or maintain their current positive behavior changes [48-51,53-56,59,60]. Some studies provided participants with suggestions for healthy lifestyle behaviors [48,49,59].

Figure 1. Literature search flow chart.



In addition to promoting PA, one study also employed SMS to provide participants with health behavior information, behavioral skills, and solutions for PA barriers to reinforce the benefits of PA and to build rapport with a virtual friend [53]. SMS text messages were also used for feedback [48,53], which in one study depended on the participant's goal attainment [48]. SMS also included statements from testimonials as well as messages targeting intrinsic motivation and reflective questioning [59]. SMS text messages were also used to reduce risk behaviors [60]. Two interventions employed SMS aiming to increase participant self-efficacy [59,60]. Three interventions sent SMS related to goal-setting, such as the participants' specific weekly challenges [55,57,59]. In addition to this, one intervention included affective SMS for encouragement and as a reminder of PA goals. In this intervention, SMS text messages sent in intervention week 2 were based on the participants' step counts from week 1 [55]. Another study sent SMS text messages regarding affective or, depending on the intervention group, instrumental gains associated with regular PA. These include messages regarding the benefits of being active, such as physical and psychological improvements [58]. Three studies used SMS text messages to remind participants to follow the intervention protocol, such as logging on to the intervention website or wearing an activity tracker [49-51,53,54,56,57].

Theory Derivation

Three studies based their interventions on the transtheoretical model (TTM) of behavior change or stage of motivational readiness for change (SOC) model [53,57,60]. One study used the SOC model to tailor intervention content and presentation, such as by adapting TM and website content according to the participant's stage of motivational readiness [53]. Participants in precontemplation, contemplation, and preparation stage were given information on benefits and barriers of PA, opportunities for PA, goal setting, as well as PA planning. Participants classed in the action stage were provided with monitoring tools and information to prevent relapse [53]. In addition to the TTM, one study also used the I-Change, Attitude-Social Influence-Self-Efficacy model and addressed attitude, social influence, and self-efficacy. They emphasized the advantages of following the recommendations and disadvantages of risk behaviors, created a healthy online social environment, and strengthened skills to avoid risk behaviors [60].

Table 2. Study characteristics of included studies—sample and outcomes.

Author, year, country	N[a]	Design	Age, mean (SD)	PA[b] and SB[c] outcomes	Assessment
Brannon et al, 2017, United States [48]	10	N-of-1 RCT[d]	16.7 (0.95)	MVPA[e] min/day, SB min/day	Objective
Chen et al, 2017, United States [49]	40	RCT	14.9 (1.7)	PA days/week, TV/computer hours/day	Self-report
Dewar et al, 2013, Australia [50]	357	Group RCT	13.2 (0.5)	Accelerometer counts/min, % MVPA, screen time min/day	PA: objective; SB: self-report
Dewar et al, 2014, Australia [51]	357	Group RCT	13.2 (0.5)	% MPA[f], VPA[g], MVPA; SB min/day	PA: objective; SB: objective + self-report
Ermetici et al, 2016, Italy [52]	487	Nonrandomized CT[h]	12.5 (0.4)	MVPA hours/week, screen time hours/day	PA: objective + self-report; SB: self-report
Lana et al, 2014, Spain and Mexico [60]	2001	RCT	Pre 13.26 (1.03); Post 12.91 (0.77)	SB (less than 360 min PA/week)	Self-report
Lau et al, 2012, Hong Kong [53]	78	Nonrandomized CT	CG[i] 13.26 (1.14); IG[j] 12.29 (0.87)	PA level last 7 days	Self-report
Lubans et al, 2012, Australia [54]	357	Group RCT	13.18 (0.45)	Accelerometer counts/min, MVPA min/day, SB min/day	PA: objective; SB: self-report
Mendoza et al, 2017, United States [55]	60	RCT	16.6 (1.5)	MVPA min/day, SB min/day	Objective
Newton et al, 2009, New Zealand [56]	78	RCT	14.4 (2.37)	Step count, MVPA min/week	Objective + self-report
Patrick et al, 2013, United States [57]	101	RCT	14.3 (1.5)	MVPA min/week, SB hours/day	Self-report
Sirriyeh et al, 2010, United Kingdom [58]	120	RCT	17.3 (0.68)	MVPA metabolic equivalent min/week	Self-report
Straker et al, 2014, Australia [59]	44	Within-subject CT	14.1 (1.6)	SB, light, moderate, vigorous PA min/day	Objective

[a]N: number of participants randomized.

[b]PA: physical activity.

[c]SB: sedentary behavior.

[d]RCT: randomized controlled trial.

[e]MVPA: moderate to vigorous physical activity.

[f]MPA: moderate physical activity.

[g]VPA: vigorous physical activity.

[h]CT: controlled trial.

[i]CG: control group.

[j]IG: intervention group.

Moreover, one study used both behavioral determinants models and TTM to guide intervention design [57]. One study employed affective and instrumental beliefs, as well as the theory of planned behavior (TPB) [58]. Two interventions were informed by social cognitive theory (SCT) [49-51,54]. One focused on self-efficacy, outcome expectation, self-monitoring, skill mastery, and self-regulation capabilities [49]. Another employed SCT by planning social support or change, providing general encouragement and information about the link between behavior and health, and identifying barriers and strategies to overcome these. Specifically, outcome expectations, social support, and self-efficacy were targeted [50,51,54]. Self-determination theory (SDT) formed the basis for 2 interventions [55,59], with one also using goal-setting theory [59]. This intervention focused on the provision of a need-supportive environment to achieve greater self-determination, autonomous motivation, and consequently greater engagement with the desired behaviors. The goal-setting theory was employed to increase autonomous and intrinsic goal setting to predict greater goal attainment and engagement with desired behaviors [59]. The other focused on psychological needs that influence motivation such as competence, autonomy, and relatedness. The Fitbit tracker and app aimed to increase competence and autonomy by providing opportunities to set personalized goals and monitor progress. The Facebook group aimed to enhance relatedness by providing support [55]. Cybernetic control theory (CCT) was used by one study, which included self-regulation strategies defined by goal-setting, self-monitoring, goal review, and feedback [48]. Two studies did not provide any information regarding theory derivation. Authors were contacted and lack of a specific theory base informing SMS was confirmed [52,56].

Table 3. Study characteristics of included studies—intervention and comparator.

Author, year	Intervention duration	TM[a] intervention	Comparators
Brannon et al, 2017 [48]	24 days	TM + mobile app	Mobile app only
Chen et al, 2017 [49]	6 months	TM + Fitbit tracker and app + online program	Online program + pedometer + diary
Dewar et al, 2013 [50]	12 months	TM + school program	Waitlist condensed intervention
Dewar et al, 2014 [51]	12 months	TM + school program	Waitlist condensed intervention
Ermetici et al, 2016 [52]	24 months	TM + school program	No information
Lana et al, 2014 [60]	9 months	TM + online program	Online intervention, limited access online intervention
Lau et al, 2012 [53]	8 weeks	TM + online program	No intervention
Lubans et al, 2012 [54]	12 months	TM + school program	Waitlist condensed intervention
Mendoza et al, 2017 [55]	10 weeks	TM + Fitbit tracker and app + Facebook group	Standard care
Newton et al, 2009 [56]	12 weeks	TM + pedometer	Standard care
Patrick et al, 2013 [57]	12 months	TM + online program	Online program, online program + group sessions + phone calls, usual care
Sirriyeh et al, 2010 [58]	2 weeks	TM only	Neutral TM
Straker et al, 2014 [59]	12 months	TM + group sessions + phone calls	No intervention

[a]TM: text messaging.

Text Message Delivery and Interactivity

In 3 studies, SMS text messages were sent weekly [55,56,60], 2 sent daily [48,58], another sent only on weekdays [53], and 2 studies sent 3 or more each week [52,57]. Two studies only sent SMS text messages during the maintenance period following the intervention [49,59]. In one, the number of SMS text messages was reduced from 3 to 1 per week and finally to 1 per month [59]. In the other, SMS text messages were sent biweekly during a 3-month maintenance phase [49]. Another intervention increased the frequency of SMS from weekly to twice per week [50,51,54]. Five studies specified the time of SMS delivery [48,50-52,54,58,59]. SMS text messages were sent at 4 pm at the end of the school day to minimize the risk of cross-contamination [58], close to meal times [52], between 7 pm and 8 pm [48] and depending on the SMS content, such as immediately after school when encouraging PA [50,51,54]. Another study sent SMS on weekday evenings at 6 pm and at 12 pm on weekends. Here, participants were able to choose on which days they wished to receive the SMS [59].

Three studies gave participants the possibility to interact with the research team and reply to the SMS [53,57,59]. Responding was optional; however, one study provided a monetary incentive to do so [53]. Another study also allowed interactivity; however, participants would only receive one reply [59].

Risk of Bias Within Studies

Five studies referred to previously published study protocols [50,51,54,59,60]. These were used to obtain missing information needed for the risk of bias assessment. The judgment of each risk of bias item across studies can be found in Figure 2. Tables 4 and 5 show the support for judgment of each item and study.

Several studies were rated as unclear selection bias with regard to random sequence allocation [48,50,51,54-57]. Three were rated high risk [52,53,59], and 3 were rated low risk [49,58,60]. Most studies also tended to be of unclear risk of selection bias with regard to allocation concealment [48-51,53-58,60]. Two studies were rated as high risk for this item [52,59]. A total of 7 studies were ranked to be of unclear risk of detection bias [20,21,23-26,30], with 4 judged as high-risk [50,54,55,59] and 2 as low-risk [56,58]. With regards to attrition bias, 7 studies were judged to be of low risk [50,51,53-56,59], whereas 5 were ranked as unclear [49,52,57,58,60] and one as high-risk [48]. Twelve studies were of low risk of reporting bias [48-57,59,60]. Only one study was classed as high risk of bias for this item [58]. Ten studies were ranked as high risk of response and recall bias [49-54,56-58,60]. Risk of compliance bias was evident in 3 studies [48,49,53]. Another study was judged to be of high risk of analytical bias [58]. Two studies appeared free of other sources of bias [55,59].

Synthesis of Results

PA and SB assessed in hours per week or hours per day were converted into min per week and min per day [52,57]. For the following, intervention group refers to those involving SMS text messages. An overview of the findings including PA and SB outcomes and outcome measures can be found in Table 6. Table 7 shows theoretical frameworks used and effectiveness of intervention groups in each study.

Figure 2. Risk of bias assessment.

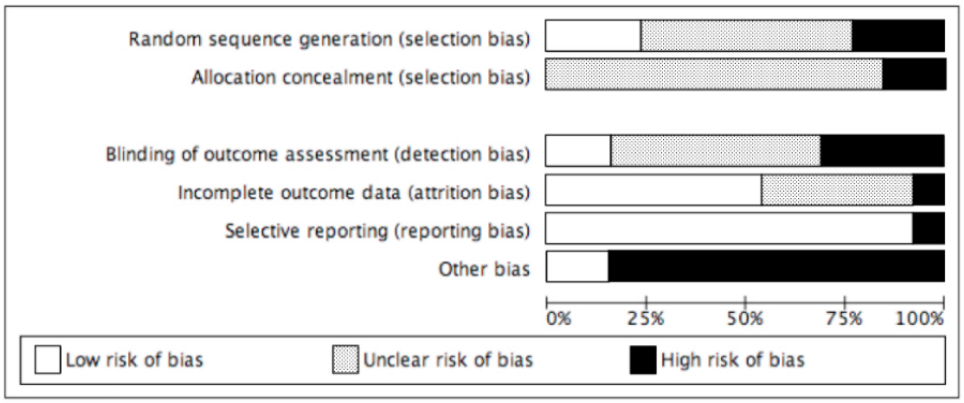


Table 4. Support for judgment of risk of bias per item and study. Random sequence generation, allocation concealment, and blinding of outcome assessment.

Author, year	Random sequence generation	Allocation concealment	Blinding of outcome assessment
Brannon et al, 2017 [48]	Unclear; Not enough information	Unclear; Not enough information	Unclear; Not enough information
Chen et al, 2017 [49]	Low; Randomization using computer program	Unclear; Not enough information	Unclear; Not enough information
Dewar et al, 2013 [50]	Unclear; Not enough information	Unclear; Not enough information	High; At baseline only. Outcomes likely to be influenced by lack of blinding
Dewar et al, 2014 [51]	Unclear; Not enough information	Unclear; Not enough information	Unclear; Not enough information
Ermetici et al, 2016 [52]	High; No randomization	High; No randomization	Unclear; Not enough information
Lana et al, 2014 [60]	Low; Randomization using computer program	Unclear; Not enough information	Unclear; Not enough information
Lau et al, 2012 [53]	High; No randomization	Unclear; Not enough information	Unclear; Not enough information
Lubans et al, 2012 [54]	Unclear; Not enough information	Unclear; Not enough information	High; At baseline only. Outcomes likely to be influenced by lack of blinding
Mendoza et al, 2017 [55]	Unclear; Not enough information	Unclear; Not enough information	High; Unblinded RCT[a]
Newton et al, 2009 [56]	Unclear; Not enough information	Unclear; Not enough information	Low; Assessors blinded at follow-up
Patrick et al, 2013 [57]	Unclear; Not enough information	Unclear; Not enough information	Unclear; Not enough information
Sirriyeh et al, 2010 [58]	Low; Randomization using random number generator	Unclear; Not enough information	Low; Assessors blinded at follow-up
Straker et al, 2014 [59]	High; Within-subject waitlist study design	High; Within-subject waitlist study design	High; Outcomes likely to be influenced by lack of blinding

[a]RCT: randomized controlled trial.

Physical Activity

Included studies assessed accelerometer counts [50,54], light PA [59], moderate or vigorous PA [48,50-59], step count [56], or the number of days when a minimum of 60 min of PA was achieved [49]. Nine studies assessed MVPA [48,50,52-58]. Three studies resulted in a decrease between baseline and longest follow-up for the intervention group [50,54,56,57]. One study, however, found an increase in MVPA between 6- and 12-month assessment [57]. In another study, MVPA of normal weight participants increased between baseline and 2-school-year follow-up for the intervention group, however, decreased for the control. For overweight or obese participants, MVPA increased in both groups [52]. Four interventions resulted in increases in MVPA for all intervention and control groups between baseline and follow-up [53,55,56,58]. Two studies assessing MVPA used different types of SMS [48,58]. TMs sent by parents were effective in increasing MVPA for 70% of participants, SMS sent by a peer for 50%, and those sent from a behavioral health specialist for 90% of participants. Overall, the intervention resulted in higher levels of PA than during the control condition [48]. Another study employed neutral, affective, instrumental, or a combination of affective and instrumental SMS [58]. Across all participants, MVPA increased during the 2-week intervention with affective SMS resulting in the highest levels of PA undertaken [58]. In 2 studies, MPA and vigorous PA (VPA) were assessed [51,59]. Total, during school, after school, and weekday MPA and VPA decreased from baseline to 12-week follow-up for both intervention and control group [51]. The other study showed increases in MPA and VPA between baseline and 12 months [59].

Table 5. Support for judgment of risk of bias per item and study. Incomplete outcome data, reporting bias, and other bias.

Author, year	Incomplete outcome data	Reporting bias	Other bias
Brannon et al, 2017 [48]	High; High amount of missing data	Low; All outcomes reported	High; Compliance bias (use of incentives)
Chen et al, 2017 [49]	Unclear; Insufficient reporting of reasons for missing data	Low; All outcomes reported	High; Response bias (use of self-report), compliance bias (use of rewards)
Dewar et al, 2013 [50]	Low; Missing outcome data balanced and similar reasons across groups	Low; All outcomes reported	High; Response bias (use of self-report)
Dewar et al, 2014 [51]	Low; Missing outcome data balanced and similar reasons across groups	Low; All outcomes reported	High; Response bias (use of self-report)
Ermetici et al, 2016 [52]	Unclear; Insufficient reporting of reasons for missing data	Low; All outcomes reported	High; Response bias (use of self-report)
Lana et al, 2014 [60]	Unclear; Insufficient reporting of attrition, exclusions, and reasons	Low; All outcomes reported	High; Response bias (use of self-report)
Lau et al, 2012 [53]	Low; Missing outcome data balanced and similar reasons across groups	Low; All outcomes reported	High; Response bias (use of self-report), compliance bias (use of incentives)
Lubans et al, 2012 [54]	Low; Missing outcome data balanced and similar reasons across groups	Low; All outcomes reported	High; Response bias (use of self-report)
Mendoza et al, 2017 [55]	Low; Missing outcome data balanced and similar reasons across groups	Low; All outcomes reported	Low; Appears free of other sources of bias
Newton et al, 2009 [56]	Low; Missing outcome data balanced and similar reasons across groups	Low; All outcomes reported	High; Response bias (use of self-report)
Patrick et al, 2013 [57]	Unclear; Insufficient reporting of reasons for exclusions and dropouts	Low; All outcomes reported	High; Response bias (use of self-report)
Sirriyeh et al, 2010 [58]	Unclear; Insufficient reporting of reasons for exclusions and dropouts	High; Missing mean and SD of MET[a] min at time point 1	High; Response bias (use of self-report), analytical bias (removal of outliers)
Straker et al, 2014 [59]	Low; Missing outcome data balanced and similar reasons across groups	Low; All outcomes reported	Low; Appears free of other sources of bias

[a]MET: metabolic equivalent.

For the intervention group, one study found an increase in PA levels between baseline and 3 months and between baseline and 6 months. PA levels decreased in the control condition [49]. Assessments of accelerometer counts, light PA, and daily step count showed decreases between baseline and follow-up [50,54,56,59].

Sedentary Behavior

Studies assessed screen time [49,50,52,54], total SB [48,51,55,57,59], and whether participants performed less than 360 min of PA per week [60]. Three interventions found a decrease in screen time between baseline and longest follow-up [49,50,52]. One study found an increase in subjectively measured screen time on weekdays, however, a decrease on weekends [54]. In one intervention [51], subjective SB decreased in the intervention group and increased in the control group between baseline and 12 months. However, objectively measured SB increased for both groups. In 2 studies [55,57], the intervention groups reduced their total SB between baseline and follow-up, whereas the usual care or control group showed an increase in SB. Another intervention found an increase in SB between baseline and 8 weeks, 3 months, 6 months, and 12 months [59]. One intervention resulted in an increase in insufficient PA in the intervention group between baseline and 9 months, although, both the control groups reduced their level of insufficient PA during the same period [60]. In another study, SB was the lowest when receiving SMS from a parent but was the highest when receiving them from a behavioral health specialist, followed by SMS from a peer [48].

Table 6. Overview of physical activity (PA) and sedentary behavior (SB) outcomes and outcome measures in intervention groups at longest follow-up.

Outcome category	Accelerometer	Pedometer	Questionnaire	Interview
Physical activity outcomes				
Accelerometer counts/min	Decrease [50]	—	—	—
Light PA min/day	Decrease [59]	—	—	—
MVPA[a] %	Decrease [50]	—	—	—
MVPA min/week		—	Increase [52,56]	Decrease [57]
MVPA min/day	Increase [48,55]; decrease [54]	—	—	—
MPA[b] %	Decrease [51]	—	—	—
MPA min/day	Increase [59]	—	—	—
VPA[c] %	Decrease [51]	—	—	—
VPA min/day	Increase [59]	—	—	—
MVPA score	—	—	Increase[d] [53]	—
4-day step count	—	Decrease [56]		—
MVPA MET[e] min/week	—	—	Increase [58]	—
PA days/week	—	—	Increase[f] [49]	—
Sedentary behavior outcomes				
Screen time min/day	—	—	Decrease [50,52]; increase and decrease [54]	—
Television/computer hours/day	—	—	Decrease[f] [49]	—
Total SB	Increase [51,59]; increase and decrease [48]; decrease [55]	—	Decrease[d] [51]; decrease [57]	—
PA less than 360 min/week	—	—	Increase [60]	—

[a]MVPA: moderate to vigorous physical activity.

[b]MPA: moderate physical activity.

[c]VPA: vigorous physical activity.

[d]Statistically significant ($P<.05$) between baseline and longest follow-up.

[e]MET: metabolic equivalent.

[f]Statistically significant ($P\leq.01$) between baseline and longest follow-up.

Table 7. Theoretical framework and intervention effectiveness for intervention group at longest follow-up for individual studies.

Outcome category	TTM[a]	TPB[b]	SCT[c]	SDT[d]	CCT[e]	N/A[f]
Physical activity						
Brannon et al, 2017 [48]	—	—	—	—	P[g]	—
Chen et al, 2017 [49]	—	—	P[h]	—	—	—
Dewar et al, 2013 [50]	—	—	N[i]	—	—	—
Dewar et al, 2014 [51]	—	—	N	—	—	—
Ermetici et al, 2016 [52]	—	—	—	—	—	P
Lau et al, 2012 [53]	P[h]	—	—	—	—	—
Lubans et al, 2012 [54]	—	—	N	—	—	—
Mendoza et al, 2017 [55]	—	—	—	P	—	—
Newton et al, 2009 [56]	—	—	—	—	—	N
Patrick et al, 2013 [57]	N	—	—	—	—	—
Sirriyeh et al, 2010 [58]	—	P	—	—	—	—
Straker et al, 2014 [59]	—	—	—	P	—	—
Sedentary behavior	—	—	—	—	—	—
Brannon et al, 2017 [48]	—	—	—	—	N, P	—
Chen et al, 2017 [49]	—	—	P[h]	—	—	—
Dewar et al, 2013 [50]	—	—	P	—	—	—
Dewar et al, 2014 [51]	—	—	N[j], P[j]	—	—	—
Ermetici et al, 2016 [52]	—	—	—	—	—	P
Lana et al, 2014 [60]	N	—	—	—	—	—
Lubans et al, 2012 [54]	—	—	P	—	—	—
Mendoza et al, 2017 [55]	—	—	—	P	—	—
Patrick et al, 2013 [57]	P	—	—	—	—	—
Straker et al, 2014 [59]	—	—	—	N	—	—

[a]TTM: transtheoretical model.

[b]TPB: theory of planned behavior.

[c]SCT: social cognitive theory.

[d]SDT: self-determination theory.

[e]CCT: cybernetic control theory.

[f]N/A: no theory framework.

[g]P: positive effect (PA increase, SB decrease).

[h]Statistically significant ($P \leq .01$) between baseline and longest follow-up.

[i]N: negative effect (PA decrease, SB increase).

[j]Statistically significant ($P < .05$) between baseline and longest follow-up.

Discussion

Summary of Evidence

This review found promising evidence regarding the effectiveness of interventions using SMS to improve PA and SBs. Out of 5 studies assessing MVPA via self-report, 4 found an increase in PA [52,53,56,58] whereas for objectively assessed MVPA, 2 interventions showed an increase [48,55] and one a decrease [50,54]. Four studies resulted in a decrease for objectively assessed accelerometer counts, light PA, MPA, VPA, and step count [50,51,56,59]. One intervention showed an increase in objectively measured MPA and VPA [59]. Five studies assessing screen time and total SB using questionnaires demonstrated improvements [49-52,57], whereas objectively measured total SB increased in 3 [48,51,59] and decreased in 2 studies [48,55]. Of 10 interventions involving PA assessment, 8 resulted in an improvement of at least one PA outcome and of 8 assessing SB outcomes, 5 showed improvements.

Most interventions included in this review focused on increasing PA, whereas elements targeting SB were scarce. Evidence

suggests that distinct assessment and approaches are required to improve PA and SB [61,62]. Previous meta-analyses have shown greater SB improvements in interventions solely targeting SB compared with PA interventions or those combining PA and SB [63,64]. To maximize intervention effectiveness, future studies should consider using distinct approaches to improve SB and PA.

The evidence presented in this review noted a variety of different outcome measures, which led to conflicting findings. For both PA and SB, more studies showed improvements when using subjective measures compared with objective measures. This is in line with previous findings showing subjective measures demonstrate greater enhancements than objective measures [65]. As self-report measures demonstrate low to moderate validity for the assessment of PA in children and adolescents, it appears that to assess effectiveness, objective measures such as accelerometers are preferred for both PA and SB [66]. For the assessment of the nature and mode of activity being undertaken, subjective measures should be used [61,66]. Further, a variety of protocols for the assessment and evaluation of participant data has been used. It has been shown that the choice of data reduction protocol when analyzing accelerometer data has a significant effect on the classification of SB and PA time in children [67]. There is a continued need for the standardization of methods when using objective measures to assess PA and SB [61], and future studies should consider following current recommendations on the assessment of both PA and SB to enhance the comparability of findings between studies and allow more distinct and unbiased conclusions to be drawn.

Identified studies also used a variety of theoretical frameworks with the more frequent use of the TTM and SCT, consistent with the findings of others [29]. Interventions informed by SDT, TPB, or CCT showed improvements in PA, whereas interventions informed by the TTM, SCT, and CCT revealed mixed results for PA and SB. Interventions employing SCT showed more positive results for SB than for PA. Nonetheless, the lack of information provided on how theory was applied within the intervention precludes our ability to confirm these assumptions with certainty. These findings are in line with those of a recent meta-analysis [44] that stated it was unclear how specific theoretical frameworks are applied or how they are linked to intervention effectiveness. Thus, our findings do not allow for a judgment on whether the ineffectiveness of some interventions included in this review is due to a lack of appropriate theory derivation and application. Furthermore, conclusions with regard to how theory relates to intervention effectiveness need to be drawn with caution, and more evidence is needed to warrant the use of specific theories when targeting PA and SB in SMS text messaging–based interventions for youth.

Evidence has shown the increased effectiveness of PA and SB interventions that include the BCTs of goal-setting, self-monitoring, and feedback [68]. In this review, 7 studies included goal-setting and monitoring, with 5 showing an increase in PA [48,49,53,55,59]. Two studies additionally included feedback and achieved improvements in PA [48,53]. Four studies that included self-monitoring and goal-setting found an improvement in SB [48,49,55,57]. These results are promising and indicate increased intervention effectiveness when including these BCTs in SMS-based interventions targeting PA and SB.

Previous reviews have shown weaknesses in the design of mHealth interventions [28,29,36,44]. Our findings were in agreement with those reviews and suggest that SMS-based interventions involving adolescents are weak in design and at a high risk of bias. The reasons for high risk of bias were attributed to the use of self-report measures (response bias), a lack of appropriate randomization method (selection bias), and a lack of blinding (detection bias).

We were also unable to infer the independent effect of SMS due to the lack of appropriate control groups. Only 4 studies employed designs that allowed for the effect of SMS text messaging alone to be assessed [48,57,58,60]. Two studies showed a positive effect of SMS on PA [48,58] and 2 on SB [48,57]. However, most studies included a variety of additional intervention components alongside SMS in the intervention and control groups. Definite conclusions with regard to the effectiveness of individual intervention designs, settings, or contents can therefore not be drawn from this review. Future research should employ study designs that allow the examination of the independent effect of SMS on PA and SB to strengthen the evidence base regarding the effectiveness of using SMS alone. Additionally, there is a need for studies exploring which specific SMS text messaging components such as content or frequency of delivery are most effective.

There is also a continued demand for studies to explore long-term intervention effects on PA and SB [24,28,32,35,37,43]. Only 4 interventions lasted for 12 months or longer [50-52,54,57,59]. Two studies assessed PA and SB after 24 months [50,52], with only one showing improvements in PA [52] but both showing decreases in SB [50,52]. It has been shown that SMS may be an effective tool to enhance participants' interest in the long term as well as to improve adherence [31,36]. Therefore, more studies should explore the effectiveness of interventions in achieving sustained behavior change.

This review shows a high heterogeneity of study designs, intervention components, outcomes, and outcome measures. Possible conclusions regarding effective intervention designs and contents are limited and should be drawn with caution. This review provides some currently limited evidence that the following approaches may result in increased effectiveness of SMS-based interventions for PA and SB in youth:

1. Specific focus on the desired behavior
2. Include self-monitoring, goal setting, and feedback components
3. Send 3 or more SMS per week for PA.

Furthermore, future research should incorporate the following methodological elements:

1. Use of objective outcome measures
2. Include long-term follow-up
3. Designs that allow assessing the independent effect of SMS.

Limitations

The authors were unable to conduct a quantitative data analysis due to high heterogeneity of included studies and a small pool of suitable data consisting of highly heterogeneous interventions and outcome measures. This review included all studies incorporating SMS text messaging as part of their intervention, which resulted in a variety of intervention designs and contents. Consequently, we were unable to draw conclusions with regard to specific intervention elements positively influencing PA and SB. To the best of our knowledge, this review provides the first account of interventions using SMS targeting PA and SB in adolescents. It provides researchers and practitioners with a database of potentially effective components crucial to the development of successful behavior change interventions.

Existing reviews have employed methods to identify and code theory-based elements such as behavior change techniques of included studies [26,28,65]. This review has refrained from following this process for studies not specifying theory base. However, the authors of those studies were contacted and a lack of theoretical foundation was confirmed. Despite the possibility that these interventions were unintentionally and unknowingly based on theory, there was no overt application of theory to study design. Therefore, it is judged to have limited contribution to intervention effectiveness.

This review does provide a detailed account of the use of theory in SMS-based interventions involving adolescents that, to the best of our knowledge, is novel and crucial for understanding current trends in intervention design and content. Moreover, a rigorous methodology was used for acquiring suitable studies, as well as during the data extraction process. This included hand-searching bibliographies, contacting authors of eligible studies, following recognized guidelines during data extraction, and pilot-testing data extraction items. Existing reviews on technology-based interventions targeting health behavior change have failed to include one or more of these components [24-31,33,35,37,43,44].

Conclusions

This review shows a high level of heterogeneity within SMS-based interventions targeting adolescent PA and SB. The evidence base consists of studies using different objective and self-report outcome measures that employ a variety of protocols, which impairs the ability to synthesize study content and results. Additionally, assessment of the risk of bias showed some limitations in the study and intervention design. Results of the individual as well as across studies should therefore be analyzed with caution. Future research should employ more rigorous research designs, more structured and coherent intervention components, as well as more appropriate and valid outcome measures. Overall, the findings of this study indicate that multicomponent interventions incorporating SMS can be effective in improving PA and SB in adolescents; however, more evidence is needed to further warrant SMS interventions to improve PA and SB.

Acknowledgments

This study was funded by the University of the West of Scotland, VP Research Fund. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' Contributions

KL performed all literature searches and all aspects of qualitative analysis. RA and DSB contributed to the design and focus of this review. HF and DSB reviewed citations obtained through the electronic literature search and provided judgment of risk of bias for eligible studies. NS reviewed available data for potential inclusion in quantitative analysis. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Alves AJ, Viana JL, Cavalcante SL, Oliveira NL, Duarte JA, Mota J, et al. Physical activity in primary and secondary prevention of cardiovascular disease: overview updated. World J Cardiol 2016 Oct 26;8(10):575-583 [FREE Full text] [doi: 10.4330/wjc.v8.i10.575] [Medline: 27847558]
2. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
3. Rhodes RE, Janssen I, Bredin SS, Warburton DE, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
4. Webcitation. 2011. Start Active, Stay Active: A report on physical activity from the four home countries' Chief Medical Officers URL: http://www.webcitation.org/723k1Lz29 [accessed 2018-08-30] [WebCite Cache ID 723nsDS1i]
5. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]

6. Arundell L, Fletcher E, Salmon J, Veitch J, Hinkley T. A systematic review of the prevalence of sedentary behavior during the after-school period among children aged 5-18 years. Int J Behav Nutr Phys Act 2016 Aug 22;13:93 [FREE Full text] [doi: 10.1186/s12966-016-0419-1] [Medline: 27549588]
7. Tenório MC, Gomes de Barros MV, Tassitano RM, Bezerra J, Tenório JM, Hallal PC. Physical activity and sedentary behavior among adolescent high school students. Rev Bras Epidemiol 2010 Mar;13(1):105-117. [doi: 10.1590/S1415-790X2010000100010]
8. Hoare E, Milton K, Foster C, Allender S. The associations between sedentary behaviour and mental health among adolescents: a systematic review. Int J Behav Nutr Phys Act 2016 Oct 8;13(1):108 [FREE Full text] [doi: 10.1186/s12966-016-0432-4] [Medline: 27717387]
9. de Rezende LF, Rodrigues Lopes M, Rey-López JP, Matsudo VK, Luiz Odo C. Sedentary behavior and health outcomes: an overview of systematic reviews. PLoS One 2014 Aug 21;9(8):e105620 [FREE Full text] [doi: 10.1371/journal.pone.0105620] [Medline: 25144686]
10. Thorp AA, Owen N, Neuhaus M, Dunstan DW. Sedentary behaviors and subsequent health outcomes in adults a systematic review of longitudinal studies, 1996-2011. Am J Prev Med 2011 Aug;41(2):207-215. [doi: 10.1016/j.amepre.2011.05.004] [Medline: 21767729]
11. Tremblay MS, LeBlanc AG, Kho ME, Saunders TJ, Larouche R, Colley RC, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. Int J Behav Nutr Phys Act 2011;8:98 [FREE Full text] [doi: 10.1186/1479-5868-8-98] [Medline: 21936895]
12. Poitras VJ, Gray CE, Borghese MM, Carson V, Chaput JP, Janssen I, et al. Systematic review of the relationships between objectively measured physical activity and health indicators in school-aged children and youth. Appl Physiol Nutr Metab 2016 Jun;41(6 Suppl 3):S197-S239 [FREE Full text] [doi: 10.1139/apnm-2015-0663] [Medline: 27306431]
13. Warburton DE, Charlesworth S, Ivey A, Nettlefold L, Bredin SS. A systematic review of the evidence for Canada's Physical Activity Guidelines for Adults. Int J Behav Nutr Phys Act 2010 May 11;7:39 [FREE Full text] [doi: 10.1186/1479-5868-7-39] [Medline: 20459783]
14. Craigie AM, Lake AA, Kelly SA, Adamson AJ, Mathers JC. Tracking of obesity-related behaviours from childhood to adulthood: a systematic review. Maturitas 2011 Nov;70(3):266-284. [doi: 10.1016/j.maturitas.2011.08.005] [Medline: 21920682]
15. Telama R, Yang X, Leskinen E, Kankaanpää A, Hirvensalo M, Tammelin T, et al. Tracking of physical activity from early childhood through youth into adulthood. Med Sci Sports Exerc 2014;46(5):955-962. [doi: 10.1249/MSS.0000000000000181] [Medline: 24121247]
16. Cobiac LJ, Vos T, Barendregt JJ. Cost-effectiveness of interventions to promote physical activity: a modelling study. PLoS Med 2009 Jul 14;6(7):e1000110 [FREE Full text] [doi: 10.1371/journal.pmed.1000110] [Medline: 19597537]
17. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013 Jan;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
18. Ofcom. 2016. Children and parents: media use and attitudes report URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0034/93976/Children-Parents-Media-Use-Attitudes-Report-2016.pdf [accessed 2018-08-30] [WebCite Cache ID 723kYctpn]
19. Chib A, van Velthoven MH, Car J. mHealth adoption in low-resource environments: a review of the use of mobile healthcare in developing countries. J Health Commun 2015;20(1):4-34. [doi: 10.1080/10810730.2013.864735] [Medline: 24673171]
20. Clark S, Birch-Jones J. Flo Simple Telehealth Evaluation Report 2013/2014. Nottinghamshire Health & Social Care AT Workstream 2014.
21. Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272. [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
22. Luxton DD, Mccann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: integrating smartphone technology in behavioral healthcare. Prof Psychol Res Pr 2011 Dec;42(6):505-512. [doi: 10.1037/a0024485]
23. Källander K, Tibenderana JK, Akpogheneta OJ, Strachan DL, Hill Z, ten Asbroek AH, et al. Mobile health (mHealth) approaches and lessons for increased performance and retention of community health workers in low- and middle-income countries: a review. J Med Internet Res 2013 Jan 25;15(1):e17 [FREE Full text] [doi: 10.2196/jmir.2130] [Medline: 23353680]
24. Chaplais E, Naughton G, Thivel D, Courteix D, Greene D. Smartphone interventions for weight treatment and behavioral change in pediatric obesity: a systematic review. Telemed J E Health 2015 Oct;21(10):822-830. [doi: 10.1089/tmj.2014.0197] [Medline: 26290954]
25. Cushing CC, Steele RG. A meta-analytic review of eHealth interventions for pediatric health promoting and maintaining behaviors. J Pediatr Psychol 2010 Oct;35(9):937-949 [FREE Full text] [doi: 10.1093/jpepsy/jsq023] [Medline: 20392790]
26. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]

27. Hieftje K, Edelman EJ, Camenga DR, Fiellin LE. Electronic media-based health interventions promoting behavior change in youth: a systematic review. JAMA Pediatr 2013 Jun;167(6):574-580 [FREE Full text] [doi: 10.1001/jamapediatrics.2013.1095] [Medline: 23568703]
28. Lau PW, Lau EY, Wong del P, Ransdell L. A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. J Med Internet Res 2011 Jul 13;13(3):e48 [FREE Full text] [doi: 10.2196/jmir.1533] [Medline: 21749967]
29. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
30. Smith AJ, Skow Á, Bodurtha J, Kinra S. Health information technology in screening and treatment of child obesity: a systematic review. Pediatrics 2013 Mar;131(3):e894-e902. [doi: 10.1542/peds.2012-2011] [Medline: 23382447]
31. Turner T, Spruijt-Metz D, Wen CK, Hingle MD. Prevention and treatment of pediatric obesity using mobile and wireless technologies: a systematic review. Pediatr Obes 2015 Dec;10(6):403-409. [doi: 10.1111/ijpo.12002] [Medline: 25641770]
32. Weihrauch-Blüher S, Koormann S, Brauchmann J, Wiegand S. Electronic media in obesity prevention in childhood and adolescence [in German]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2016 Nov;59(11):1452-1464. [doi: 10.1007/s00103-016-2455-z] [Medline: 27757512]
33. 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]
34. Loescher LJ, Rains SA, Kramer SS, Akers C, Moussa R. A systematic review of interventions to enhance healthy lifestyle behaviors in adolescents delivered via mobile phone text messaging. Am J Health Promot 2018 May;32(4):865-879 [FREE Full text] [doi: 10.1177/0890117116675785] [Medline: 27923885]
35. Militello LK, Kelly SA, Melnyk BM. Systematic review of text-messaging interventions to promote healthy behaviors in pediatric and adolescent populations: implications for clinical practice and research. Worldviews Evid Based Nurs 2012 Apr;9(2):66-77. [doi: 10.1111/j.1741-6787.2011.00239.x] [Medline: 22268959]
36. Wei J, Hollin I, Kachnowski S. A review of the use of mobile phone text messaging in clinical and healthy behaviour interventions. J Telemed Telecare 2011;17(1):41-48. [doi: 10.1258/jtt.2010.100322] [Medline: 21097565]
37. Siopis G, Chey T, Allman-Farinelli M. A systematic review and meta-analysis of interventions for weight management using text messaging. J Hum Nutr Diet 2015 Feb;28 Suppl 2:1-15. [doi: 10.1111/jhn.12207] [Medline: 24480032]
38. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
39. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
40. Bluethmann SM, Bartholomew LK, Murphy CC, Vernon SW. Use of theory in behavior change interventions. Health Educ Behav 2017 Apr;44(2):245-253 [FREE Full text] [doi: 10.1177/1090198116647712] [Medline: 27226430]
41. Taylor N, Conner M, Lawton R. The impact of theory on the effectiveness of worksite physical activity interventions: a meta-analysis and meta-regression. Health Psychol Rev 2012 Mar;6(1):33-73. [doi: 10.1080/17437199.2010.533441]
42. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
43. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
44. Orr JA, King RJ. Mobile phone SMS messages can enhance healthy behaviour: a meta-analysis of randomised controlled trials. Health Psychol Rev 2015 May;9(4):397-416. [doi: 10.1080/17437199.2015.1022847] [Medline: 25739668]
45. Cochrane Consumers and Communication. 2016. Data Extraction Template for Included Studies URL: https://tinyurl.com/y7ftd2va [accessed 2018-01-29]
46. Handbook-5-1.cochrane. Cochrane Handbook for Systematic Reviews of Interventions URL: http://handbook-5-1.cochrane.org/ [accessed 2018-08-30] [WebCite Cache ID 723owZgzd]
47. Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. J Clin Epidemiol 2004 Jun;57(6):543-550. [doi: 10.1016/j.jclinepi.2003.12.010] [Medline: 15246122]
48. Brannon EE, Cushing CC, Walters RW, Crick C, Noser AE, Mullins LL. Goal feedback from whom? A physical activity intervention using an N-of-1 RCT. Psychol Health 2018;33(6):701-712. [doi: 10.1080/08870446.2017.1385783] [Medline: 28988493]
49. Chen JL, Guedes CM, Cooper BA, Lung AE. Short-term efficacy of an innovative mobile phone technology-based intervention for weight management for overweight and obese adolescents: pilot study. Interact J Med Res 2017 Aug 2;6(2):e12 [FREE Full text] [doi: 10.2196/ijmr.7860] [Medline: 28768612]

50. Dewar DL, Morgan PJ, Plotnikoff RC, Okely AD, Collins CE, Batterham M, et al. The nutrition and enjoyable activity for teen girls study: a cluster randomized controlled trial. Am J Prev Med 2013 Sep;45(3):313-317. [doi: 10.1016/j.amepre.2013.04.014] [Medline: 23953358]
51. Dewar DL, Morgan PJ, Plotnikoff RC, Okely AD, Batterham M, Lubans DR. Exploring changes in physical activity, sedentary behaviors and hypothesized mediators in the NEAT girls group randomized controlled trial. J Sci Med Sport 2014 Jan;17(1):39-46. [doi: 10.1016/j.jsams.2013.02.003] [Medline: 23506657]
52. Ermetici F, Zelaschi RF, Briganti S, Dozio E, Gaeta M, Ambrogi F, et al. Association between a school-based intervention and adiposity outcomes in adolescents: the Italian "EAT" project. Obesity (Silver Spring) 2016 Mar;24(3):687-695 [FREE Full text] [doi: 10.1002/oby.21365] [Medline: 26833570]
53. Lau EY, Lau PW, Chung PK, Ransdell LB, Archer E. Evaluation of an internet-short message service-based intervention for promoting physical activity in Hong Kong Chinese adolescent school children: a pilot study. Cyberpsychol Behav Soc Netw 2012 Aug;15(8):425-434. [doi: 10.1089/cyber.2012.0161] [Medline: 22897473]
54. Lubans DR, Morgan PJ, Okely AD, Dewar D, Collins CE, Batterham M, et al. Preventing obesity among adolescent girls: one-year outcomes of the nutrition and enjoyable activity for teen girls (NEAT Girls) cluster randomized controlled trial. Arch Pediatr Adolesc Med 2012 Sep 01;166(9):821-827. [doi: 10.1001/archpediatrics.2012.41] [Medline: 22566517]
55. Mendoza JA, Baker KS, Moreno MA, Whitlock K, Abbey-Lambertz M, Waite A, et al. A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: a pilot study. Pediatr Blood Cancer 2017 Dec;64(12). [doi: 10.1002/pbc.26660] [Medline: 28618158]
56. Newton KH, Wiltshire EJ, Elley CR. Pedometers and text messaging to increase physical activity: randomized controlled trial of adolescents with type 1 diabetes. Diabetes Care 2009 May;32(5):813-815 [FREE Full text] [doi: 10.2337/dc08-1974] [Medline: 19228863]
57. Patrick K, Norman GJ, Davila EP, Calfas KJ, Raab F, Gottschalk M, et al. Outcomes of a 12-month technology-based intervention to promote weight loss in adolescents at risk for type 2 diabetes. J Diabetes Sci Technol 2013;7(3):759-770. [Medline: 23759410]
58. Sirriyeh R, Lawton R, Ward J. Physical activity and adolescents: an exploratory randomized controlled trial investigating the influence of affective and instrumental text messages. Br J Health Psychol 2010 Nov;15(Pt 4):825-840. [doi: 10.1348/135910710X486889] [Medline: 20156396]
59. Straker LM, Howie EK, Smith KL, Fenner AA, Kerr DA, Olds TS, et al. The impact of Curtin University's activity, food and attitudes program on physical activity, sedentary time and fruit, vegetable and junk food consumption among overweight and obese adolescents: a waitlist controlled trial. PLoS One 2014;9(11):e111954 [FREE Full text] [doi: 10.1371/journal.pone.0111954] [Medline: 25375109]
60. Lana A, Faya-Ornia G, López ML. Impact of a web-based intervention supplemented with text messages to improve cancer prevention behaviors among adolescents: results from a randomized controlled trial. Prev Med 2014 Feb;59:54-59. [doi: 10.1016/j.ypmed.2013.11.015] [Medline: 24287124]
61. Hardy LL, Hills AP, Timperio A, Cliff D, Lubans D, Morgan PJ, et al. A hitchhiker's guide to assessing sedentary behaviour among young people: deciding what method to use. J Sci Med Sport 2013 Jan;16(1):28-35. [doi: 10.1016/j.jsams.2012.05.010] [Medline: 22749939]
62. Pearson N, Braithwaite RE, Biddle SJ, van Sluijs EM, Atkin AJ. Associations between sedentary behaviour and physical activity in children and adolescents: a meta-analysis. Obes Rev 2014 Aug;15(8):666-675 [FREE Full text] [doi: 10.1111/obr.12188] [Medline: 24844784]
63. Martin A, Fitzsimons C, Jepson R, Saunders DH, van der Ploeg HP, Teixeira PJ, EuroFIT consortium. Interventions with potential to reduce sedentary time in adults: systematic review and meta-analysis. Br J Sports Med 2015 Aug;49(16):1056-1063. [doi: 10.1136/bjsports-2014-094524] [Medline: 25907181]
64. Prince SA, Saunders TJ, Gresty K, Reid RD. A comparison of the effectiveness of physical activity and sedentary behaviour interventions in reducing sedentary time in adults: a systematic review and meta-analysis of controlled trials. Obes Rev 2014 Nov;15(11):905-919 [FREE Full text] [doi: 10.1111/obr.12215] [Medline: 25112481]
65. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
66. Loprinzi PD, Cardinal BJ. Measuring children's physical activity and sedentary behaviors. J Exerc Sci Fit 2011;9(1):15-23. [doi: 10.1016/S1728-869X(11)60002-6]
67. Ojiambo R, Cuthill R, Budd H, Konstabel K, Casajús JA, González-Agüero A, IDEFICS Consortium. Impact of methodological decisions on accelerometer outcome variables in young children. Int J Obes (Lond) 2011 Apr;35 Suppl 1:S98-103. [doi: 10.1038/ijo.2011.40] [Medline: 21483428]
68. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

Abbreviations

BCT: behavior change technique
CCT: cybernetic control theory
MET: metabolic equivalent
mHealth: mobile health
MPA: moderate physical activity
MVPA: moderate to vigorous physical activity
PA: physical activity
SB: sedentary behavior
SCT: social cognitive theory
SDT: self-determination theory
SMS: short message service
SOC: stage of motivational readiness for change
TPB: theory of planned behavior
TTM: transtheoretical model
VPA: vigorous physical activity

Edited by G Eysenbach; submitted 17.04.18; peer-reviewed by F Mitchell, C Cushing, G Kok; comments to author 31.05.18; revised version received 13.06.18; accepted 16.06.18; published 17.09.18.

Please cite as:
Ludwig K, Arthur R, Sculthorpe N, Fountain H, Buchan DS
Text Messaging Interventions for Improvement in Physical Activity and Sedentary Behavior in Youth: Systematic Review
JMIR Mhealth Uhealth 2018;6(9):e10799
URL: http://mhealth.jmir.org/2018/9/e10799/
doi:10.2196/10799
PMID:30224335

Original Paper

The Remote Food Photography Method and SmartIntake App for the Assessment of Alcohol Use in Young Adults: Feasibility Study and Comparison to Standard Assessment Methodology

Tera L Fazzino[1,2], PhD; Corby K Martin[3], PhD; Kelsie Forbush[1], PhD

[1]Department of Psychology, University of Kansas, Lawrence, KS, United States

[2]Cofrin Logan Center for Addiction Research and Treatment, University of Kansas, Lawrence, KS, United States

[3]Pennington Biomedical Research Center, Louisiana State University System, Baton Rouge, LA, United States

Corresponding Author:
Tera L Fazzino, PhD
Department of Psychology
University of Kansas
Fraser Hall, 4th Floor
1415 Jayhawk Boulevard
Lawrence, KS, 66045
United States
Phone: 1 7858640062
Email: tfazzino@ku.edu

Abstract

Background: Heavy drinking is prevalent among young adults and may contribute to obesity. However, measurement tools for assessing caloric intake from alcohol are limited and rely on self-report, which is prone to bias.

Objective: The purpose of our study was to conduct feasibility testing of the Remote Food Photography Method and the SmartIntake app to assess alcohol use in young adults. Aims consisted of (1) quantifying the ability of SmartIntake to capture drinking behavior, (2) assessing app usability with the Computer System Usability Questionnaire (CSUQ), (3) conducting a qualitative interview, and (4) comparing preference, usage, and alcohol use estimates (calories, grams per drinking episode) between SmartIntake and online diet recalls that participants completed for a parent study.

Methods: College students (N=15) who endorsed a pattern of heavy drinking were recruited from a parent study. Participants used SmartIntake to send photographs of all alcohol and food intake over a 3-day period and then completed a follow-up interview and the CSUQ. CSUQ items range from 1-7, with lower scores indicating greater usability. Total drinking occasions were determined by adding the number of drinking occasions captured by SmartIntake plus the number of drinking occasions participants reported that they missed capturing. Usage was defined by the number of days participants provided food/beverage photos through the app, or the number of diet recalls completed.

Results: SmartIntake captured 87% (13/15) of total reported drinking occasions. Participants rated the app as highly usable in the CSUQ (mean 2.28, SD 1.23). Most participants (14/15, 93%) preferred using SmartIntake versus recalls, and usage was significantly higher with SmartIntake than recalls (42/45, 93% vs 35/45, 78%; P=.04). Triple the number of participants submitted alcohol reports with SmartIntake compared to the recalls (SmartIntake 9/15, 60% vs recalls 3/15, 20%; P=.06), and 60% (9/15) of participants reported drinking during the study.

Conclusions: SmartIntake was acceptable to college students who drank heavily and captured most drinking occasions. Participants had higher usage of SmartIntake compared to recalls, suggesting SmartIntake may be well suited to measuring alcohol consumption in young adults. However, 40% (6/15) did not drink during the brief testing period and, although findings are promising, a longer trial is needed.

(JMIR Mhealth Uhealth 2018;6(9):e10460) doi:10.2196/10460

KEYWORDS

alcohol consumption; alcohol college students; alcohol assessment; dietary assessment; self report; mobile phone; mobile health; ehealth; photography; young adults

Introduction

Alcohol use is prevalent among young adults [1]. Most (78%) US adults aged 18-24 report drinking alcohol and 40% report heavy drinking (5+ drinks on one occasion) at least once in the previous month [1]. Heavy drinking during young adulthood is associated with a host of negative consequences, from increased risk of accidents and injuries to the development of alcohol use disorder symptoms [2]. In addition to these well-known consequences, recent evidence suggests that heavy episodic drinking during young adulthood increases the risk of excess weight gain and the transition to obesity 5 years later [3]. Drinking may disrupt energy balance directly through ingestion of calories in alcoholic beverages and indirectly through effects on alcohol-related eating [4,5]. It is important to understand the direct and indirect effects of alcohol use on energy balance and obesity risk to develop relevant obesity prevention programs.

Researchers' ability to delineate the direct and indirect contributions of alcohol intake on energy balance, however, is limited by available measurement tools. Gold standard alcohol assessments involve asking participants to self-report the total number of drinks they consumed each day in the past 3-6 months [6]. While validity data indicate that this method may be sufficient to identify number of drinks consumed [6], it does not provide enough detail to reliably ascertain the precise caloric, nutritional, and alcoholic content of drinks. Information on the drink type, size in ounces, all alcoholic and nonalcoholic drink contents, and the amount consumed would be required to determine caloric intake from alcoholic beverages [7]. All of the aforementioned information is collected with the multiple pass 24-hour diet recall method [7]. The 24-hour diet recall method involves an iterative process through which individuals are asked to identify, for all food and beverages consumed in the past 24 hours, the food or drink type, the portion size, all contents of the food/beverage, and the amount they consumed [7]. Diet recalls have been applied to estimate caloric intake from alcohol as a component of overall energy intake in the general population [8-11]. Using data from the National Health and Nutrition Examination Survey (NHANES), researchers found that alcohol intake estimates were similar between the NHANES Alcohol Use Questionnaire, a standardized questionnaire that assesses typical quantity and frequency of alcohol use, as compared to alcohol intake estimated using diet recall data [12]. In addition, evidence suggested that 24-hour diet recalls performed similarly in measuring low to moderate levels of typical alcohol intake when compared to a 7-day retrospective recall of alcohol use, and 7-day prospectively recorded alcohol use with a food diary [13].

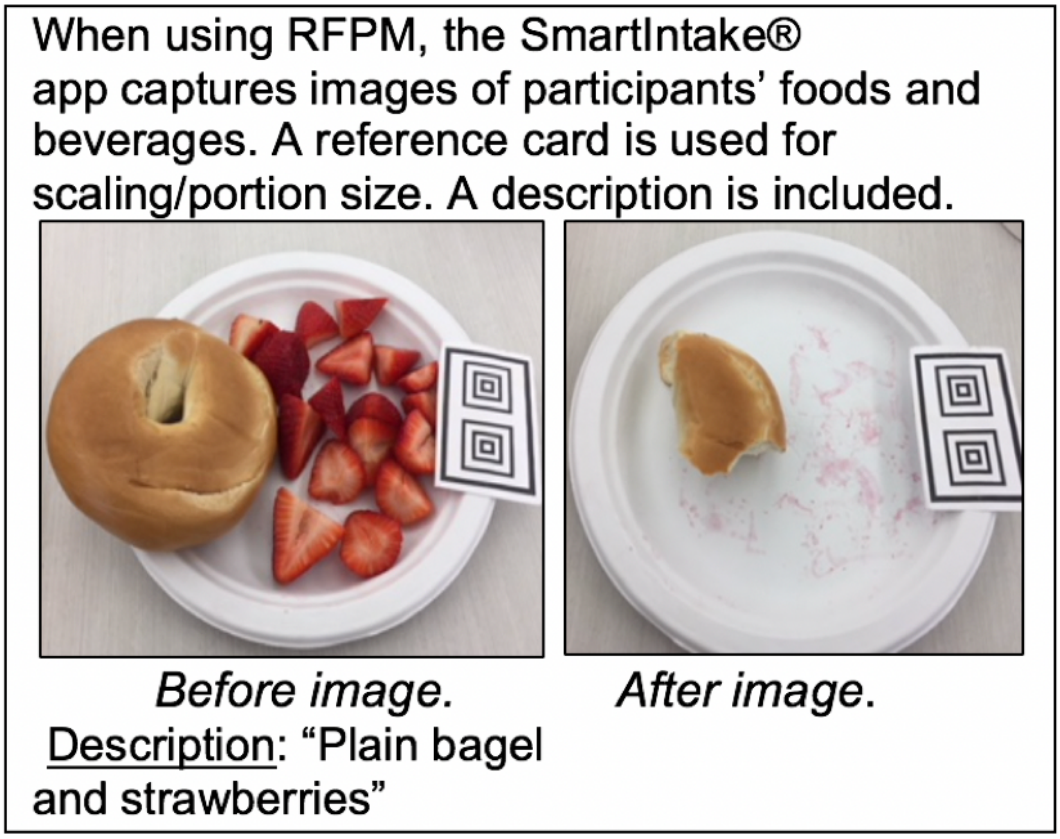
Despite their utility, assessments that rely on self-report are vulnerable to reporting biases due to memory inaccuracies from retrospective recall, social desirability, and inaccuracies in portion size estimates [14-16]. For example, researchers recently found that NHANES participants underestimated their intake in diet recalls by up to 800 calories per day [14], and Beasley et al [17] found that approximately 50% of the error in self-reported food intake was due to the inability of participants to accurately estimate portion size. Self-reported alcohol use suffers similar problems in underestimation [15]. A recent study of daily alcohol use found that alcoholic drink size and strength were underreported by at least 20% compared to daily alcohol use data recorded by transdermal alcohol sensors [15].

The Remote Food Photography Method (RFPM) was developed to address concerns regarding food and drink portion size estimation, to minimize participant burden, and to obtain accurate estimates of food and beverage intake [18-20]. With RFPM, participants capture photo images of their food selection and plate and drink waste using a mobile device in near real-time in their natural environments. Photos are analyzed by nutrition experts to estimate energy and nutrient content using standardized methods [20,21], eliminating the need for participants to accurately recall and report portion sizes. RFPM has excellent evidence for validity in measuring energy intake in the general adult population; RFPM estimates had only a 3.7% error rate when compared to energy expenditure estimates from doubly labeled water in weight-stable adults [18]. RFPM was developed prior to smartphones and has been used with various forms of mobile technology as advances have become available. RFPM was originally deployed using cellular-connected personal digital assistants, followed by camera-enabled flip phones, BlackBerry phones, and finally smartphones. For the past few years, RFPM has been deployed through a mobile phone app, SmartIntake, which can be downloaded directly onto participants' personal mobile phones and streamlines the RFPM data collection process. Figure 1 depicts the data collection process with the RFPM and SmartIntake app.

The RFPM and SmartIntake app can be adapted to measure alcohol use in young adults to address potential inaccuracies in self-reported drink size and content. The purpose of this pilot study, therefore, was to conduct feasibility testing of the RFPM and SmartIntake app via the following four aims:

1. Quantify the ability of SmartIntake to capture drinking behavior, defined as (1) the percent of total drinking occasions captured with SmartIntake, and (2) the percent of participants who submitted alcoholic drink photos through SmartIntake. The total number of drinking occasions was determined by adding the total number of drinking occasions captured by SmartIntake plus the total number of drinking occasions participants self-reported that they failed to capture through the app.
2. Use a standard technology usability questionnaire to collect usability data for the RFPM/SmartIntake.
3. Conduct a qualitative interview to assess acceptability and feasibility of using the SmartIntake app during drinking occasions.
4. Compare preference, usage, and alcohol use estimates per drinking occasion between SmartIntake and online diet recalls, the latter of which were completed by participants for a parent study. Usage was defined by the number of days participants provided food/beverage photo data through the app, and number of diet recalls completed.

Figure 1. The Remote Food Photography Method (RFPM) applied using the SmartIntake app.



Methods

Ethics and Data Security

The research was approved by the Institution Review Boards at the University of Kansas and Pennington Biomedical, Louisiana State University System. All participants provided written informed consent. Due to the sensitive nature of the data collected, participants were protected under a Certificate of Confidentiality issued by the National Institutes of Health. All photos submitted through SmartIntake were not linked with participant-identifying information.

Participants

Participants in the current study were recruited from a larger parent study. Below we first describe the parent study and then describe participant recruitment and enrollment into the current study.

Parent Study

The parent study was designed to examine the effects of heavy alcohol use and alcohol-related eating behavior on weight gain in the first year of college. At the beginning of the academic year, interested freshmen completed an online screening that consisted of a demographics questionnaire and the Alcohol Use Disorders Identification Test–Consumption questions (AUDIT-C) to assess a pattern of heavy alcohol use [22]. A random sample of study-eligible freshmen stratified by sex (52% male), race/ethnicity (44% racial or ethnic minority), and heavy drinking status (45% endorsing a heavy drinking pattern) were enrolled (N=103).

Participants attended three study visits at the beginning, middle, and end of the 2016-2017 academic year during which they completed an alcohol assessment and provided anthropometric measurements. Following each visit, participants completed a series of three online diet recalls using the Automated Self-Administered 24-Hour Diet Recall [23], the Web-based version of the United States Department of Agriculture 5-step diet recall [24], to report their dietary intake and alcohol consumption. Diet recalls were completed on 3 days randomly selected by study staff at each assessment point—one on a weekday and two on weekend days. Participants were required to complete all diet recalls within a 1-week window and could complete recalls late if they were still within the assessment window. Participants were compensated US $15 per completed recall.

Study Sample

The current study enrolled a convenience sample of 15 students selected from the parent study. When students attended a visit for the parent study, they were invited to participate in the current study if they endorsed a pattern of heavy drinking on the AUDIT-C at baseline or if they reported multiple (3+) recent heavy drinking episodes in the alcohol assessment. This procedure was in place to increase the likelihood that we would capture drinking episodes during the SmartIntake testing period and diet recalls. Students were also required to complete at least one diet recall (for the parent study) before starting the current study—a criterion that was met by the vast majority of participants in the parent sample. Most parent sample participants completed 1+ recall at baseline (96/103, 93%), 83% (85/103) completed 1+ recall at Visit 2, and 72% (74/103)

completed 1+ recall at Visit 3. Students were consented only for the current study when their 1-week window to complete the diet recalls for the parent study had passed (to avoid overlap in assessment methods). Enrollment was conducted on a rolling basis until we reached our target (N=15).

Procedure

Students attended an initial visit during which they provided informed consent and completed a training session to learn how to use the SmartIntake app. Participants were asked to use the app to report their food and alcohol intake for 3 consecutive days. Figure 2 depicts the RFPM and SmartIntake app process applied to alcoholic beverages.

SmartIntake testing days consisted of one weekday (Thursday) and two weekend days (Friday and Saturday). Participants returned the following week to complete a standardized app usability questionnaire and a qualitative interview about their experience using SmartIntake. Participants were not provided feedback or information about the photos they submitted (eg, alcohol calories consumed), as feedback could have altered their consumption and/or SmartIntake reporting behavior during the study.

Participants could earn up to US $60 for participating in the study. Participants were compensated US $15 per day for using SmartIntake, for a total of US $45 possible over 3 testing days. Independent of participants' usage with app testing, participants were compensated an additional US $15 for completing the follow-up interview. The compensation structure was explained to participants during the consent process. We matched compensation for 3 days of app testing (US $15 per day; US $45 total) directly to compensation for 3 recalls (US $15 per recall; US $45 total) to facilitate comparisons between the methods.

Measures

The Computer System Usability Questionnaire (CSUQ) is a widely used standardized questionnaire that was originally designed to measure computer program usability in field-testing studies at IBM [25,26]. The CSUQ has since been applied to studying the usability of websites [27] and mobile phone apps, including mHealth apps for adults [28-30] and adolescents [31]. This 19-item questionnaire uses a 7-point Likert scale ranging from 1 (strongly agree) to 7 (strongly disagree) and yields an overall score representing overall satisfaction with the program and three scale scores for System Usefulness, Information Quality (quality of instructions in the program and utility of error messages), and Interface Quality [32,33]. Items are averaged to obtain scores, with lower scores indicating greater usability. Evidence indicates the CSUQ has strong internal consistency across scale items and a replicable structure across tests of different types of computer programs (eg, computer, voice activated programs, Web apps) [25,26,34].

Figure 2. The Remote Food Photography Method (RFPM) and SmartIntake app applied to measuring alcohol intake.

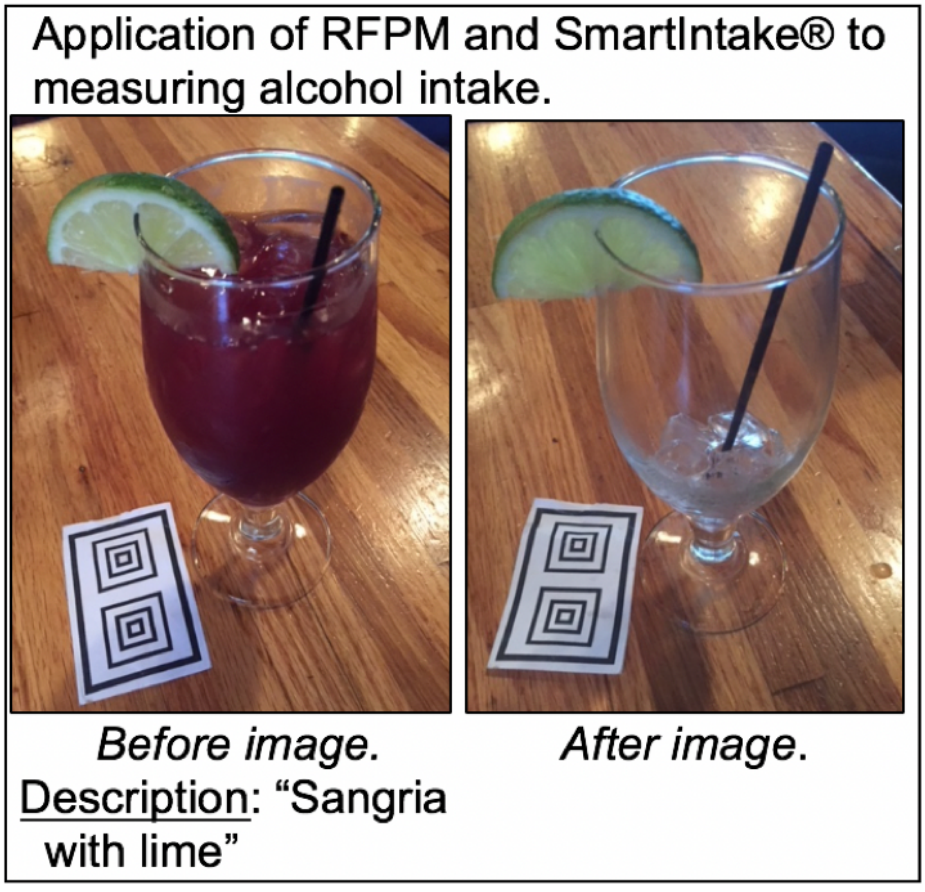


Table 1. Information used to calculate outcomes for Aims 1 and 4.

Dependent variables	SmartIntake app testing (3 days)	Qualitative interview	Diet recalls[a] (3 days)
Aim 1			
• Percentage of drinking occasions captured by SmartIntake • Percentage of participants who submitted alcohol photos	• Number of total drinking occasions captured with SmartIntake • Number of participants who submitted and did not submit alcohol photos	• Number of total drinking occasions not captured with SmartIntake (self-reported)	• N/A[b]
Aim 4			
• Preference	• N/A	• Number of participants who preferred SmartIntake and number who preferred diet recalls for reporting alcohol and food intake	• N/A
• Usage	• Number of days out of 3 that each participant completed SmartIntake testing	• N/A	• Number of recalls out of 3 that each participant completed
• Alcohol use estimates • Percentage of participants who reported alcohol use	• Alcohol consumption in grams and calories per drinking occasion; number of heavy drinking episodes • Number of participants who submitted and did not submit alcohol photos	• N/A	• Alcohol consumption in grams and calories per drinking occasion; number of heavy drinking episodes • Number of participants who reported and did not report alcohol use

[a]Diet recalls were completed during the parent study and used for comparisons with SmartIntake in the current study, as described in the Parent Study section of the Methods.

[b]N/A: not applicable.

The qualitative interview assessed participants' likes and dislikes about using the app, the utility of the reminders sent from the app (these remind participants to capture images), and their experiences using SmartIntake while drinking alcohol. Participants were asked directly about instances during which they forgot or almost forgot to take photos of alcohol or food and to identify situations in which using SmartIntake might be difficult. Participants were asked to describe any circumstances during which they felt uncomfortable using the app. All questions were open-ended. Finally, participants were asked about their preference for using SmartIntake or the online diet recalls to report their alcohol and food intake.

All interviews were conducted individually with participants by the study's principal investigator (PI). To minimize the potential for social desirability responding, the interview was framed as an opportunity for the PI to understand participants' experiences using the app, with the purpose of working together to identify things that worked and did not work, and to hear their suggestions for improving the app and data collection methods. Participants were asked to describe times they drank alcohol and forgot to report it with the app, so that the PI could understand the circumstances under which this type of reporting did not seem feasible. Similarly, when the PI inquired about participants' preferred method for reporting alcohol and food intake, participants were asked to explain what about the method worked best for them, so that she could understand circumstances in which one method might be preferred or work better than the other.

Outcomes

Information used to calculate dependent variables (DV) for Aims 1 and 4 was derived from multiple sources, as detailed in Table 1. DV calculations for Aims 1-4 are presented following Table 1.

Feasibility and usability outcomes were calculated using the following metrics. For Aim 1a), the percentage of total drinking occasions captured with SmartIntake was calculated as N captured / N captured + N missed, as reported by participants. For 1b), the percentage of participants who submitted alcohol photos was calculated as N participants who submitted alcohol photos / N submitted + N who did not submit alcohol photos. For Aim 2, the CSUQ overall satisfaction score was calculated as the mean of all CSUQ items. Three scale scores for System Usefulness, Information Quality, and Interface Quality were determined by calculating the mean of items in each scale. For Aim 3, common themes were identified regarding acceptability and feasibility for using SmartIntake overall and during drinking episodes. For Aim 4, we used a repeated-measures, within-subjects design to compare preference, usage, and alcohol use estimates per drinking occasion with SmartIntake and diet recalls. Usage was defined by the number of days participants provided food/beverage photo data through the app, or number of diet recalls submitted. Alcohol use estimates per drinking

occasion were calculated for SmartIntake and the diet recalls because both provide grams of alcohol consumed and caloric contents of the alcoholic beverages. Heavy drinking occasions captured through SmartIntake and the diet recalls were defined as 4+ drinks for females or 5+ for males on one occasion, in excess of low-risk drinking guidelines from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) [35]. The NIAAA defines a standard drink as 14 grams of pure alcohol [35]. The total number of participants who submitted alcohol photos through SmartIntake and the total number of participants who reported alcohol use in the diet recalls were summed for comparison.

Diet recalls from the parent study that were completed at the same assessment point as SmartIntake testing were used for comparison. Because the diet recalls were completed before the current study, we did not inquire about whether participants missed reporting alcohol use in the recalls; thus, we were unable to calculate the percentage of total drinking occasions captured in diet recalls as we were for SmartIntake.

Dependent *t* tests were used to compare usage and alcohol use estimates per drinking occasion between SmartIntake and the diet recalls. Fisher exact tests were used to test the difference between number of heavy drinking episodes reported between SmartIntake and the diet recalls, and the number of participants who reported alcohol use in each method.

Results

Participant Characteristics

Participants (N=15) provided informed consent, tested the app, and completed the follow-up visit. Participant characteristics are presented in Table 2. On the AUDIT-C, 93% (14/15) of participants endorsed drinking alcohol 2+ times per week and one endorsed drinking 2-4 times per month. Most (13/15, 87%) reported that they engaged in weekly heavy episodic drinking.

Aim 1: Quantifying the Ability of SmartIntake to Capture Drinking Behavior

SmartIntake captured 87% of reported drinking occasions (Aim 1a; Figure 3). Participants submitted a total of 15 alcohol photos during 13 drinking episodes. There were two instances in which participants reported that they drank alcohol but forgot to submit photos. Both missed occasions occurred among participants who submitted other alcohol photos through SmartIntake.

Sixty percent (9/15) of participants submitted alcohol use photos through SmartIntake (Aim 1b). Of the 40% (6/15) who did not send alcohol photos through SmartIntake, all reported that they did not drink during the days they used SmartIntake.

Aim 2: Usability

Results of the CSUQ indicated that participants were highly satisfied with SmartIntake overall (mean 2.52 on a 7-point scale, SD 1.13) and that the app was highly usable (mean 2.28, SD 1.23), provided good quality information and instructions for use (mean 2.36, SD 1.14), and had acceptable interface quality (mean 3.10, SD 1.68).

Aim 3: Qualitative Interview to Assess Acceptability and Feasibility

Overall Feedback on SmartIntake

Themes from the follow-up interview largely mirrored responses to the CSUQ. Participants liked that the app was quick and easy to use and that they could report their food and beverage intake in real-time. The majority of participants indicated the reminders to submit photos were mistimed on weekends because their eating schedules were less consistent and reliable than on weekdays, despite the reminder system accommodating different schedules on the weekends. Many participants also stated they often did not notice the notifications because they were sent via email and not text message, even though the notifications showed up on their phones when their screens were locked.

Table 2. Participant characteristics.

Variable	Value
Age (years), mean (SD)	18.1 (0.3)
Male, n (%)	9 (60)
White, non-Hispanic, n (%)	13 (87)
AUDIT-C score[a,b], mean (SD)	
Males	7.0 (0.7)
Females	6.3 (1.2)
Body Mass Index (BMI), mean (SD)	26.3 (6.5)
Weight class[c], n (%)	
Healthy weight	9 (60)
Overweight	3 (20)
Obese	3 (20)

[a]AUDIT-C: Alcohol Use Disorder Identification Test–Consumption Questions.

[b]AUDIT-C score of 5+ for females or 7+ for males indicates a pattern of heavy drinking in college students [22].

[c]Healthy weight: BMI<25 and >19; overweight: BMI=25-29.9; obese: BMI≥30.

Figure 3. Drinking occasions captured by SmartIntake.

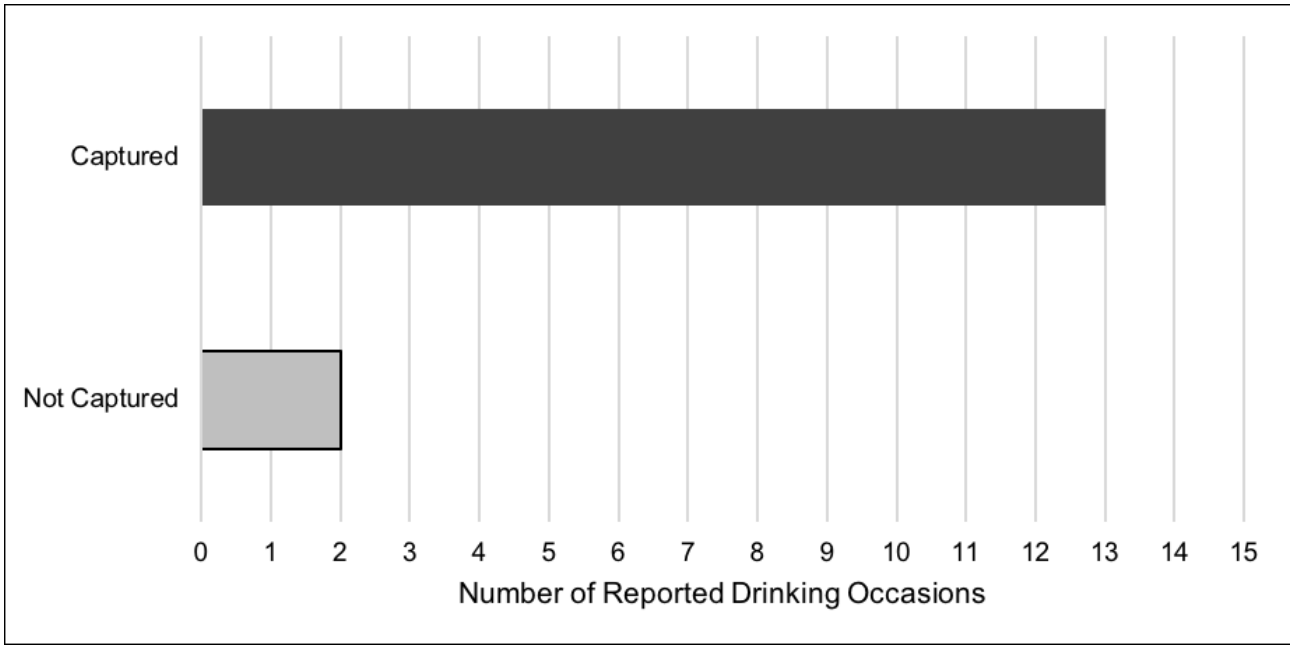


Table 3. Alcohol use estimates from SmartIntake and 24-hour online diet recalls.

Alcohol estimates	SmartIntake (N=13)[a]	Diet recalls (N=7)[b]	*P* value[c]	95% CI
Alcohol grams per drinking occasion, mean (SD); range	40.0 (32.1); 11.2-95.4	40.2 (23.6); 14.0-74.9	.25	-4.03 to 15.15
Alcohol calories per drinking occasion, mean (SD)	357.0 (254.0)	375.8 (228.3)	.26	-35.43 to 128.94
Heavy drinking episodes[d], n (%)	4 (31)	2 (29)	.99	0.10 to 16.41

[a]Alcohol reported by 60% of participants.

[b]Alcohol reported by 20% of participants.

[c]*P* values for continuous outcomes refer to within-subjects *t* tests; *P* value for count of heavy drinking episodes refers to Fisher exact test.

[d]Consumption of 4+ drinks for females, 5+ for males on one occasion, in excess of low-risk drinking guidelines from the National Institute on Alcohol Abuse and Alcoholism, which considers 14 grams of alcohol as one standard drink [35].

Acceptability and Feasibility of Using SmartIntake During Drinking Episodes

Most participants reported that it was feasible to take individual photos of alcoholic beverages if they were drinking with a meal. Participants reported that when they were drinking at parties or in social gatherings, it was more difficult to capture individual drink photos due to low lighting and social distractions. However, participants were trained to use the method flexibly and this appeared to facilitate data completeness. For example, during social events/parties, most participants sent summary photos of the number of drinks they consumed in one or two images. Some participants took before and after photos of liquor bottles to indicate how much they consumed. Others stacked solo cups and sent photos of all of their empty cups in one after-drinking image, along with a text description. In their interviews, participants reported that these methods helped them send data while minimizing the impact of sending photos on their social interactions.

SmartIntake Use in Social Situations

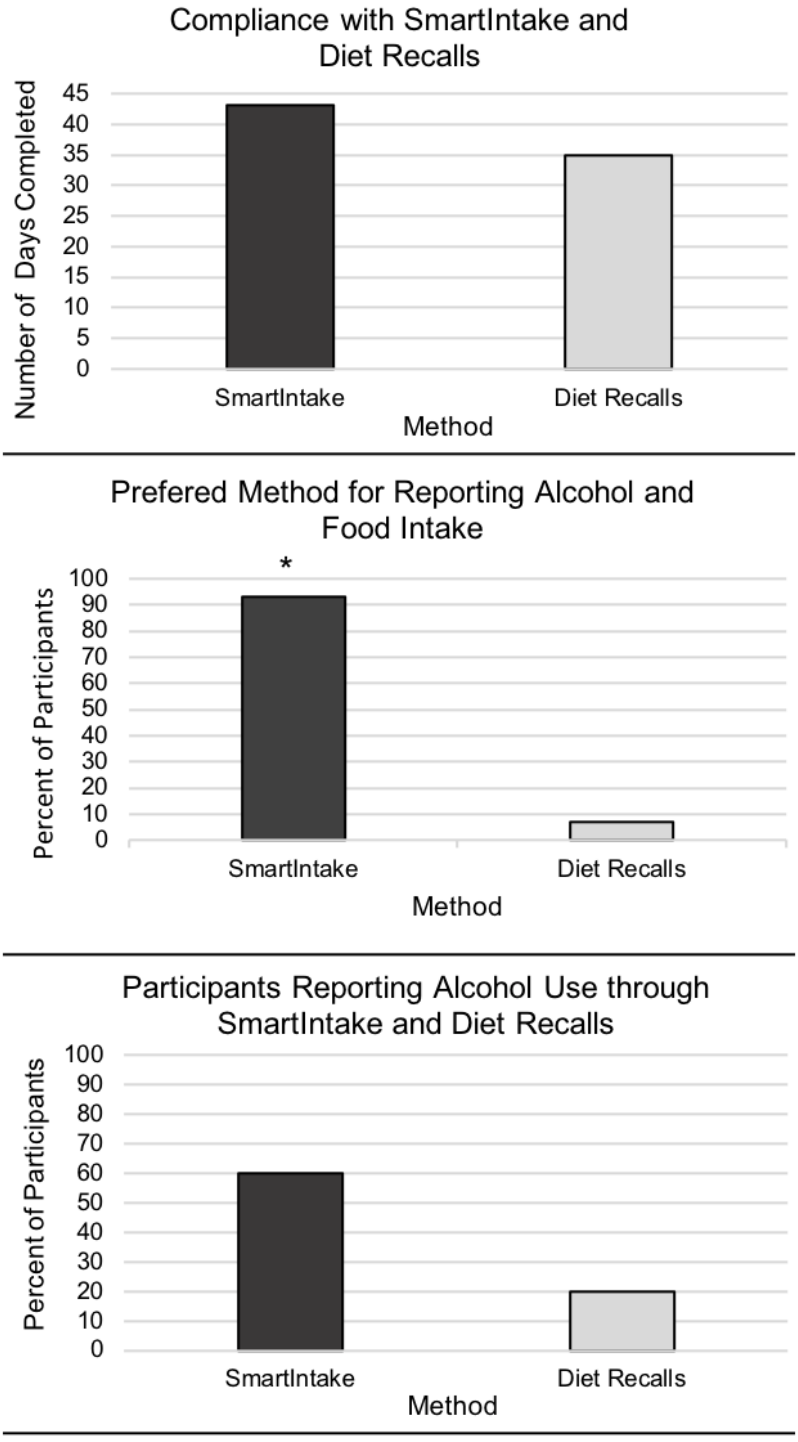
When asked to describe a time in which they forgot or almost forgot to take a food or drink photo, the vast majority of participants reported this happened while they were distracted in social situations and on weekends when they were not in normal routine. Both drinking occasions that participants reported they missed capturing with SmartIntake occurred in social drinking situations and were heavy drinking episodes. In addition, one third of participants (5/15) reported forgetting to submit a food photo while eating out with friends (n=4) or when eating on the run (n=1). The majority of participants (12/15, 80%) reported that using the app to record their alcohol and food intake did not make them feel uncomfortable. Three participants described feeling slightly awkward in social situations when they first started using the app due to taking out the reference card for each photo, but all reported this feeling diminished by the second or third day of app use.

Aim 4: Within-Subjects Comparisons of SmartIntake and Diet Recalls

Usage, preference, and alcohol use estimates are presented in Table 3 and Figure 4. Usage was significantly higher with SmartIntake versus diet recalls (t_{14}=2.26, *P*=.04, 95% CI 0.03-1.04; Figure 2). All but one participant preferred SmartIntake over the diet recalls because it was easier to use and took less time to complete (SmartIntake, 14/15 vs diet recalls, 1/15; odds ratio [OR] 121.78; *P*<.001, 95% CI 8.67-8055.30; Figure 4). Estimates of grams and calories

consumed from alcoholic drinks were not significantly different from SmartIntake estimates when alcohol was reported (Table 3). The number of participants who submitted alcohol photos using SmartIntake was triple compared to the number of participants who reported alcohol intake in the diet recalls, although the difference missed statistical significance (SmartIntake, 9/15 vs diet recalls, 3/15; OR 5.61; *P*=.06, 95% CI 0.94-44.93; Figure 4). Across all participants, total alcohol grams reported through SmartIntake was nearly double the total grams reported in recalls (SmartIntake=520.4 g vs recalls=281.3 g).

Figure 4. Within-subjects comparisons of SmartIntake and online diet recalls for usage, preference, and alcohol use reports. Significance test of compliance refers to within-subjects *t* test. Significance test for method preference refers to Fisher exact test. A significance of *P*<.05 is indicated by an asterisk.



Discussion

Principal Considerations

The current study demonstrated that using the RFPM and SmartIntake mobile app to measure alcohol intake was feasible and well accepted by college students who endorse a pattern of heavy alcohol use. This pilot was the first to measure alcohol use via mobile photography in real-time, thus circumventing the potential for biases in participant-estimated drink size and content. Our findings indicated that SmartIntake captured the majority of reported drinking occasions. Additionally, participants preferred using SmartIntake compared to standard 24-hour diet recalls administered online due to the convenience and immediacy in submitting alcohol and food data that SmartIntake afforded. Usage with SmartIntake was significantly higher than with the diet recalls, despite the procedural advantage that existed for the diet recalls, in that they could be completed later. Alcohol use estimates per drinking occasion were similar between methods when alcohol was reported. However, the number of participants who submitted alcohol photos with SmartIntake was triple compared to the number of participants who reported alcohol use in the diet recalls. Thus, our findings suggest SmartIntake assessment may be preferable as a way to gather detailed alcohol use data from young adults.

While SmartIntake methods captured the majority of reported drinking occasions, alcohol use, and heavy drinking episodes occurred less frequently than expected, based on the drinking patterns that participants endorsed at screening. Thus, our ability to test SmartIntake for assessing a full range of drinking behavior was limited, likely in part due to our brief 3-day testing period, even though it spanned the weekend. For example, 40% of participants did not drink on the days they tested SmartIntake, although most reported typically drinking multiple times per week. In addition, while most participants endorsed a pattern of weekly heavy episodic drinking, only four drinking occasions captured through SmartIntake were heavy drinking episodes and both occasions in which participants forgot to report their alcohol use via SmartIntake were heavy drinking episodes. Thus, further work and a longer testing period is needed to comprehensively evaluate the utility of SmartIntake in assessing heavy drinking episodes and a broader range of drinking behavior.

Although SmartIntake usage was high, qualitative interviews indicated that participants did occasionally forget to send photos of alcohol and food in social situations when they were distracted. In addition, participants indicated that reminder prompts were easy to miss or disregard, even though they showed on participants' phone screens, because they were not sent as text messages (this has been rectified in the more recent version of the SmartIntake app, version 3). However, our findings did indicate that participants found that the flexible approach to reporting alcohol use with the app was most acceptable and less disruptive in social drinking situations. Given that most drinking episodes among young adults do occur in social settings [36], our future work will be focused on further developing methods that facilitate participant response in social situations and in times of heightened distraction, while minimizing impact on their social interactions.

Mobile photo-based assessment of alcohol and food intake may be particularly well suited to young adults due to similarities with young adults' use of mobile phones, social media, and food and drink photography. For example, the vast majority of young adults (85/103, 83%) use photo-based social media apps such as Instagram regularly [37,38], and they often use social media–based apps to display their food and beverage intake [39,40]. Further, young adults commonly use photo-based social media apps during drinking episodes, including at parties and festivals [41,42]. Thus, SmartIntake assessment may be a natural extension of young adults' existing behavior with mobile photography of food and beverage intake. In this way, preference for and higher usage with SmartIntake as compared to diet recalls may have been influenced by participants' greater familiarity with photographing alcohol and food intake using their smartphones.

Strengths and Limitations

The study had several limitations. First, participants were college students and it is unclear to what degree findings would generalize to the general population or clinical populations. Second, we asked participants to self-report whether they missed capturing drinking occasions with SmartIntake, which could be subject to retrospective recall bias. However, participants attended the follow-up interview the day after they completed SmartIntake testing; thus, their memories of drinking over the past 3 days were likely sufficiently reliable for identifying number of drinks consumed [6]. We also structured the qualitative interview in a manner to limit socially desirable responding. In addition, we compared SmartIntake to diet recalls that were completed in the parent study, which resulted in all participants completing the diet recalls first, followed by SmartIntake. Thus, it is possible that the differences in alcohol report rates across the two methods may be due to other factors, such as timing in the semester. However, if semester timing did contribute to differences in alcohol use estimates, we would expect that the diet recalls would have captured more frequent alcohol reports. Drinking among college students is usually higher early and mid-semester, and lower around final exams [43]. Diet recalls were conducted earlier in the semester, while SmartIntake testing was conducted towards the end of the semester. Additionally, we did not ask participants about whether they missed reporting alcohol use in the recalls, so we do not have this information to compare directly with SmartIntake data on percentage of drinking episodes captured. Finally, our requirement that potential participants completed 1+ diet recall in the parent study may limit the generalizability of the findings to participants who did not complete recalls. However, the vast majority of participants in the parent study did complete 1+ recall at each time point (96/103, 93% at Visit 1; 85/103, 85% at Visit 2; 74/103, 72% at Visit 3); thus, our findings should generalize to the majority of the parent study sample. However, future research is needed to test the level of SmartIntake usage among individuals who do not engage with standard diet recall assessment methods.

Strengths of the study included the use of a sample that endorsed a pattern of heavy drinking, assessment of app usability via a standardized questionnaire specific to computer/app technology, and within-subjects comparison between SmartIntake and standardized assessment methodology.

Conclusions

Photo-based mobile assessment of alcohol use with the SmartIntake app may provide a scalable, objective measure of drinking behavior that captures data in near real-time and can be remotely delivered. This methodology provides fine-grained data on caloric and nutritional content of alcoholic beverages, which will afford future opportunities to assess caloric contributions from alcohol and alcohol-related eating to weight gain and obesity in young adults. This method could also facilitate the development of future interventions that rely on real-time treatment delivery using Ecological Momentary Intervention and Just-In-Time Adaption Intervention principles.

Acknowledgments

This work was partially supported by the Nutrition Obesity Research Center grant P30 DK072476, entitled “Nutrition and Metabolic Health Through the Lifespan,” sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases, by grant F32 AA024669-01A1 sponsored by the NIAAA, and by grant U54 GM104940 from the National Institute of General Medical Sciences, which funds the Louisiana Clinical and Translational Science Center.

Conflicts of Interest

The intellectual property for the Remote Food Photography Method and SmartIntake app are owned by Pennington Biomedical / Louisiana State University, and CKM is an inventor of the technology.

References

1. Dawson DA, Goldstein RB, Saha TD, Grant BF. Changes in alcohol consumption: United States, 2001-2002 to 2012-2013. Drug Alcohol Depend 2015 Mar 01;148:56-61 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.12.016] [Medline: 25620731]
2. White A, Hingson R. The burden of alcohol use: excessive alcohol consumption and related consequences among college students. Alcohol Res 2013;35(2):201-218 [FREE Full text] [Medline: 24881329]
3. Fazzino TL, Fleming K, Sher K, Sullivan D, Befort C. Heavy Drinking in Young Adulthood Increases Risk of Transitioning to Obesity. Am J Prev Med 2017 Aug;53(2):169-175 [FREE Full text] [doi: 10.1016/j.amepre.2017.02.007] [Medline: 28365088]
4. Nelson MC, Lust K, Story M, Ehlinger E. Alcohol use, eating patterns, and weight behaviors in a university population. Am J Health Behav 2009;33(3):227-237. [Medline: 19063644]
5. Yeomans MR. Alcohol, appetite and energy balance: is alcohol intake a risk factor for obesity? Physiol Behav 2010 Apr 26;100(1):82-89. [doi: 10.1016/j.physbeh.2010.01.012] [Medline: 20096714]
6. Sobell LC, Sobell MB. Timeline Follow-Back: A Technique for Assessing Self-Reported Alcohol Consumption. In: Litten RZ, Allen JP, editors. Measuring alcohol consumption: Psychosocial and biochemical methods. Totowa, NJ: Humana Press; 1992:41-72.
7. Buzzard IM, Faucett CL, Jeffery RW, McBane L, McGovern P, Baxter JS, et al. Monitoring dietary change in a low-fat diet intervention study: advantages of using 24-hour dietary recalls vs food records. J Am Diet Assoc 1996 Jun;96(6):574-579. [doi: 10.1016/S0002-8223(96)00158-7] [Medline: 8655904]
8. Breslow RA, Chen CM, Graubard BI, Jacobovits T, Kant AK. Diets of drinkers on drinking and nondrinking days: NHANES 2003-2008. Am J Clin Nutr 2013 May;97(5):1068-1075 [FREE Full text] [doi: 10.3945/ajcn.112.050161] [Medline: 23535109]
9. Serra MC, Treuth MS, Ryan AS. Dietary prescription adherence and non-structured physical activity following weight loss with and without aerobic exercise. J Nutr Health Aging 2014 Dec;18(10):888-893 [FREE Full text] [doi: 10.1007/s12603-014-0481-9] [Medline: 25470804]
10. Sieri S, Krogh V, Saieva C, Grobbee DE, Bergmann M, Rohrmann S, et al. Alcohol consumption patterns, diet and body weight in 10 European countries. Eur J Clin Nutr 2009 Nov;63 Suppl 4:S81-100. [doi: 10.1038/ejcn.2009.76] [Medline: 19888282]
11. Butler L, Poti JM, Popkin BM. Trends in Energy Intake from Alcoholic Beverages among US Adults by Sociodemographic Characteristics, 1989-2012. J Acad Nutr Diet 2016 Jul;116(7):1087-1100.e6 [FREE Full text] [doi: 10.1016/j.jand.2016.03.008] [Medline: 27132721]
12. Agarwal S, Fulgoni VL, Lieberman HR. Assessing alcohol intake & its dose-dependent effects on liver enzymes by 24-h recall and questionnaire using NHANES 2001-2010 data. Nutr J 2016 Dec 22;15(1):62 [FREE Full text] [doi: 10.1186/s12937-016-0180-y] [Medline: 27334005]
13. Serra-Majem L, Santana-Armas JF, Ribas L, Salmona E, Ramon JM, Colom J, et al. A comparison of five questionnaires to assess alcohol consumption in a Mediterranean population. Public Health Nutr 2002 Aug;5(4):589-594. [doi: 10.1079/PHN2001323] [Medline: 12186667]

14. Archer E, Hand GA, Blair SN. Validity of U.S. nutritional surveillance:National Health and Nutrition Examination Survey caloric energy intake data, 1971-2010. PLoS One 2013 Oct;8(10):e76632 [FREE Full text] [doi: 10.1371/journal.pone.0076632] [Medline: 24130784]
15. Bond JC, Greenfield TK, Patterson D, Kerr WC. Adjustments for drink size and ethanol content: new results from a self-report diary and transdermal sensor validation study. Alcohol Clin Exp Res 2014 Dec;38(12):3060-3067 [FREE Full text] [doi: 10.1111/acer.12589] [Medline: 25581661]
16. Dhurandhar NV, Schoeller D, Brown AW, Heymsfield SB, Thomas D, Sørensen TIA, Energy Balance Measurement Working Group. Energy balance measurement: when something is not better than nothing. Int J Obes (Lond) 2015 Jul;39(7):1109-1113 [FREE Full text] [doi: 10.1038/ijo.2014.199] [Medline: 25394308]
17. Beasley J, Riley WT, Jean-Mary J. Accuracy of a PDA-based dietary assessment program. Nutrition 2005 Jun;21(6):672-677. [doi: 10.1016/j.nut.2004.11.006] [Medline: 15925290]
18. Martin CK, Correa JB, Han H, Allen HR, Rood JC, Champagne CM, et al. Validity of the Remote Food Photography Method (RFPM) for estimating energy and nutrient intake in near real-time. Obesity (Silver Spring) 2012 Apr;20(4):891-899 [FREE Full text] [doi: 10.1038/oby.2011.344] [Medline: 22134199]
19. Martin CK, Nicklas T, Gunturk B, Correa JB, Allen HR, Champagne C. Measuring food intake with digital photography. J Hum Nutr Diet 2014 Jan;27 Suppl 1:72-81 [FREE Full text] [doi: 10.1111/jhn.12014] [Medline: 23848588]
20. Martin CK, Han H, Coulon SM, Allen HR, Champagne CM, Anton SD. A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method. Br J Nutr 2009 Feb;101(3):446-456 [FREE Full text] [doi: 10.1017/S0007114508027438] [Medline: 18616837]
21. Williamson D, Allen H, Martin P, Alfonso A, Gerald B, Hunt A. Digital photography: A new method for estimating food intake in cafeteria settings. Eat Weight Disord 2013 Jul 26;9(1):24-28. [doi: 10.1007/BF03325041]
22. Demartini KS, Carey KB. Optimizing the use of the AUDIT for alcohol screening in college students. Psychol Assess 2012 Dec;24(4):954-963 [FREE Full text] [doi: 10.1037/a0028519] [Medline: 22612646]
23. 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]
24. Conway JM, Ingwersen L, Vinyard B, Moshfegh A. Effectiveness of the US Department of Agriculture 5-step multiple-pass method in assessing food intake in obese and nonobese women. Am J Clin Nutr 2003 May;77(5):1171-1178. [doi: 10.1093/ajcn/77.5.1171] [Medline: 12716668]
25. Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
26. Lewis JR. Usability: Lessons Learned ... and Yet to Be Learned. International Journal of Human-Computer Interaction 2014 Jun 18;30(9):663-684. [doi: 10.1080/10447318.2014.930311]
27. Tullis TS, Stetson JN. A comparison of questionnaires for assessing website usability. 2004 Presented at: Usability Professionals Association Annual Conference; 2004; Minneapolis, USA p. 7-11.
28. Al Ayubi SU, Parmanto B, Branch R, Ding D. A Persuasive and Social mHealth Application for Physical Activity: A Usability and Feasibility Study. JMIR Mhealth Uhealth 2014 May;2(2):e25 [FREE Full text] [doi: 10.2196/mhealth.2902] [Medline: 25099928]
29. English LL, Dunsmuir D, Kumbakumba E, Ansermino JM, Larson CP, Lester R, et al. The PAediatric Risk Assessment (PARA) Mobile App to Reduce Postdischarge Child Mortality: Design, Usability, and Feasibility for Health Care Workers in Uganda. JMIR Mhealth Uhealth 2016 Feb 15;4(1):e16 [FREE Full text] [doi: 10.2196/mhealth.5167] [Medline: 26879041]
30. Rospo G, Valsecchi V, Bonomi AG, Thomassen IW, van Dantzig S, La Torre A, et al. Cardiorespiratory Improvements Achieved by American College of Sports Medicine's Exercise Prescription Implemented on a Mobile App. JMIR Mhealth Uhealth 2016 Jun 23;4(2):e77 [FREE Full text] [doi: 10.2196/mhealth.5518] [Medline: 27339153]
31. Sheehan B, Lee Y, Rodriguez M, Tiase V, Schnall R. A comparison of usability factors of four mobile devices for accessing healthcare information by adolescents. Appl Clin Inform 2012 Dec;3(4):356-366 [FREE Full text] [doi: 10.4338/ACI-2012-06-RA-0021] [Medline: 23227134]
32. Lewis JR. Psychometric Evaluation of the Post-Study System Usability Questionnaire: The PSSUQ. 2016 Aug 06 Presented at: Human Factors and Ergonomics Society Annual Meeting; Oct. 12-16, 1992; Atlanta, GA p. 1259-1260. [doi: 10.1177/154193129203601617]
33. Lewis JR. Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. International Journal of Human-Computer Interaction 2002 Sep;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
34. Erdinç O, Lewis JR. Psychometric Evaluation of the T-CSUQ: The Turkish Version of the Computer System Usability Questionnaire. International Journal of Human-Computer Interaction 2013 Apr;29(5):319-326. [doi: 10.1080/10447318.2012.711702]
35. US Department of Health and Human Services, National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism. Helping Patients Who Drink Too Much: A Clinician's Guide Updated Edition. 2005. URL: https://pubs.niaaa.nih.gov/publications/Practitioner/CliniciansGuide2005/guide.pdf [accessed 2018-09-04] [WebCite Cache ID 72CAu23W2]

36. Beck KH, Arria AM, Caldeira KM, Vincent KB, O'Grady KE, Wish ED. Social context of drinking and alcohol problems among college students. Am J Health Behav 2008;32(4):420-430 [FREE Full text] [doi: 10.5555/ajhb.2008.32.4.420] [Medline: 18092902]
37. Roberts JA, Yaya LHP, Manolis C. The invisible addiction: cell-phone activities and addiction among male and female college students. J Behav Addict 2014 Dec;3(4):254-265 [FREE Full text] [doi: 10.1556/JBA.3.2014.015] [Medline: 25595966]
38. Villanti AC, Johnson AL, Ilakkuvan V, Jacobs MA, Graham AL, Rath JM. Social Media Use and Access to Digital Technology in US Young Adults in 2016. J Med Internet Res 2017 Jun 07;19(6):e196 [FREE Full text] [doi: 10.2196/jmir.7303] [Medline: 28592394]
39. Hu Y, Manikonda L, Kambhampati S. What we instagram: A first analysis of instagram photo content and user types. 2014 Presented at: International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media; 2014; Ann Arbor, Michigan p. A URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewFile/8118/8087
40. Holmberg C, Hillman T, Berg C. Adolescents' presentation of food in social media: An explorative study. Appetite 2016 Apr 01;99:121-129. [doi: 10.1016/j.appet.2016.01.009] [Medline: 26792765]
41. Whitehill JM, Pumper MA, Moreno MA. Emerging adults' use of alcohol and social networking sites during a large street festival: A real-time interview study. Subst Abuse Treat Prev Policy 2015 May 20;10:21 [FREE Full text] [doi: 10.1186/s13011-015-0016-3] [Medline: 25990109]
42. Whitehill J, Rataj S, Moreno M, Pumper M. College students' use of mobile phones during episodes of alcohol and marijuana use. 2015 Presented at: 143rd American Public Health Association Annual Meeting; Oct. 31, 2015; Chicago, IL.
43. Tremblay PF, Graham K, Wells S, Harris R, Pulford R, Roberts SE. When do first-year college students drink most during the academic year? An internet-based study of daily and weekly drinking. J Am Coll Health 2010 Mar;58(5):401-411. [doi: 10.1080/07448480903540465] [Medline: 20304752]

Abbreviations

AUDIT-C: Alcohol Use Disorder Identification Test–Consumption Questions
CSUQ: Computer System Usability Questionnaire
NHANES: National Health and Nutrition Examination Survey
NIAAA: National Institute on Alcohol Abuse and Alcoholism
RFPM: Remote Food Photography Method

Edited by G Eysenbach; submitted 22.03.18; peer-reviewed by F Sartor, M Wiens; comments to author 26.04.18; revised version received 21.06.18; accepted 05.07.18; published 24.09.18.

Please cite as:
Fazzino TL, Martin CK, Forbush K
The Remote Food Photography Method and SmartIntake App for the Assessment of Alcohol Use in Young Adults: Feasibility Study and Comparison to Standard Assessment Methodology
JMIR Mhealth Uhealth 2018;6(9):e10460
URL: http://mhealth.jmir.org/2018/9/e10460/
doi:10.2196/10460
PMID:30249590

Original Paper

A Mobile Health Intervention for HIV Prevention Among Racially and Ethnically Diverse Young Men: Usability Evaluation

Hwayoung Cho[1], PhD, RN; Dakota Powell[1], MPH; Adrienne Pichon[1], MPH; Jennie Thai[2], BA; Josh Bruce[3], BA; Lisa M Kuhns[2,4], PhD, MPH; Robert Garofalo[2,4], MPH, MD; Rebecca Schnall[1], PhD, MPH, RN

[1]School of Nursing, Columbia University, New York, NY, United States

[2]Division of Adolescent Medicine, Ann & Robert H Lurie Children's Hospital of Chicago, Chicago, IL, United States

[3]Birmingham AIDS Outreach, Birmingham, AL, United States

[4]Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Corresponding Author:
Hwayoung Cho, PhD, RN
School of Nursing
Columbia University
560 West 168th Street
New York, NY, 10032
United States
Phone: 1 2123056201
Email: hc2787@columbia.edu

Abstract

Background: Mobile health (mHealth) apps have the potential to be a useful mode of delivering HIV prevention information, particularly for young men (13-24 years) who account for 21% of new HIV diagnoses in the United States. We translated an existing evidence-based, face-to-face HIV prevention curriculum into a portable platform and developed a mobile Web app: MyPEEPS Mobile.

Objective: The purpose of this study was to assess the usability of MyPEEPS Mobile from both expert and end user perspectives.

Methods: We conducted a heuristic evaluation with five experts in informatics to identify violations of usability principles and end user usability testing with 20 young men aged 15 to 18 years in New York, NY, Birmingham, AL, and Chicago, IL to identify potential obstacles to their use of the app.

Results: Mean scores of the overall severity of the identified heuristic violations rated by experts ranged from 0.4 and 2.6 (0=no usability problem to 4=usability catastrophe). Overall, our end users successfully completed the tasks associated with use case scenarios and provided comments/recommendations on improving usability of MyPEEPS Mobile. The mean of the overall Post-Study System Usability Questionnaire scores rated by the end users was 1.63 (SD 0.65), reflecting strong user acceptance of the app.

Conclusions: The comments made by experts and end users will be used to refine MyPEEPS Mobile prior to a pilot study assessing the acceptability of the app across diverse sexual minority young men in their everyday lives.

(JMIR Mhealth Uhealth 2018;6(9):e11450) doi:10.2196/11450

KEYWORDS

mobile apps; mobile health; information technology; health information technology; usability evaluation; adolescents; HIV prevention; men who have sex with men

Introduction

Background

With the rapid proliferation of mobile phone ownership across the world, use of mobile technologies in health care has expanded [1]. More than 325,000 mobile health (mHealth) apps were available globally on Apple iTunes and Google Play in 2017, and the number of mHealth apps continues to increase [2]. The ubiquitous nature of mobile phones brings convenience to everyday lives and creates opportunities to deliver health interventions in a portable format with enhanced privacy, increasing accessibility to the health interventions particularly tailored for stigmatized and disenfranchised populations [3-5].

Figure 1. Four YMSM (young men who have sex with men) avatars on the MyPEEPS Mobile app.

mHealth technology specifically has the potential to be a useful delivery mode of health information because it allows for the dissemination of information quickly and broadly [6-8].

For the success of health information technologies, usability must be considered from the start of system development [9], yet few mHealth apps have undergone rigorous usability evaluation prior to their dissemination [10]. Usability factors remain one of the major obstacles to adoption of mHealth technologies because mHealth apps produced with poor quality are difficult to use, or are misused, which can lead to unintended consequences [11-13]. Therefore, usability evaluations are necessary to identify usability violations, guide system modification, and enhance technology acceptance by end users [14]. To provide the most effective and thorough usability evaluation results, a combination of usability evaluation techniques, including both experts and intended end users, during the evaluations is recommended [15,16].

Study Context: MyPEEPS Intervention

In 2016, of the 39,782 people in the United States newly infected with HIV, 21% were youth ages 13 to 24 years and 81% of incident cases among these youth were diagnosed in young men who have sex with men (YMSM), disproportionately occurring in African-American/black and Latino/Hispanic men [17]. An original Male Youth Pursuing Education, Empowerment & Prevention around Sexuality (MyPEEPS) intervention is a theory-driven (ie, social cognitive theory) [18], manualized HIV prevention curriculum developed for racially and ethnically diverse YMSM to address the need for an evidence-based HIV prevention intervention for this population [19]. The group-based, in-person intervention was found to be efficacious on reducing sexual risk, specifically sexual risk while under the influence of alcohol or drugs, in a 12-week feasibility trial. Nonetheless, participant engagement proved challenging due to travel distance and logistics around scheduling a group-based intervention.

The use of mobile apps has been a popular way, particularly for YMSM, to get health information, connect with gay friends, and seek sex partners [20]. With the great promise of mHealth technology, we translated the existing face-to-face intervention into a mobile platform using an iterative design process [21,22]. The mobile Web app, MyPEEPS Mobile, was implemented by software developers at Little Green Software. MyPEEPS Mobile is guided by four YMSM avatars (ie, Philip aka P, Artemio, Nico, and Tommy; Figure 1) who manage their sexual health against a backdrop of personal, family-based, and relational challenges, and deliver the HIV prevention information to the end users. MyPEEPS Mobile consists of 21 activities divided into four modules or "PEEPScapades." The activities include didactic content, graphical reports, videos, and true/false and multiple-choice quizzes. A user is required to complete the activities in consecutive order. On completing each activity, the user receives a trophy as a reward to promote continued participation. The purpose of this study was to assess the usability of the mHealth intervention, MyPEEPS Mobile, from the perspectives of experts and end users.

Methods

Overview

We conducted two types of rigorous usability evaluations of MyPEEPS Mobile. First, we conducted a heuristic evaluation with informatics experts to identify violations of usability principles. Next, end user usability testing was conducted with target users, young men who are attracted to other men, to identify obstacles to their use of MyPEEPS Mobile. The Institutional Review Board (IRB) of Columbia University Medical Center in New York, NY, served as the central IRB for this study and approved all study activities.

Heuristic Evaluation

Sample Selection

Five informaticians were invited via email to participate in a heuristic evaluation of MyPEEPS Mobile. The sample size was chosen in accordance with Nielsen's recommendation to include three to five heuristic evaluators, as no additional information is likely to be produced with a larger sample [23]. Qualifications of the experts included (1) at least a Master's degree in the field of informatics and (2) training in human-computer interaction. These qualifications were essential since the quality of the heuristic evaluation is dependent on the skills and experience of the usability experts [24].

Procedures

Heuristic evaluators were given a description of the full functionality of MyPEEPS Mobile. Each heuristic evaluator completed each of the 21 activities within the app (Textbox 1) at least once.

Textbox 1. Summary of the 21 activities on MyPEEPS Mobile.

I. Intro

1. Welcome to MyPEEPS

- Introduction to the app explaining what the user is to expect. User inputs name, telephone number, email address, and how they prefer to get notifications.

2. BottomLine

- User is asked the farthest they will go with a one-time hookup in a number of sexual scenarios (when I give head, when I top, etc) and given a selection of responses about what they will and won't do and how they will do it (always use a condom, won't use a condom, will never do this).

3. Underwear Personality Quiz

- User completes a personality quiz and is introduced to the avatars that they will be seeing in the app. Avatars' personality traits and identities are shared with "gossip."

4. My Bulls-I

- User is asked to think about their important identity traits and create a list of their top five favorite or best identity traits after seeing an example of the activity done by one of the app avatars, P.

II. #realtalk

5. P's On-Again Off-Again BottomLine

- Video of a text conversation between two avatars, P and Nico, about P's new relationship and P ignoring his BottomLine. The user is asked to complete questions about why P should be concerned about his BottomLine with a new partner. There are two videos with two sets of questions (video → questions → video → questions).

6. Sexy Settings

- User is presented with a setting in which sex could be taking place and is given one potential threat to a BottomLine and are asked to select another potential threat for the given setting.

7. Goin' Downhill Fast

- User is presented with information about drugs and alcohol and how they can affect a BottomLine. Resources for additional information about drugs and alcohol are provided. After reading through the information, users complete a set of questions about the potential impact of drugs and alcohol on their BottomLine.

8. Step Up, Step Back

- User is introduced to identity traits that may identify them as a VIP (privileged)/non-VIP (nonprivileged) and then asked a series of identity-related questions. An avatar representing the user moves back and forth in a line for a night club, relative to the avatars in the app, as questions are answered.

9. HIV True/False

- User completes a series of true/false questions related to HIV, with information following a correct answer.

10. Checking in on Your BottomLine

- User is given the opportunity to review and make changes to their BottomLine, taking into consideration any information that they may have learned from completing the activities prior to this check-in.

III. Woke Up Like This

11. P Gets Woke About Safer Sex

- User is presented a scenario about P trying to make his way to the clinic to get tested. P experiences difficulties and rude behavior, and the user is presented with recommendations for managing anger and frustration.

12. Testing With Tommy

- User watches a video about a character's (Tommy) experience with getting tested for HIV for the first time. The video presents a clinic scenario and a discussion with the HIV testing and prevention counselor. Information about accessing HIV testing services is provided.

13. Well Hung

- User is introduced to the association of HIV transmission risk with different sexual behaviors categorized into no risk, low, medium, and high risk. The user completes an activity dragging and dropping a given sexual activity onto the risk category associated with the sex act.

14. Ordering Steps to Effective Condom Use

- User is presented with 12 steps for effective condom use and must correctly order the steps by selecting them chronologically from a list of all the steps.

15. Checking in on Your BottomLine Again

- User is again given the opportunity to review and make changes to their BottomLine, taking into consideration any information that they may have learned from completing the activities prior to this check-in.

IV. Making Tough Situations LITuations

16. Peep in Love

- User is presented a scene where P is with his partner and wanting to engage in sexual activity without protection, a violation of P's BottomLine. The user is then asked about possible feelings and emotions that P might be having in the scene. An overview of the feelings is given at the end so the user can see the possible "swirl of emotions" from the scenario. User is then given information about how to communicate effectively with sex partners so that they can maintain their BottomLine.

17. 4 Ways to Manage Stigma

- User is presented with four stigma management strategies, then a scene for each of the four app avatars and asked to answer which strategy each character is using in the scene.

18. Rubber Mishap

- User is asked to complete a series of questions relating to condom usage as the screen shakes to mimic being under the influence of drugs or alcohol.

19. Get a Clue!

- Jumbled scenarios are created using either a shake of the phone or press of a button. User answers from given options how they would act in the scenario, keeping the BottomLine and communication strategies in mind.

20. Last Time Checking in on Your BottomLine

- User is again given the opportunity to review and make changes to their BottomLine, taking into consideration any information that they may have learned from completing the activities prior to this check-in.

21. BottomLine Overview

- User is presented with a list of their BottomLine selections since the initial activity and subsequent check-ins.

Experts were instructed to think-aloud as they evaluated the app. The process was recorded using a TechSmith Morae Recorder [25], which enables the researcher to record and analyze the audio recording and screenshots captured during the heuristic evaluation. Following completion of the tasks, heuristic evaluators were asked to rate the severity of the violations using an online version of the Heuristic Evaluation Checklist developed by Bright et al [26], based on Nielsen's 10 heuristics [27]. Each heuristic was evaluated by one or more items and the overall severity of the identified heuristic violations were rated into five categories: no problem (0), cosmetic problem only (1), minor problem (2), major problem (3), and usability catastrophe (4). The evaluators were also asked to provide additional comments regarding the user interface. After the surveys were completed, evaluators received US $150 as compensation for their time.

Data Analysis

All experts' comments about usability problems on the evaluation form and from the Morae recordings were compiled and reviewed by two research team members. Discrepancies in coding the data according to the usability factors of Nielsen's 10 heuristics were discussed until consensus was achieved. Mean severity scores were calculated for each heuristic principle.

End User Usability Testing

Sample Selection

For end user usability testing, potential participants were recruited from local community organizations through the use of passive and active methods (ie, convenience sampling; flyers, posting on social media, and direct outreach at community-based organizations) in New York, NY; Birmingham, AL; and Chicago, IL. Eligibility criteria were (1) between 13 and 18 years of age, (2) self-identified as male, (3) male sex assigned at birth, (4) understand and read English, (5) living within the metropolitan area of one of the three cities, (6) ownership of a mobile phone, (7) sexual interest in men and having either kissed another man or plans on having sex with a man in the next year, and (8) self-reported HIV-negative or unknown status. A sample of 20 participants was anticipated to be sufficient because prior research suggests an increasing benefit with samples up to 20 in usability testing (ie, the minimum percentage of problems identified rose from 82% up to 95% when the number of users was increased from 10 to 20) [28].

Procedures

All participants were given a brief explanation of MyPEEPS Mobile. The first 10 participants were provided with use case scenario version 1 (Textbox 2); once data saturation was achieved, the remaining 10 participants were provided with use case scenario version 2 (Textbox 3). Participants were asked to complete tasks using MyPEEPS Mobile on an iOS simulator for Windows computers. While participants were doing the tasks, the computer screen was video recorded using iMotions software (iMotions Biometric Research Platform 6.0, iMotions A/S, Copenhagen), which enables researchers to present images or screen/scene recordings and synchronize data from a variety of hardware platforms, if needed (eg, eye-tracking data), simultaneously. After the participants completed the tasks, they were then asked to watch a recording of their task performance on the computer screen. Participants were encouraged to retrospectively think-aloud and asked to verbalize their thoughts about the tasks they completed while watching a replay of the screen recordings. The process, including participants' verbal comments, was audio recorded using Morae [25]. As part of the usability assessment, participants were asked to rate the app's usability using the third version of the Post-Study System Usability Questionnaire (PSSUQ) [29] administered via Qualtrics (Provo, UT, USA) following the testing of MyPEEPS Mobile. The third version of the PSSUQ is a 16-item survey instrument to assess system usability on a scale ranging from 1 (strongly agree) to 7 (strongly disagree) including a neutral midpoint. A lower score on the PSSUQ indicates higher perceived usability of the app. The study visit took approximately 2 hours and participants were compensated US $40 to US $50 for their time. Interested individuals were consented for participation with a waiver of parental permission for minors.

Textbox 2. Use case scenario version 1 (N=10).

1. Log in to the MyPEEPS Mobile
   - Click on activity #1 "Welcome to MyPEEPS!" to begin
   - Collect the trophy from activity #2, "BottomLine"
   - Collect the trophy from #3, "Underwear Personality Quiz"
   - Collect the trophy from #4, "My Bulls-I"
   - Collect the trophy from #5, "P's On-Again Off-Again BottomLine"
   - Collect the trophy from #7, "Goin' Downhill Fast"
   - Collect the trophy from #8, "Step Up, Step Back"
   - Collect the trophy from #9, "HIV True/False"
   - Collect the trophy from #10, "Checking in on Your BottomLine"
   - Collect the trophy from #13, "Well Hung??"
   - Collect the trophy from #18, "Rubber Mishap"
2. View Settings
3. Log Out

Textbox 3. Use case scenario version 2 (N=10).

1. Log in to the MyPEEPS Mobile
   - Click on activity #1 "Welcome to MyPEEPS!" to begin
   - Collect the trophy from activity #2, "BottomLine"
   - Collect the trophy from #6, "Sexy Settings"
   - Collect the trophy from #10, "Checking in on Your BottomLine"
   - Collect the trophy from #13, "Well Hung??"
   - Collect the trophy from #15, Checking in on Your BottomLine Again"
   - Collect the trophy from #17, "4 Ways to Manage Stigma"
   - Collect the trophy from #19, "Get a Clue!"
   - Collect the trophy from #20, "Last Time Checking in on Your BottomLine"
   - Collect the trophy from #21, "BottomLine Overview"
2. View Settings
3. Log Out

Data Analysis

Data analysis was based on the audio/video recordings collected by Morae [25] and iMotions software. Participants' verbalizations from the audio recordings were transcribed verbatim. Notes of critical incidents, characterized by comments, silence, repetitive actions, and error messages, were compiled from the recordings. Content analysis, a technique for making replicative and valid inferences from data, was performed by two research team members by reviewing the transcripts and critical incidents to identify common usability concerns. A third reviewer consulted in instances of uncertainty or discrepancy in the content analysis. Results from the PSSUQ were analyzed using Stata SE 14 (StataCorp LP, College Station, TX, USA) to calculate the descriptive statistics to complement the findings from the usability assessment.

Results

Heuristic Evaluation

The mean age of the heuristic evaluators was 46.2 (SD 8.9) years, and mean years of experience in informatics that they had was 13.0 (SD 4.5) years. All the heuristic evaluators were female, 60% (n=3) were Asian, and 40% (n=2) were white. Mean scores and sample comments from the heuristic evaluation were organized into Nielsen's 10 usability heuristics (Table 1) [27,30]. The mean scores of the overall severity of the identified heuristic violations ranged from 0.4 and 2.6, in which scores closest to 0 indicate a more usable app.

The heuristic principle identified as the most in need of refinement was "user control and freedom" (mean 2.60, SD 1.14). Experts pointed out that MyPEEPS Mobile did not allow users the ability to move forward and backward: the "Back to Map" button only appears at the beginning of each activity. The second heuristic most identified for improvement was visibility of system status (mean 2.20, SD 0.45). A total of 21 activities divided into four PEEPScapades (Textbox 1) were displayed along a virtual "map" within MyPEEPS Mobile (Figure 2). The heuristic evaluators indicated that it was unclear which PEEPscapade they were in on the map, and the app should keep users informed about what was going on. Moreover, heuristic evaluators identified that the white navigation arrows used to advance through the "Testing with Tommy" activity and related comics (eg, illustrating what symptoms to look for when it comes to getting tested and treated for sexually transmitted diseases) were not clearly visible (Figure 3).

In response to the usability factor "help and documentation" (mean 1.60, SD 0.89), one expert pointed to the lack of an instruction manual on how to navigate the app as a major concern. To improve "match between system and the real world" (mean 1.40, SD 0.89), experts recommended that an individual's five important identity traits in the "My Bulls-I" activity (ie, move most important at the top followed by second through fifth; Figure 4) and response options for risk level in the "Well Hung" activity (ie, move no risk to far left side, followed by greater levels of risk to the right; Figure 5) be listed in a natural/logical order.

End User Usability Testing

The mean age of the end users was 17.4 (SD 0.88, range 15-18) years. Demographic characteristics including race, ethnicity, current student status, and education level are reported in Table 2. Sexual orientation was characterized using a gradient scale ranging from exclusively gay/homosexual to exclusively heterosexual, to capture the fluidity of sexuality at the time of the survey. Descriptive statistics on technology use, including type of mobile phone, social media sites, and apps are reported in Table 3. The majority of participants (85%, n=17) reported almost constant internet use (more than several times a day). The same percentage (85%, n=17) of participants reported using mobile devices (eg, mobile phone, tablet, and cell phone) as opposed to using laptop/desktop (15%, n=3) to access the internet in the past month. The mean duration of participants' use of mobile apps on a mobile phone per day was 9.40 (SD 5.52) hours.

Table 1. Mean severity scores[a] and sample comments from the heuristic evaluation.

Nielsen's 10 usability heuristics	Mean (SD)	Sample comments
Visibility of system status	2.20 (0.45)	Unclear where I am on the map
Match between system and the real world	1.40 (0.89)	Five identity traits should be listed from top-most important to bottom-least important
User control and freedom	2.60 (1.14)	Unavailable "Back to Map" (only available at the beginning of each activity)
Consistency and standards	0.40 (0.89)	Hints should be consistently provided for both incorrect/correct answers
Help users recognize, diagnose, and recover from errors	1.00 (1.00)	Error messages should provide with additional information of incorrect answers
Error prevention	1.00 (1.41)	Data entry boxes should contain default values (eg, email address/phone number)
Recognition rather than recall	1.00 (1.00)	Need instructions on how to answer; vertical compression to see buttons
Flexibility and efficiency of use	0.80 (1.30)	Have an option of directly texting a link to friends
Esthetic and minimalist design	0.80 (0.84)	Visual layout of "BottomLine Overview" should be redesigned for simplicity
Help and documentation	1.60 (0.89)	No manual on how to navigate the app

[a]Rating score from 0=best to 5=worst; no usability problem (0), cosmetic problem only (1), minor usability problem (2), major usability problem (3), and usability catastrophe (4).

Figure 2. Map on MyPEEPS Mobile.

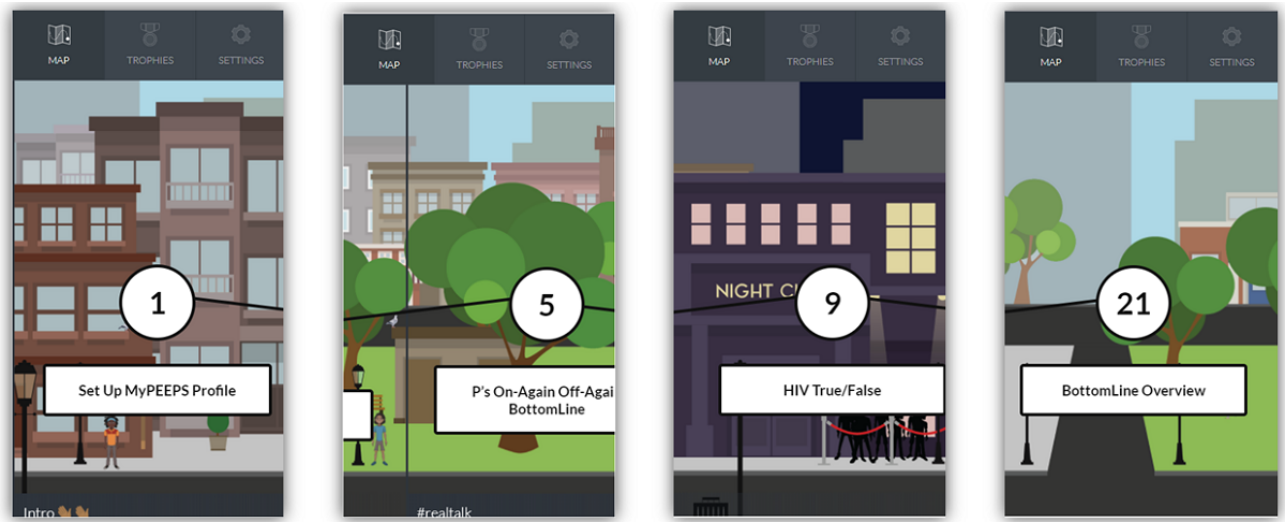


Figure 3. Comics with unclear navigation.

Positive Comments

Overall, our end users successfully completed the tasks associated with the use case scenarios (Textboxes 2 and 3) and provided positive comments regarding the use of MyPEEPS Mobile. For example, participants liked the design and layout of MyPEEPS Mobile. One participant stated:

> *The basic structure is that there's pretty much an outline and you have four boys/men who are kind of the characters that kind of take you along this journey to HIV prevention and sexual health for MSM. This is sort of like fun. I like cartoons, videos, and quizzes.*

Also, participants liked the ease of the overall app use. One participant stated:

> *I think it was pretty easy once I got the hang of moving to the side. I would just click the number and then I would start the quiz. It was simple. It was pretty quick.*

Recommendations

General App Use

Participants provided recommendations to improve usability by expressing their frustrations in general use of the app. For example, several participants commented that receiving error messages over and over frustrated them. They suggested that error messages may be more helpful if an explanation is provided for wrong answers, and they preferred to be provided with a correct answer after two wrong attempts. Moreover, participants identified several terms might be difficult to understand for younger participants (eg, ages 13-15 years). They suggested that explanations be provided for potentially unfamiliar terms (eg, related to sex work, such as "client/tricks"). In addition to the supplementary explanation, participants recommended that the key sexual health terms/keywords (eg, give head/get head/top/bottom) be bolded for emphasis. Several participants reported an issue with the "Previous" button at the end of an activity; instead of taking them to the previous page, it would erroneously take them back to the beginning of the activity.

Figure 4. My Bulls-I activity.

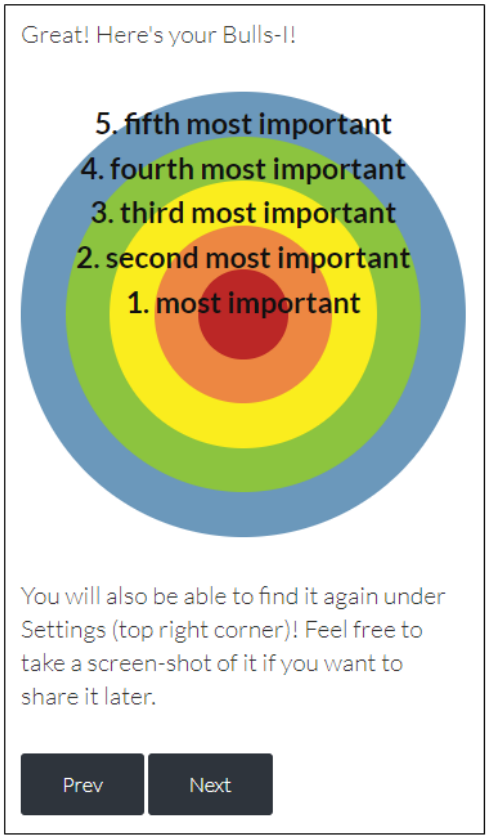


Figure 5. Well Hung activity.

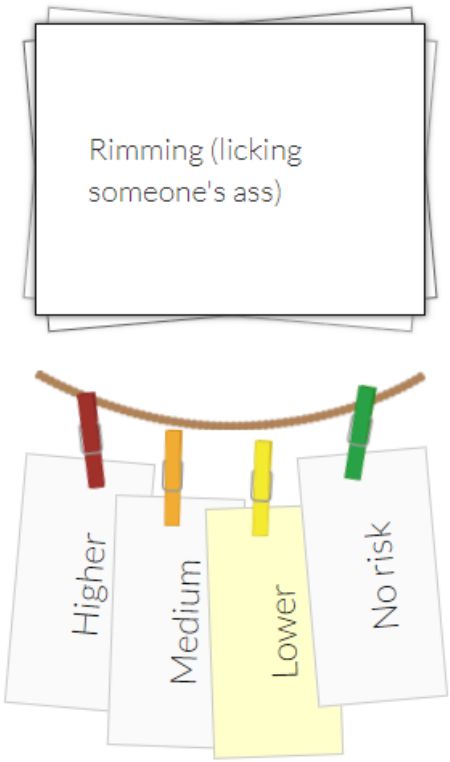


Table 2. Demographic characteristics of participants (N=20).

Characteristic	n (%)	New York (n=9), n (%)	Chicago (n=7), n (%)	Birmingham (n=4), n (%)
Gender identity				
Male	20 (100)	9 (100)	7 (100)	4 (100)
Sexual orientation				
Only gay/homosexual	14 (70)	7 (78)	5 (71)	2 (50)
Mostly gay/homosexual	2 (10)	2 (22)	0 (0)	0 (0)
Bisexual	3 (15)	0 (0)	2 (29)	1 (25)
Something else	1 (5)	0 (0)	0 (0)	1 (25)
Race				
White	9 (45)	2 (22)	4 (57)	3 (75)
Black or African-American	4 (20)	2 (22)	1 (14)	1 (25)
Hispanic or Latino/Latinx	4 (20)	2 (22)	2 (29)	0 (0)
Asian or Asian American	2 (10)	2 (22)	0 (0)	0 (0)
Multiracial	1 (5)	1 (11)	0 (0)	0 (0)
Ethnicity				
Hispanic	9 (45)	6 (67)	3 (43)	0 (0)
Non-Hispanic	11 (55)	3 (33)	4 (57)	4 (100)
Current student status				
Currently a student	16 (80)	8 (89)	5 (71)	3 (75)
Highest level of education completed				
Grade 8	2 (10)	2 (22)	0 (0)	0 (0)
Some high school	7 (35)	3 (33)	3 (43)	1 (25)
High school diploma (GED[a])	6 (30)	1 (11)	3 (43)	2 (50)
Some college	5 (25)	3 (33)	1 (14)	1 (25)

[a]GED: General Equivalency Diploma.

Specific Activities Within the App

Participants also provided comments on specific activities within MyPEEPS Mobile. For instance, several participants noted an issue with the features and functions on an activity, Underwear Personality Quiz, in which participants were introduced to four YMSM avatars, and the four avatars' personality traits were shared with a "gossip" link (Figure 6). Participants had difficulty recalling if they had viewed each avatar's gossip page. They recommended that an indication mark (eg, checkmark) on the top right corner of each avatar be shown when the avatar's gossip has been viewed.

A running theme of the app was sexual risk reduction and goal-setting through an activity called the BottomLine. In this activity, participants were challenged to articulate how much risk they were willing to accept for different sexual acts. They were asked to continually reconsider these limits as they progressed through the app (Textbox 1; the activity appears four times throughout the app). At the end, participants were presented with the activity #21, BottomLine Overview, which shows them a chronological overview of how their BottomLine changed as they progressed through the app. The participants were then encouraged to continue to stick to their sexual health goals. Many participants felt that the activity required too much reading (ie, full history of the end user's BottomLine changes for each sexual act), and/or they did not recognize that the display reflected their own changes (ie, selected responses when they previously completed the four BottomLine activities). One participant commented, "It was like a receipt. It was so long! Oh, I thought it was other people's [BottomLines]." Rather than a full report, participants suggested that we show only the most current BottomLine responses and highlight where they made changes. One participant stated:

> *I need only the most current bottom line. I feel like if you click on the bottom line link, you should show like the most current one, and then below that maybe like a link to see the previous ones.*

Another participant commented:

> *Just divide it. Put like all of what I first answered my bottom like, and have it separate in the first square, and then, read it down in a different square, like point out some changes I've made to my bottom line since the last time. So, I can compare to the ones on top to see how different they are so they're not all like mixed.*

Table 3. Technology use by participants (N=20).

Question	Participants
Frequency of internet use, n (%)	
Almost constantly	17 (85)
Several times a day	3 (15)
Devices used in the past month to access the internet, n (%)	
Mobile phone/smartphone/mobile handheld device	17 (85)
Laptop/desktop	3 (15)
Model/type of mobile phone used, n (%)	
iPhone	16 (80)
Android phone	2 (10)
Windows phone	1 (5)
Other (unknown)	1 (5)
Frequency of using social media sites in the past month, n (%)	
Several times a day	17 (85)
About once a day	2 (10)
Once every few weeks	1 (5)
Top sites or apps used for social networking, n (%)	
Snapchat	18 (90)
Instagram	17 (85)
Facebook	9 (45)
Twitter	6 (30)
YouTube	2 (10)
Tumblr	2 (10)
LinkedIn	1 (5)
Daily text messages sent and received on cell phones, mean (SD)	183.05 (204.03)
App use on mobile phones (hours/day), mean (SD)	9.40 (5.52)

Figure 6. Underwear Personality Quiz.

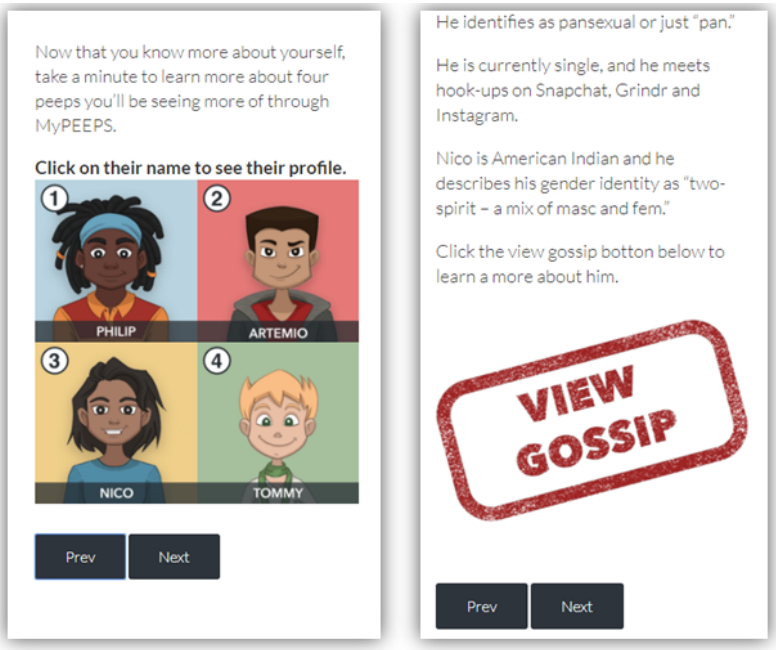


Table 4. Post-Study System Usability Questionnaire (PSSUQ) scores[a] (N=20).

Construct	Mean (SD)
System quality	1.48 (0.59)
Information quality	1.87 (0.86)
Interface quality	1.55 (0.77)
PSSUQ overall	1.63 (0.65)

[a]Rating score from 1=best to 7=worst (16 items).

Participants' perceived usability scores rated using PSSUQ [29] are reported in Table 4. Mean of the overall PSSUQ score was 1.63 (SD 0.65), reflecting strong user acceptance of MyPEEPS Mobile.

Discussion

Principal Findings

Despite the proliferative use of mobile technologies in health care, few mHealth apps have been released with consideration of their quality through comprehensive and rigorous usability evaluations. Given a lack of evidence-based mHealth apps for HIV risk reduction in high-risk populations, we developed a mobile HIV prevention app for young men ages 13 to 18 years and assessed the app's usability through a heuristic evaluation with informatics experts and end user testing to identify potential usability issues.

Usability factors remain obstacles to mobile technology adoption. Usability evaluations are foundational to the success of achieving systems that meet human-computer interaction principles and in many cases improve their application in a real-world setting [31]. In the context of rigorous usability testing of the systems, it is critical to choose the most appropriate evaluation methods that meet the study aims and ultimately achieve the goals of the systems. We employed two usability evaluation methods most commonly used in usability studies (ie, a heuristic evaluation and end user testing) to capture different usability perspectives from experts and end users [32]. Similar to prior research, usability experts were more likely to identify usability problems related to general interface features working in a natural and logical order [33,34], whereas end users identified those related to impact on task performance interacting with the app [16]. For example, contrary to feedback received from experts in the heuristic evaluation regarding the usability factor "match between system and the real world," our end users did not identify the natural/logical order of response options as a problem. Given the natural/logical order matched to the real world, no risk (0) should start on the left ending with higher risk (3) on the right in the Well Hung activity (Figure 5). An example of the attitude of end users regarding the ordering issues was expressed by one participant, who stated, "It was something that I didn't really pay attention to. I would say like I don't mind it. It doesn't matter." Inclusion of intended end users in the usability testing, in addition to the heuristic experts, enabled us to identify both logic and flow issues as well as functionality most important to end users to support overall engagement with the app [35].

Use case scenarios play an important role in usability evaluations, impacting the quality of usability testing [36]. The use case scenarios should be formulated to facilitate determination of system usability by researchers. Guided by the objectives of usability testing, use case scenarios should include key tasks that can provide valid usability data related to users' experience with the app use. In this study, we utilized two versions of use case scenarios to capture specific aspects of representative tasks as well as "big picture" issues related to the goals of the app. For example, in the first version of use case scenarios we included tasks that examine end users' performance on every type of learning activity including comics, animated videos, and games. Given that the running theme throughout the app was sexual risk reduction and goal-setting via the BottomLine activity, in the next version of use case scenarios we included all tasks to test end users' engagement in BottomLine activities in addition to the learning activities we included in the first version. Using these different versions of robust use case scenarios in end user usability testing enabled us to identify areas where usability was a potential concern and to obtain users' valuable comments on their overall app use (ie, version 1) as well as specific task performance (ie, version 2), which is a strength of our usability study.

Although findings from our usability evaluations yielded specific feedback regarding ways to improve users' experience, the overall usability scores rated by the PSSUQ were high, which indicated that our app was perceived as highly usable. The results support the potential for high acceptability of MyPEEPS Mobile. The promising usability of the app provides a foundation for user satisfaction in the planned randomized controlled trial which aims to reduce sexual risk behavior in a high-risk population.

Limitations

The generalizability of the results may be limited by the study sample, settings, and inclusion/exclusion criteria. Our targeted population was diverse YMSM living within the metropolitan area in New York, Chicago, and Birmingham, and those who had either kissed another man or planned on having sex with a man in the next year. Results may differ in transgender groups, other groups who live in rural areas, or those who have more/less experience in sexual activities. Although the age range for inclusion in this study was between 13 and 18 years, our participants were between 15 and 18 years of age, which may differ in younger adolescent MSM (eg, 13-14 years of age). A limitation of this study was related to the participants' self-reported data such as end users' perceived usability scores, which may be influenced and could bias the results.

A limitation of this study was that we conducted the usability evaluations on a computer as opposed to a mobile device. We chose to conduct the evaluations on a computer so that we would be able to analyze the data more effectively using Morae and iMotions. Understanding that there may be differences in computer and mobile device user interactions, we employed an iOS simulator on the computer so that the app can be utilized in the same manner as on a mobile phone. Although still a limitation, the use of the iOS simulator minimized its impact.

Conclusions

We tested the usability for an evidence-based HIV prevention mobile app intended for diverse YMSM through a heuristic evaluation with informatics experts and end user testing. The use of the two usability assessment methods for a mHealth app added value to this study by producing reliable results of a user interface from experts as well as user interaction with the app from end users. Findings from our rigorous usability evaluations will be used to refine the content, organization, and workflow of MyPEEPS Mobile. Following these refinements, we will conduct a 6-week pilot study to assess end users' acceptability of the app before beginning a multicity, 12-month efficacy study. Our work highlights the importance of utilizing a rigorous usability approach to refine a mHealth app before it is deployed in a high stakes environment.

Acknowledgments

This research was supported by the National Institute of Minority and Health Disparities of the National Institutes of Health (NIH) under award number U01MD11279 (multiple principal investigators including RS and RG). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflicts of Interest

None declared.

References

1. US Department of Health and Human Services. US Food and Drug Administration. 2018. Mobile medical applications URL: https://www.fda.gov/MedicalDevices/DigitalHealth/MobileMedicalApplications/default.htm [accessed 2018-08-23] [WebCite Cache ID 71tC4sld5]
2. Research2Guidance. mHealth App Economics 2017: Current Status and Future Trends in Mobile Health. 2017 Nov. URL: https://www.uzelf.org/wp-content/uploads/2017/12/R2G-mHealth-Developer-Economics-2017-Status-And-Trends.pdf [accessed 2018-08-23] [WebCite Cache ID 71tCV7jsy]
3. Amanda L, Rich L, Scott C, Kristen P. Teens and Mobile Phones. Washington, DC: Pew Internet & American Life Project; 2010 Apr 10. URL: http://www.pewinternet.org/files/old-media//Files/Reports/2010/PIP-Teens-and-Mobile-2010-with-topline.pdf [accessed 2018-08-23] [WebCite Cache ID 71tCY9tT4]
4. Kleinke JD. Dot-gov: market failure and the creation of a national health information technology system. Health Aff (Millwood) 2005;24(5):1246-1262. [doi: 10.1377/hlthaff.24.5.1246] [Medline: 16162569]
5. Devi BR, Syed-Abdul S, Kumar A, Iqbal U, Nguyen P, Li YJ, et al. mHealth: an updated systematic review with a focus on HIV/AIDS and tuberculosis long term management using mobile phones. Comput Methods Programs Biomed 2015 Nov;122(2):257-265. [doi: 10.1016/j.cmpb.2015.08.003] [Medline: 26304621]
6. Roess A. The promise, growth, and reality of mobile health - another data-free zone. N Engl J Med 2017 Nov 23;377(21):2010-2011. [doi: 10.1056/NEJMp1713180] [Medline: 29116869]
7. Sullivan PS, Grey JA, Simon RB. Emerging technologies for HIV prevention for MSM: what we have learned, and ways forward. J Acquir Immune Defic Syndr 2013 Jun 1;63 Suppl 1:S102-S107 [FREE Full text] [doi: 10.1097/QAI.0b013e3182949e85] [Medline: 23673879]
8. Kay M. mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. Geneva: World Health Organization; 2011.
9. Goldberg L, Lide B, Lowry S, Massett HA, O'Connell T, Preece J, et al. Usability and accessibility in consumer health informatics current trends and future challenges. Am J Prev Med 2011 May;40(5 Suppl 2):S187-S197. [doi: 10.1016/j.amepre.2011.01.009] [Medline: 21521594]
10. Zapata B, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
11. Zahra F, Hussain A, Mohd H. Usability factors of mobile health application for chronic diseases. In: AIP Conference Proceedings 1761. 2016 Aug 12 Presented at: ICAST'16 International Conference on Applied Science and Technology; April 11-13, 2016; Kedah, Malaysia p. 020108 URL: https://aip.scitation.org/doi/pdf/10.1063/1.4960948
12. Yusof MM, Stergioulas L, Zugic J. Health information systems adoption: findings from a systematic review. Stud Health Technol Inform 2007;129(Pt 1):262-266. [Medline: 17911719]
13. Schnall R, Mosley JP, Iribarren SJ, Bakken S, Carballo-Diéguez A, Brown IW. Comparison of a user-centered design, self-management app to existing mHealth apps for persons living with HIV. JMIR Mhealth Uhealth 2015 Sep 18;3(3):e91 [FREE Full text] [doi: 10.2196/mhealth.4882] [Medline: 26385783]

14. Issa T, Isaias P. Usability and human computer interaction (HCI). In: Sustainable Design. London: Springer; 2015:19-36.
15. Yen P, Bakken S. A comparison of usability evaluation methods: heuristic evaluation versus end-user think-aloud protocol - an example from a web-based communication tool for nurse scheduling. AMIA Annu Symp Proc 2009 Nov 14;2009:714-718 [FREE Full text] [Medline: 20351946]
16. Jaspers MWM. 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]
17. Centers for Disease Control and Prevention. 2018. HIV among youth URL: https://www.cdc.gov/hiv/group/age/youth/index.html [accessed 2018-08-23] [WebCite Cache ID 71tEyu4o5]
18. Bandura A. Social cognitive theory exercise control of HIV infection. In: DiClemente R, Peterson J, editors. Preventing AIDS: Theories and Methods of Behavioral Interventions. New York: Plenum Press; 1994:25-59.
19. Hidalgo MA, Kuhns LM, Hotton AL, Johnson AK, Mustanski B, Garofalo R. The MyPEEPS randomized controlled trial: a pilot of preliminary efficacy, feasibility, and acceptability of a group-level, HIV risk reduction intervention for young men who have sex with men. Arch Sex Behav 2015 Feb;44(2):475-485 [FREE Full text] [doi: 10.1007/s10508-014-0347-6] [Medline: 25135064]
20. Macapagal K, Moskowitz DA, Li DH, Carrión A, Bettin E, Fisher CB, et al. Hookup app use, sexual behavior, and sexual health among adolescent men who have sex with men in the United States. J Adolesc Health 2018 Jun;62(6):708-715. [doi: 10.1016/j.jadohealth.2018.01.001] [Medline: 29784114]
21. Schnall R, Kuhns L, Hidalgo M, Hirshfield S, Pearson C, Radix A, et al. Development of MyPEEPS Mobile: a behavioral health intervention for young men. Stud Health Technol Inform 2018;250:31. [Medline: 29857362]
22. Schnall R, Kuhns L, Bullock K, Powell D, Thai J, Pearson C, et al. Participatory end-user feedback to update MyPEEPS: a theory-driven evidence based intervention for YMSM. 2018 Nov 10 Presented at: American Public Health Association (APHA) Annual Meeting & Expo; Nov 10-14, 2018; San Diego, CA.
23. Nielsen J, Molich R. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1990 Apr 01 Presented at: CHI '90 SIGCHI Conference on Human Factors in Computing Systems; April 1-5, 1990; Seattle, WA p. 249-256.
24. Po S, Howard S, Vetere F, Skov M. Heuristic evaluation and mobile usability: bridging the realism gap. In: Mobile Human-Computer Interaction - MobileHCI 2004. Berlin: Springer; 2004:49-60.
25. TechSmith. Morae Usability and Web Site Testing. Michigan, USA: TechSmith Corporation; 1995.
26. Bright TJ, Bakken S, Johnson SB. Heuristic evaluation of eNote: an electronic notes system. AMIA Annu Symp Proc 2006:864 [FREE Full text] [Medline: 17238484]
27. Nielsen J. Heuristic evaluation. In: Usability Inspection Methods. New York: John Wiley & Sons; 1994:25-62.
28. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [Medline: 14587545]
29. Lewis JR. Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. Int J Hum-Comput Int 2002 Sep;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
30. Nielsen J. Finding usability problems through heuristic evaluation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1992 May 03 Presented at: CHI '92 SIGCHI Conference on Human Factors in Computing Systems; May 3-7, 1992; Monterey, CA p. 373-380.
31. Norgaard M, Hornb K. What do usability evaluators do in practice? An explorative study of think-aloud testing. In: Proceedings of the 6th Conference on Designing Interactive Systems. 2006 Presented at: DIS '06 6th conference on Designing Interactive systems; June 26-28, 2006; University Park, PA p. 209-218.
32. Lai T, Larson E, Rockoff ML, Bakken S. User acceptance of HIV TIDES--Tailored Interventions for Management of Depressive Symptoms in persons living with HIV/AIDS. J Am Med Inform Assoc 2008;15(2):217-226 [FREE Full text] [doi: 10.1197/jamia.M2481] [Medline: 18096915]
33. Lathan CE, Sebrechts MM, Newman DJ, Doarn CR. Heuristic evaluation of a web-based interface for internet telemedicine. Telemed J 1999;5(2):177-185. [doi: 10.1089/107830299312140] [Medline: 10908430]
34. Lai T. Iterative refinement of a tailored system for self-care management of depressive symptoms in people living with HIV/AIDS through heuristic evaluation and end user testing. Int J Med Inform 2007 Oct;76 Suppl 2:S317-S324 [FREE Full text] [doi: 10.1016/j.ijmedinf.2007.05.007] [Medline: 17616431]
35. Schnall R, Bakken S, Rojas M, Travers J, Carballo-Dieguez A. mHealth technology as a persuasive tool for treatment, care and management of persons living with HIV. AIDS Behav 2015 Jun;19 Suppl 2:81-89 [FREE Full text] [doi: 10.1007/s10461-014-0984-8] [Medline: 25572830]
36. Russ A, Saleem J. Ten factors to consider when developing usability scenarios and tasks for health information technology. J Biomed Inform 2018 Feb;78:123-133. [doi: 10.1016/j.jbi.2018.01.001] [Medline: 29329702]

Abbreviations

GED: General Equivalency Diploma
IRB: Institutional Review Board

mHealth: mobile health
MyPEEPS: Male Youth Pursuing Education, Empowerment & Prevention around Sexuality
PSSUQ: Post-Study System Usability Questionnaire
YMSM: young men who have sex with men

Edited by G Eysenbach; submitted 29.06.18; peer-reviewed by C Adewole, S Canidate, K Bond, D Levine; comments to author 25.07.18; revised version received 03.08.18; accepted 07.08.18; published 07.09.18.

Please cite as:
Cho H, Powell D, Pichon A, Thai J, Bruce J, Kuhns LM, Garofalo R, Schnall R
A Mobile Health Intervention for HIV Prevention Among Racially and Ethnically Diverse Young Men: Usability Evaluation
JMIR Mhealth Uhealth 2018;6(9):e11450
URL: http://mhealth.jmir.org/2018/9/e11450/
doi:10.2196/11450
PMID:30194060

Original Paper

A Dietary Assessment App for Hospitalized Patients at Nutritional Risk: Development and Evaluation of the MyFood App

Mari Mohn Paulsen[1,2], MS; Martina Lovise Lindhart Hagen[2], MS; Marte Hesvik Frøyen[3], MS; Rikke Julie Foss-Pedersen[3], MS; Dagfinn Bergsager[3], BSc; Randi Julie Tangvik[1,4], PhD; Lene Frost Andersen[1,2], PhD

[1]National Advisory Unit on Disease-Related Malnutrition, Division of Cancer Medicine, Oslo University Hospital, Oslo, Norway

[2]Institute of Basic Medical Sciences, Department of Nutrition, University of Oslo, Oslo, Norway

[3]The University Center for Information Technology, University of Oslo, Oslo, Norway

[4]Department of Clinical Medicine, Faculty of Medicine, University of Bergen, Bergen, Norway

Corresponding Author:
Mari Mohn Paulsen, MS
Institute of Basic Medical Sciences
Department of Nutrition
University of Oslo
PO Box 1046
Blindern
Oslo, 0317
Norway
Phone: 47 22851103
Email: m.m.paulsen@medisin.uio.no

Abstract

Background: Disease-related malnutrition is a common challenge among hospitalized patients. There seems to be a lack of an effective system to follow-up nutritional monitoring and treatment of patients at nutritional risk after risk assessment. We identify a need for a more standardized system to prevent and treat disease-related malnutrition.

Objective: We aimed to develop a dietary assessment app for tablets for use in a hospital setting and to evaluate the app's ability to measure individual intake of energy, protein, liquid, and food and beverage items among hospitalized patients for two days. We also aimed to measure patients' experiences using the app.

Methods: We have developed the MyFood app, which consists of three modules: 1) collection of information about the patient, 2) dietary assessment function, and 3) evaluation of recorded intake compared to individual needs. We used observations from digital photography of the meals, combined with partial weighing of the meal components, as a reference method to evaluate the app's dietary assessment system for two days. Differences in the intake estimations of energy, protein, liquid, and food and beverage items between MyFood and the photograph method were analyzed on both group and individual level.

Results: Thirty-two patients hospitalized at Oslo University Hospital were included in the study. The data collection period ran from March to May 2017. About half of the patients had ≥90% agreement between MyFood and the photograph method for energy, protein, and liquid intake on both recording days. Dinner was the meal with the lowest percent agreement between methods. MyFood overestimated patients' intake of bread and cereals and underestimated fruit consumption. Agreement between methods increased from day 1 to day 2 for bread and cereals, spreads, egg, yogurt, soup, hot dishes, and desserts. Ninety percent of participants reported that MyFood was easy to use, and 97% found the app easy to navigate.

Conclusions: We developed the MyFood app as a tool to monitor dietary intake among hospitalized patients at nutritional risk. The recorded intake of energy, protein, and liquid using MyFood showed good agreement with the photograph method for the majority of participants. The app's ability to estimate intake within food groups was good, except for bread and cereals which were overestimated and fruits which were underestimated. The app was well accepted among study participants and has the potential to be a dietary assessment tool for use among patients in clinical practice.

(JMIR Mhealth Uhealth 2018;6(9):e175) doi:10.2196/mhealth.9953

KEYWORDS

decision support system; disease-related malnutrition; eHealth; mHealth; dietary assessment; validation study

Introduction

Disease-related malnutrition is a common challenge in patients with chronic or severe diseases [1] with a prevalence of 30%-50% in hospitals [1-7]. Malnutrition has several health-related consequences for patients. It increases morbidity and mortality [1,2,8-10], length of stay [2,3,9,11,12], and readmission rates [2]. Disease-related malnutrition has significant economic consequences for the health care system [8,12-13].

According to the Norwegian "National guidelines for prevention and treatment of malnutrition" [14] and European guidelines recommended by the European Society of Clinical Nutrition and Metabolism [15], all patients should be screened for nutritional risk upon admission to hospital and weekly thereafter. Information on the patient's nutritional status and treatment should be documented in medical records and communicated to the next level of care. All patients at nutritional risk should have an individual nutrition plan including documentation of nutritional status, needs, dietary intake, and recommended actions. Hospitals, nursing homes, and home care services are responsible for integrating nutrition in the care and treatment of all patients [14].

European data from the nutritionDay survey indicates that dietary assessment is only performed for a small number of patients at nutritional risk and that documentation of food intake is rarely done [16]. Norwegian studies have reported that about half [17] or fewer than half [7] of patients identified to be at nutritional risk receive nutritional treatment. A barrier to adequate nutritional care for malnourished patients in hospitals is the absence of routines, as demonstrated in qualitative studies among nurses in Norway and Sweden [18,19]. Nurses report a lack of tools to estimate patients' needs and the content of energy and protein in hospital menus [11]. They also report insufficient knowledge and skills to identify and treat malnourished patients [18,20].

In the next decade, the need for healthcare will increase and, there will be a shortage of labor. This should be met with more effective, less people-demanding services and increased use of welfare technology [21]. There seems to be a lack of an effective system to follow up nutritional treatment in the healthcare system. We have identified a need for a more standardized system for prevention and treatment of disease-related malnutrition. To the best of our knowledge, no studies regarding development of an electronic decision support system for prevention and treatment of disease-related malnutrition among hospitalized patients at nutritional risk have been performed.

We developed an app, MyFood (MinMat), for mini tablet computers as part of a decision support system to prevent and treat disease-related malnutrition. Assessment in the app is based on self-reported dietary intake where the patient (or a nurse) records consumption of food and beverages. The memory of intake, ability to estimate portion sizes, and perceptions of socially desirable responses are well-known challenges associated with self-reported dietary intake [22]. Self-reported methods for assessment of dietary intake have been found to underestimate energy intake by approximately 20% when compared to doubly labeled water [23-25]; dietary assessment methods should always be validated because of these methodological challenges [22]. Therefore, evaluation of MyFood's ability to track the patients' dietary intake is of crucial importance.

The aim of this study was to develop a dietary assessment app for tablets for use in a hospital setting and to evaluate the app's ability to measure individual intake of energy, protein, liquid, and food and beverages for two days compared to photograph observations combined with partial weighing as the reference method. We also aimed to measure the patients' experiences using the app.

Methods

Development of the MyFood App

My Food was developed by researchers at the University of Oslo and Oslo University Hospital (OUH) and by interaction designers and developers at the University Center for Information Technology (USIT).

Nurses and patients were involved in the design process. Paper sketches of MyFood were developed and explored with three nurses and three patients at the Department of Gastrointestinal Surgery at OUH, Rikshospitalet. The feedback we received was used to modify the design and content of the app before the technical development process began. A prototype of MyFood was then developed and tested by four patients and two nurses. Their feedback was used for additional modifications of MyFood before the evaluation study was performed.

MyFood consisted of the following three modules:

Module 1: Collection of Information About the Patient

In the first module, the nurse, or other healthcare professional, recorded information about the patient. This information included: Norwegian patient registry (NPR) number, gender, date of birth, height (in centimeters), weight (in kilograms), whether the patient had a fever (and, if so, the number of degrees, and whether the patient was following a special diet or had any special preferences with regard to food or beverages.

Module 2: Dietary Assessment Function

Figure 1 shows the main menu in the dietary assessment function in MyFood.

Recording of food intake was done by first selecting the relevant meal category and then selecting the category for the food or beverage item. The food and beverage categories included pictures of the different items. Pictures could also be found using free text search. After selecting the food or beverage item consumed, the item amount was recorded. Portion size could be selected with a precision of a half unit. Figure 2 is a flowchart of dietary recording in the app. Multimedia Appendix 1 shows some selected print screens from MyFood.

Figure 1. The main menu of dietary recording in the MyFood app.

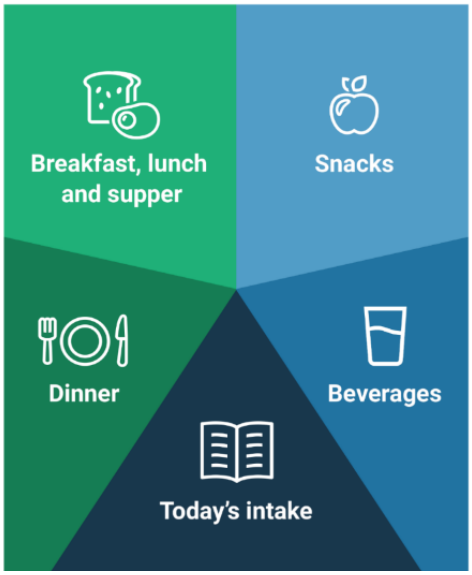
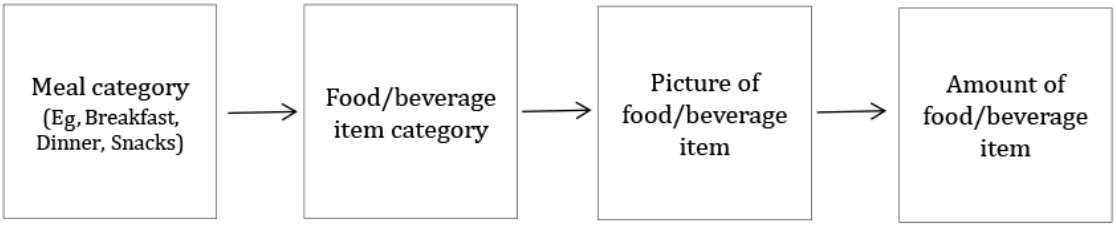


Figure 2. Flowchart on the dietary recording function in module 2.



Intake of energy, protein, and liquid was calculated based on information of the nutrient content in standard units (eg, 1 slice of bread, 1/2 glass of milk). The app included prompting questions (eg, regarding the use of spreads when recording intake of bread or regarding intake of beverages together with meals). Hot dishes were recorded by selecting an icon depicting the portion consumed (full, three quarters, half, one quarter; Figure 2). If only components of the meal were consumed (eg, 1 potato) this could be recorded by choosing the "ate only components" function shown in Figure 3. Portion sizes for beverages were recorded by selecting an icon depicting sections of a glass/cup (full, three quarters, half, one quarter) or by inputting the number of deciliters consumed.

The app included pictures of all food and beverages served at OUH, Rikshospitalet. It also included pictures of different groceries, food, and beverages that may be brought by relatives or friends from outside the hospital as well as advanced medical nutrition products. Nutritional information in the app was given for the intake of energy (kcal), protein (grams), and liquid (milliliters). Nutritional data were retrieved from an in-house data program (KBS version 7.0), based on the Norwegian food composition table [26], and from manufacturers.

Module 3: Evaluation of Recorded Intake Compared to Individual Needs

The third module automatically compared dietary intake with individual requirements for energy, protein, and liquids. This module was developed by including several algorithms in the app. The algorithms estimated the patients' daily requirements for energy, protein, and liquids and were based on recommendations from the Norwegian Directorate of Health [14,27].

Technological Features

The data flow in the app used a Web form and secure storage in "Services for sensitive data" (or TSD, Tjenester for Sensitive Data) [28] hosted by USIT (Figure 4). TSD meets the stringent requirements for the processing and storage of sensitive research data and is included in NorStore, the Norwegian national infrastructure for handling and storage of scientific data [29]. All recorded data were sent to TSD continuously during the data collection period. The recorded data were also stored locally on the iPad and visible in the app until 3 am the following day. This made it possible for the respondents to edit their recordings of dietary intake and the app was able to give the users feedback on their intake of energy, protein, and liquid during the current day. If the iPad was not able to send the data to TSD (eg, missing internet connection), the data were encrypted, temporarily queued, and resent as soon as the iPad was online again. All iPads were "clean" every morning and could possibly be given to a new patient. The data were later retrieved from TSD for data analysis in the evaluation study.

The Mobile Device Management System, AirWatch, was used to control the iPads during the data collection period. If tablets disappeared, we were able to clean the disappeared tablet remotely and make it impossible to use until reopened via AirWatch. It was possible to maintain total control of sensitive data stored on the tablets using this system.

Figure 3. Recording of hot dishes in MyFood.

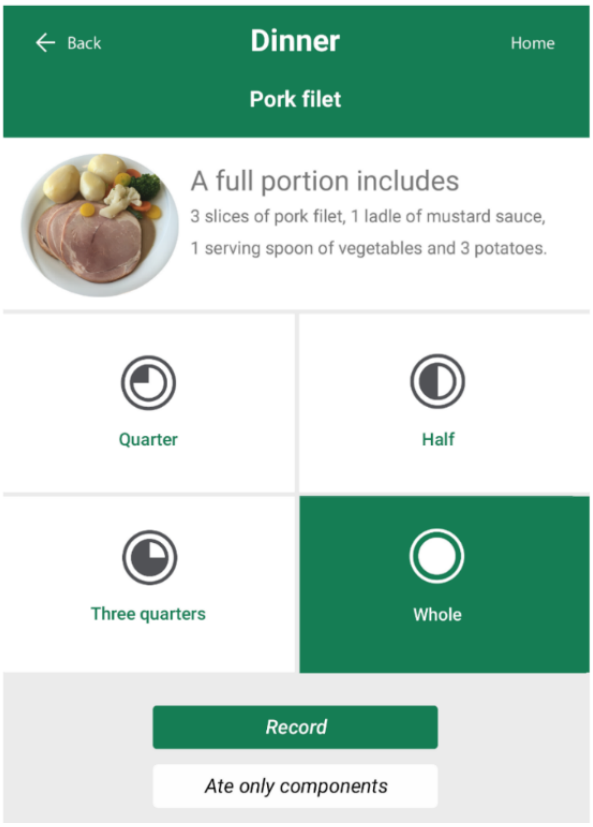
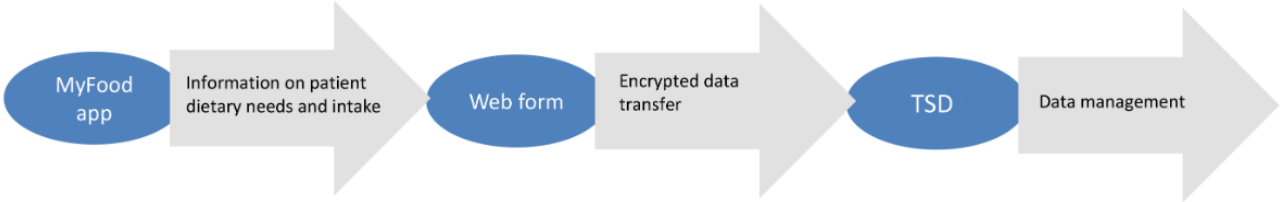


Figure 4. Data flow in the MyFood app. TSD: services for sensitive data.



Evaluation of the MyFood App

Participants

The evaluation study was performed at OUH, Rikshospitalet in the Departments of Gastrointestinal Surgery and Hematology.

Inclusion criteria were:

- ≥18 years of age
- ≥2 days of expected stay

Exclusion criteria were:

- Pregnancy
- Special infection precautions
- Psychiatric patients
- Critically ill patients
- Patients not able to read the Norwegian language

Even though MyFood is designed to be used by patients at nutritional risk, this was not an inclusion criterion as we wanted to include patients eating various amounts of food to evaluate the app. Based on a power of 0.8, a significance level of P=.05 and a calculated standardized difference of 1.0, 32 patients were included in the evaluation study.

Ethics

The study was performed in accordance with the Helsinki declaration and was acknowledged by the Norwegian Regional ethical committee (2016/1464), the Data protection officer at OUH, and the Chief Information Security Officer at the University of Oslo. Informed written consent was collected from all participating patients.

Performance

Information about the study, including the nurses' responsibilities in the data collection period, was sent to all nurses via e-mail. In the Department of Hematology, a ten-minute presentation by the project workers was held for the nurses during the morning meetings of the first two days of the data collection period. The responsible nurse in each department identified patients who met all of the inclusion criteria and none of the exclusion criteria.

The patients were registered in the app by a nurse or by one of the project workers before breakfast. Information including the

NPR number, height, weight, presence of fever, special diet, or special preferences was registered in the app before patient use.

Written instructions on how to record dietary intake in MyFood were given to the patients and the nurses. Once included in the study, patients answered a form with information about education, living conditions, and level of experience with apps and tablets or smartphones.

Included patients were given a tablet (iPad mini 32GB) and were asked to use MyFood for two days to record their intake of food and beverages for the breakfast, lunch, and dinner meals. If patients were not able to or did not want to record information themselves, a nurse performed the dietary recording for them. The patients were instructed to record dietary intake as soon as possible after the meals in order to get as precise recordings as possible. They were also informed both verbally and in writing to record the intake for the breakfast, lunch, and dinner meals only, not snacks or beverages consumed between the respective meals. If patients did not find exactly what they had consumed in the app, they were instructed to record something similar.

After two days of using MyFood, participants were asked to answer a form regarding assumptions about comprehension, content, and perceived value and usability of the app.

Reference Method

We used observations from digital photography of the meals combined with partial weighing of meal components as the reference method to evaluate the dietary assessment function in MyFood. The reference method is further described as the photograph method. A digital system camera (Sony A500/16-50mm PZ objective) was mounted to a removable trolley (85 cm * 50 cm) on an adjustable and pivotable tripod. The camera lens was approximately 0.6 m above the trolley. A researcher photographed the trays with the patients' meal before and after consumption. The trays were marked with the study participant number. The numbered trays were placed on a marked area on the trolley and a 30-cm ruler was placed on the tray as a reference size. The photographs were taken at an angle of 45° to the trays so that in-depth images could be taken for more convenient meal content estimation. In addition to the observations from photographs, partial weighing of meal components was performed by the researchers. Plates, glasses, cups, and food items in separate packaging were weighed on an electronic scale before and after the meal. In cases where determining the type of food or beverage from the photographs were challenging (eg, whole fat or skimmed milk, sugar-sweetened or light soft drinks, butter or margarine), the patients were asked about what specific type of foods or beverages they included in the meal.

Training of Project Workers

Two project workers underwent practice in photographing and estimating portion sizes before the data collection, to secure a standardized method and higher level of agreement. Thirteen meals (both bread-based meals and hot dishes) were prepared by a third person. The meals were prepared to illustrate the portion size before consumption and after consumption, by removing all or parts of the food. The meals were photographed before and after some or all the food were removed from the tray. Glasses, plates, cups, and food items in separate packaging were weighed. Both project workers observed the photographs and calculated the weights to estimate the consumption of food and beverages. The interobserver reliability (IOR) between the two project workers was calculated to be 0.92 for energy content. The project workers' estimations of energy content matched with the known energy content by 0.94. This was considered satisfactory, based on criteria in other studies [30].

Data Handling and Statistical Analyses

The food and beverage intake observed from the photographs and estimated from partly weighing in the evaluation study were compared with the intake recorded in MyFood. Observed and weighed intake was estimated separately by the two project workers, before recording the data in an in-house diet calculation system (KBS version 7.0). The project workers estimations were compared with the requirement of an IOR above 0.85 for energy, protein, and liquids in each meal. If IOR was <0.85 the calculations were repeated and recompared. In cases with obvious typing mistakes, this was corrected by the respective project worker. If the project workers had estimated different amounts, the pictures were re-evaluated, and the project workers agreed on where to adjust the estimated amounts (in grams). After corrections, the total IOR was 0.97 for energy, 0.98 for protein, and 0.98 for liquid. A final data file with estimated consumption based on the photograph method was created by averaging the estimations of the two project workers.

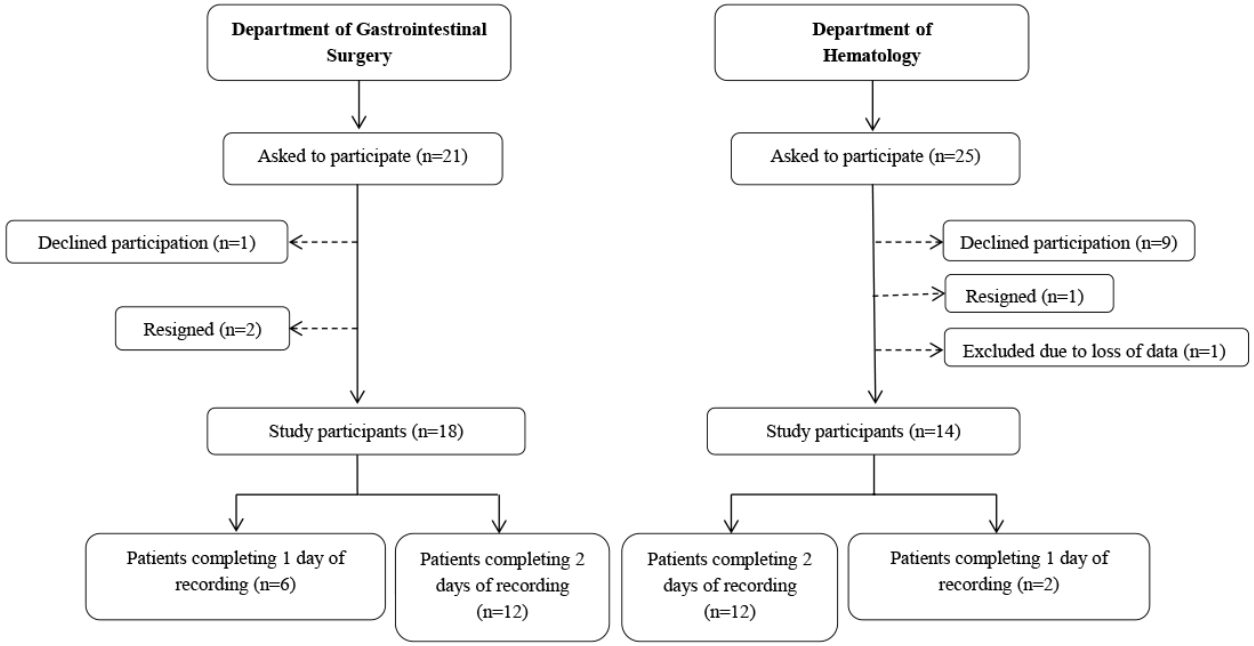
Statistical analyses were performed using the statistical software package IBM SPSS Statistics 24. All tests were two-sided with a 5% level of significance. The data were analyzed on both group and individual level. Differences in the intake estimations of energy, protein, liquid, and food groups, between MyFood and the photograph method, were analyzed with Wilcoxon Signed Ranks Test due to nonnormally distributed variables. Multiple scatter plots of consumption of energy, protein, liquid, and selected food groups were used to illustrate the difference between the estimated intake in MyFood and from the photograph method for each individual subject. The differences between the methods were assessed in total and divided into the breakfast, lunch, and dinner meals for recording days 1 and 2 separately. To calculate omitted food items, one omission was counted as an item observed from photographs in a meal but not recorded in MyFood.

Results

Participants

The study sample consisted of 32 patients at OUH, Rikshospitalet; 18 from the Department of Gastrointestinal Surgery and 14 from the Department of Hematology. The data collection period ran from March to May 2017, and the participants were recruited continuously during the period. A flowchart describing the recruitment process is illustrated in Figure 5.

Figure 5. Flowchart of the recruitment process of study participants.



Characteristics of the study participants are illustrated in Table 1. More than two-thirds were men and the age distribution was from 17 to 77 years. About 40 percent of the participants were characterized as normal weight, according to body mass index (BMI) and more than half as overweight or obese. The majority of the participants had some or a lot of experience with apps and smartphones or tablets.

Estimations of Energy, Protein, and Liquid Consumption in MyFood Compared to the Photograph Method on Group Level

Table 2 shows the intake of energy, protein, and liquids estimated in MyFood and the photograph method. The results are presented for the total of breakfast, lunch, and dinner, and separately for each meal.

The median intake of energy was not significantly different between the methods the first day, except for lunch where median recorded intake in MyFood was significantly higher compared to the photograph method. The second day a significantly lower median total energy intake was found in MyFood compared to the photograph method. The opposite was observed for the lunch and dinner meal.

The recorded median protein intake in MyFood was significantly lower for total intake, breakfast, and lunch, compared to the photograph method on day 1. The second day the median intake of protein for breakfast was significantly lower in MyFood, compared to the photograph method. No other statistically significant differences were found for median protein intake on day 2.

The median liquid intake showed relatively good agreement between the methods on the group level. Only for breakfast the first day the median recorded intake was significantly lower in MyFood compared to the photograph method.

Estimations of Energy, Protein, and Liquid Consumption in MyFood Compared to the Photograph Method on Individual Level

Table 3 shows the percentage of the patients who had 90 and 80 percent agreement between their recordings in MyFood compared to the photograph method, in total and separately for the breakfast, lunch, and dinner meal.

About half of the patients had ≥90% agreement in total for energy, protein, and liquid intake, somewhat lower for protein and higher for liquids both recording days. The breakfast meal had the highest proportion of participants with ≥80% agreement between the methods for all nutrient components both days, except for protein intake the first recording day. The agreement between the methods was lowest for the dinner meal.

Energy intake

Recorded individual energy intake in MyFood and intake estimated from the photograph method are illustrated in Figure 6, which shows individual drop-plots from the first and second recording days.

MyFood estimated the energy consumption relatively accurate for the majority of the patients. On average for the two days, approximately 70% of the participants had less than 20% disagreement between the two methods, and approximately 50% had less than 10% disagreement (Table 3). For some participants, the intake was overestimated in MyFood compared to photograph observations (Figure 6). This overestimation was more pronounced on day 1 than day 2. The largest discrepancies with regard to energy consumption at the individual level were found for the dinner meal the first day (Table 3).

Table 1. Characteristics of participants (n=32) in the evaluation study of MyFood.

Characteristic	n (%)
Hospital department	
Gastrointestinal surgery	18 (56)
Hematology	14 (44)
Gender	
Men	22 (69)
Women	10 (31)
Age (years)[a]	
<30	3 (10)
30-39	3 (10)
40-49	7 (23)
50-59	8 (26)
60-69	8 (26)
70-80	2 (7)
Body mass index (kg/m^2)	
<18.5	1 (3)
18.5-24.9	13 (41)
25-29.9	14 (44)
>30	4 (13)
Education	
Primary and secondary schools	4 (13)
Comprehensive school/high school	16 (50)
College/university ≤4 years	6 (19)
College/university >4 years	6 (19)
Earlier experiences with apps and smartphones/tablets	
None/little	3 (9)
Some (use sometimes)	9 (28)
A lot (use often/daily)	20 (63)

[a]Missing n=1.

Protein Intake

The individual protein consumption recorded in MyFood, compared to the photograph method showed relatively coinciding agreement. The agreement was most coinciding on day 2 (Multimedia Appendix 2). On average for the two days, about 70% of the participants had less than 20% disagreement between the two methods, and just below half of the participants had less than 10% disagreement (Table 3). The discrepancy between the methods was largest for the dinner meal (Table 3).

Liquid Intake

The agreement between the methods for low and medium liquid intake was good, with a tendency to increased deviations for higher intakes. This was seen on both recording days (Multimedia Appendix 3). On average for the two days, about 60%-70% of the participants had less than 20% disagreement in liquid intake between the two methods, and about 50% had less than 10% disagreement (Table 3).

Estimations of Food Intake in MyFood Compared to the Photograph Method on Group Level

The consumption (grams) within food groups are shown in Table 4. No statistically significant differences were seen between the methods, except for bread and cereals, and fruits. The median recorded intake of bread and cereals was significantly higher in MyFood compared to the photograph method, both recording days. Median fruit intake was significantly lower in MyFood the first recording day.

Table 2. Energy, protein, and liquid consumption recorded in MyFood compared to the photograph method. The data are presented as a total of the breakfast, lunch, and dinner meals, and separately for each meal.

Energy, protein and liquid	Mean	Median (25-75 percentile)	*P* value
Energy			
Day 1			
Breakfast (n=28)			.73[a]
MyFood	471	398 (244-616)	
Photograph method	458	373 (222-373)	
Lunch (n=27)			.04[a]
MyFood	408	389 (262-494)	
Photograph method	382	308 (201-308)	
Dinner (n=27)			.58[a]
MyFood	476	468 (210-711)	
Photograph method	461	477 (226-575)	
Total (n=32)			.11[a]
MyFood	1157	1039 (556-1541)	
Photograph method	1102	951 (446-1495)	
Day 2			
Breakfast (n=29)			.74[a]
MyFood	400	374 (223-527)	
Photograph method	407	367 (175-630)	
Lunch (n=20)			.02[a]
MyFood	454	501 (258-608)	
Photograph method	394	418 (245-514)	
Dinner (n=20)			.01[a]
MyFood	489	413 (134-820)	
Photograph method	425	368 (105-696)	
Total (n=29)			.009[a]
MyFood	1050	928 (380-1876)	
Photograph method	972	957 (308-1720)	
Protein (g)			
Day 1			
Breakfast (n=28)			.02
MyFood	16.2	13.5 (6.4-23.5)	
Photograph method	18.2	14.3 (6.9-27.7)	
Lunch (n=27)			.001
MyFood	13.0	10.0 (8.0-18.0)	
Photograph method	14.6	13.1 (7.8-20.2)	
Dinner (n=27)			.22
MyFood	15.2	14.5 (3.0-20.5)	
Photograph method	16.8	14.3 (6.1-22.8)	
Total (n=32)			.046
MyFood	38.0	35.0 (17.4-45.6)	

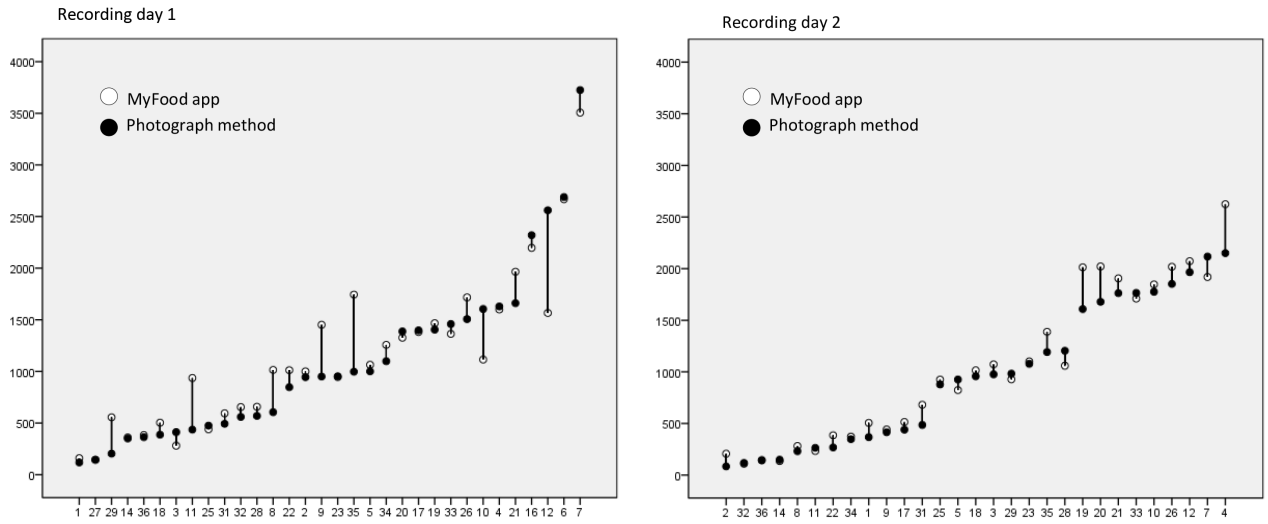
Energy, protein and liquid	Mean	Median (25-75 percentile)	*P* value
Photograph method	42.2	38.4 (14.0-62.1)	
Day 2			
Breakfast (n=29)			<.001
MyFood	14.1	11.0 (5.0-19.8)	
Photograph method	16.1	15.8 (5.7-22.0)	
Lunch (n=20)			.31
MyFood	14.2	13.3 (5.9-20.3)	
Photograph method	12.9	11.5 (7.8-16.7)	
Dinner (n=20)			.97
MyFood	17.6	15.3 (5.3-30.4)	
Photograph method	17.6	16.5 (3.9-28.7)	
Total (n=29)			.15
MyFood	36.1	28.0 (8.5-61.5)	
Photograph method	37.1	34.6 (9.1-61.4)	
Liquid (ml)			
Day 1			
Breakfast (n=28)			.02
MyFood	292	272 (158-412)	
Photograph method	339	320 (202-466)	
Lunch (n=27)			.72
MyFood	287	256 (194-409)	
Photograph method	285	257 (159-374)	
Dinner (n=27)			.33
MyFood	336	304 (189-304)	
Photograph method	332	326 (222-445)	
Total (n=32)			.71
MyFood	781	696 (479-1047)	
Photograph method	808	643 (461-1227)	
Day 2			
Breakfast (n=29)			.97
MyFood	301	287 (169-429)	
Photograph method	303	312 (162-435)	
Lunch (n=20)			.87
MyFood	275	251 (142-405)	
Photograph method	260	256 (154-345)	
Dinner (n=20)			.06
MyFood	311	269 (84-540)	
Photograph method	273	245 (58-487)	
Total (n=29)			.11
MyFood	706	587 (313-1077)	
Photograph method	670	559 (318-1029)	

[a]Differences between MyFood and the photograph method for the breakfast, lunch and dinner meal. The totals of these meals are tested with Wilcoxon Signed Rank test separately for each recording day.

Table 3. Proportions with 90% and 80% agreement between MyFood and the photograph method in estimated intake of energy, protein, and liquids.

Energy, protein and liquid	Percent agreement	
	90%	80%
Day 1		
Energy		
Total (n=32)	47	69
Breakfast (n=28)	46	79
Lunch (n=27)	48	59
Dinner (n=27)	30	52
Protein		
Total (n=32)	44	66
Breakfast (n=28)	32	64
Lunch (n=27)	44	74
Dinner (n=27)	19	41
Liquids		
Total (n=32)	53	63
Breakfast (n=28)	39	68
Lunch (n=27)	48	63
Dinner (n=27)	26	44
Day 2		
Energy		
Total (n=29)	55	76
Breakfast (n=29)	45	66
Lunch (n=20)	40	55
Dinner (n=20)	25	50
Protein		
Total (n=29)	48	83
Breakfast (n=29)	52	69
Lunch (n=20)	35	55
Dinner (n=20)	40	50
Liquids		
Total (n=29)	59	72
Breakfast (n=29)	52	69
Lunch (n=20)	45	65
Dinner (n=20)	45	55

Figure 6. Drop plots illustrating individual intake of energy recording day 1 (n=32) and recording day 2 (n=29). Y-axis represents energy intake (kcal). X-axis represents participant number ranged with increasing energy intake according to the photograph method. Equal energy intake from app and photograph observations is presented with only black dots.



Estimations of Food Intake in MyFood Compared to the Photograph Method on Individual Level

Table 5 shows the percentage of the participants who had 90 and 80 percent agreement between their recordings in MyFood compared to the photograph method within food groups. Egg was the food group with the best agreement between MyFood and the photograph method with the majority of the estimations ≥90% agreement. The food groups with the lowest agreement were fruit and vegetables. The agreement between the methods increased from day 1 to day 2 for bread and cereals, spreads, egg, yogurt, soup, hot dishes, and desserts.

Estimated bread and cereal consumption was, in most cases, higher in MyFood compared to estimations from the photograph method (Multimedia Appendix 4). On average for the two days, about 60% of the participants had less than 20% disagreement in estimated bread and cereal intake between the two methods, and about 25% had less than 10% disagreement (Table 5).

Recordings of spreads tended to be lower in MyFood compared to the photograph method when the intake increased (Multimedia Appendix 4). About 70% of the participants had less than 20% disagreement between MyFood and the photograph method in estimated intake of spreads on day 2, compared to 50% on day 1 (Table 5).

The food group with the largest deviations between the methods was hot dishes. The discrepancies were highest the first day (Multimedia Appendix 5). About 30%-40% of the participants had ≥80% agreement between the methods (Table 5).

On average for the two days, about 70% of the participants had less than 20% disagreement in estimated intake of cold beverages between the two methods, and about 50% had less than 10% disagreement (Table 5). No particular pattern in discrepancies of cold beverages between the methods was seen (Multimedia Appendix 5).

Omitted Food Items in MyFood Recordings Compared to the Photograph Method

The number of food and beverage items recorded in MyFood and observed from photographs was calculated (Multimedia Appendices 6 and 7). The first day the number of medical nutrition drinks, cheese, fish-based spreads, and meat-based spreads recorded had 100% matches between the methods (Multimedia Appendix 6). The second day 100% matches were found for the recordings of hot dishes, medical nutrition drinks, vegetables, and meat-based spreads (Multimedia Appendix 7). Butter, margarine, and mayonnaise (27% omissions both days), fruit (27% omissions on day 1), vegetables (28% omissions on day 1), yogurt (27% omissions on day 2) and meal condiments (29% omissions on day 1, 33% omissions on day 2) were the food groups most often omitted among participants (Multimedia Appendices 6 and 7). Five participants had duplicate recordings of some meal components the first day and one participant the second day.

Patients' Experiences Using the MyFood App

Ninety percent of the participants reported that MyFood was easy to use. All but one (97%) of the participating patients found the app easy to navigate in. Most of the patients (87%) experienced to record correct amount of foods and beverages. Thirteen percent had to acquire new knowledge to use the app. Seventy-one percent reported to be become more aware of the amount of foods and beverages needed, after using MyFood.

Table 4. Food and beverage intake (grams) recorded in MyFood and estimated in the photograph method. Significant *P* values (<.05) are in italics.

Food and beverages	MyFood (grams)		Photograph methods (grams)		*P* value[a]
	Median (25-75 percentile)	Mean (SD)	Median (25-75 percentile)	Mean (SD)	
Day 1					
Bread and cereals (n=23)	110 (50-180)	134 (106)	93 (44-150)	110 (80)	*<.001*
Spreads[b] (n=22)	72 (22-120)	77 (59)	89 (20-135)	90 (74)	.17
Egg (n=11)	56 (56-72)	61 (34)	56 (56-96)	68 (29)	.18
Yogurt (n=9)	150 (100-223)	154 (81)	190 (101-292)	186 (94)	.09
Cold beverages (n=30)	350 (200-463)	360 (216)	400 (203-534)	383 (185)	.35
Hot beverages (n=9)	200 (149-350)	244 (147)	200 (145-394)	261 (175)	.24
Oral nutritional supplements (n=3)	200 (0-400)	200 (200)	200 (150-400)	250 (132)	.32
Soup (n=9)	225 (225-270)	235 (114)	202 (147-307)	219 (119)	.37
Hot dishes (n=20)	217 (0-419)	222 (212)	238 (80-344)	245 (165)	.58
Desserts (n=12)	58 (10-80)	49 (32)	57 (27-78)	63 (50)	.86
Fruit (n=10)	30 (0-94)	44 (49)	91 (57-148)	101 (47)	*.04*
Vegetables[c] (n=10)	35 (5-60)	51 (64)	60 (38-103)	76 (50)	.18
Day 2					
Bread and cereals (n=24)	83 (40-159)	100 (67)	72 (30-137)	86 (62)	*<.001*
Spreads[b] (n=17)	44 (23-60)	56 (47)	54 (24-81)	64 (50)	.13
Egg (n=8)	56 (50-56)	67 (60)	56 (50-56)	68 (59)	.32
Yogurt (n=8)	150 (75-150)	118 (63)	145 (90-170)	131 (50)	.78
Cold beverages (n=25)	300 (200-600)	362 (247)	335 (183-559)	358 (239)	.88
Hot beverages (n=9)	200 (170-300)	221 (133)	182 (112-282)	192 (124)	.16
Oral nutritional supplements (n=3)	125 (50-225)	135 (96)	200 (220-225)	210 (22)	.11
Soup (n=5)	135 (0-225)	117 (113)	111 (79-226)	144 (80)	.35
Hot dishes (n=14)	390 (74-425)	296 (221)	329 (150-384)	296 (142)	.62
Desserts (n=11)	80 (55-80)	75 (42)	55 (45-126)	80 (54)	.77
Fruit (n=6)	13 (0-98)	43 (61)	63 (55-104)	80 (46)	.09
Vegetables[c] (n=10)	13 (0-71)	36 (47)	13 (0-71)	36 (18)	.88

[a]Differences between the methods are tested with Wilcoxon Signed Rank test.

[b]Includes butter/margarine/mayonnaise, sugary-based spreads, meat-based spreads, mayonnaise-based spreads, fish-based spreads, and cheese.

[c]Does not include vegetables as part of hot dishes.

Table 5. Proportions with 90% and 80% agreement between MyFood and the photograph method in estimated intake within food groups.

Food and beverage items	Percent agreement	
	90%	80%
Day 1		
Bread and cereals (n=23)	22	57
Spreads[a] (n=22)	23	50
Egg (n=11)	82	82
Yogurt (n=9)	33	67
Cold beverages (n=30)	57	77
Hot beverages (n=9)	44	67
Oral nutritional supplements (n=3)	67	67
Soup (n=9)	22	56
Hot dishes (n=20)	15	30
Desserts (n=12)	8	42
Fruit (n=10)	10	20
Vegetables[b] (n=10)	20	40
Day 2		
Bread and cereals (n=24)	29	63
Spreads[a] (n=17)	53	71
Egg (n=8)	88	88
Yogurt (n=8)	50	75
Cold beverages (n=25)	44	68
Hot beverages (n=9)	44	67
Oral nutritional supplements (n=3)	67	67
Soup (n=5)	40	60
Hot dishes (n=14)	14	36
Desserts (n=11)	27	55
Fruit (n=6)	0	17
Vegetables[b] (n=10)	10	20

[a]Includes butter/margarine/mayonnaise, sugary-based spreads, meat-based spreads, mayonnaise-based spreads, fish-based spreads, and cheese

[b]Does not include vegetables as part of hot dishes.

Discussion

Principal Findings

The MyFood app is developed for use among hospitalized patients at nutritional risk. According to the Norwegian Safety Program: "In Safe Hands" [31] all patients at nutritional risk should have a nutritional assessment, including dietary recording to compare intake against individual needs of energy, protein, and liquids. Further, nutrition-related measures should be performed and an individual nutrition plan created, before performing a reassessment after 3 days [31]. We found that 60%-80% of the participants had less than 20% disagreement between estimated intake of energy, protein, and liquids in MyFood and the photograph method. The agreement between the methods was higher the second recording day, compared to the first, and for the breakfast and lunch meal compared to the dinner. Recorded consumption of bread and cereals was higher in MyFood compared to the photograph method both days. Spreads; particularly butter, margarine, and mayonnaise, fruit, vegetables, and meal condiments were the food groups most often omitted by the patients. The majority of the patients rated MyFood as easy to use.

To our knowledge, no similar study has been conducted in patients to allow direct comparison of results.

The Accuracy of MyFood's Estimations of Energy, Protein, and Liquid on Group Level

Even though the main objective of the study was to evaluate recorded intake in MyFood on the individual level, estimated intake on the group level was analyzed to investigate if overall disagreement was present. The median total energy intake was

not different between the methods the first recording day, however, a lower median intake in MyFood compared to the photograph method was found the second day. Underestimation of energy intake is often seen in validation studies of self-reporting dietary assessment tools among healthy adults [32] and among hospitalized patients [33]. This is also found for technology-based records [34,35]. MyFood's target population is patients at nutritional risk who often have reduced food intake compared to needs [16]. Intentional underreporting of food intake may not be as relevant for our target population compared to healthy populations.

Median recorded protein intake was lower in MyFood compared to estimations from the photograph method in total, for breakfast, and for lunch recording day 1, and for breakfast recording day 2. This deviates from results in other validation studies on electronic dietary assessment tools. Raatz and coworkers [36] and Fukuo and colleagues [37] did not find a different recording of protein intake among healthy subjects in a personal digital assistant and a Web app, compared to 24-hour dietary recall and paper-based food records, respectively. Sliced bread with different types of spreads typically constitutes Norwegian breakfast and lunch meals, also in hospitals. Half of the participants had up to 20% disagreement in consumption of spreads between MyFood and the photograph method, and the individual drop-plots demonstrated that MyFood estimated lower intake of spreads compared to the photograph method for several participants on day 1. Spreads are often a protein source in bread-based meals. The agreement between the methods in intake of spreads was better the second day.

Recorded liquid intake in MyFood showed generally good agreement with the photograph method. However, recorded intake to the breakfast meal the first day was significantly lower compared to the photograph method. Several of the participants consumed both cold and hot beverages for the breakfast meal. This may have increased the chance of not remember to record all types of beverages consumed.

The Accuracy of MyFood's Estimations of Energy, Protein, and Liquid on Individual Level

The main aim of the present study was to evaluate MyFood's ability to estimate the patients' dietary intake on an individual level. This contrasts most other validation studies which focus on mean intake on group level and cross-classification, but not on absolute intakes. We evaluated MyFood compared to the photograph method for two separate days. A comparison of one-day and three-day calorie counts to estimate dietary intake by Breslow and Sorkin [38] suggested that 1-day calorie counts may be a valid alternative to the more labor-intensive 3-day count commonly performed in hospitalized patients. Førli and coworkers argue, however, that one day may be too short to estimate dietary intake among hospitalized patients [33]. The MyFood app is intended for use over several days, to follow-up dietary intake. This is in line with the common recommendation at Oslo University Hospital of using paper-based dietary assessment forms on a daily basis for patients at nutritional risk. The dietary recording in MyFood is more detailed than the paper-based forms used today by including a higher differentiation between type of meals and meal items. MyFood also includes more alternatives for portion sizes and provides the possibility to only record components of composite dishes. By these means there are reasons to assume that MyFood will provide a higher accuracy of the patient's diet than the paper-based forms, if used correctly.

The individual drop-plots presented in the present study showed an overestimation of energy intake in MyFood, compared to the photograph method for some participants and underestimation for others. An explanation may be that both duplicate recordings and omissions of food items were observed. Five participants had duplicate recordings of some meal components the first day and one participant the second day. The largest discrepancies in energy intake between the methods on the individual level were found for the dinner meal the first day. This may be explained by inaccurate estimation of portion sizes for hot dishes. We found that several participants selected a full portion, even though not consuming a whole plate. The discrepancies in individual energy intake between the methods were wider the first recording day, and more coinciding the second day. We also observed fewer duplicate recordings in the app the second, compared to the first recording day. This may be due to a learning effect where the patients became more familiar with the app after one day of recording. A tendency to such a learning effect was observed in general in the evaluation study. A potential learning effect with the repeated use of a computerized dietary assessment tool was also found in a validation study among 41 adults with type 2 diabetes mellitus. The authors argued that the patients became more familiar with the website with repeated use [39].

We found a tendency to lower recorded protein intake in MyFood compared to the photograph method for participants with higher protein intake. This may be explained by the omission of typical protein-rich spreads with higher intakes due to recall bias. Cheese was omitted by three participants, ham by two participants, and egg by one participant. The second recording day the recorded protein intake in MyFood was more coinciding with the estimated protein intake in the photograph method.

Liquid consumption on individual level showed a tendency to increased deviations between MyFood and the photograph method among participants with a higher liquid intake. The first day the patient with the largest deviation had omitted both coffee and milk from the app recordings. The second day the intake recorded in MyFood was higher compared to the photograph method for some of the participants due to the recording of a full glass in the app, even though only consuming half or three-quarters of a full glass size.

The proportion of participants having less than 20% disagreement in MyFood and the photograph method was 69% for energy intake, 66% for protein intake, and 63% for liquid intake the first day, whereas the corresponding proportions were 76% for energy, 83% for protein and 72% for liquid intake the second day. This may be due to a learning effect as discussed above.

The Accuracy of MyFood's Estimation Within Food Groups

The majority of the food groups showed good agreement between MyFood and the photograph method on the group level. Good agreement in recording of food groups is consistent with findings from a validation study on a dietary assessment app for smartphones compared to repeated 24 h recall interviews [40] and Foodbook24; a Web-based dietary assessment tool [41]. The median intake of bread and cereals was higher in MyFood compared to the photograph method both recording days. Based on photograph observations and partial weighing this was found to be due to too large portion sizes of sliced bread and bread rolls in the app compared to the actual sizes served at the hospital. Recorded fruit intake was significantly lower in MyFood than consumption observed from photographs the first recording day. A possible explanation is that fruit intake was omitted by 27% of the participants on day 1. Medin and coworkers also found a high omission rate of fruits in a validation study among school children [30]. About 80% of the participants had more than 20% disagreement between estimated fruit consumption in MyFood and the photograph method both recording days. The majority of the participants' fruit consumption was preprepared fruit boxes with sliced fruits. Based on the photograph method including observations and partial weighing we found the fruit boxes in the app to be disproportionately lower than the size most often observed and weighed. Revision of portion sizes for bread and cereals, and fruit cups, will probably lead to more accurate recordings of these food groups in the MyFood app.

In the present study, some of the standard portion sizes of hot dishes included in MyFood seemed to be too large compared to actual size served to the patients. A full portion size in MyFood was based on information from the hospital kitchen at OUH, Rikshospitalet on how standard portion sizes should be constituted when served. The visibility and description of what constitutes a full portion size in MyFood (Figure 2) may not have been clear enough for the patient. Several patients may have assumed eating a whole standard portion if the plates seemed full. In addition, studies have shown that small portion sizes tend to be overestimated and large portion sizes to be underestimated [42,43], and the former may have occurred in our study.

Twelve percent of foods and beverages were omitted in MyFood the first day, and 11% the second day. The food group most often omitted both recording days was butter, margarine, and mayonnaise. When recording several types of spreads, butter and margarine are typically easy to forget. Spreads were found to be among the food items most often omitted in a validation study of a Web-based dietary assessment tool among 117 school children [30]. The omission of food items in meals consisting of several secondary ingredients, like sandwiches, has been argued to be more common than in less composite meals [44]. Frequent omission of margarine was also found by Førli and colleagues [33] in a validation study of a self-administered dietary assessment form among 45 patients at OUH, Rikshospitalet. Prompting of questions related to the use of butter/margarine will be included in the further development of MyFood.

Acceptance of Use by the Patients

The majority of the patients found MyFood easy to use and more than 70% became more aware of own nutritional needs. Electronic dietary assessment tools are generally well accepted and preferred over conventional methods among healthy subjects [46]. Our study population included patients ranging from 17 to 77 years with a mean age of 51 years. It is possible that use of an app for tablets is a larger barrier among older patients. However, qualitative studies among older persons have demonstrated that elderly persons often are positive to using tablets and eager to learn, even though cognitive deficits increase by age and low self-efficacy may limit the potential for use [47,48].

Strengths and Limitations

The development process of the MyFood app involved nurses and patients, which is considered an important strength. The evaluation of recorded dietary intake in MyFood was compared to observations from meal photographs. The photograph method is a validated tool for assessment of dietary intake, compared to weighed records [45,49]. In addition, we combined the photograph observations with partial weighing of meal components which probably strengthened the method. Our photograph method is associated with different measurement errors than the dietary assessment functionality in MyFood, which is also considered an important strength. In addition to validate the MyFood app with regard to the accuracy of dietary recording we also investigated the users' experiences with the tool. A recent scoping review on the use of technology in identifying hospital malnutrition highlighted the importance of establishing usability rating to determine the app's actual usefulness in practical settings [50].

A limitation of the study is that only dietary intake to breakfast, lunch, and dinner was evaluated. Energy consumption for the breakfast, lunch, and dinner meals together have been reported to account for 85% of patients' total intake [33]. By not including the total intake we do not know how accurate the dietary assessment function in MyFood estimate intake to snack meals, evening meals, and beverage intake in-between meals. Another potential limitation is that the patients knew that the researchers were taking photographs of their meals before and after consumption. This may have influenced their recording in the app by acting as a reminder. The evaluation study was performed among patients at a hematology and a gastrointestinal surgery department. We do not know whether our findings are representative for other groups of patients. The included patients were all sick, some quite severe. The presence of disease and fatigue may have influenced the precision of the recordings. MyFood is intended for use among patients at nutritional risk. Nutritional risk was, however, not an inclusion criterion in the present study, as we wanted to evaluate the app for patients with both small and larger food intake. Only patients with a certain food intake orally were included and we, therefore, do not know how the dietary assessment function in MyFood measures the intake for patients with tube feeding or parenteral nutrition.

MyFood's Potential for Use as a Dietary Assessment Tool Among Hospitalized Patients at Nutritional Risk

Based on the results in the present evaluation study, we consider MyFood as having good potential for use as a dietary assessment tool among hospitalized patients at nutritional risk. MyFood may provide support to health care workers in their tasks related to the nutritional treatment of patients at nutritional risk. This support may contribute to prevent development of disease-related malnutrition among at-risk patients. Corrections of some of the portion sizes in the app and prompting related to use of butter/margarine and portion size of dinner may increase the accuracy of the app further. An evaluation study among other patient groups may be valuable to amplify the potential for use of MyFood in the hospital setting.

Conclusion

We have developed an app for tablets for use among hospitalized patients at nutritional risk. The app includes dietary assessment functionality for evaluation of patients' dietary intake compared to individual needs of energy, protein, and liquids. The recorded intake of energy, protein, and liquids in MyFood showed good agreement with the photograph method for the majority of the participants. The app's ability to estimate intake within food groups was good, except for bread and cereals which were overestimated, and fruit which was underestimated. MyFood was well accepted among the study participant and has the potential to be a dietary assessment tool for use among patients in clinical practice.

Acknowledgments

The development of the MyFood app was performed on the TSD facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Selected screenshots from the MyFood app.

[MP4 File (MP4 Video), 4MB - mhealth_v6i9e175_app1.mp4]

Multimedia Appendix 2

Drop plots illustrating individual intake of protein recording day 1 (n=32) and recording day 2 (n=29). Y-axis represents protein intake (grams). X-axis represents participant number ranged with increasing protein intake according to the photograph method. Equal protein intake from app and photograph observations is presented with only black dots.

[PNG File, 177KB - mhealth_v6i9e175_app2.png]

Multimedia Appendix 3

Drop plots illustrating individual intake of liquid recording day 1 (n=32) and recording day 2 (n=29). Y-axis represents liquid intake (grams). X-axis represents participant number ranged with increasing liquid intake according to the photograph method. Equal liquid intake from app and photograph observations is presented with only black dots.

[PNG File, 190KB - mhealth_v6i9e175_app3.png]

Multimedia Appendix 4

Drop plots illustrating individual intake of bread and cereals, and spreads recording day 1 and recording day 2. The Y-axis represents grams of food item. X-axis represents participant number ranged with increasing intake according to the photograph method. Equal food intake from app and photograph observations is presented with only white dots.

[PNG File, 316KB - mhealth_v6i9e175_app4.png]

Multimedia Appendix 5

Drop plots illustrating individual intake of hot dishes and cold beverages recording day 1 and recording day 2. The Y-axis represents grams of food or beverage item. X-axis represents participant number ranged with increasing intake according to photograph observations. Equal food intake from app and photograph observations is presented with only white dots.

[PNG File, 308KB - mhealth_v6i9e175_app5.png]

Multimedia Appendix 6

Number of food and beverage items observed from photographs and recorded in MyFood the first recording day. One item means one type of food or beverage in each meal.

[PNG File, 63KB - mhealth_v6i9e175_app6.png]

Multimedia Appendix 7

Number of food and beverage items observed from photographs and recorded in MyFood the second recording day. One item means one type of food or beverage in each meal.

[PNG File, 69KB - mhealth_v6i9e175_app7.png]

References

1. Norman K, Pichard C, Lochs H, Pirlich M. Prognostic impact of disease-related malnutrition. Clin Nutr 2008 Feb;27(1):5-15. [doi: 10.1016/j.clnu.2007.10.007] [Medline: 18061312]
2. Agarwal E, Ferguson M, Banks M, Batterham M, Bauer J, Capra S, et al. Malnutrition and poor food intake are associated with prolonged hospital stay, frequent readmissions, and greater in-hospital mortality: results from the Nutrition Care Day Survey 2010. Clin Nutr 2013 Oct;32(5):737-745. [doi: 10.1016/j.clnu.2012.11.021] [Medline: 23260602]
3. Bavelaar JW, Otter CD, van BAA, Thijs A, van BVDSMAE. Diagnosis and treatment of (disease-related) in-hospital malnutrition: the performance of medical and nursing staff. Clin Nutr 2008 Jun;27(3):431-438. [doi: 10.1016/j.clnu.2008.01.016] [Medline: 18387718]
4. Imoberdorf R, Meier R, Krebs P, Hangartner PJ, Hess B, Stäubli M, et al. Prevalence of undernutrition on admission to Swiss hospitals. Clin Nutr 2010 Feb;29(1):38-41. [doi: 10.1016/j.clnu.2009.06.005] [Medline: 19573958]
5. Sorensen J, Kondrup J, Prokopowicz J, Schiesser M, Krähenbühl L, Meier R, EuroOOPS study group. EuroOOPS: an international, multicentre study to implement nutritional risk screening and evaluate clinical outcome. Clin Nutr 2008 Jun;27(3):340-349. [doi: 10.1016/j.clnu.2008.03.012] [Medline: 18504063]
6. Tangvik RJ, Tell GS, Guttormsen AB, Eisman JA, Henriksen A, Nilsen RM, et al. Nutritional risk profile in a university hospital population. Clin Nutr 2015 Aug;34(4):705-711 [FREE Full text] [doi: 10.1016/j.clnu.2014.08.001] [Medline: 25159298]
7. Henriksen C, Gjelstad IM, Nilssen H, Blomhoff R. A low proportion of malnourished patients receive nutrition treatment - results from nutritionDay. Food Nutr Res 2017;61(1):1391667 [FREE Full text] [doi: 10.1080/16546628.2017.1391667] [Medline: 29151831]
8. Correia MITD, Waitzberg DL. The impact of malnutrition on morbidity, mortality, length of hospital stay and costs evaluated through a multivariate model analysis. Clin Nutr 2003 Jun;22(3):235-239. [Medline: 12765661]
9. Kyle UG, Schneider SM, Pirlich M, Lochs H, Hebuterne X, Pichard C. Does nutritional risk, as assessed by Nutritional Risk Index, increase during hospital stay? A multinational population-based study. Clin Nutr 2005 Aug;24(4):516-524. [doi: 10.1016/j.clnu.2005.04.003] [Medline: 15916837]
10. Tangvik RJ, Tell GS, Eisman JA, Guttormsen AB, Henriksen A, Nilsen RM, et al. The nutritional strategy: four questions predict morbidity, mortality and health care costs. Clin Nutr 2014 Aug;33(4):634-641 [FREE Full text] [doi: 10.1016/j.clnu.2013.09.008] [Medline: 24094814]
11. Kondrup J, Johansen N, Plum LM, Bak L, Larsen IH, Martinsen A, et al. Incidence of nutritional risk and causes of inadequate nutritional care in hospitals. Clin Nutr 2002 Dec;21(6):461-468. [Medline: 12468365]
12. Kyle UG, Genton L, Pichard C. Hospital length of stay and nutritional status. Curr Opin Clin Nutr Metab Care 2005 Jul;8(4):397-402. [Medline: 15930964]
13. Freijer K, Tan SS, Koopmanschap MA, Meijers JMM, Halfens RJG, Nuijten MJC. The economic costs of disease related malnutrition. Clin Nutr 2013 Feb;32(1):136-141. [doi: 10.1016/j.clnu.2012.06.009] [Medline: 22789931]
14. The NDOH. National guidelines for prevention and treatment of malnutrition (Nasjonale faglige retningslinjer for forebygging og behandling av underernæring). Oslo: The Norwegian Directorate of Health; 2009.
15. Kondrup J, Allison SP, Elia M, Vellas B, Plauth M. ESPEN guidelines for nutrition screening 2002. Clin Nutr 2003 Aug;22(4):415-421. [Medline: 12880610]
16. Schindler K, Pernicka E, Laviano A, Howard P, Schütz T, Bauer P, et al. How nutritional risk is assessed and managed in European hospitals: a survey of 21,007 patients findings from the 2007-2008 cross-sectional nutritionDay survey. Clin Nutr 2010 Oct;29(5):552-559. [doi: 10.1016/j.clnu.2010.04.001] [Medline: 20434820]
17. Tangvik RJ, Guttormsen AB, Tell GS, Ranhoff AH. Implementation of nutritional guidelines in a university hospital monitored by repeated point prevalence surveys. Eur J Clin Nutr 2012 Mar;66(3):388-393 [FREE Full text] [doi: 10.1038/ejcn.2011.149] [Medline: 21863042]
18. Eide HD, Halvorsen K, Almendingen K. Barriers to nutritional care for the undernourished hospitalised elderly: perspectives of nurses. J Clin Nurs 2015 Mar;24(5-6):696-706 [FREE Full text] [doi: 10.1111/jocn.12562] [Medline: 24646060]

19. Khalaf A, Berggren V, Westergren A. Caring for undernourished patients in an orthopaedic setting. Nurs Ethics 2009 Jan;16(1):5-18. [doi: 10.1177/0969733008097986] [Medline: 19103687]
20. Ross L, Mudge A, Young A, Banks M. Everyone's problem but nobody's job: staff perceptions and explanations for poor nutritional intake in older medical patients. Nutr Diet 2011;68(1):41-46. [doi: 10.1111/j.1747-0080.2010.01495.x]
21. The NG. HealthCare21 (HelseOmsorg21). In: HealthCare21 (HelseOmsorg21). Bergen: Fagbokforlaget; Jun 2014.
22. Biltoft-Jensen A, Bysted A, Trolle E, Christensen T, Knuthsen P, Damsgaard CT, et al. Evaluation of Web-based Dietary Assessment Software for Children: comparing reported fruit, juice and vegetable intakes with plasma carotenoid concentration and school lunch observations. Br J Nutr 2013 Jul 14;110(1):186-195. [doi: 10.1017/S0007114512004746] [Medline: 23181984]
23. Freedman LS, Commins JM, Moler JE, Arab L, Baer DJ, Kipnis V, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. Am J Epidemiol 2014 Jul 15;180(2):172-188. [doi: 10.1093/aje/kwu116] [Medline: 24918187]
24. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. Am J Epidemiol 2003 Jul 01;158(1):1-13. [Medline: 12835280]
25. Trabulsi J, Schoeller DA. Evaluation of dietary assessment instruments against doubly labeled water, a biomarker of habitual energy intake. Am J Physiol Endocrinol Metab 2001 Nov;281(5):E891-E899. [doi: 10.1152/ajpendo.2001.281.5.E891] [Medline: 11595643]
26. Norwegian Food Safety Authority, The Norwegian Directorate of Health, University of Oslo. The Norwegian Food Composition Table 2016 URL: http://www.matvaretabellen.no/?language=en [accessed 2018-01-26] [WebCite Cache ID 6wkkYJtvO]
27. The NDOH. The dietary planner: supervisor in nutrition in health and care services (Kosthåndboken: veileder i ernæringsarbeid i helse- og omsorgstjenesten). Oslo: The Norwegian Directorate of Health; 2012.
28. University of Oslo. Services for sensitive data (TSD). 2015. Services for sensitive data (TSD) URL: https://www.uio.no/english/services/it/research/sensitive-data/index.html [WebCite Cache ID 6wkl0EEvK]
29. University of Oslo. System description. Oslo: Universty of Oslo; 2017. System description URL: http://www.uio.no/english/services/it/research/sensitive-data/about/description-of-the-system.html [WebCite Cache ID 6wknk9x7E]
30. Medin AC, Astrup H, Kåsin BM, Andersen LF. Evaluation of a Web-Based Food Record for Children Using Direct Unobtrusive Lunch Observations: A Validation Study. J Med Internet Res 2015 Dec 07;17(12):e273 [FREE Full text] [doi: 10.2196/jmir.5031] [Medline: 26680744]
31. The Norwegian Patient Safety Programme. In Safe Hands 2016 URL: http://www.pasientsikkerhetsprogrammet.no/om-oss/english [accessed 2018-01-26] [WebCite Cache ID 6wkohBi3c]
32. Subar AF, Freedman LS, Tooze JA, Kirkpatrick SI, Boushey C, Neuhouser ML, et al. Addressing Current Criticism Regarding the Value of Self-Report Dietary Data. J Nutr 2015 Dec;145(12):2639-2645 [FREE Full text] [doi: 10.3945/jn.115.219634] [Medline: 26468491]
33. Førli L, Oppedal B, Skjelle K, Vatn M. Validation of a self-administered form for recording food intake in hospital patients. Eur J Clin Nutr 1998 Dec;52(12):929-933. [Medline: 9881889]
34. Hutchesson MJ, Rollo ME, Callister R, Collins CE. Self-monitoring of dietary intake by young women: online food records completed on computer or smartphone are as accurate as paper-based food records but more acceptable. J Acad Nutr Diet 2015 Jan;115(1):87-94. [doi: 10.1016/j.jand.2014.07.036] [Medline: 25262244]
35. Yon BA, Johnson RK, Harvey-Berino J, Gold BC. The use of a personal digital assistant for dietary self-monitoring does not improve the validity of self-reports of energy intake. J Am Diet Assoc 2006 Aug;106(8):1256-1259. [doi: 10.1016/j.jada.2006.05.004] [Medline: 16863723]
36. Raatz SK, Scheett AJ, Johnson LK, Jahns L. Validity of electronic diet recording nutrient estimates compared to dietitian analysis of diet records: randomized controlled trial. J Med Internet Res 2015 Jan 20;17(1):e21 [FREE Full text] [doi: 10.2196/jmir.3744] [Medline: 25604640]
37. Fukuo W, Yoshiuchi K, Ohashi K, Togashi H, Sekine R, Kikuchi H, et al. Development of a hand-held personal digital assistant-based food diary with food photographs for Japanese subjects. J Am Diet Assoc 2009 Jul;109(7):1232-1236. [doi: 10.1016/j.jada.2009.04.013] [Medline: 19559141]
38. Breslow RA, Sorkin JD. Comparison of one-day and three-day calorie counts in hospitalized patients: a pilot study. J Am Geriatr Soc 1993 Sep;41(9):923-927. [Medline: 8409178]
39. Probst YC, Faraji S, Batterham M, Steel DG, Tapsell LC. Computerized dietary assessments compare well with interviewer administered diet histories for patients with type 2 diabetes mellitus in the primary healthcare setting. Patient Educ Couns 2008 Jul;72(1):49-55. [doi: 10.1016/j.pec.2008.01.019] [Medline: 18325720]
40. Rangan AM, Tieleman L, Louie JCY, Tang LM, Hebden L, Roy R, et al. Electronic Dietary Intake Assessment (e-DIA): relative validity of a mobile phone application to measure intake of food groups. Br J Nutr 2016 Jun;115(12):2219-2226. [doi: 10.1017/S0007114516001525] [Medline: 27121045]

41. Timon CM, Blain RJ, McNulty B, Kehoe L, Evans K, Walton J, et al. The Development, Validation, and User Evaluation of Foodbook24: A Web-Based Dietary Assessment Tool Developed for the Irish Adult Population. J Med Internet Res 2017 May 11;19(5):e158 [FREE Full text] [doi: 10.2196/jmir.6407] [Medline: 28495662]
42. De KW, Huybrechts I, De MM, Ocké M, Slimani N, van ', et al. Food photographs in nutritional surveillance: errors in portion size estimation using drawings of bread and photographs of margarine and beverages consumption. Br J Nutr 2011 Apr;105(7):1073-1083. [doi: 10.1017/S0007114510004551] [Medline: 21092383]
43. Lucas F, Niravong M, Villeminot S, Kaaks R, Clavel-Chapelon F. Estimation of food portion size using photographs: validity, strengths, weaknesses and recommendations. J Hum Nutr Diet 1995 Feb;8(1):65-74. [doi: 10.1111/j.1365-277X.1995.tb00296.x]
44. Kirkpatrick SI, Subar AF, Douglass D, Zimmerman TP, Thompson FE, Kahle LL, et al. Performance of the Automated Self-Administered 24-hour Recall relative to a measure of true intakes and to an interviewer-administered 24-h recall. Am J Clin Nutr 2014 Jul;100(1):233-240 [FREE Full text] [doi: 10.3945/ajcn.114.083238] [Medline: 24787491]
45. Williamson DA, Allen HR, Martin PD, Alfonso AJ, Gerald B, Hunt A. Comparison of digital photography to weighed and visual estimation of portion sizes. J Am Diet Assoc 2003 Sep;103(9):1139-1145. [doi: 10.1053/jada.2003.50567] [Medline: 12963941]
46. Lieffers JRL, Hanning RM. Dietary assessment and self-monitoring with nutrition applications for mobile devices. Can J Diet Pract Res 2012;73(3):e253-e260. [Medline: 22958633]
47. Alvseike H, Brønnick K. Feasibility of the iPad as a hub for smart house technology in the elderly; effects of cognition, self-efficacy, and technology experience. J Multidiscip Healthc 2012;5:299-306 [FREE Full text] [doi: 10.2147/JMDH.S35344] [Medline: 23226024]
48. Vaportzis E, Clausen MG, Gow AJ. Older Adults Perceptions of Technology and Barriers to Interacting with Tablet Computers: A Focus Group Study. Front Psychol 2017 Oct 04;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
49. Dahl LA, Poulsen S, Ernst L, Kaae AK, Biltoft-Jensen A, Tetens I. Evaluation of a digital method to assess evening meal intake in a free-living adult population. Food Nutr Res 2010;54 [FREE Full text] [doi: 10.3402/fnr.v54i0.5311] [Medline: 21085516]
50. Trtovac D, Lee J. The Use of Technology in Identifying Hospital Malnutrition: Scoping Review. JMIR Med Inform 2018 Jan 19;6(1):e4 [FREE Full text] [doi: 10.2196/medinform.7601] [Medline: 29351894]

Abbreviations

BMI: body mass index
IOR: interobserver reliability
NPR: Norwegian patient registry
OUH: Oslo University Hospital
TSD: services for sensitive data
USIT: University Center for Information Technology

Edited by G Eysenbach; submitted 27.01.18; peer-reviewed by A Rangan, N Unciano, M Boaz; comments to author 31.03.18; revised version received 03.05.18; accepted 08.05.18; published 07.09.18.

Please cite as:
Paulsen MM, Hagen MLL, Frøyen MH, Foss-Pedersen RJ, Bergsager D, Tangvik RJ, Andersen LF
A Dietary Assessment App for Hospitalized Patients at Nutritional Risk: Development and Evaluation of the MyFood App
JMIR Mhealth Uhealth 2018;6(9):e175
URL: http://mhealth.jmir.org/2018/9/e175/
doi:10.2196/mhealth.9953
PMID:30194059

Original Paper

The Effectiveness of Near-Field Communication Integrated with a Mobile Electronic Medical Record System: Emergency Department Simulation Study

Kwang Yul Jung[1], MD; Taerim Kim[1], MD, PhD; Jaegon Jung[2], PhD; JeanHyoung Lee[3], DM; Jong Soo Choi[4], PhD; Kang Mira[4], MD, PhD; Dong Kyung Chang[4], MD; Won Chul Cha[1,4], MD

[1]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic Of Korea

[2]Department of Computer Engineering, Seoul Digital University, Seoul, Republic Of Korea

[3]Department of Information Strategy, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic Of Korea

[4]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic Of Korea

Corresponding Author:
Won Chul Cha, MD
Department of Emergency Medicine
Samsung Medical Center
Sungkyunkwan University School of Medicine
81, Irwon-ro
Gangnam-gu
Seoul, 06351
Republic Of Korea
Phone: 82 1053866597
Fax: 82 221487899
Email: docchaster@gmail.com

Abstract

Background: Improved medical practice efficiency has been demonstrated by physicians using mobile device (mobile phones, tablets) electronic medical record (EMR) systems. However, the quantitative effects of these systems have not been adequately measured.

Objective: This study aimed to determine the effectiveness of near-field communication (NFC) integrated with a mobile EMR system regarding physician turnaround time in a hospital emergency department (ED).

Methods: A simulation study was performed in a hospital ED. Twenty-five physicians working in the ED participated in 2 scenarios, using either a mobile device or personal computer (PC). Scenario A involved randomly locating designated patients in the ED. Scenario B consisted of accessing laboratory results of an ED patient at the bedside. After completing the scenarios, participants responded to 10 questions that were scored using a system usability scale (SUS). The primary metric was the turnaround time for each scenario. The secondary metric was the usability of the system, graded by the study participants.

Results: Locating patients from the ED entrance took a mean of 93.0 seconds (SD 34.4) using the mobile scenario. In contrast, it only required a mean of 57.3 seconds (SD 10.5) using the PC scenario (P<.001). Searching for laboratory results of the patients at the bedside required a mean of only 25.2 seconds (SD 5.3) with the mobile scenario, and a mean of 61.5 seconds (SD 11.6) using the PC scenario (P<.001). Sensitivity analysis comparing only the time for login and accessing the relevant information also determined mobile devices to be significantly faster. The mean SUS score of NFC-mobile EMR was 71.90 points.

Conclusions: NFC integrated with mobile EMR provided for a more efficient physician practice with good usability.

(JMIR Mhealth Uhealth 2018;6(9):e11187) doi:10.2196/11187

KEYWORDS

near-field communication; electronic medical records; emergency department; mobile health; mHealth

Introduction

Background

An emergency department (ED) is often characterized by chaos and inefficiency [1]. It is where the severity of a patient's injury or distress varies and changes unpredictably. The location of patients also changes based on their clinical process or test results, which may become available at different times. Physicians need to check changed locations and laboratory results frequently by walking back and forth between personal computer (PC) stations and patients' beds. In a fast-paced ED, such interruptions cause physicians to waste a substantial amount of time and ultimately result in patient dissatisfaction [2]. As the installation of PCs to all bedsides is costly and ineffective regarding space utilization, better alternatives must be considered.

Related Technologies

The ED providers strive to improve the efficiency of the workflow by exploiting advanced technologies. With the emergence of electronic medical records (EMR), mobile EMR systems are receiving increasing attention as mobile devices (ie, mobile phones, tablets), and mobile apps are becoming more common [3,4]. The portability and near ubiquity of mobile EMR allow health care providers to access patient records wherever they are needed [5].

Near-field communication (NFC) is widely used in various communication apps. In the field of health care, usage scenarios including patient identification [6], blood transfusion [7], drug administration [8], medical staff tracking [9], and medical record access [10] have already been proven. The wave of NFC technology in the health care field has been combined with the internet of things technologies [11]. Through a combination of mobile EMR systems, NFC technology can improve workflow using bedside technology [12].

Study Objectives

Numerous studies have investigated the qualitative and quantitative benefits of each technology separately [12-14]. However, the efficacy of the combined technologies for improved physician productivity in health care has not been investigated to date. More rigorous quantitative studies investigating usability estimates are required to develop and eventually adopt such systems in practice. Additionally, it is essential to determine a system's effectiveness in clinical settings. This study aims to determine the effectiveness of an NFC-integrated mobile EMR system regarding physician turnaround time in an ED.

Methods

Study Setting

This simulation study took place in an academic ED in Seoul, South Korea. The study was reviewed and approved by the Samsung Medical Center Institutional Review Board (IRB no. SMC 2018-01-144-001).

The ED is part of a tertiary academic teaching hospital with approximately 9000 daily outpatients and 2000 inpatient beds [15]. The ED holds 69 treating beds. The number of annual visits is approximately 79,000. Although the ED is heavily equipped with PCs at each station (84 PCs total), there are no PCs at the bedsides. Most of the beds are not in private rooms but are open to stations except isolation beds.

The hospital developed the proposed mobile EMR system. It operates on the institution's EMR system, which was also developed internally. The overall system had a significant update in July 2016. Figure 1 shows a schematic of the architecture of the EMR system.

The mobile EMR uses an Android app that gives physicians access to inpatient, outpatient, and emergency department information. Users can log into the system with their fingerprint and search for locations, clinical notes, vital signs, laboratory results, and medical images. The NFC function was implemented in April 2017. When a physician links the EMR mobile device to an NFC tag which contains information about its location, the mobile app is automatically initiated. Using fingerprint authentication, physicians can log in and view the EMR of patients at the location that corresponds with their NFC location. When tagging NFC tags at the entrance of a specific zone, the list of patients at the tagged NFC zone who are in charge of mobile device users is popped up. Figure 2 shows the access process.

Study Participants

Physicians who worked in the ED during the study were asked to participate. Physician participants were recruited between April 1 to April 20, 2018. Among the 35 ED physicians, 25 (71%) agreed to participate in this study.

Study Scenarios and Sensitivity Analysis

After a brief introduction, the physicians went through 2 sequential scenarios. The first scenario (scenario A) involved locating patients in the ED from the ED gate. Physicians were given the name of a patient and were required to locate them using either a PC EMR at the nearest site in the ED gate or mobile EMR. After locating the patient, the participant was guided to reach their bedside. The second scenario (scenario B) involved looking up a laboratory result from the bedside. Physicians were brought to a patient's bedside and were required to determine a specific laboratory result using either a mobile device or PC interface. As there were no PCs at the bedside, physicians had to perform a few steps to identify available PCs and return with a report. The steps in each scenario are shown in Figure 3, and the flow of each scenario is shown in Figure 4.

Physicians were randomly assigned to follow either scenario A or B using either a mobile device (mobile case) or a PC (PC case) as described in Multimedia Appendix 1. An independent observer recorded the activities with a camera and completed a case report form with time stamps during the process. Patients were not simulated. Real patients were accessed in the emergency department. However, since we used only partial patient data such as name, location, and laboratory data which

was already available, the clinical condition did not influence the study's outcome.

Figure 1. Overall schematic description of the hospital information system architecture relationship at the Samsung Medical Center. DARWIN: data analytics and research window for integrated knowledge; CPOE: computerized physician order entry; MIS: management information system; MDM: master data management; CRM: customer relationship management.

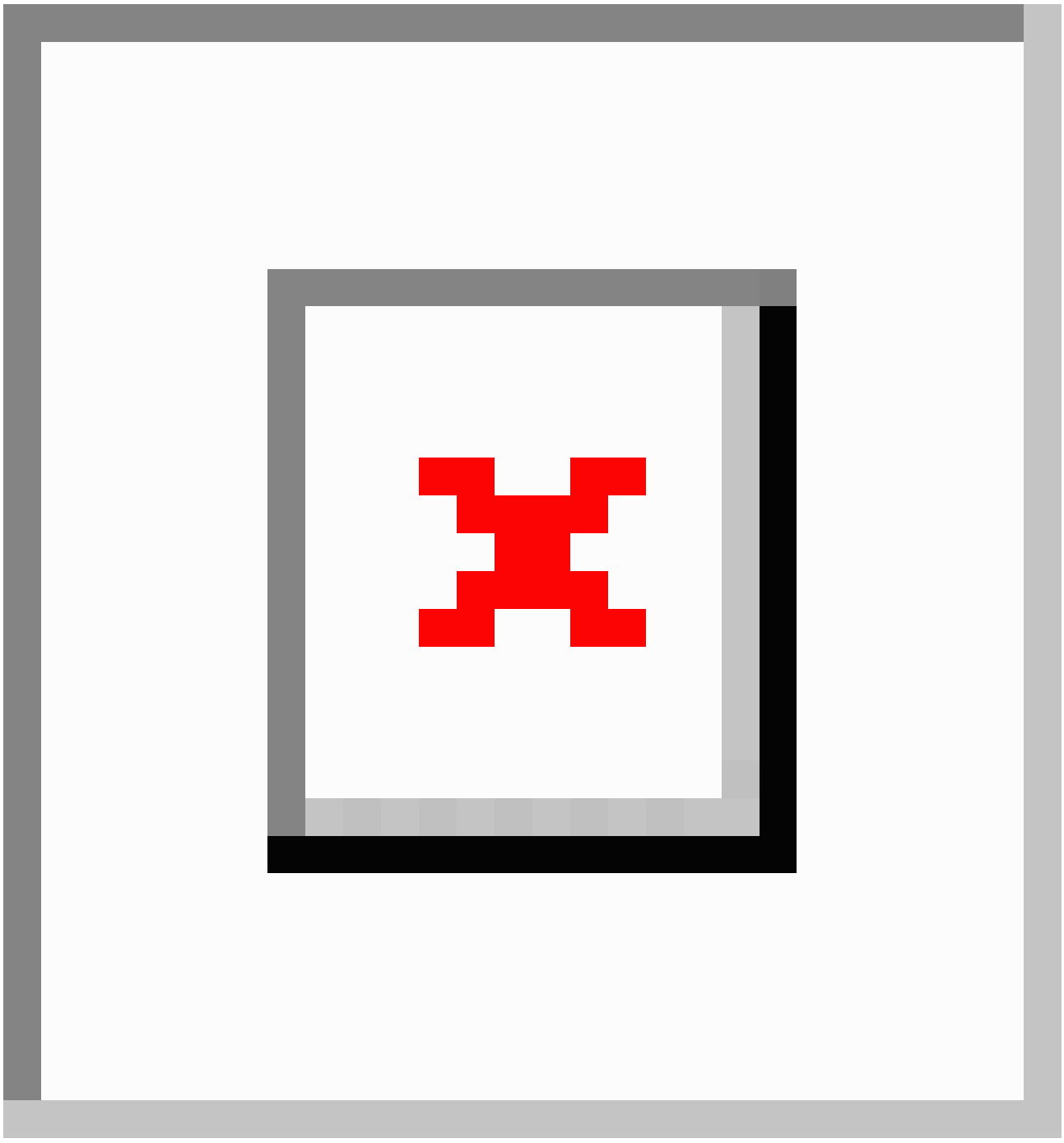


Figure 2. Usage scene when the mobile electronic medical records (EMR) communicate with the near-field communication (NFC) system and the display of the mobile EMR progression after tagging NFC. V/S: vital sign.

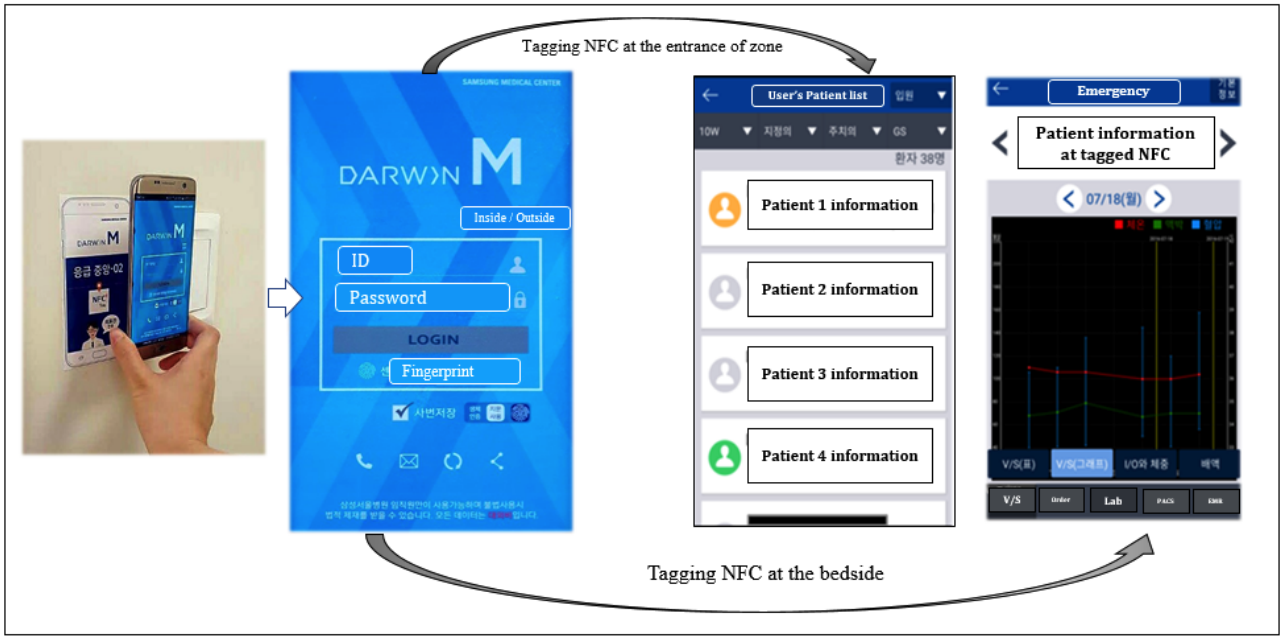


Figure 3. Schematic view of simulation scenarios. (a) Locating the patient. (b) Looking up laboratory results for the patient. ED: emergency department; PC: personal computer.

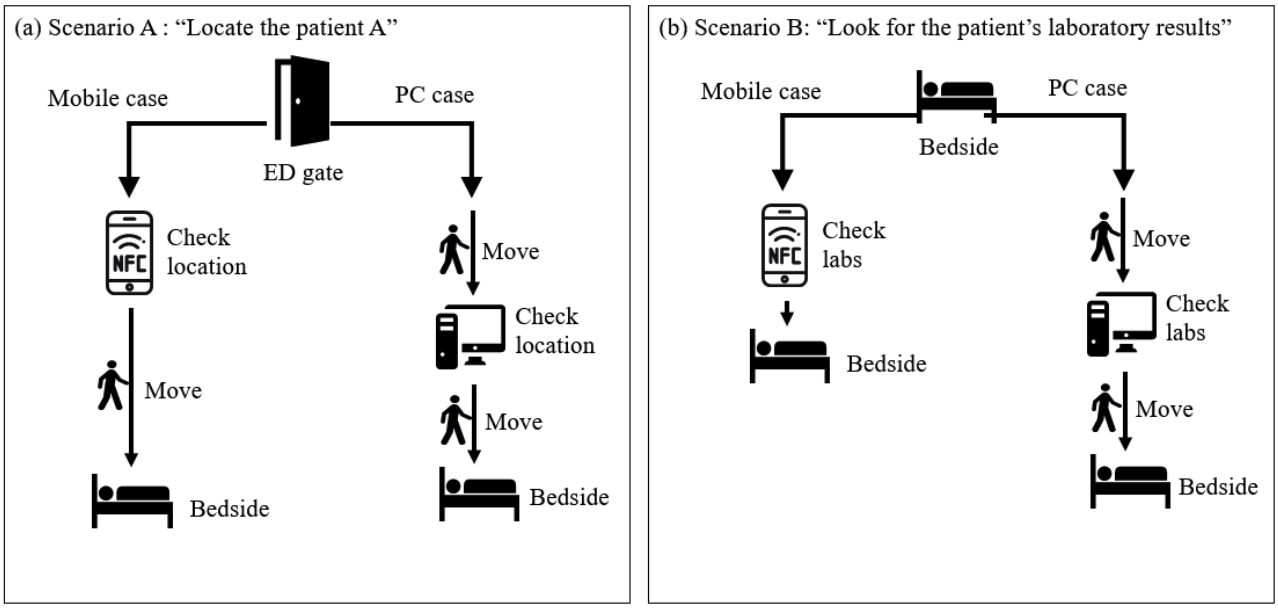
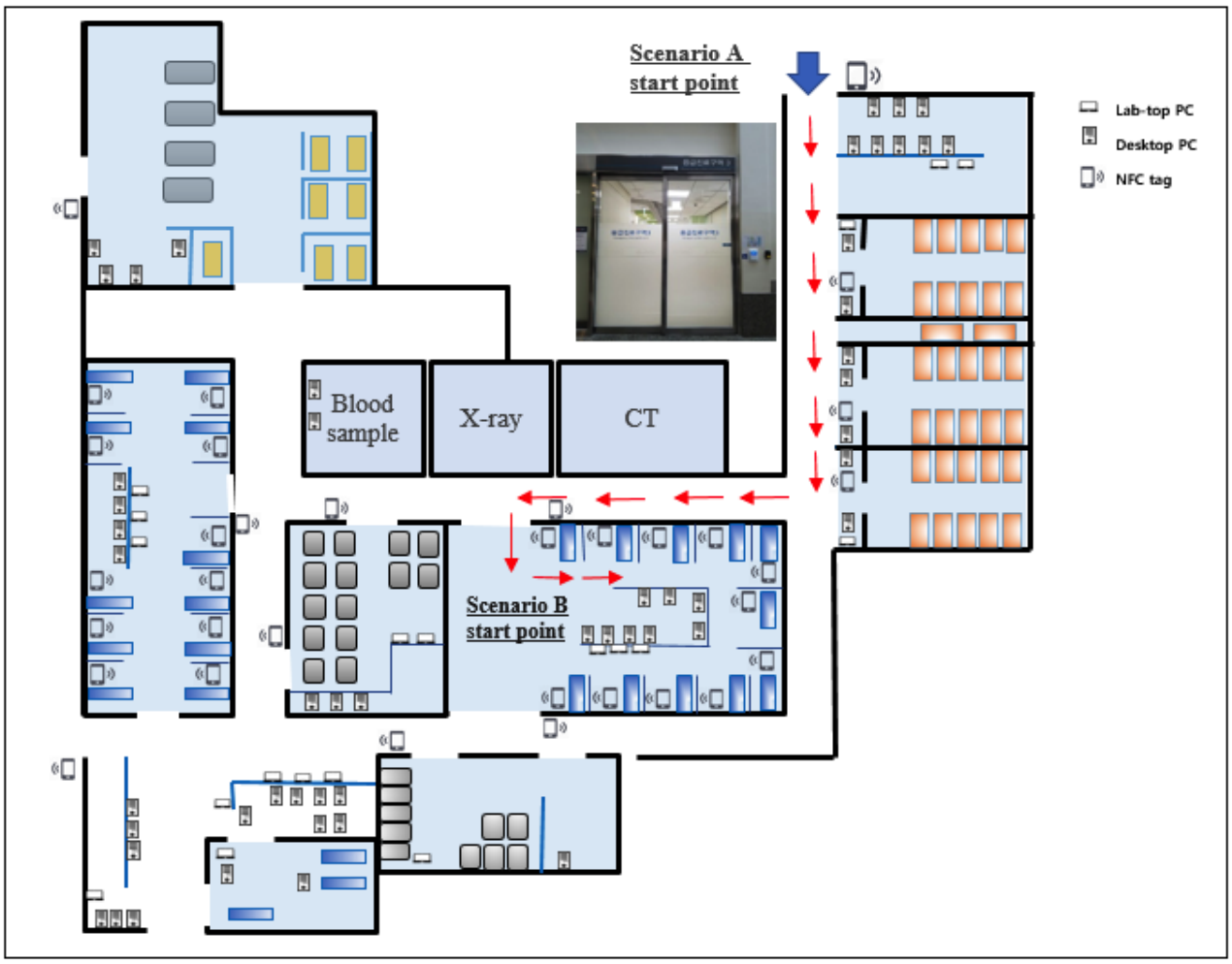


Figure 4. The flow of locating the patient and the scenario place in the emergency department. CT: computed tomography; NFC: near-field communication; PC: personal computer.



We performed a sensitivity analysis using the data without considering movement intervals. This test was performed to determine whether or not there was a consistent outcome if the condition allowed for more available PCs, which are at the gate and the bedside.

Survey

After completing all scenarios, physician participants responded to 10 questions using the system usability scale (SUS). The SUS is composed of a 5-point Likert scale rated from 1 (strongly disagree) to 5 (strongly agree) that investigates the usability of the NFC-integrated mobile EMR system [16]. The SUS score calculation formula is as follows:

Measurement and Outcome

The primary metric was the length of turnaround time for each scenario. The secondary metric was the usability of the system, as graded by the study physician participants. We collected demographic data from each participant and recorded the time intervals of each step of the process for both scenarios. We also analyzed time intervals among groups sorted by age, gender, and occupation. Afterward, the SUS questionnaires were collected and analyzed.

Statistical Analysis

Continuous variables are expressed in terms of mean and standard deviations (SD), whereas categorical variables are expressed in frequencies and percentages. The time mean difference was examined using a paired *t*-test. A value of $P<.05$ was considered to be statistically significant. As descriptive statistics could not confirm a normal distribution of participants between the 2 dependent groups divided by age, gender, and occupation, the Mann-Whitney U test was applied for time interval difference analysis.

Results

Main Outcome

Among 25 physician participants, 14 (56%) were male, and 11 (44%) were female. The general characteristics of the participants are shown in Table 1.

It required a mean of 93.0 seconds (SD 34.4) to locate the patient from the entrance of the ED in the PC case but only a mean of 57.3 seconds (SD 10.5) in the mobile case, which was significantly faster ($P<.001$). Accessing laboratory results at

the patient's bedside required a mean of only 25.2 seconds (SD 5.3) in the mobile case compared to a mean of 61.5 seconds (SD 11.6) in the PC case. These data were statistically significant (P<.001). A schematic comparison is shown in Figure 5.

Sensitivity Analysis

We compared the time required for login with the time for finding relevant information. Login using the mobile device EMR required a mean of 13.1 seconds (SD 2.9) for scenario A and a mean of 12.5 seconds (SD 2.1) for scenario B. Login by PC took longer with a mean of 36.2 seconds (SD 15.2) for scenario A and a mean of 30.5 seconds (SD 7.7) for scenario B. The differences in time were statistically significant (P<.001). Finding the location of patients after login required a mean of only 6.8 seconds (SD 3.6) using the mobile device, whereas it took a mean of 18.9 seconds (SD 16.9) using a PC. Accessing a specific laboratory test result required a mean of 12.8 seconds (SD 5.3) using the mobile device and a mean of 26.5 seconds (SD 8.0) using a PC. These data were statistically significant (P<.001). The results are shown in Table 2.

Survey

The mean SUS score of NFC-mobile EMR was 71.90 points. The results are shown in Table 3.

Table 1. Characteristics of the physician participants.

Participant characteristic	Value
Age (years), mean (SD)	30.6 (4.9)
Age groups (years), n (%)	
≥30	12 (48)
<30	13 (52)
Gender, n (%)	
Male	14 (56)
Female	11 (44)
Occupation, n (%)	
Intern	4 (16)
Resident	15 (60)
Specialist	6 (24)
Time worked at hospital (years), mean (SD)	4.6 (4.0)

Figure 5. Graphical view of main results. (a) Locating the patient. (b) Looking up laboratory results for the patient. NFC: near-field communication; PC: personal computer.

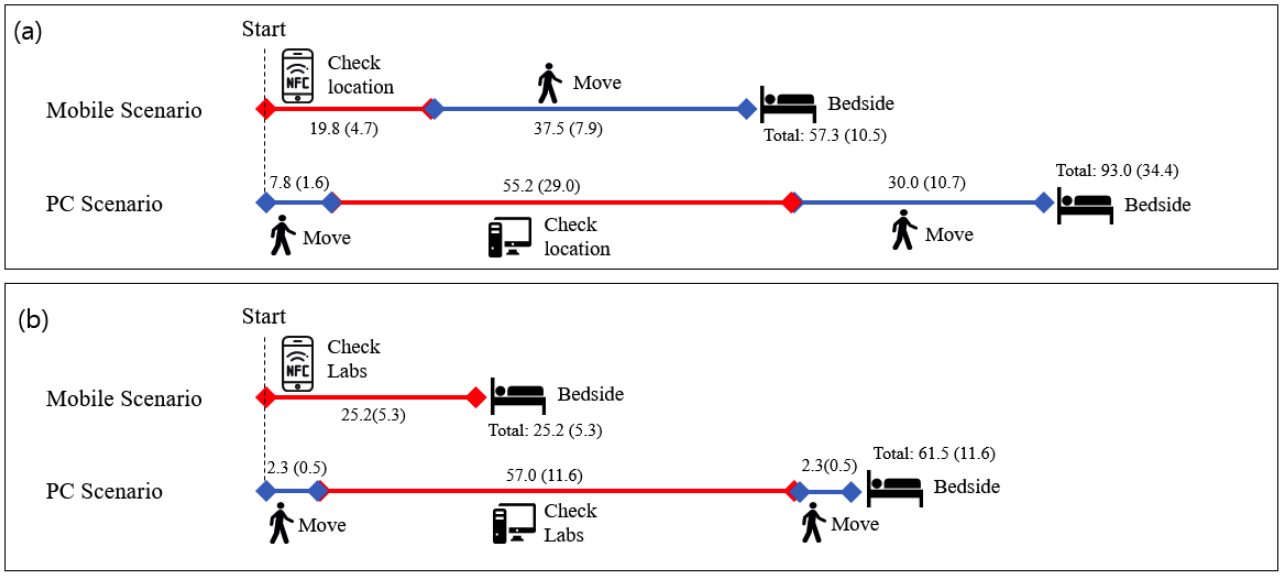


Table 2. A comparison between the 2 scenarios of time spent on specific tasks.

Task	Time (seconds), mean (SD)		*P* value
	Mobile case	Personal computer case	
Scenario A			
Login	13.1 (2.9)	36.2 (15.2)	<.001
Accessing relevant information	6.8 (3.6)	18.9 (16.9)	<.001
Total	19.8 (4.7)	55.2 (29.0)	<.001
Scenario B			
Login	12.5 (2.1)	30.5 (7.7)	<.001
Accessing relevant information	12.8 (5.3)	26.5 (8.0)	<.001
Total	25.2 (5.3)	57.0 (11.6)	<.001

Table 3. Score results (n=25) from the system usability scale (SUS) to assess the near-field communication mobile emergency medical record (NFC-mobile EMR).

Question	Mean (SD)
1. I think that I would like to use this NFC-mobile EMR frequently.	3.92 (0.95)
2. I found the NFC-mobile EMR unnecessarily complex.	1.76 (0.83)
3. I thought the NFC-mobile EMR was easy to use.	4.40 (0.50)
4. I think that I would need the support of a technical person to be able to use the NFC-mobile EMR.	2.72 (1.10)
5. I found that the various functions in the NFC-mobile EMR were well-integrated.	4.24 (0.72)
6. I thought there was too much inconsistency in the NFC-mobile EMR.	4.20 (0.64)
7. I would imagine that most people would learn to use the NFC-mobile EMR very quickly.	4.48 (0.59)
8. I found the NFC-mobile EMR very cumbersome to use.	1.56 (0.51)
9. I felt very confident using the NFC-mobile EMR.	3.88 (0.78)
10. I needed to learn a lot of things before I could get going with the NFC-mobile EMR.	1.92 (0.57)
Total score	71.90 (7.61)

Discussion

Principal Findings

This study aimed to improve physician efficiency by reducing the time spent walking to check patient information with the aid of the technological integration between NFC and mobile device EMR. To the best of our knowledge, this is the first study to examine the efficiency of this system and comparing it with the PC EMR. The mobile total turnaround time for performing tasks was significantly reduced in both scenarios. Sensitivity analysis showed that mobile device EMR incorporated with NFC was significantly faster than PC-integrated EMR regarding login time and accessing laboratory results.

As the familiarity of mobile device use could be different among the demographic groups, we compared the total time interval difference between PC and mobile cases. Multimedia Appendix 2 shows that the mobile case was consistently faster for all groups. However, there were significant differences in the time interval between age and occupation during scenario B. These findings are contrary to the general belief that the younger generation is more familiar with newer technology [17]. A further study on mobile device familiarity is needed because the simulation was done with a small sample size.

We also evaluated usability with the SUS questionnaire. The SUS was used after the physician participant had an opportunity to use the system being evaluated. A score over 70 on the questionnaire (range 0-100) indicated that the NFC-integrated mobile device EMR was “acceptable,” and the adjective rating was “good” [18]. There was no significant statistical difference among groups based on age, gender, and occupation.

Advantages and Disadvantages

Various measures have been implemented to address ED inadequacies. Improving the ED work efficiency is one crucial in-hospital factor. Ideal physical structures for work have already been demonstrated [19]. Several studies have shown the positive effect of developing clinical guidelines and protocols for effective evaluation of efficiency [20,21]. Newer technologies such as radio frequency identification-integrated point-of-care testing [22], triage kiosks [23], and dashboards [24] have been well studied. Ubiquitous near patient access to EMR via NFC is determined to be useful in this regard. Compared to installing new structures in an already heavily equipped ED, implementing an NFC tag system is a relatively easy way to improve workflow regarding cost and space utilization.

As most mobile EMR functions are more readily accessible with PCs, our study paid attention to superiority only available in mobile EMR. Mobile device systems outperform PC systems concerning mobility and personalization (at the provider level, and patient level). We have measured the turnaround time as the primary outcome of these merits. Thus, we have shown that physicians can gain access to information without physically moving the location of their patients.

Portability of mobile EMR could be improved by incorporating accessibility through NFC. Our study revealed a statistically significant difference in login time which was more effective by mobile EMR than by PC EMR (Table 2). A previous study by Holden [25] demonstrated that the issue of accessibility to EMR such as system login and system response time could negatively impact the usability of mobile EMR. NFC integrated response and fingerprint login at the location of interest using a mobile device could be beneficial because the process is simplified and less time-consuming. This system also appears to reduce security concerns from failed logouts or departure without logging out, by using the individual's mobile device.

An increase in the length of time physicians spend at the bedside is likely to increase patient satisfaction [26]. With this bedside technology, the physician can show radiologic results or laboratory results to patients who cannot ambulate.

Inconsistent loading time due to varying network coverage could be a disadvantage for this technology. For example, mobile devices without NFC function cannot be used. Physicians might routinely tend to use PC EMR because PC EMR covers mobile EMR. A previous study by Duhm et al [14] demonstrated that a physician usually underestimates actual time savings during their professional capacity. The results of this study make a compelling argument and provide preliminary evidence in support of adequately addressing this tendency, particularly concerning reduced workflow using mobile EMR with NFC functionality.

However, to enhance emergency physician performance, a multidimensional approach is required, rather than a single tool. ED processes are complicated, with multiple steps from various providers often originating from outside the ED.

Limitations

First among the limitations of this study is that this investigation was conducted at a single center. Additional studies conducted at multiple centers or EDs are needed to improve the generalizability of our conclusions.

Secondly, participants had different levels of familiarity with mobile devices and NFC tags. Only some participants were familiar with NFC because the system was built over a year ago, which might cause bias.

Thirdly, each participant encountered various encumbrances because this study was conducted in an actual emergency room. For example, when attempting to locate a patient in the middle of a scenario, the nearest PC may have been occupied by another staff member, which led to the physician being forced to use a PC that was further away. Also, while moving to a patient's bedside, there was an occasion when a participant was forced to stop because a moving stretcher cart or medical staff member blocked the aisle. In addition, some of the PCs used were comparatively slow. As mentioned above, unpredictable circumstances might influence the overall time measured for each scenario. As shown in Multimedia Appendix 3, the variability of turnaround time fluctuated. However important, these events could not be systemically quantified.

Finally, the usability assessment for NFC-mobile EMR via SUS could be overrated because responses were filled out immediately after performing scenarios, which in most cases, resulted in the superiority of NFC-mobile EMR. Further studies could investigate usability over a more extended period of the physician's working practice.

Conclusion

NFC-integrated mobile EMR is effective for reducing the turnaround time of physicians when practicing in the field and has excellent usability.

Acknowledgments

The authors would like to acknowledge the support provided by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant number 2018R1C1B6002877).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Randomly allocated scenario quest for each participant. PC: personal computer.

[PDF File (Adobe PDF File), 19KB - mhealth_v6i9e11187_app1.pdf]

Multimedia Appendix 2

Comparison of time spent among age, gender, and occupation for each scenario.

[PDF File (Adobe PDF File), 24KB - mhealth_v6i9e11187_app2.pdf]

Multimedia Appendix 3

Graphical comparison between mobile and personal computer scenario time of all participants.

[PPTX File, 51KB - mhealth_v6i9e11187_app3.pptx]

References

1. Berger E. Breaking point: Report calls for Congressional rescue of hospital emergency departments. Ann Emerg Med 2006 Aug;48(2):140-142. [Medline: 16953525]
2. Weigl M, Beck J, Wehler M, Schneider A. Workflow interruptions and stress atwork: a mixed-methods study among physicians and nurses of a multidisciplinary emergency department. BMJ Open 2017 Dec 22;7(12):e019074 [FREE Full text] [doi: 10.1136/bmjopen-2017-019074] [Medline: 29275350]
3. Kim J, Lee Y, Lim S, Lee J. How Are Doctors Using Mobile Electronic Medical Records? An In-Depth Analysis of the Usage Pattern. Stud Health Technol Inform 2017;245:1231. [Medline: 29295318]
4. Yeo K, Lee K, Kim J, Kim T, Choi Y, Jeong W, et al. Pitfalls and Security Measures for the Mobile EMR System in Medical Facilities. Healthc Inform Res 2012 Jun;18(2):125-135 [FREE Full text] [doi: 10.4258/hir.2012.18.2.125] [Medline: 22844648]
5. Burdette SD, Herchline TE, Oehler R. Surfing the web: practicing medicine in a technological age: using smartphones in clinical practice. Clin Infect Dis 2008 Jul 01;47(1):117-122. [doi: 10.1086/588788] [Medline: 18491969]
6. Aguilar A, van der Putten W, Kirrane F. Positive patient identification using RFID and wireless networks. 2006 Nov Presented at: 11th Annual Conference and Scientific Symposium (HISI); November 2006; Dublin, Ireland.
7. Gutierrez A, Levitt J, Reifert D, Raife T, Diol B, Davis R, et al. Tracking blood products in hospitals using radio frequency identification: Lessons from a pilot implementation. VOXS 2013 May 31;8(1):65-69. [doi: 10.1111/voxs.12015]
8. Houliston B. Integrating RFID technology into a drug administration system. In: Journal of Applied Computing and Information. 2005 Presented at: HIC 2005 and HINZ 2005: Proceedings; 2005; Brunswick East, Vic.: Health Informatics Society of Australia p. 453-461.
9. Halamka J. Early experiences with positive patient identification. J Healthc Inf Manag 2006;20(1):25-27. [Medline: 16429955]
10. Ajami S, Rajabzadeh A. Radio Frequency Identification (RFID) technology and patient safety. J Res Med Sci 2013 Sep;18(9):809-813 [FREE Full text] [Medline: 24381626]
11. Lee BM, Ouyang J. Intelligent Healthcare Service by using Collaborations between IoT Personal Health Devices. IJBSBT 2014 Feb 28;6(1):155-164. [doi: 10.14257/ijbsbt.2014.6.1.17]
12. Köstinger H, Gobber M, Grechenig T, Tappeiner B, Schramm W. Developing an NFC based patient identification and ward round system for mobile devices using the android platform. In: Point-of-Care. 2013 Jan 16 Presented at: 2013 IEEE Point-of-Care Healthcare Technologies (PHT); 16-18 Jan. 2013; Bangalore, India p. 2013.
13. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010 Oct 06;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
14. Duhm J, Fleischmann R, Schmidt S, Hupperts H, Brandt SA. Mobile Electronic Medical Records Promote Workflow: Physicians' Perspective From a Survey. JMIR Mhealth Uhealth 2016 Jun 06;4(2):e70 [FREE Full text] [doi: 10.2196/mhealth.5464] [Medline: 27268720]
15. Cha WC, Ahn KO, Shin SD, Park JH, Cho JS. Emergency Department Crowding Disparity: a Nationwide Cross-Sectional Study. J Korean Med Sci 2016 Aug;31(8):1331-1336 [FREE Full text] [doi: 10.3346/jkms.2016.31.8.1331] [Medline: 27478347]
16. Jordan P, Thomas B, Weerdmeester B, McClelland I. SUS-A quick and dirty usability scale. In: Usability Evaluation in Industry. Bristol, PA, USA: Florence, Kentucky, U.S.A.: Taylor & Francis; 1996:189-194.
17. McAlearney AS, Schweikhart SB, Medow MA. Doctors' experience with handheld computers in clinical practice: qualitative study. BMJ 2004 May 15;328(7449):1162 [FREE Full text] [doi: 10.1136/bmj.328.7449.1162] [Medline: 15142920]
18. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Studies 2009 May;4(3):114-123.
19. Zusman E. Form facilitates function: innovations in architecture and design drive quality and efficiency in healthcare. Neurosurgery 2010 Jun;66(6):N24. [doi: 10.1227/01.neu.0000375281.91778.45] [Medline: 20495418]
20. Butti L, Bierti O, Lanfrit R, Bertolini R, Chittaro S, Delli CS, et al. Evaluation of the effectiveness and efficiency of the triage emergency department nursing protocol for the management of pain. J Pain Res 2017;10:2479-2488 [FREE Full text] [doi: 10.2147/JPR.S138850] [Medline: 29081670]
21. Wright S, Trott A, Lindsell C, Smith C, Gibler W. Evidence-based emergency medicine. Creating a system to facilitate translation of evidence into standardized clinical practice: a preliminary report. Ann Emerg Med 2008 Jan;51(1):80-6, 86.e1. [doi: 10.1016/j.annemergmed.2007.04.009] [Medline: 17719134]
22. Rao A, Dighe A. Radiofrequency Identification and Point-of-Care Testing. Point of Care: The Journal of Near-Patient Testing & Technology 2004;3(3):130-134. [doi: 10.1097/01.poc.0000138646.56228.65]

23. Ackerman S, Tebb K, Stein J, Frazee B, Hendey G, Schmidt L, et al. Benefit or burden? A sociotechnical analysis of diagnostic computer kiosks in four California hospital emergency departments. Soc Sci Med 2012 Dec;75(12):2378-2385. [doi: 10.1016/j.socscimed.2012.09.013] [Medline: 23063214]
24. Swartz J, Cimino J, Fred M, Green RA, Vawdrey DK. Designing a clinical dashboard to fill information gaps in the emergency department. AMIA Annu Symp Proc 2014;2014:1098-1104 [FREE Full text] [Medline: 25954420]
25. Holden RJ. Physicians' beliefs about using EMR and CPOE: in pursuit of a contextualized understanding of health IT use behavior. Int J Med Inform 2010 Feb;79(2):71-80 [FREE Full text] [doi: 10.1016/j.ijmedinf.2009.12.003] [Medline: 20071219]
26. Lin CT, Albertson GA, Schilling LM, Cyran EM, Anderson SN, Ware L, et al. Is patients' perception of time spent with the physician a determinant of ambulatory patient satisfaction? Arch Intern Med 2001 Jun 11;161(11):1437-1442. [Medline: 11386893]

Abbreviations

ED: emergency department
EMR: emergency medical record
NFC: near-field communication
PC: personal computer
SUS: system usability scale

Edited by G Eysenbach; submitted 31.05.18; peer-reviewed by S Lim, J Wu; comments to author 09.07.18; revised version received 31.07.18; accepted 31.08.18; published 21.09.18.

Please cite as:
Jung KY, Kim T, Jung J, Lee J, Choi JS, Mira K, Chang DK, Cha WC
The Effectiveness of Near-Field Communication Integrated with a Mobile Electronic Medical Record System: Emergency Department Simulation Study
JMIR Mhealth Uhealth 2018;6(9):e11187
URL: http://mhealth.jmir.org/2018/9/e11187/
doi:10.2196/11187
PMID:30249577

Original Paper

Using a ResearchKit Smartphone App to Collect Rheumatoid Arthritis Symptoms From Real-World Participants: Feasibility Study

Michelle Crouthamel[1], MSc; Emilia Quattrocchi[2], MD; Sarah Watts[3], MSc; Sherry Wang[1], MSc; Pamela Berry[1], MSc; Luis Garcia-Gancedo[3], PhD; Valentin Hamy[2], PhD; Rachel E Williams[1], MS, PhD

[1]GlaxoSmithKline, Collegeville, PA, United States

[2]GlaxoSmithKline, Stockley Park, Uxbridge, United Kingdom

[3]GlaxoSmithKline, Stevenage, United Kingdom

Corresponding Author:
Michelle Crouthamel, MSc
GlaxoSmithKline
1000 Black Rock Road
Collegeville, PA, 19426
United States
Phone: 1 215 756 5101
Email: ming-chih.h.crouthamel@gsk.com

Abstract

Background: Using smartphones to enroll, obtain consent, and gather self-reported data from patients has the potential to enhance our understanding of disease burden and quantify physiological impact in the real world. It may also be possible to harness integral smartphone sensors to facilitate remote collection of clinically relevant data.

Objective: We conducted the Patient Rheumatoid Arthritis Data From the Real World (PARADE) observational study using a customized ResearchKit app with a bring-your-own-device approach. Our objective was to assess the feasibility of using an entirely digital approach (social media and smartphone app) to conduct a real-world observational study of patients with rheumatoid arthritis.

Methods: We conducted this observational study using a customized ResearchKit app with a bring-your-own-device approach. To recruit patients, the PARADE app, designed to guide patients through a series of tasks, was publicized via social media platforms and made available for patients in the United States to download from the Apple App Store. We collected patient-reported data, such as medical history, rheumatoid arthritis-related medications (past and present), and a range of patient-reported outcome measures. We included in the assessment a joint-pain map and a novel objective assessment of wrist range of movement, measured by the smartphone-embedded gyroscope and accelerometer.

Results: Within 1 month of recruitment via social media campaigns, 399 participants self-enrolled, self-consented, and provided complete demographic data. Joint pain was the most frequently reported rheumatoid arthritis symptom to bother study participants (344/393, 87.5%). Severe patient-reported wrist pain appeared to be inversely linked with the range of wrist movement measured objectively by the app. At study entry, 292 of 399 participants (73.2%) indicated a preference for participating in a mobile app–based study. The number of participants in the study declined to 45 of 399 (11.3%) at week 12.

Conclusions: Despite the declining number of participants over time, the combination of social media and smartphone app with sensor integration was a feasible and cost-effective approach for the collection of patient-reported data in rheumatoid arthritis. Integral sensors within smartphones can be harnessed to provide novel end points, and the novel wrist range of movement test warrants further clinical validation.

(JMIR Mhealth Uhealth 2018;6(9):e177) doi:10.2196/mhealth.9656

KEYWORDS

rheumatoid arthritis; smartphone; mobile phone; patient-reported outcome measures; mobile applications

Introduction

The Need for Novel Data Collection Methods

Traditionally, clinical research and data collection is primarily conducted through a site-based approach. Patients are screened and recruited by clinical research centers, followed by a series of clinical visits scheduled for procedures and clinical assessments. Other methods include data collection via telephone, mail, or electronic surveys. These site-based approaches can be complex, costly, and time consuming; limitations associated with these approaches include investigator bias [1], parking lot syndrome [2], and white coat syndrome [3]. Since health status involves multiple dimensions and is a continuum, clinicians' assessments during infrequent clinical visits may not be sufficient to fully evaluate health status. Novel data collection methods to capture continuous data from the patient perspective are needed.

Mobile phone ownership is becoming more widespread globally, with the number of people using mobile phones in 2017 estimated to be 4.8 billion worldwide [4]. Following the decline in the popularity of other forms of communication (postal mail and landline phones) the mobile phone has recently emerged as a potentially useful and successful technology for health measurement and health management [5-7], a clear shift that appears to be welcomed by patients [8]. When paired with specialized apps, designed to provide standardization of the captured data, mobile phones may be easily used to both engage the patient with their treatment and provide a conduit through which patient-reported data can be gathered and transmitted [9]. The use of smartphones to collect real-world data directly from patients has been shown to be cost effective and fast, and, importantly, it empowers the patients [5,7,10]. Although smartphones have been shown to be effective at gathering data, high rates of attrition have been reported in studies using smartphones alone [5,11,12], highlighting a need to improve and optimize current methods of promoting participant retention.

Rheumatoid Arthritis as a Case Study

Patients with long-term chronic and physically disabling conditions, such as rheumatoid arthritis (RA), who need to manage daily care activities and fluctuations in their RA symptoms at home, may be a population for whom electronic data collection could be of particular use. RA is a chronic, progressive autoimmune disease affecting the mobile joints of the body, which may result in substantial and irreversible disability; symptoms include joint pain or tenderness, joint swelling, morning stiffness, reduction in joint range of movement (ROM), muscle pain, and fatigue [9,13]. Various chronic and progressive autoimmune conditions, including RA, are typically characterized by periodic disease flares followed by periods of relative quiescence that are unpredictable. Assessment of disease status during prescheduled physician visits, whether for routine purposes or for gathering data for clinical research purposes, can easily miss these periods of disease exacerbation, leading to an incorrect interpretation of the patient's disease status. The use of smartphone apps in clinical research means that medical data can be efficiently collected at a greater number of time points, compared with more traditional research methods [12]. Thus, the ability of the patient to self-report data relating to the subjective signs and symptoms of their disease may pave the way for more holistic clinical research datasets, able to better capture the experience of patients living with chronic conditions, and may provide immediate feedback to their health care providers, ultimately improving clinical benefit overall [11].

Data Capture Using ResearchKit

ResearchKit (Apple Inc, Cupertino, CA, USA) is an iOS-based, open source framework for mobile medical research released in 2015. It is an app composed of preconstructed modules, which can be adapted according to the research requirements. To date, ResearchKit has been used in observational, real-world studies of cardiovascular health, asthma, Parkinson disease, type 2 diabetes, and cancer [5,7,11,12]. So far, to our knowledge, it has not been used to conduct an entire scientific research study in patients with RA enrolled remotely, including obtaining patients' consent and data capture.

Several observational studies have assessed the use of smartphones to measure subjective RA disease activity through adoption of validated questionnaire-based patient-reported outcome (PRO) measures [14-16]. Additionally, creative use of the integrated sensor platforms present within most smartphones could facilitate collection of a large number of objective variables relevant to RA. For example, monitoring the impact of physical activity on cardiovascular health, by gathering objective data relating to patient mobility and activity levels through the use of the integral smartphone global positioning system, was shown to be successful [12]. An attempt to capture data relating to gait analysis in patients with RA has been reported, but this method awaits full validation [15,17].

Objectives

The Patient Rheumatoid Arthritis Data From the Real World (PARADE) study was a siteless, prospective, real-world observational study in which patients with RA could self-recruit, provide consent, enroll, and report their medical data entirely via a customized ResearchKit app downloaded to their own smartphone. The key objectives were to assess the feasibility of using ResearchKit to enroll patients into a study and of collecting patient data using this app, including subjective assessments (ie, self-reported symptoms, validated health-related quality of life surveys, and a joint-pain map), as well as a novel, objective wrist ROM test. To determine the feasibility of the app and siteless approach to recruitment, the aims of the study were to enroll a minimum of 300 participants in 1 month, demonstrate participant demographics similar to those of an existing RA registry, and evaluate algorithms developed to support the objective measurement of RA symptoms via the app.

Methods

App Design and Testing

The PARADE app was created and developed by GlaxoSmithKlein (GSK; Brentford, UK) and Possible Mobile (Denver, CO, USA) using the ResearchKit platform, which is open source and available on GitHub [18]. Two sets of user

acceptance testing were performed during the app's development. During user acceptance testing, users tested the software to ensure that it could handle required tasks in real-world scenarios. The first user acceptance testing was performed with 3 volunteers with RA who provided feedback on screen functionality and the electronic PROs. The second user acceptance testing was performed with study team members 1 month before the app launch as the final approval testing. Multimedia Appendix 1 provides screenshots from the app.

Recruitment Process and Ethics

This study (study number 205718) was approved by the Quorum institutional review board (Quorum Review Inc, Seattle, WA, USA). The study was an observational platform pilot; it was not registered within the clinical trial registry. The app was made available in the United States via the Apple App Store. We identified California, Texas, New York, Pennsylvania, and Florida as the top 5 states with a high prevalence of RA; therefore, we launched targeted digital patient recruitment campaigns via social media platforms, such as Facebook (Facebook, Inc, Menlo Park, CA, USA) and Twitter (Twitter, Inc, San Francisco, CA, USA), in those states. We also targeted HealthUnlocked (HealthUnlocked, London, UK), Inspire (Inspire, Arlington, VA, USA), and users on Facebook who were followers of the Arthritis Foundation [19] and Creaky Joints [20] to increase awareness and drive interest to download the PARADE app. Prospective participants downloaded the app using their own App Store credentials and self-navigated through elements of the app, including a study video, eligibility screen, electronic informed consent screen, and data collection screen, via the smartphone touchscreen interface. The involvement of GSK was made clear on the Welcome screen of the app and in the informed consent process.

We set the recruitment time frame for 1 month with a target enrollment of 300 patients. The consenting process was compliant with Title 21 of the Code of Federal Regulations Part 11 (21CP11) to ensure that it was trustworthy, reliable, and equivalent to paper records according to the US Food and Drug Administration.

Once prospective participants had downloaded and opened the app, they were presented with information about the study and an inclusion and exclusion criteria questionnaire. Inclusion criteria were being 21 years of age or over, being English speaking, living in the United States, and having a physician's diagnosis of RA. Participants who met the eligibility criteria proceeded to electronic informed consent. Once they completed this, they could receive a copy of the consent form by email for their records. Collection of personally identifiable information (including email address) was restricted to the consent process and was housed separately from the study data on secure servers (Medidata Solutions Inc, New York, NY, USA). Throughout, participants navigated the app using the smartphone touchscreen.

Study Design

We conducted the study entirely via the app with no human interaction, medical intervention, or financial incentives.

Possible Mobile programmed the app to automatically randomly allocate participants, without stratification, into 2 groups with differing degrees of access provided to their personal data. Group A could access their personal symptom data dashboard every day, whereas group B could access it only at the end of the study after 12 weeks of data collection. The informed consent form notified participants that they would be assigned by chance into 1 of 2 study groups; however, participants were not informed about the purpose of the randomization.

Assessments

Data captured at enrollment (week 1) were baseline demographics, health history information, retention, and medications. Participants were encouraged to complete several subjective and objective study tasks at different time points, as described below.

Subjective Study Tasks

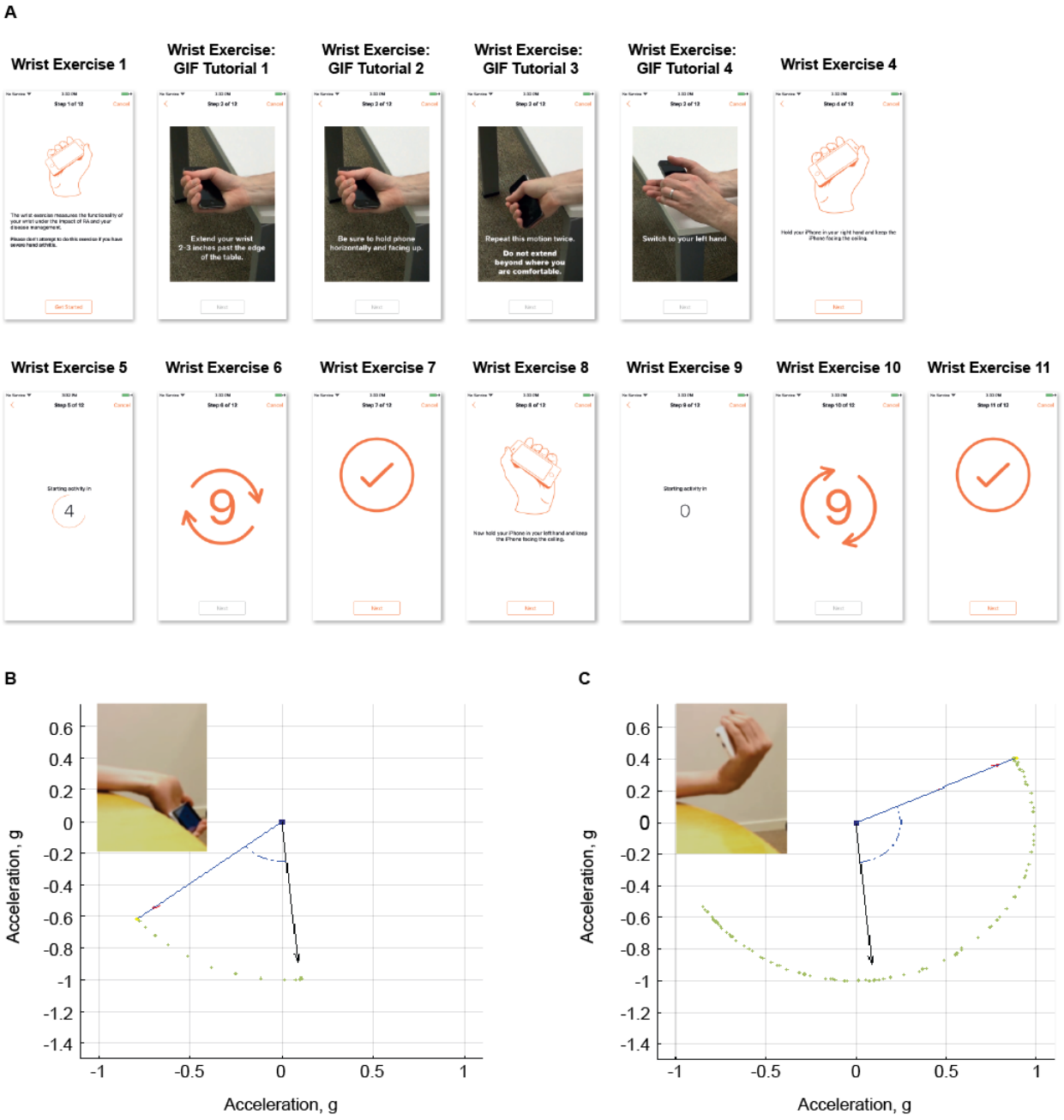
Each week, participants received a reminder via the app to complete a variety of tasks accessed via their personalized dashboard and designed to evaluate their disease status. Weekly reported outcome measures were the Rheumatoid Arthritis Severity Scale (patient global assessment) [21] and a series of semiquantitative scales, including a pain scale, morning stiffness scale, and mood scale (not reported here). In addition, participants were encouraged to complete a weekly survey to monitor their satisfaction with the app. At weeks 1, 4, 8, and 12, participants were encouraged to complete additional validated PRO measures: a health status survey (5-level version of the EuroQoL, 5 dimensions) [22], a physical function assessment (Health Assessment Questionnaire-Disability Index) [23] and a fatigue scale (Functional Assessment of Chronic Illness Therapy-Fatigue) [24].

An interactive joint-pain map was designed specifically for the app to record the number and severity of painful joints (from 55 prespecified joints). At weeks 1 and 8, participants were asked to score the pain in each joint (joints were presented in a body map) as 0 (no pain), 1 (mild pain), 2 (moderate pain), or 3 (severe pain). At the final assessment (week 12), a participant satisfaction survey was deployed.

Objective Study Task: Wrist Range of Movement Test

Given the shortage of validated techniques available to assess wrist ROM and the lack of a reference standard comparator, we developed a novel objective wrist ROM exercise for this study. The exercise was to be completed at weeks 1 and 12 (Multimedia Appendix 1). Briefly, participants were instructed to sit down and, in turn, to place their forearm at the edge of a standard-sized table, holding the smartphone in their hand, and to flex and extend their wrist joint to its maximum ROM (Figure 1, part A). We used raw sensor data from the smartphone gyroscope and accelerometers, captured by the app during this exercise, to assess the extent of flexo-extension ROM of each wrist joint as an objective measure of disease activity. We developed mathematical algorithms in Matlab 2016 (The Mathworks Inc) to convert the raw sensor data into ROM data. Figure 1 parts B and C illustrate the process for ROM extraction based on the phone's orientation computed by the algorithm. Images from experimental data acquisition are also displayed for reference.

Figure 1. Wrist range of movement (ROM) exercise. (A) Instructions for the wrist ROM exercise provided to participants via the app. (B, C) Wrist ROM preliminary validation based on experimental test data. (B) Relative change in phone orientation (blue arrow) at full wrist extension with respect to initial orientation (black arrow). (C) Relative change in phone orientation at full wrist flexion. In both examples, the video frame from the task performance at the moment of measurement is displayed as a reference for visual inspection. Green dots correspond to phone orientation at previous measurements.



Data Handling and Analysis

Data collected from the PARADE app were stored in secure 21CP11-compliant servers within the Medidata clinical research platform. Data were converted into JavaScript Object Notation files. Completed questionnaire data were subsequently converted into statistical analysis system datasets. All study data were anonymized, and no personally identifying information was obtained as part of the study data. The study was primarily a feasibility study of the Apple ResearchKit app, and the data analyses focused on descriptive information. We included data from participants who discontinued the study when analyzable. No missing data imputation was performed. As this was a pilot study to investigate the feasibility of using a smartphone app to conduct an observational study, we anticipated no specific results.

Results

Study Enrollment and Data Collection

We conducted the PARADE real-world observational study in the United States between July and November 2016. Within 1 month of launch, 1170 downloads of the PARADE app were completed, the majority in response to Facebook advertisements (1018/1170, 87.01%). Of these, 428 proceeded to consent; however, 29 individuals consented outside the 1-month window and therefore we did not include their data in the evaluation.

Figure 2. Patient Rheumatoid Arthritis Data From the Real World (PARADE) app study recruitment. *Defined as those completing all demographic questions.

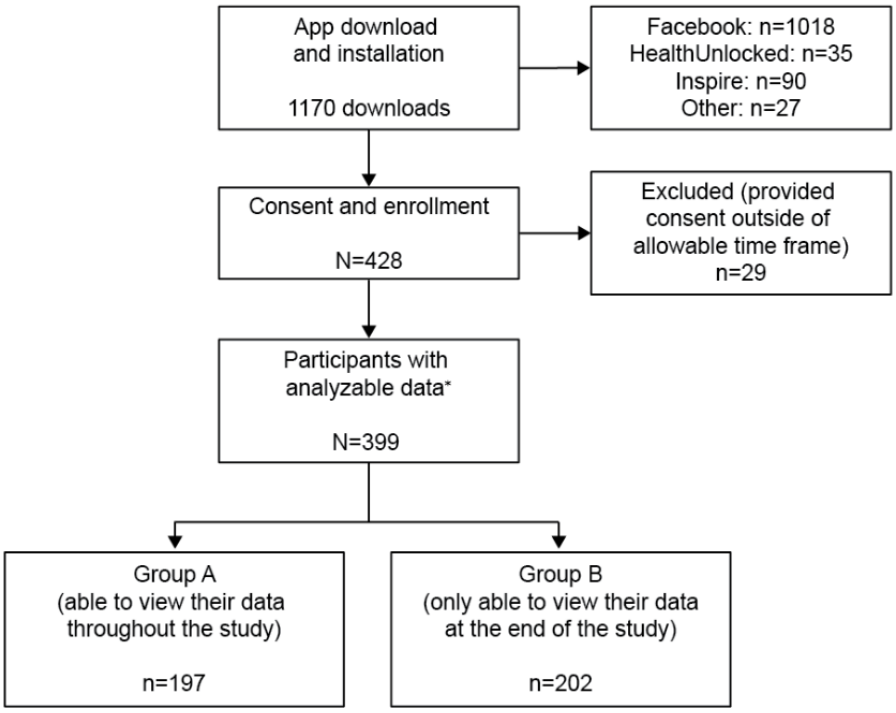


Table 1. Demographics and baseline clinical characteristics.

Characteristics	Group A (n=197)	Group B (n=202)	Total (N=399)
Female, n (%)	158 (80.2)	164 (81.2)	322 (80.7)
Age (years), mean (SD)	49.2 (12.48)	46.8 (12.00)	47.9 (12.28)
Ethnicity, n (%)			
White	154 (78.2)	168 (83.2)	322 (80.7)
African American	5 (2.5)	11 (5.4)	16 (4.0)
Hispanic	27 (13.7)	13 (6.4)	40 (10.0)
Asian	7 (3.6)	4 (2.0)	11 (2.8)
Other	4 (2.0)	6 (3.0)	10 (2.5)
Body mass index (kg/m^2), mean (SD)	29.3 (7.03)	29.7 (7.55)	29.5 (7.29)
Education, n (%)			
Middle school or below	4 (2.0)	1 (0.5)	5 (1.3)
High school	44 (22.3)	42 (20.8)	86 (21.6)
College	91 (46.2)	111 (55.0)	202 (50.6)
Graduate school	58 (29.4)	48 (23.8)	106 (26.6)
Smoking history, n (%)			
Current	18 (9.1)	19 (9.4)	37 (9.3)
Previous	61 (31.0)	60 (29.7)	121 (30.3)
Never	118 (59.9)	123 (60.9)	241 (60.4)
Duration (years) since diagnosis, n (%)			
<2	60 (30.5)	62 (30.7)	122 (30.6)
2-5	44 (22.3)	47 (23.3)	91 (22.8)
5-10	45 (22.8)	46 (22.8)	91 (22.8)
>10	48 (24.4)	47 (23.3)	95 (23.8)

The remaining 399 (34.10%) of the 1170 participants enrolled, consented, and provided complete demographic data, and hence we considered them to have contributed analyzable data (Figure 2).

Table 1 lists participant demographics and clinical characteristics. Table 2 lists the participants' current medications. The study population was predominately female (322/399, 80.7%) and white (322/399, 80.7%) with a mean age of 47.9 (SD 12.28) years; 77.2% (308/399) of participants were educated to college or graduate school level (Table 1). Disease duration was less than 2 years in 30.6% (122/399) of participants, between 2 and 5 years in 22.8% (91/399), between 5 and 10 years in 22.8% (91/399), and over 10 years in 23.8% (95/399). Differences between group A and group B were minimal, suggesting that the app-programmed randomization worked effectively. Patients were enrolled from a wide geographic distribution across the United States (Figure 3 [25]). Consistent with our geotargeting approach, California, Texas, Pennsylvania, New York, and Florida were among the states with the highest number of downloads.

Subjective Study Tasks

Joint pain was the most frequently reported RA symptom to bother study participants (344/393, 87.5%). Consistent with RA symptoms, painkillers, nonsteroidal anti-inflammatory drugs, corticosteroids, and methotrexate were the most commonly used medications (Table 2). Participant self-reported joint pain from the joint-pain map assessment was completed by 83.7% (334/399) of participants at week 1. Responses included identification of painful joints and the severity of pain in each individual joint scored on a scale of 0 to 3. At week 1, the right wrist (195/336, 58.0%) and left knee (194/336, 57.7%) were the most frequently cited locations of mild, moderate, or severe joint pain (Figure 4). Joint pain was most commonly reported to be mild and was similar for the left and right joints.

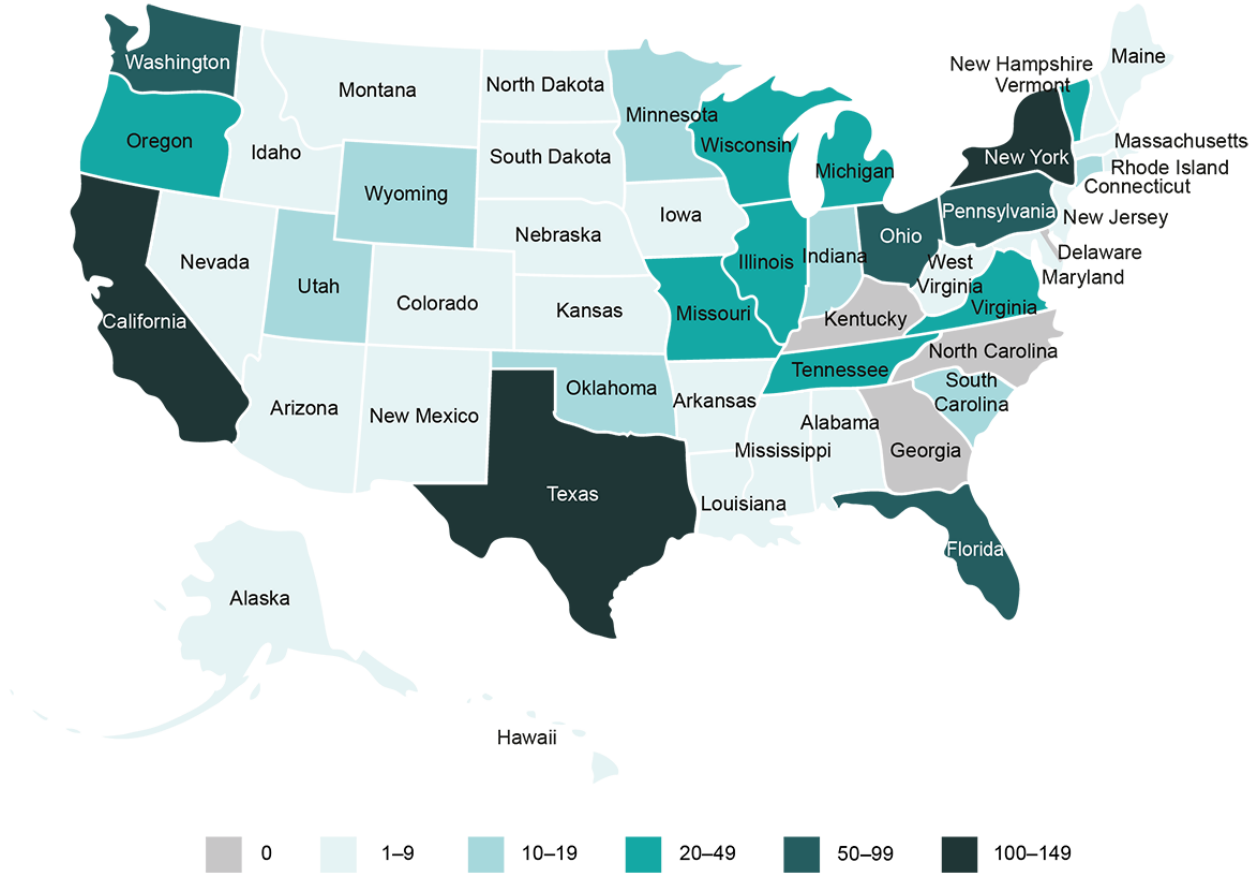
Data from PRO measures collected via the app showed no substantial changes in any of the scales throughout the study; however, the number of participants decreased throughout the study (Multimedia Appendix 2).

Table 2. Current rheumatoid arthritis medications.

Medication	Study population, n (%)		
	Group A (n=194)	Group B (n=194)	Total (n=388)
Painkillers	89 (45.9)	81 (41.8)	170 (43.8)
NSAIDs[a]	111 (57.2)	83 (42.8)	194 (50.0)
Corticosteroids	57 (29.4)	60 (30.9)	117 (30.2)
Methotrexate	76 (39.2)	79 (40.7)	155 (39.9)
Azathioprine	3 (1.5)	2 (1.0)	5 (1.3)
Auranofin	1 (0.5)	0	1 (0.3)
Chloroquine	1 (0.5)	0	1 (0.3)
Hydroxychloroquine	46 (23.7)	61 (31.4)	107 (27.6)
Leflunomide	13 (6.7)	17 (8.8)	30 (7.7)
Mycophenolate	3 (1.5)	0	3 (0.8)
Sulfasalazine	11 (5.7)	18 (9.3)	29 (7.5)
Abatacept	14 (7.2)	15 (7.7)	29 (7.5)
Adalimumab	23 (11.9)	19 (19.8)	42 (10.8)
Certolizumab	7 (3.6)	3 (1.5)	10 (2.6)
Etanercept	14 (7.2)	20 (10.3)	34 (8.8)
Golimumab	4 (2.1)	7 (3.6)	11 (2.8)
Infliximab	6 (3.1)	10 (5.2)	16 (4.1)
Rituximab	4 (2.1)	7 (3.6)	11 (2.8)
Tocilizumab	7 (3.6)	6 (3.1)	13 (3.4)
Tofacitinib	9 (4.6)	9 (4.6)	18 (4.6)
Others	19 (9.8)	21 (10.8)	40 (10.3)

[a]NSAID: nonsteroidal anti-inflammatory drug.

Figure 3. Geographic distribution of participants within the United States.



Objective Study Task: Wrist Range of Movement Test

We developed a novel tool to objectively evaluate the participants' wrist joint ROM and tested it for the first time in this study. The novel wrist ROM task was carried out by 71.4% (285/399) of participants at week 1. To evaluate potential associations between this objective test and patient-reported joint pain, we compared the wrist ROM data with the wrist joint pain scores from the joint-pain map assessment of each participant. Severe patient-reported wrist pain appeared to be inversely linked with the wrist ROM measured by the app (Figure 5).

App Evaluation and Participant Retention

At the beginning of the study, we asked participants whether they would prefer to participate in a study conducted at a clinic or using a mobile app; 73.2% (292/399) expressed a preference for a mobile app, 3.0% (12/399) preferred a clinic-based study, and 13.8% (55/399) answered both, with the remainder having no preference.

At week 2, 162 of 399 (40.6%) participants completed at least one study assessment; this decreased to 45 of 399 (11.3%) participants at week 12. We did not collect reasons for attrition. The percentage of participants remaining in the study was slightly greater among those who had daily access to their data than among those who did not (26/197, 13.2% vs 19/202, 9.4%, respectively; Figure 6); however, the number of participants remaining in both groups was low. We did not statistically compare retention rates due to the high rate of attrition.

Figure 4. Joint pain map. Percentage of patients reporting any pain in each of 55 joints at week 1 (n=336).

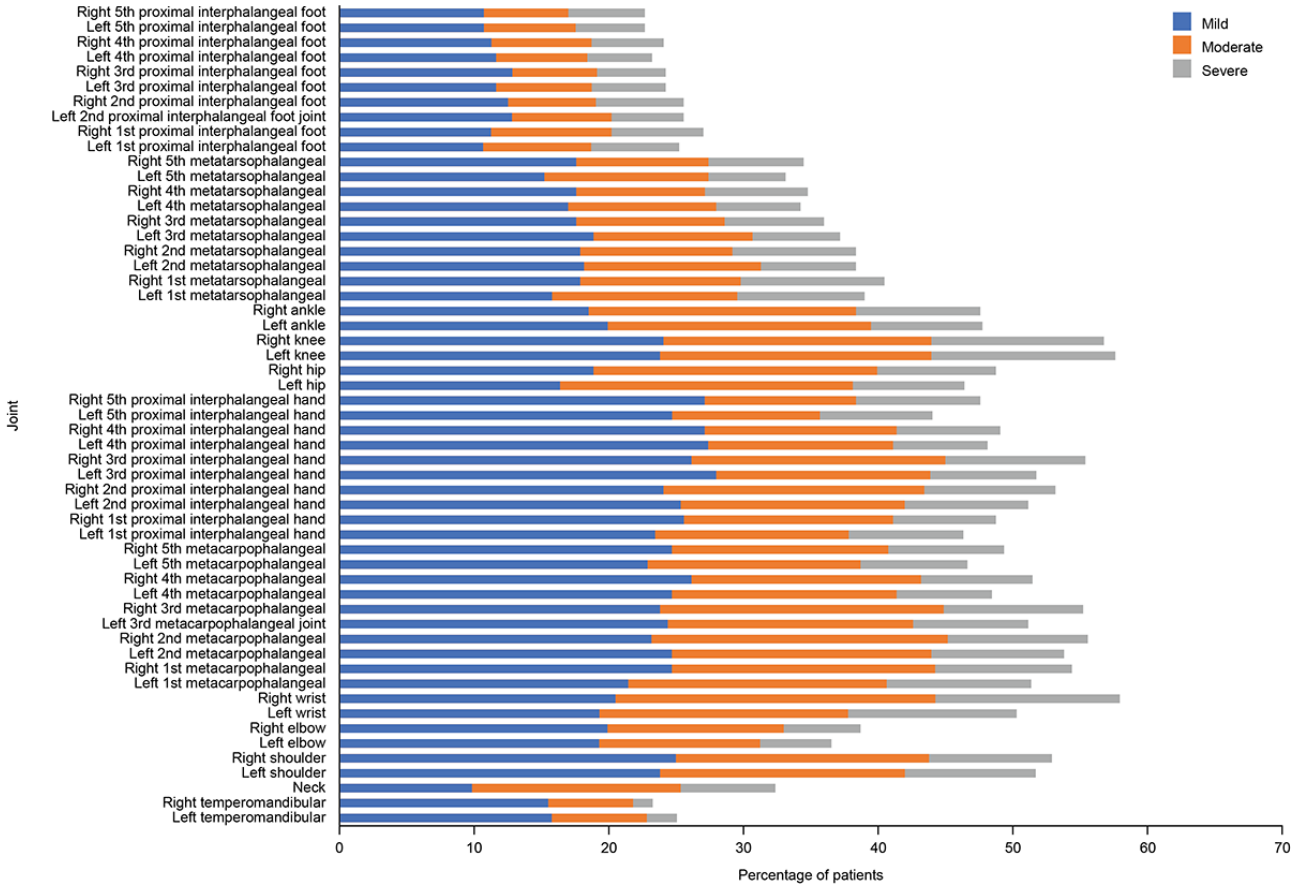


Figure 5. Association between patient wrist range of movement and reported level of wrist pain from the joint pain map assessment at week 1. Boxes represent the upper and lower quartiles; the line inside each box represents the median; the whiskers extending vertically from the boxes represent the range.

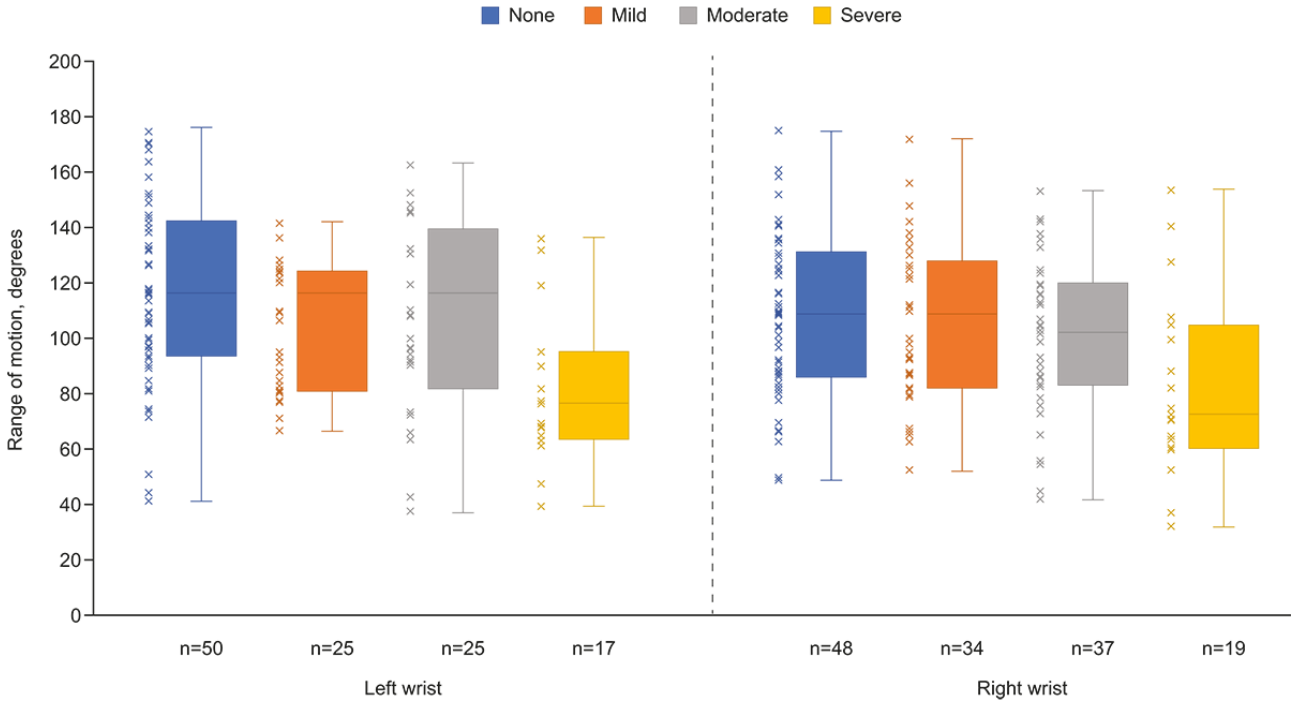
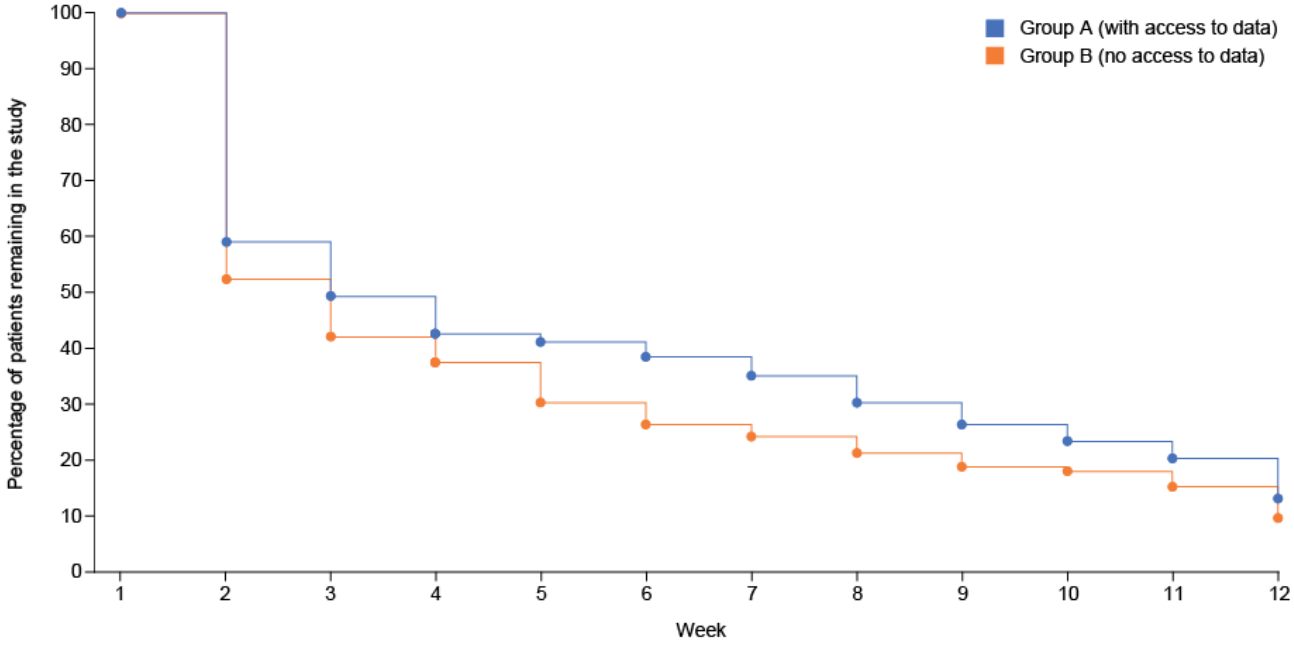


Figure 6. Proportion of patients retained within the study over 12 weeks.



Discussion

Principal Findings

The PARADE study was, to our knowledge, the first industry-sponsored study in which patients with RA could self-recruit, consent, enroll, and report data entirely via their iPhone using a ResearchKit app. This study demonstrated the feasibility of using smartphones to conduct a real-world study. In particular, the enrollment approach was successful in obtaining participants from a wide geographic distribution across the United States, as well as a wide ethnic diversity and a demography representative of patients with RA. Within 30 days, we exceeded our enrollment target of 300, demonstrating that the use of digital platforms to reach a large RA population can result in rapid study enrollment. The use of a smartphone was well received by the participants, with 73.2% (292/399) reporting a preference for participating in a mobile app–based study over a site-based study; however, this finding may not be representative of the general RA population.

To reduce any inherent enrollment bias and to ensure the validity of future studies using smartphones for data collection, it is particularly important to ensure that a representative sample of patients can access and enroll in the study. The PARADE study population was comparable with an RA population from a study that used more traditional methods of data collection, the Consortium of Rheumatology Researchers of North America (CORRONA) RA registry, believed to be representative of the US RA population [25]. Similarly, high proportions of women participated, consistent with the higher prevalence of RA in women; however, we found that the PARADE study population were younger and were more educated than those in the CORRONA registry (Table 3). Presumably this reflects the increasing likelihood of a younger and more highly educated demographic owning a smartphone [26]. Participants in the PARADE study had a greater ethnic diversity (19% other ethnicities) compared with those represented in the CORRONA registry (11%), suggesting that using downloadable smartphone apps to engage patients may be effective in recruiting a broadly representative ethnic population but may favor participation by individuals with higher education and younger age.

The app enabled collection of a range of RA-related data, including medications, symptoms, quality of life, joint-pain map data, and wrist ROM measures. The app has the potential for data collection at a higher frequency (eg, multiple points per day) than would be logistically possible in standard clinical studies; therefore, it can provide a more holistic view of disease exacerbation and remission. Furthermore, a previous study demonstrated that capture of Disease Activity Score in 28 joints data using a dedicated smartphone app correlated well with monthly clinical assessments of RA disease activity [15].

Table 3. Demographic profile of Patient Rheumatoid Arthritis Data From the Real World (PARADE) participants compared with representative data from the Consortium of Rheumatology Researchers of North America (CORRONA) registry of US patients with rheumatoid arthritis [25].

Demographic	PARADE	CORRONA
Age (years), mean (SD)	47.9 (12.3)	58.9 (13.4)
Female, n (%)	322 (81)	19,242 (77)
White, n (%)	322 (81)	22,240 (89)
Other ethnicities, n (%)	77 (19)	2749 (11)
College/graduate school educated, n (%)	308 (77)	13,744 (55)

Whereas other studies have reported issues with storage and transmission of data files from the phone, due to file size [14], we encountered no such challenges or difficulties in this study.

The design of studies in which patients self-report via electronic interfaces has the potential to revolutionize how clinical research is conducted [27] but is limited by a relative lack of simple and validated objective measures that can be captured electronically. This has led to a predominance of subjective assessments, typically patient questionnaires or PRO measures. Some studies have observed that such self-reporting can be prone to bias, with patients over- or underestimating the true situation [12]. Development of methods that can provide objective as well as subjective data are needed to improve the breadth and quality of results. However, some previous attempts to obtain objective measures remotely (eg, by connecting a hand dynamometer to a smartphone) have required additional instrumentation [28]. We explored the use of a smartphone app to capture objective data directly relevant to disease activity via the smartphone-embedded sensors to record ROM in the wrist. Validated tools to evaluate wrist ROM are lacking, and there is no reference standard comparator available. Previous work, however, has demonstrated the reliability of smartphone apps for the goniometric evaluation of joint ROM [29,30]. In our study, we saw a link between participants' subjective assessment of severe wrist pain and functional assessment of wrist ROM, suggesting that combining the use of questionnaires and sensor-based recordings from the smartphone may provide a valuable combination to monitor and quantify patient symptoms and disease impacts. Further research is required to validate the extent to which wrist ROM correlates with RA management and remission.

One common trend observed with the use of smartphones in clinical research is that, while engagement may be initially high, the rate of attrition is also high [31,32]. Although only 41% of our participants provided data at week 2, overall, retention was slightly better among participants who could view their data throughout the study than among those who could view it only at the final assessment (88/197, 44.7% vs 74/202, 36.6%, respectively), although we did not statistically evaluate these results and did not monitor whether participants actually accessed their data. This is in line with the findings of other studies (eg, [14]) where the ability to access personalized data could act as an incentive for patients to continue engagement. For example, in a study using computer-based technology to support a weight-loss program, patients who frequently used smartphone technology to view their progress lost more weight than those who did not [31]. For personalized data to provide an incentive to boost participation, the data provided must be meaningful and valuable to the patient, for example, by tracking improvement or progress toward goals, or flagging potential issues. Previous studies have explored various forms of incentivization to boost engagement, most notably the ability to visualize a medical benefit and payment [32,33]. Other ideas include data sharing, participation in an online participant forum, and a strong initial understanding and belief in the objective of the study [14]. It is likely that a combination of different incentives, tailored to suit each specific population, may be required to obtain maximal engagement.

Limitations

Our study had several limitations. The proportion of participants completing the study was low, and few people completed the evaluation of the app. Given that there was no medical intervention or other variables introduced that might drive a change in PRO measures over time, we expected little change in data from the PROs over the course of the study. However, the high level of attrition precludes the possibility of longitudinal assessments. Differences in retention rates between groups should be considered with caution. The results from this study are descriptive only, and we did not collect data on whether those patients who were able to access their personal data actually did so. A separate study would be required to further investigate retention rates. The app relied on participants' self-motivation and accurate self-reporting with no way to authenticate the data. It is important to ensure that, in the design of future studies, data shared with patients must add real value, including clinical value, so that any smartphone data that patients share with their physicians will facilitate clinical decisions, not just exacerbate clinician information overload [33].

Conclusion

This study successfully demonstrated the feasibility of using a smartphone coupled with ResearchKit to obtain patient-reported data in RA from a real-world perspective. It reports the first use of the smartphone gyroscope to measure wrist joint ROM, which was linked with patient-reported joint pain. We created a bespoke algorithm to derive clinically meaningful information on wrist ROM from raw sensor data. Further details on the methodology and accuracy assessment may be presented in a separate publication. This may lead to development and validation of other novel objective end points using smartphone-integrated sensors and may lead to an expansion of the objective data that can be captured from electronic patient-reported clinical research.

Acknowledgments

This study (study number 205718) was sponsored by GSK. GSK designed the study; contributed to the collection, analysis, and interpretation of the data; and supported the authors in the development of the manuscript. GSK is committed to publicly disclosing the results of GSK - sponsored clinical research that evaluates GSK medicines and, as such, was involved in the decision to submit the manuscript for publication.

The authors wish to thank Katie White, PhD, and Jennie McLean, PhD, from Fishawack Indicia Ltd, UK, funded by GSK, who provided assistance with developing this manuscript. The authors also wish to acknowledge the contributions of Possible Mobile, who developed the PARADE app, and Medidata, who developed algorithms for handling activity data and provided a validated

secure database for data collection. The team would also like to thank Julian Jenkins and Rob Dicicco for their sponsorship of the project.

Conflicts of Interest

At the time of the study, all authors were employees of GlaxoSmithKline and held stock options; at the time of publication, PB is no longer with GlaxoSmithKline.

Multimedia Appendix 1

Screenshots from the PARADE app.

[PDF File (Adobe PDF File), 18MB - mhealth_v6i9e177_app1.pdf]

Multimedia Appendix 2

Mean patient-reported outcome responses.

[PDF File (Adobe PDF File), 162KB - mhealth_v6i9e177_app2.pdf]

References

1. DeLucca P, Raghavarao D. Effect of investigator bias on the significance level of the Wilcoxon rank-sum test. Biostatistics 2000 Mar;1(1):107-111. [doi: 10.1093/biostatistics/1.1.107] [Medline: 12933528]
2. Lee P. Why go ePRO? Eur Pharm Contractor 2006;Spring:72-77 [FREE Full text]
3. Den Hond E, Celis H, Vandenhoven G, O'Brien E, Staessen JA, THOP investigators. Determinants of white-coat syndrome assessed by ambulatory blood pressure or self-measured home blood pressure. Blood Press Monit 2003 Feb;8(1):37-40. [doi: 10.1097/01.mbp.0000057015.67622.10] [Medline: 12604935]
4. Statista. Number of mobile phone users worldwide from 2013 to 2019 (in billions). 2017. URL: https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide [accessed 2018-05-09] [WebCite Cache ID 6zHtL95YR]
5. Chan YY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Apr;35(4):354-362 [FREE Full text] [doi: 10.1038/nbt.3826] [Medline: 28288104]
6. Schatz BR. National surveys of population health: big data analytics for mobile health monitors. Big Data 2015;3(4):219-229 [FREE Full text] [doi: 10.1089/big.2015.0021] [Medline: 26858915]
7. Dorsey ER, Yvonne CY, McConnell MV, Shaw SY, Trister AD, Friend SH. The use of smartphones for health research. Acad Med 2017;92(2):157-160. [doi: 10.1097/ACM.0000000000001205] [Medline: 27119325]
8. Azevedo R, Bernardes M, Fonseca J, Lima A. Smartphone application for rheumatoid arthritis self-management: cross-sectional study revealed the usefulness, willingness to use and patients' needs. Rheumatol Int 2015;35(10):1675-1685. [doi: 10.1007/s00296-015-3270-9] [Medline: 25903352]
9. Piga M, Cangemi I, Mathieu A, Cauli A. Telemedicine for patients with rheumatic diseases: systematic review and proposal for research agenda. Semin Arthritis Rheum 2017;47(1):121-128. [doi: 10.1016/j.semarthrit.2017.03.014] [Medline: 28420491]
10. Jardine J, Fisher J, Carrick B. Apple's ResearchKit: smart data collection for the smartphone era? J R Soc Med 2015;108(8):294-296. [doi: 10.1177/0141076815600673] [Medline: 26268915]
11. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
12. McConnell MV, Shcherbina A, Pavlovic A, Homburger JR, Goldfeder RL, Waggot D, et al. Feasibility of obtaining measures of lifestyle from a smartphone app: the myHeart counts cardiovascular health study. JAMA Cardiol 2017;2(1):67-76. [doi: 10.1001/jamacardio.2016.4395] [Medline: 27973671]
13. Banderas B, Skup M, Shields AL, Mazar I, Ganguli A. Development of the Rheumatoid Arthritis Symptom Questionnaire (RASQ): a patient reported outcome scale for measuring symptoms of rheumatoid arthritis. Curr Med Res Opin 2017;33(9):1643-1651. [doi: 10.1080/03007995.2017.1338562] [Medline: 28574727]
14. Reade S, Spencer K, Sergeant JC, Sperrin M, Schultz DM, Ainsworth J, et al. Cloudy with a chance of pain: engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. JMIR Mhealth Uhealth 2017;5(3):e37 [FREE Full text] [doi: 10.2196/mhealth.6496] [Medline: 28341616]
15. Nishiguchi S, Ito H, Yamada M, Yoshitomi H, Furu M, Ito T, et al. Self-assessment of rheumatoid arthritis disease activity using a smartphone application. Development and 3-month feasibility study. Methods Inf Med 2016;55(1):65-69. [doi: 10.3414/ME14-01-0106] [Medline: 26391694]

16. Nishiguchi S, Ito H, Yamada M, Yoshitomi H, Furu M, Ito T, et al. Self-assessment tool of disease activity of rheumatoid arthritis by using a smartphone application. Telemed J E Health 2014;20(3):235-240. [doi: 10.1089/tmj.2013.0162] [Medline: 24404820]
17. Yamada M, Aoyama T, Mori S, Nishiguchi S, Okamoto K, Ito T, et al. Objective assessment of abnormal gait in patients with rheumatoid arthritis using a smartphone. Rheumatol Int 2012;32(12):3869-3874. [doi: 10.1007/s00296-011-2283-2] [Medline: 22193221]
18. Apple Inc. ResearchKit. San Francisco, CA: GitHub, Inc; 2018. URL: https://github.com/researchkit/researchkit [accessed 2018-08-22] [WebCite Cache ID 71rq6I5XT]
19. Arthritis Foundation. Atlanta, GA: Arthritis Foundation; 2018. URL: https://www.arthritis.org/ [accessed 2018-08-22] [WebCite Cache ID 71rr0mHkI]
20. Global Health Living Foundation. GHLF creakyjoints. Upper Nyack, NY: CreakyJoints; 2017. URL: https://creakyjoints.org/ [accessed 2018-08-22] [WebCite Cache ID 71rrGg7HD]
21. Nikiphorou E, Radner H, Chatzidionysiou K, Desthieux C, Zabalan C, van Eijk-Hustings Y, et al. Patient global assessment in measuring disease activity in rheumatoid arthritis: a review of the literature. Arthritis Res Ther 2016;18(1):251 [FREE Full text] [doi: 10.1186/s13075-016-1151-6] [Medline: 27793211]
22. Jolles BM, Bogoch ER. Quality of life after TKA for patients with juvenile rheumatoid arthritis. Clin Orthop Relat Res 2008;466(1):167-178 [FREE Full text] [doi: 10.1007/s11999-007-0010-9] [Medline: 18196390]
23. Bruce B, Fries JF. The Health Assessment Questionnaire (HAQ). Clin Exp Rheumatol 2005;23(5 Suppl 39):S14-S18. [Medline: 16273780]
24. Hewlett S, Dures E, Almeida C. Measures of fatigue: Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAF MDQ), Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (BRAF NRS) for severity, effect, and coping, Chalder Fatigue Questionnaire (CFQ), Checklist Individual Strength (CIS20R and CIS8R), Fatigue Severity Scale (FSS), Functional Assessment Chronic Illness Therapy (Fatigue) (FACIT-F), Multi-Dimensional Assessment of Fatigue (MAF), Multi-Dimensional Fatigue Inventory (MFI), Pediatric Quality Of Life (PedsQL) Multi-Dimensional Fatigue Scale, Profile of Fatigue (ProF), Short Form 36 Vitality Subscale (SF-36 VT), and Visual Analog Scales (VAS). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S263-S286 [FREE Full text] [doi: 10.1002/acr.20579] [Medline: 22588750]
25. Solomon DH, Reed GW, Kremer JM, Curtis JR, Farkouh ME, Harrold LR, et al. Disease activity in rheumatoid arthritis and the risk of cardiovascular events. Arthritis Rheumatol 2015;67(6):1449-1455 [FREE Full text] [doi: 10.1002/art.39098] [Medline: 25776112]
26. Poushter J. Smartphones are common in advanced economies, but digital divides remain. Washington, DC: Pew Research Centre; 2017 Apr 21. URL: https://tinyurl.com/lj3vvyl/ [accessed 2018-05-09] [WebCite Cache ID 6zHoA95Lc]
27. 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;1375(1):3-18. [doi: 10.1111/nyas.13117] [Medline: 27384501]
28. Espinoza F, Le Blay P, Coulon D, Lieu S, Munro J, Jorgensen C, et al. Handgrip strength measured by a dynamometer connected to a smartphone: a new applied health technology solution for the self-assessment of rheumatoid arthritis disease activity. Rheumatology (Oxford) 2016;55(5):897-901. [doi: 10.1093/rheumatology/kew006] [Medline: 26867731]
29. Santos C, Pauchard N, Guilloteau A. Reliability assessment of measuring active wrist pronation and supination range of motion with a smartphone. Hand Surg Rehabil 2017;36(5):338-345. [doi: 10.1016/j.hansur.2017.06.005] [Medline: 28754335]
30. Shin SH, Ro du DH, Lee O, Oh JH, Kim SH. Within-day reliability of shoulder range of motion measurement with a smartphone. Man Ther 2012;17(4):298-304. [doi: 10.1016/j.math.2012.02.010] [Medline: 22421186]
31. Anton SD, LeBlanc E, Allen HR, Karabetian C, Sacks F, Bray G, et al. Use of a computerized tracking system to monitor and provide feedback on dietary goals for calorie-restricted diets: the POUNDS LOST study. J Diabetes Sci Technol 2012;6(5):1216-1225 [FREE Full text] [Medline: 23063049]
32. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
33. Say P, Stein D, Ancker J, Hsieh C, Pollak J, Estrin D. Smartphone data in rheumatoid arthritis - what do rheumatologists want? AMIA Annu Symp Proc 2015;2015:1130-1139 [FREE Full text] [Medline: 26958252]

Abbreviations

21CP11: Title 21 of the Code of Federal Regulations Part 11
CORRONA: Consortium of Rheumatology Researchers of North America
GSK: GlaxoSmithKline
PARADE: Patient Rheumatoid Arthritis Data From the Real World
PRO: patient-reported outcome
RA: rheumatoid arthritis
ROM: range of movement

Edited by G Eysenbach; submitted 14.12.17; peer-reviewed by K Okamoto, A Silva Almodovar; comments to author 25.01.18; revised version received 09.05.18; accepted 10.07.18; published 13.09.18.

Please cite as:
Crouthamel M, Quattrocchi E, Watts S, Wang S, Berry P, Garcia-Gancedo L, Hamy V, Williams RE
Using a ResearchKit Smartphone App to Collect Rheumatoid Arthritis Symptoms From Real-World Participants: Feasibility Study
JMIR Mhealth Uhealth 2018;6(9):e177
URL: http://mhealth.jmir.org/2018/9/e177/
doi:10.2196/mhealth.9656
PMID:30213779

Original Paper

Exploration of Users' Perspectives and Needs and Design of a Type 1 Diabetes Management Mobile App: Mixed-Methods Study

Yiyu Zhang[1,2,3*], MMed; Xia Li[1,2,3*], MD; Shuoming Luo[1,2,3], MD; Chaoyuan Liu[4], MMed; Fang Liu[1], RN, MN; Zhiguang Zhou[1,2,3], MD

[1]Department of Metabolism and Endocrinology, The Second Xiangya Hospital, Central South University, Changsha, China

[2]Key Laboratory of Diabetes Immunology, Ministry of Education, Changsha, China

[3]National Clinical Research Center for Metabolic Diseases, Changsha, China

[4]Department of Oncology, The Second Xiangya Hospital, Central South University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Zhiguang Zhou, MD
Department of Metabolism and Endocrinology
The Second Xiangya Hospital
Central South University
No 139, Renmin Road
Changsha, 410011
China
Phone: 86 073185292154
Fax: 86 073185367220
Email: zhouzhiguang@csu.edu.cn

Abstract

Background: With the popularity of mobile phones, mobile apps have great potential for the management of diabetes, but the effectiveness of current diabetes apps for type 1 diabetes mellitus (T1DM) is poor. No study has explored the reasons for this deficiency from the users' perspective.

Objective: The aims of this study were to explore the perspectives and needs of T1DM patients and diabetes experts concerning a diabetes app and to design a new T1DM management mobile app.

Methods: A mixed-methods design combining quantitative surveys and qualitative interviews was used to explore users' needs and perspectives. Experts were surveyed at 2 diabetes conferences using paper questionnaires. T1DM patients were surveyed using Sojump (Changsha ran Xing InfoTech Ltd) on a network. We conducted semistructured, in-depth interviews with adult T1DM patients or parents of child patients who had ever used diabetes apps. The interviews were audio-recorded, transcribed, and coded for theme identification.

Results: The expert response rate was 63.5% (127/200). The respondents thought that the reasons for app invalidity were that patients did not continue using the app (76.4%, 97/127), little guidance was received from health care professionals (HCPs; 73.2%, 93/127), diabetes education knowledge was unsystematic (52.8%, 67/127), and the app functions were incomplete (44.1%, 56/127). A total of 245 T1DM patient questionnaires were collected, of which 21.2% (52/245) of the respondents had used diabetes apps. The reasons for their reluctance to use an app were limited time (39%, 20/52), complicated operations (25%, 13/52), uselessness (25%, 13/52), and cost (25%, 13/52). Both the experts and patients thought that the most important functions of the app were patient-doctor communication and the availability of a diabetes diary. Two themes that were useful for app design were identified from the interviews: (1) problems with patients' diabetes self-management and (2) problems with current apps. In addition, needs and suggestions for a diabetes app were obtained. Patient-doctor communication, diabetes diary, diabetes education, and peer support were all considered important by the patients, which informed the development of a prototype multifunctional app.

Conclusions: Patient-doctor communication is the most important function of a diabetes app. Apps should be integrated with HCPs rather than stand-alone. We advocate that doctors follow up with their patients using a diabetes app. Our user-centered method explored comprehensively and deeply why the effectiveness of current diabetes apps for T1DM was poor and what T1DM patients needed for a diabetes app and provided meaningful guidance for app design.

(JMIR Mhealth Uhealth 2018;6(9):e11400) doi:10.2196/11400

KEYWORDS

diabetes mellitus, type 1; mobile applications; qualitative research; surveys and questionnaires

Introduction

Background

The incidence of type 1 diabetes mellitus (T1DM) has been increasing worldwide [1,2]. An estimated 13,000 new T1DM cases occur every year in China [3]. Failure of islet beta-cell function occurs in the early stage of T1DM [4]; thus, controlling blood glucose is difficult. Despite the development of therapeutic drugs and treatment techniques, the blood sugar of T1DM patients is still poorly controlled [5]. The 3C study in China showed that the average glycosylated hemoglobin (HbA_{1c}) of T1DM patients in Beijing and Shantou was 8.5% [6], far higher than the guideline recommendations [7], and a clear gap existed between China and developed countries. Poor glycemic control can cause various complications [8] and place heavy financial burdens on the country and patients.

For T1DM patients, self-management ability is very important [9]. Increasing communication with doctors and strengthening blood sugar monitoring are beneficial for glycemic control [10,11]. The following challenges are present in outpatient clinics: inconvenience because of time and space limitations; limited ability to gain diabetes self-management knowledge in a short period of time; and compliance with a diabetes diary is often poor, which prevents doctors from providing effective treatment guidance [12]. Due to the imbalance of medical resources in China [13], patients flock to tertiary hospitals in large cities to seek medical resources, but they receive an outpatient consultation lasting just a few minutes. Continuity of care is a challenge in traditional outpatient settings as T1DM patients usually do not return to the same hospital or at regular intervals [6]. Mobile apps can record, transmit, and receive feedback anytime and anywhere. Mobile phones have been integrated into individuals' personal lives because of their popularity [14]. Thus, an app has great potential for the management of diabetes [15], especially for patients from remote areas.

However, people do not continue using health apps because of data entry burden and loss of interest [16]. Pernille's study revealed that the use of a diabetes self-management app by young T1DM patients decreased gradually after the first few weeks [17]. The majority of diabetes apps contain only a few functions [18]. The number of functions offered by apps influences HbA_{1c} levels [19]. Diabetes apps achieve different results in terms of glycemic control [15,20]; the effects in T1DM patients are poor [21].

App development must be closely integrated with clinical guidelines, and they must work closely with health care professionals (HCPs) and patients [22]. Most apps are developed by software engineers without medical backgrounds [21]. Thus, the developed apps are not well integrated with guidelines and clinical needs [21,23]. For example, despite the emphasis by diabetes guidelines for the need for ongoing patient education [24], very few studies used mobile apps that have education as a functionality [23]. Personalized education is an under-represented feature in diabetes mobile apps [25], and the role of HCPs is missing in most apps [15].

T1DM is different from type 2 diabetes mellitus (T2DM) in many aspects [26]. For example, T1DM patients are younger. They are insulin-dependent, whereas most T2DM patients do not require insulin treatment. Insulin dose and carbohydrate calculation and self-monitoring of blood sugar are more important for T1DM patients. However, although numerous diabetes apps have been developed, few are specific for T1DM [27]; thus, the developed apps might not be suitable for T1DM patients.

Few diabetes apps have been introduced with the methodology of their development [23]. Gaining a deep understanding of the perspectives of patients is important when developing a mobile app for their use [28,29]. Qualitative research methodology has become more recognized and valued in diabetes behavioral research. By exploring patients' motivations, perspectives, and expectations, this approach can answer questions that cannot be addressed using a quantitative study. A mixed-methods study can combine qualitative and quantitative results to provide a more comprehensive and deeper understanding of user perspectives [30].

Objectives

No study has explored the reasons for poor effects of current diabetes apps in T1DM patients from the users' perspectives. To improve glycemic control in Chinese T1DM patients, we used a mixed-methods study to explore users' perspectives and needs and cooperated with a software team to develop a mobile app for T1DM management.

Methods

Part 1: Questionnaire Survey

Questionnaire Design

An expert panel consisting of 3 diabetologists (YZ, SL, and XL) and a diabetes education nurse (FL) from our hospital designed the questionnaires according to the functions of current diabetes apps [18,21,25,31-33], the problems they encountered during clinical practice, and diabetes guidelines [26]. The questions were presented in a choice format. If responders did not agree with the listed options, they could select the option "other" and write their answers in the remarks column. The expert questions covered their use of and perspectives about diabetes apps. The patient questions covered their use of, perspectives about, and needs for diabetes apps; demographic information; and basic disease information. Before the questionnaires were administered, we performed pilot tests with 10 diabetologists in our hospital and 20 diabetes patients from our outpatient department.

Samples and Survey Methods

The expert questionnaires (Multimedia Appendix 1) were administered using a paper format at 2 national diabetes conferences held in October 2017 and December 2017, with a total of 200 diabetologists attending. From 23rd January to 1st March 2018, the T1DM patient questionnaires (Multimedia Appendix 2) were administered using the Web-based questionnaire tool Sojump on the WeChat network [34]. The questionnaire links were spread among the first author's WeChat friends circle and WeChat groups of diabetes patients. The questions were answered by adult patients or the parents of child patients. No compensation was given for participation in the study.

Data Analysis

Descriptive statistics were used to characterize the samples. Frequencies and percentages were used to describe categorical variables. Incomplete responses were included in the analysis.

Part 2: Qualitative Study

Data Collection

After administering the questionnaire surveys, semistructured one-on-one in-depth interviews were conducted by a diabetologist (YZ). T1DM patients who had previously used diabetes apps were contacted. First, we introduced the objective of the study to establish trust. Adult patients or the parents of child patients were invited for a one-on-one interview. The interview environments were quiet, and interruptions were minimized. An interview guideline (Multimedia Appendix 3) was created by the expert panel and covered questions about the patients' daily diabetes management behavior, problems with apps they had used in the past, and their needs and suggestions for an app. The questions were open-ended. Each interview lasted approximately 30 to 60 min. Data collection ended when data saturation was achieved [35]. All interviews were audio-recorded, and all participants gave written informed consent.

Data Analysis

The data analysis was ongoing during the data collection process to ensure data saturation. Records were transcribed verbatim by the interviewer (YZ) and were verified by the interviewees. Data analysis was managed using NVivo 11.0 (QSR International Pty Ltd). Using inductive thematic analysis [36], the transcripts were independently read and coded by 2 investigators (YZ and XL). Disagreements and emerging themes were discussed with the expert panel.

Part 3: App Prototype Design and Development

On the basis of the results of the questionnaires and interviews, the expert panel combined their clinical experiences and clinical guidelines [9,24,26] to determine the modules and contents of the app and held discussions with the software team at least once a week in the form of workshops. The software team developed the app iteratively using an agile software development methodology. During each workshop, the software team introduced their app design and the prototype developed in the last iteration. The expert panel operated the prototype and proposed some modifications and new requirements for the app according to their expertise. One patient was invited to share their user experience in each workshop. The workshop members discussed the layout, design, and contents of the app. Brainstorming was adopted in this process. The software team developed the app accordingly in the next iteration.

Ethical Approval

The study was approved by the ethics committee of the Second Xiangya Hospital, Central South University.

Results

Part 1: Questionnaire Survey

Expert Survey

Factors Influencing Experts' Use and Recommendation of Diabetes Management Apps

The response rate for the expert survey was 63.5% (127/200). Overall, 52.8% (67/127) of the experts had recommended diabetes apps to their patients. Figure 1 shows the factors influencing their recommendations for a diabetes app. A total of 34.6% (44/127) of the experts had used diabetes apps to manage diabetes patients. These experts thought that the biggest obstacle to their use of apps to manage diabetes patients was limited time (57.6%, 68/118; see Figure 2). A total of 57.5% (73/127) of the experts did not know whether using an app to manage patients was legal, 26.7% (34/127) thought that using an app for this purpose was legal, and 15.0% (19/127) thought that it was illegal.

Experts' Perceptions of Diabetes Management Apps

The experts' proposed reasons for app invalidity were that patients did not continue using them (76.4%, 97/127), patients received little guidance from HCPs (73.2%, 93/127), diabetes knowledge on the app was unsystematic (52.8%, 67/127), and the apps' functions were incomplete (44.1%, 56/127). The experts thought that the most important functions of an app were patient-doctor communication (42.4%, 53/125), the diabetes diary (39.2%, 49/125), diabetes education (10.4%, 13/125), and abnormal blood sugar reminders (6.4%, 8/125). Most experts did not recommend or were opposed to insulin calculators (62.0%, 75/121) because 78.2% (97/124) thought that these tools were dangerous or very dangerous. Overall, 82.5% (104/126) of the experts thought that the prospect for diabetes apps was good or very good.

Patient Survey

Factors Influencing Patients' Use of Diabetes App

A total of 245 T1DM patient questionnaires were collected. Table 1 shows the characteristics of the respondents. Overall, 61.2% (150/245) of the responders did not know about the existence of diabetes apps, and only 21.2% (52/245) had ever used diabetes apps. Only 8% (4/52) of the apps were recommended by HCPs. Most of the apps were recommended by patients (38%, 20/52) or selected randomly (37%, 19/52) because the respondents did not know which app was the best.

Figure 1. Factors influencing experts' recommendation of diabetes apps (n=127).

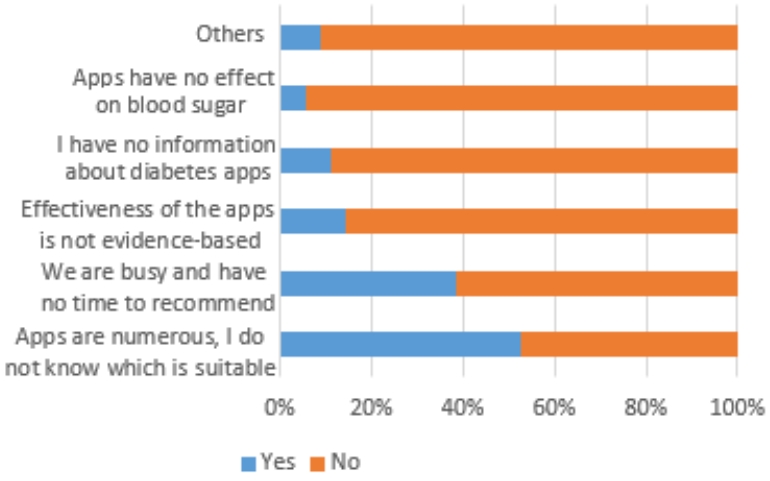


Figure 2. The biggest obstacles to experts' use of apps to manage diabetes patients (n=118).

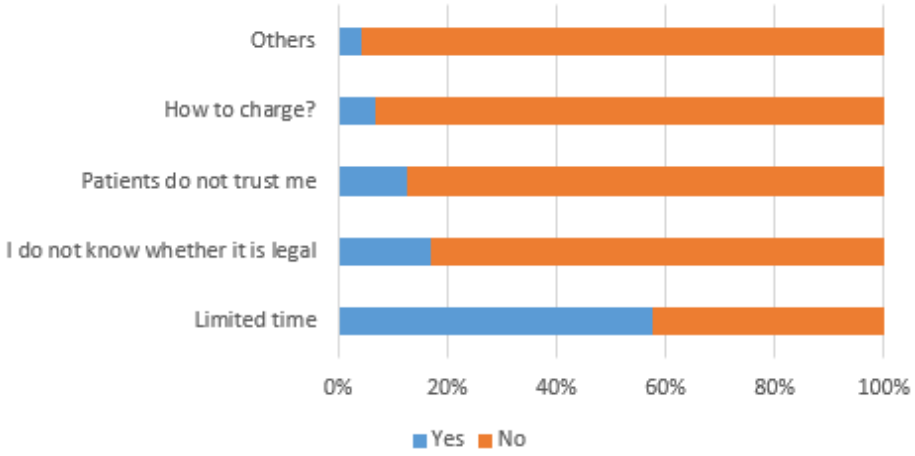


Table 1. Characteristics of patients responding to the surveys.

Characteristics	Total (N=245)	Adolescent (n=115)	Adults (n=130)
Gender, n (%)			
Male	98 (40.0)	49 (42.6)	49 (37.7)
Female	147 (60.0)	66 (57.4)	81 (62.3)
Age in years, median (IQR[a])	18 (11-30)	11 (8-14)	29 (23-35.3)
Disease duration in years, median (IQR)	3 (1-9)	2 (1-4)	5 (1.75-15)
Treatment type, n (%)			
Insulin pump	65 (26.5)	27 (23.5)	38 (29.2)
Insulin injection	280 (73.5)	88 (76.5)	92 (70.8)

[a]IQR: interquartile range.

The reasons for their reluctance to use an app were limited time (39%, 20/52), complicated operations (25%, 13/52), uselessness (25%, 13/52), and cost (25%, 13/52). The most common functions of their apps were diabetes knowledge (92%, 48/52) and blood sugar record (90%, 47/52; see Figure 3). A total of 70% (33/47) of the patients thought manual input of blood sugar was troublesome or a little troublesome. A total of 58% (30/52) of the apps could consult HCPs, but only 30% (9/30) of the patients had ever used this function.

Patients' Needs for a Diabetes App

The patients thought the most important functions of the apps were consulting HCPs (33.9%, 83/245), the diabetes diary (24.4%, 55/245), diabetes knowledge (12.7%, 31/245), the insulin calculator (11.8%, 29/245), abnormal blood sugar reminders (10.6%, 26/245), peer support (2.9%, 7/245), and blood sugar test reminders (1.2%, 3/245). Almost all patients thought the above functions were important or very important

(see Figure 4). A total of 65.3% (160/245) of the patients thought that they were in need or in great need (32.7%, 80/245) of a good app to manage their diabetes.

Part 2: Qualitative Study

Participants

The final sample consisted of 18 participants (12 adult patients and 6 parents of young patients; see Table 2).

Themes

Two themes including 10 subthemes that were helpful for our app design were identified.

Theme 1: Problems in Patients' Diabetes Self-Management Conduct

Diabetes self-management education (DSME), diet, exercise, and self-monitoring of blood sugar are 4 important parts of diabetes self-management. Understanding the problems with self-management helped refine the design of our app.

Diabetes Self-Management Education

Most patients did not receive DSME programs in the hospital. DSME in the hospital had many shortcomings, including inconvenience, reluctance of young people to go to the hospital, lack of individualization, and low efficiency. Compared with receiving DSME in the hospital, receiving information on mobile apps was preferable. The patients could select subjects that they were interested in, learn repeatedly, and learn when they had time. Additionally, the time and economic costs were lower.

Two patients stated the following:

> *From Monday to Friday, there is no time. Secondly, I think sometimes we will select contents to learn after we have mastered some knowledge. Because we have mastered some basic knowledge, if lectures are about such contents, we will not go to learn.* [P5, 30-year-old female]
>
> *Both are fine. But if I go to the hospital, I feel I have no time. Because if I learn on a mobile app, videos can be saved; I can learn when I have time. I think the app is better.* [P10, 24-year-old female]

Figure 3. Proportions of different functions of patients' diabetes apps (n=52). HCP: health care professional.

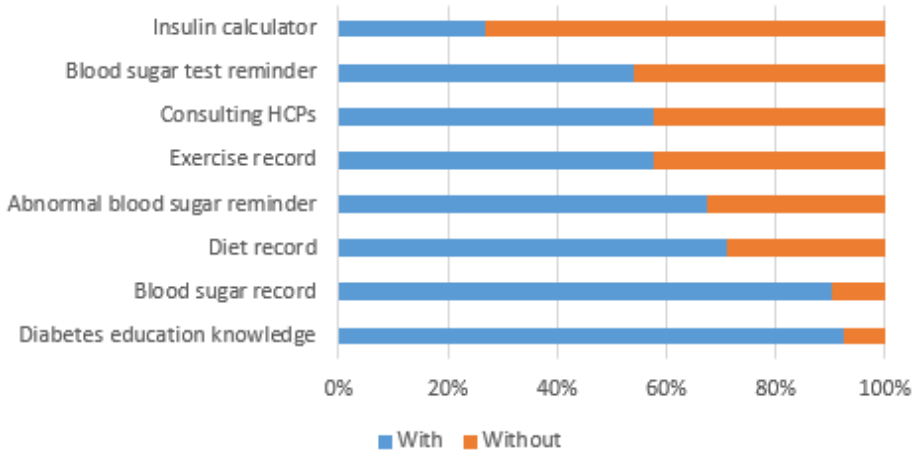


Figure 4. Usefulness of app functions reported as important by patients (n=245).

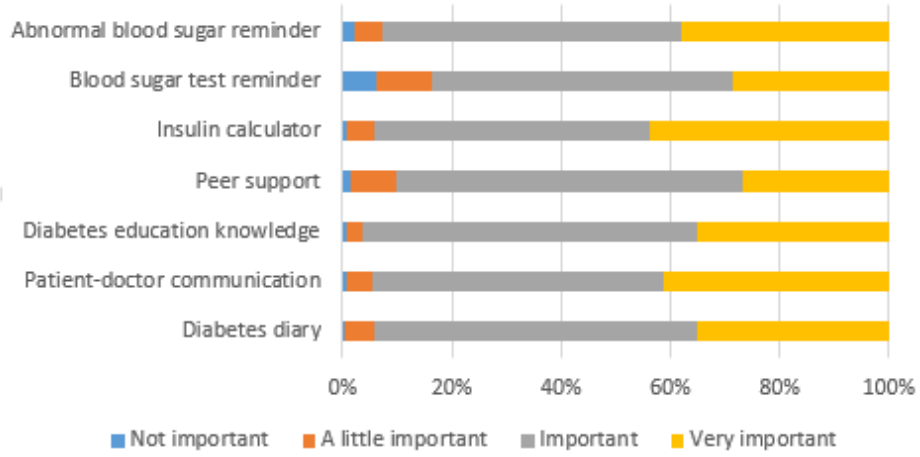


Table 2. Characteristics of the interviewees.

Characteristics	Adult patients (n=12)	Parents of young patients (n=6)
Gender, n (%)		
Male	1 (8)	3 (50)
Female	11 (92)	3 (50)
Patients' age in years, mean (range)	26.8 (20-33)	10.5 (6-16)
Patients' disease duration in years, mean (range)	4.0 (1-12)	3.3 (0-9)
Treatment type, n (%)		
Insulin pump	3 (25)	3 (50)
Insulin injection	9 (75)	3 (50)
Education, n (%)		
Postgraduate	1 (8)	—
University	8 (67)	4 (67)
High school	3 (25)	1 (17)
Junior middle school	—	1 (17)
Residence, n (%)		
Urban	9 (75)	5 (83)
Rural	3 (25)	1 (17)

Self-Monitoring of Blood Sugar

Space, time, economy, and pain were all factors that influenced blood sugar tests. Some patients forgot to test because they did not form a habit of testing their blood sugar, or they were doing other things. Some patients did not know when they needed to test, and some were not aware of the importance of testing. One patient stated the following:

> *It is not as important as insulin injection. If you don't inject insulin, your blood sugar will surely be high. But if you pay attention to your food, you have a sense of your blood sugar level, so you don't attach much importance to it...* [P11, 31-year-old female]

Diet

Some patients had incorrect diet conceptions. Calculating carbohydrates and calories is important for blood sugar control, but most patients do not perform these calculations for their daily diets. They thought that the calculation process was complicated and troublesome. One patient stated the following:

> *I don't know. At the beginning, they told me to calculate. It is complicated. In a WeChat group, some people told me how to calculate, and when I came to the nutrition department, they told me how to calculate. But after that I will say, I try to eat vegetables as much as possible.* [P15, 27-year-old female]

Exercise

Most patients knew the importance of exercise for glycemic control, but many of them lacked the time and will. Some patients selected the wrong time to exercise. Some were afraid to exercise because they were worried about hypoglycemia, as illustrated in the following quote:

> *Blood sugars fluctuate greatly. I dare not exercise. I'd rather have higher blood sugar. I'd rather give a bolus. I'm not willing to exercise.* [P12, 26-year-old male]

Theme 2: Problems With the Functions of Current Apps and Patients' Needs and Suggestions for a New App

Diabetes Diary

Although they thought a mobile diabetes diary was more convenient than a paper diary, most of them thought manual input was burdensome (see Table 3 for the problems with current diabetes app). The patients wanted glucose data to be transmitted to the apps automatically. Diet and exercise records were even more troublesome. Some of the patients thought that these types of records were useless and that their display was not as intuitive as that of a paper diary. Most of the patients only recorded blood sugar.

One patient stated the following:

> *If input manually, when you are outside, testing blood sugar is inconvenient, but you have to record...you will think it doesn't matter. They are just in the glucose meter. It's very burdensome. But if it can be transmitted to app automatically, it is convenient.* [P7, 33-year-old female]

Most patients reported that the greatest problem with diabetes diaries was the lack of feedback from HCPs. As the diary was useless for glycemic control, they did not continue to use the apps. They hoped to obtain feedback after recording and to have a doctor analyze their data.

Table 3. Problems with current diabetes apps and needs for a new app.

Modules of current app	Problems and needs
Patient-doctor communication	• Distrust • Responses are not timely • Inconvenient • Cost
Diabetes diary	• Burdensome • Lack of feedback • Display is not as intuitive as a paper diary • Food database is needed
Diabetes education knowledge	• Unsystematic • Unprofessional • Avoid irrelevant knowledge interference • Find materials of interest easily • Update in a timely manner • Interaction is needed • Tend to learn pop-up knowledge • Different learning habits
Peer-support	• Inconvenient • Avoid excessive information interference • Peer leader is needed • Privacy protection
Psychological module	• Most apps lack this module
Electronic health records	• Access to hospital electronic medical records

One parent stated the following:

> *It is meaningless if you record there. But if these data, I think, let me think, if after these data are submitted, an online doctor analyzes them for you, I think people will like it.* [P6, mother of a 10-year-old patient]

Patient-Doctor Communication

Some diabetes apps had a function for consulting HCPs. However, most users did not consult HCPs using the app because they did not trust unfamiliar doctors. App communication in the form of typing words was inconvenient, and the communication efficiency was low. Consultations needed to be charged, feedback was not timely, and the consultation effect was low. These factors hampered consultations with doctors by the patients using the apps.

Two patients stated the following:

> *I tried once to make an appointment with a doctor in the weltang app. But for his few minutes he needed to charge, so I exited. An unfamiliar doctor, you consult him, but you need to pay. Maybe you have a sense of...* [P11, 31-year-old female]

> *I consulted once. Because the doctor was busy, the response was not timely. Describing our condition by typing words, maybe it is not so good to meet the needs of patients. After all, they are not our familiar doctors, they don't know our condition. I hope to communicate directly with the doctor.* [P4, 30-year-old female]

Most patients want to consult doctors on the app. However, doctors approached via the internet are not familiar with the patients' conditions. The patients wanted their outpatient doctors to continue to follow them up. Doctors from primary hospitals lack experience with managing T1DM. Moreover, the patients do not trust doctors from primary hospitals and only trust doctors from large tertiary hospitals. One patient stated the following:

> *Yes, unless he is your outpatient doctor. I think it can be set on that app, for example, you consult your outpatient doctor and have good effects.* [P7, 33-year-old female]

One parent stated the following:

> *There are only two type 1 diabetes patients in our county. When I went to the county hospital to ask the doctors, they never heard of this disease...* [P8, father of a 12-year-old patient]

Diabetes Education

Most patients hoped to gain diabetes knowledge on the app. They were most concerned about the latest progress in diabetes, knowledge about complications, nutrition, exercise, and insulin dose calculation. Some patients thought diabetes knowledge on apps was unsystematic and unprofessional. Patients did not know whether the diabetes knowledge was accurate. Patients hoped for the inclusion of authoritative and practical knowledge. One patient stated the following:

> *It's too miscellaneous. You can't tell which is right. Because now most of us get information through the*

> *internet, I think accuracy is important for information about disease.* [P7, 33-year-old female]

The patients liked different modes of educational materials. Some liked to watch videos, whereas others liked to read articles. They hoped diabetes knowledge could be classified according to categories and that knowledge about T1DM could be separated from that about T2DM, which would enable the patients to learn pertinent information and avoid excessive information interference, as illustrated in the following quote:

> *Because I'm type 1, so it is more targeted...we are all type 1. It is not mixed with type 2. Because other apps were mixed with type 2 diabetes, gestational diabetes, and so on, it's really very chaotic. There is lots of information. You need to screen which is useful, which is useless.* [P7, 33-year-old female]

Peer Support

Almost all the patients wanted to communicate with similar diabetes patients. Some patients said they had no way to find such patients after the onset of diabetes. They thought peer support could help them exchange glycemic control strategies and emotional experiences. Some of them even thought that patient experiences were more important than consulting doctors because patient experiences were person-specific and practical, as illustrated in the following quote:

> *There are a lot of these patients in our group. Their disease durations are many years. Their own experiences may be better than that of doctors because they are more practical. What the doctor said is theoretical. Some diabetic friends, based on their own experiences, may be more practical.* [P6, mother of a 10-year-old patient]

Many patients believed that having a peer leader was very important. Patients with a long disease duration and rich experience in glycemic control can act as peer leaders. Peer leaders can play a leading, interactive, and cohesive role and drive the atmosphere of a peer support module, as illustrated in the following quote:

> *For example, the key is, like a family, there is no backbone. There is no person with comprehensive knowledge. His knowledge is comprehensive; whatever questions you put forward, he can help you to solve it. Like that teacher, his prestige is high. He is willing to listen to others, and then he is willing to help others.* [P8, father of a 12-year-old patient]

The patients hoped to have different types of peer support modes. However, all peer communications in the diabetes apps took place in the form of forums. Most patients thought that communicating in that way was inconvenient and that responses were not timely. Very few patients chatted in the diabetes apps, as illustrated in the following quote:

> *[WeChat] Group chat is timely. Questions you ask can be answered immediately. But on the forum, you will wait a few days. I think feedback in group chat is more timely. It is better. I don't use forums now...* [P5, 30-year-old female]

Psychological Module

Mental health specialists are recommended as a part of diabetes management by diabetes guidelines. Almost all patients said diabetes brought negative emotions to them to varying degrees. Some patients indicated that the apps had no psychological module, and they hoped we could pay attention to their mental health, as illustrated in the following quote:

> *Another is psychological, a psychological module for patients. I have lots of apps on my mobile phone. Almost all are about knowledge, how to control blood sugar. Attention to children's mental health, a psychological module doesn't exist.* [P8, father of a 12-year-old patient]

Electronic Medical Records

The patients hoped to access their hospital electronic medical records (EMRs) through the app (eg, to view their test results and their diagnostic and treatment records and to register for outpatient visits). This possibility would be convenient, allow them to build health records in the app, and motivate them to continue using the app. One patient stated the following:

> *Connect to hospital health records systems directly. You can register for outpatient visits, and whenever you have problems, you can consult your outpatient doctor. Maybe these can be included.* [P5, 30-year-old female]

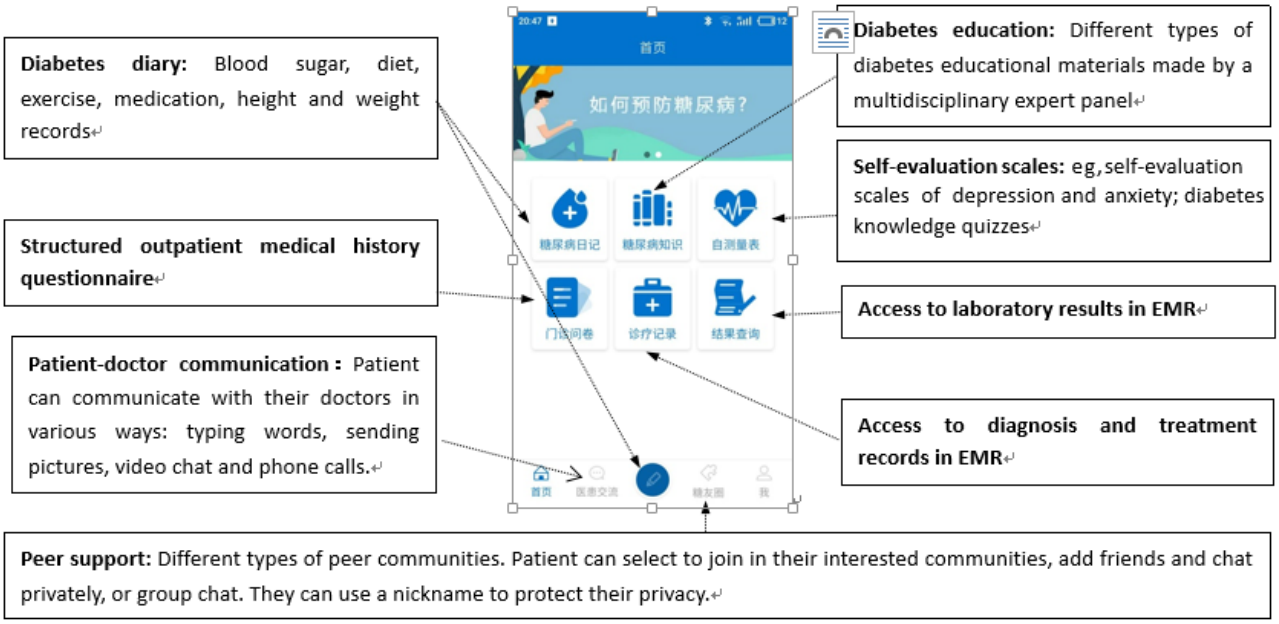
Part 3: App Prototype

The final solution consists of a patient-end app and a doctor-end app. Both of them are based on the iOS and Android platforms. Modules of the patient app are shown in Figure 5.

All the following functions are included: patient-doctor communication, diabetes diary (blood sugar, diet, exercise, and medication records), diabetes education, peer support, blood sugar test reminder, and abnormal blood sugar reminder. According to the expert panel's clinical experiences, it is important to know patients' former diagnosis and laboratory results if they are to give treatment recommendations for their patients, and the qualitative results suggested that access to EMRs would motivate patients to continue using the app. This function is also included in our app.

The qualitative results suggested that inconvenience and lack of a timely response were problems with the patient-doctor communication function of current apps, and according to our expert panel's clinical experience, it is difficult for patients to describe their condition clearly just by typing words. Thus, patient-doctor communication in our app employs various types of communication modes to ensure this convenience: typing words, sending pictures, video chat, and phone calls. Doctors can view their patients' diabetes diaries, diagnosis, treatment records, and laboratory results in their EMRs and the lengths of patients' study times with the diabetes education materials through their own app and can give tailored feedback or send tailored education materials to them. Notifications are automatically sent to patients or their doctors if there are new messages for them.

Figure 5. Homepage screenshot of the patient app. EMR: electronic medical record.



Data entry burden, lack of feedback, and a display that is not as intuitive as a paper diary discourage patients from maintaining diabetes diaries. According to our solution, patients can link the app with their glucose meter, and blood sugar results can automatically be transferred from glucose meters to the app by Bluetooth or General Packet Radio Service. Our app obtains daily step count data from step counter software in mobile phones and records the daily step counts automatically. Patients can take photos of foods to record their diet using the built-in camera. Due to the limitation of the mobile phone page, displays of blood sugar, diet, exercise, and medication records in most apps are scattered across different pages, and it is inconvenient to combine them together. By brainstorming, we constructed a design that enables specific diet, exercise, and medication information to be viewed on the blood sugar display page, thus allowing a comprehensive analysis of the causes of blood sugar changes. Blood sugar history graphs and statistics make blood sugar clear at a glance. For feedback on the diabetes diary, we included the following solutions: (1) blood sugar targets can be established collaboratively by patients and their doctors; (2) patients are alerted to off-target blood sugar with warning colors and messages; (3) if blood sugar levels are dangerous (lower than 3.9 mmol/L or higher than 20 mmol/L), a reminder message will be sent automatically to the patient's doctor's app; and (4) a patient's doctor can view the patient's diabetes diary and give tailored feedback.

More information about our app functions and design based on quantitative and qualitative results is shown in Multimedia Appendix 4.

Discussion

Principal Findings

Our study established the reasons that the effects of current diabetes apps for T1DM are poor and investigated patient requirements from the users' perspective.

The questionnaire surveys suggested that patient-doctor communication and the availability of a diabetes diary were the most important functions of a diabetes app. Detailed records of blood sugar, diet, exercise, and medication can help doctors analyze the factors influencing blood sugar. The blood sugar record was the most used function of the apps [37], but the users did not continue to use this feature. The in-depth interviews revealed that the greatest problem with the diabetes diaries was the absence of feedback from HCPs. Automatic feedback could not meet patients' needs. Patients thought that the diary was useless; thus, they gradually stopped using it. Most experts thought 1 important reason for app invalidity was that patients received little guidance from HCPs. Our study suggests that the role of the doctor is central for a diabetes app. A recent meta-analysis revealed that the effects of diabetes apps were explained by the frequency of HCP feedback. HCP functionality is important for achieving clinical effectiveness [38], but few apps offer an integrated function for communication and feedback from HCPs [39]. The questionnaire surveys showed that some diabetes apps had an HCP consultation function, but only a small number of patients had used this function. Through in-depth interviews, we identified the reasons for patients not using this function. One important reason was that patients did not trust unfamiliar doctors on the app, and doctors from primary hospitals in China lacked experience in managing T1DM patients.

We advocate that doctors follow up with their outpatients or inpatients using this app. Internet hospitals are developing rapidly in China. A cross-sectional survey determined that 43 internet hospitals were established in 2017, and patients accessed outpatient service delivery via app in 43% of these hospitals [40]. However, doctors from primary hospitals in China need training to enhance their expertise [41]. Many doctors did not know whether using an app to guide patients' medication was legal, and doctors in China are overloaded [42]. These issues discourage doctors from using an app to manage patients. Health insurance coverage and charge systems are also needed to

encourage HCPs to use an app to manage their patients in the long term. Fortunately, standardization of residents' training and the hierarchical medical system is underway in China, which will reduce the burden of doctors from tertiary hospitals and will enhance the expertise of doctors from primary hospitals. The Chinese Government is energetically advocating internet medical treatment [43], which will enable doctors to follow up with their patients using an app.

DSME is an important part of diabetes management according to diabetes guidelines. Several studies have shown the benefits of DSME [24]. However, few patients in China receive DSME programs in hospitals [6]. Our in-depth interview found some problems in patients' self-management conduct and suggested that a mobile app was preferable to education in a hospital for DSME. Digital health interventions can help overcome some of the barriers to self-management posed by the limitations of existing health care systems [44]. The questionnaire surveys suggested that both experts and patients thought DSME was very important for a diabetes app. The experts believed 1 important reason for app invalidity was that diabetes education knowledge on apps was unsystematic. Many diabetes apps do not have sound educational quality [27]. Different modes of systematic diabetes education knowledge created by a multidisciplinary expert panel are needed for the app.

The effectiveness of peer support for diabetes outcomes is ambiguous because of the availability of different support modes [45]. Our study showed that most patients thought peer support was an important function of a diabetes app, and most patients hoped to communicate with similar patients. Peer support can help patients exchange glycemic control strategies and emotional experiences. They considered the role of peer leaders very important. Peer leaders can play leading, interactive, and cohesive roles and can improve the atmosphere of a peer support module. Internet-based mentoring programs can increase the frequency of blood sugar monitoring [46], and studies have demonstrated that peer leaders can provide effective diabetes self-management support [47,48]. However, exchanges in current diabetes apps all take the form of forums, which is inconvenient. Thus, few patients exchanged information with others in diabetes apps.

The expert survey suggested that 1 important reason for app invalidity was that diabetes apps lacked comprehensive functions. A meta-analysis revealed that the number of functions offered by apps influences HbA_{1c} levels [19]. Therefore, modules such as patient-doctor communication, diabetes diary, diabetes education, and peer support are all included in our app. However, diabetes apps offering a wider range of functions performed worse in terms of usability [49], and our study suggested that lack of time and complicated operations were factors influencing patients' use of an app. Most patients considered the manual input of diabetes diary data burdensome. To increase app usability and patients' adherence to complete a diabetes diary, blood sugar readings and daily steps can be recorded automatically in our app. Of course, feedback from HCPs will encourage patients to adhere to diabetes diaries. Our app design principle was that the operations should be simple and clear, and the use of clear navigation in our app will enable its usability.

The patient survey suggested that patients greatly needed an insulin calculator, but they did not know whether the calculator was accurate. The expert surveys suggested that most experts did not recommend or were opposed to an insulin calculator, and most of them thought insulin calculators were dangerous or very dangerous. Similar results were found in a New Zealand survey [37]. As these algorithms were found to have limited efficacy and were incorrect [50], we did not include an insulin calculator in our app. Artificial intelligence may have potential use in this area [51].

Our study revealed that the awareness and utilization rates of diabetes apps in China were low. Only a small subset of the patients' apps was recommended by HCPs. One important reason was that the effects of the apps were not evidence-based; thus, they did not know which app was better. Only 1 Chinese diabetes app was tested in a short-term randomized controlled trial (RCT) [52]. Thus, high-quality RCTs are needed [39]. We are planning a multicenter RCT to test the long-term efficacy of our app. We hope we can provide evidence for patients to choose a valid diabetes app.

Limitations

We did not interview child or adolescent patients as their needs and diabetes management models are slightly different from those of adults. However, in child patients, disease management is always performed by their parents; thus, our app is also suitable for this population. However, children occasionally manage their disease independently. In particular, adolescents of transitioning age gradually withdraw support from their parents and take over management tasks. Our app can help these patients through this transitioning period. We can set a family member account to supervise them, and we included some peer communities of their age in our app to help them solve problems specific to this period. However, further improvement is needed to satisfy the specific needs of this population. We also did not interview diabetes experts. As our expert panel consisted of diabetes experts with rich experience in T1DM management, we did not think that interviewing diabetes experts was necessary. The security of young HCPs regarding making incorrect medical suggestions should be taken into consideration. In our solution, we employed qualification certification: registered doctors were from tertiary hospitals with years of experience in T1DM management. A test for doctors' specialties may also be needed to ensure the qualifications of registered doctors.

Comparison With Prior Work

The effects of current diabetes apps on T1DM are poor. No study has explored the reasons for this ineffectiveness from a user's perspective, and very few diabetes apps have shared their methodology [23]; thus, app developers do not know how to choose a valid method. App development should be based on thorough knowledge of user needs [53]. Two studies designed diabetes apps by exploring users' needs though in-depth interviews with young patients and their parents [54,55]. However, because the interviewees had never used diabetes

apps, their understanding of diabetes apps was abstract, and they had difficulty describing their needs accurately. In addition, a purely qualitative study may not provide a comprehensive understanding of user needs. Castensoe-Seidenfaden et al first introduced a mixed-method study to design an app for improving self-management of young patients [56]. However, their quantitative and qualitative prestudies did not investigate patients' and doctors' perspectives of diabetes management apps. Our app design was led by diabetes experts. First, we conducted a quantitative survey to grasp the perspectives of patients and diabetes experts about diabetes apps from a macro level. Second, in-depth interviews with experienced patients supplemented and deepened the results of the questionnaire survey and gave us a better understanding of the problems of current apps and the need for a new diabetes app.

Conclusions

Patient-doctor communication is the most important function of a diabetes app. A mobile app is the preferable method for patients to receive DSME compared with studying in a hospital, but apps should be integrated with HCPs rather than stand-alone. We advocate that doctors follow up with their patients using diabetes apps. Our mixed-method study combined qualitative and quantitative data to comprehensively and deeply explore why the effects of current diabetes apps in T1DM are poor and what T1DM patients need for a diabetes app from the user perspective, which provided meaningful guidance for our app design. This study has significance as a reference for the development of similar apps in the future.

Acknowledgments

The authors would like to thank all participants for sharing their perspectives in the course of this study. They would like to thank the software team of Sinocare Inc for providing technical support. They would also like to thank American Journal Experts for their assistance in language editing. This work was supported by the National Key R&D Program of China (2017YFC1309604), the National Science and Technology Infrastructure Program (2015BAI12B13), and the National Key R&D Program of China (2016YFC1305000, 2016YFC1305001).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey for diabetologists.
[PDF File (Adobe PDF File), 40 KB - mhealth_v6i9e11400_app1.pdf]

Multimedia Appendix 2
Survey for type 1 diabetes mellitus (T1DM) patients.
[PDF File (Adobe PDF File), 65 KB - mhealth_v6i9e11400_app2.pdf]

Multimedia Appendix 3
Interview guideline.
[PDF File (Adobe PDF File), 15 KB - mhealth_v6i9e11400_app3.pdf]

Multimedia Appendix 4
App functions and design.
[PDF File (Adobe PDF File), 32 KB - mhealth_v6i9e11400_app4.pdf]

References

1. You WP, Henneberg M. Type 1 diabetes prevalence increasing globally and regionally: the role of natural selection and life expectancy at birth. Br Med J Open Diabetes Res Care 2016;4(1):e000161 [FREE Full text] [doi: 10.1136/bmjdrc-2015-000161] [Medline: 26977306]
2. Onkamo P, Väänänen S, Karvonen M, Tuomilehto J. Worldwide increase in incidence of Type I diabetes--the analysis of the data on published incidence trends. Diabetologia 1999 Dec;42(12):1395-1403. [doi: 10.1007/s001250051309] [Medline: 10651256]
3. Weng J, Zhou Z, Guo L, Zhu D, Ji L, Luo X, T1D China Study Group. Incidence of type 1 diabetes in China, 2010-13: population based study. Br Med J 2018 Dec 03;360:j5295 [FREE Full text] [Medline: 29298776]
4. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. Lancet 2014 Jan 04;383(9911):69-82 [FREE Full text] [doi: 10.1016/S0140-6736(13)60591-7] [Medline: 23890997]
5. Ryan EA, Holland J, Stroulia E, Bazelli B, Babwik SA, Li H, et al. Improved A1C levels in type 1 diabetes with smartphone app use. Can J Diabetes 2017 Feb;41(1):33-40. [doi: 10.1016/j.jcjd.2016.06.001] [Medline: 27570203]

6. McGuire HC, Ji L, Kissimova-Skarbek K, Whiting D, Aguirre F, Zhang P, et al. Type 1 diabetes mellitus care and education in China: the 3C study of coverage, cost, and care in Beijing and Shantou. Diabetes Res Clin Pract 2017 Jul;129:32-42. [doi: 10.1016/j.diabres.2017.02.027] [Medline: 28500868]
7. American Diabetes Association. 6. Glycemic targets: standards of medical care in diabetes-2018. Diabetes Care 2018 Dec;41(Suppl 1):S55-S64. [doi: 10.2337/dc18-S006] [Medline: 29222377]
8. Devries JH, Snoek FJ, Heine RJ. Persistent poor glycaemic control in adult type 1 diabetes. A closer look at the problem. Diabet Med 2004 Dec;21(12):1263-1268. [doi: 10.1111/j.1464-5491.2004.01386.x] [Medline: 15569126]
9. Reddy M, Rilstone S, Cooper P, Oliver NS. Type 1 diabetes in adults: supporting self management. Br Med J 2016 Mar 10;352:i998. [Medline: 26965473]
10. Landau Z, Mazor-Aronovitch K, Boaz M, Blaychfeld-Magnazi M, Graph-Barel C, Levek-Motola N, et al. The effectiveness of internet-based blood glucose monitoring system on improving diabetes control in adolescents with type 1 diabetes. Pediatr Diabetes 2012 Mar;13(2):203-207. [doi: 10.1111/j.1399-5448.2011.00800.x] [Medline: 21848925]
11. Garg SK, Shah VN, Akturk HK, Beatson C, Snell-Bergeon JK. Role of mobile technology to improve diabetes care in adults with type 1 diabetes: the remote-T1D study iBGStar in type 1 diabetes management. Diabetes Ther 2017 Aug;8(4):811-819 [FREE Full text] [doi: 10.1007/s13300-017-0272-5] [Medline: 28555339]
12. Given JE, O'Kane MJ, Bunting BP, Coates VE. Comparing patient-generated blood glucose diary records with meter memory in diabetes: a systematic review. Diabet Med 2013 Aug;30(8):901-913. [doi: 10.1111/dme.12130] [Medline: 23324062]
13. Chen Y, Yin Z, Xie Q. Suggestions to ameliorate the inequity in urban/rural allocation of healthcare resources in China. Int J Equity Health 2014;13:34 [FREE Full text] [doi: 10.1186/1475-9276-13-34] [Medline: 24884614]
14. Baron JS, Hirani S, Newman SP. A randomised, controlled trial of the effects of a mobile telehealth intervention on clinical and patient-reported outcomes in people with poorly controlled diabetes. J Telemed Telecare 2016 Feb 15;23(2):207-216. [doi: 10.1177/1357633X16631628] [Medline: 26880694]
15. Istepanian RS. Mobile applications for diabetes management: efficacy issues and regulatory challenges. Lancet Diabetes Endocrinol 2015 Dec;3(12):921-923. [doi: 10.1016/S2213-8587(15)00414-3] [Medline: 26545304]
16. 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]
17. Castensøe-Seidenfaden P, Husted GR, Jensen AK, Hommel E, Olsen B, Pedersen-Bjergaard U, et al. Testing a smartphone app (Young with Diabetes) to improve self-management of diabetes over 12 months: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jun 26;6(6):e141 [FREE Full text] [doi: 10.2196/mhealth.9487] [Medline: 29945861]
18. Chavez S, Fedele D, Guo Y, Bernier A, Smith M, Warnick J, et al. Mobile apps for the management of diabetes. Diabetes Care 2017 Oct;40(10):e145-e146. [doi: 10.2337/dc17-0853] [Medline: 28774944]
19. Bonoto BC, de Araújo VE, Godói IP, de Lemos LL, Godman B, Bennie M, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2017 Mar 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
20. 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]
21. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. JMIR Mhealth Uhealth 2017 Mar 14;5(3):e35 [FREE Full text] [doi: 10.2196/mhealth.6522] [Medline: 28292740]
22. Powell AC, Landman AB, Bates DW. In search of a few good apps. J Am Med Assoc 2014 May 14;311(18):1851-1852. [doi: 10.1001/jama.2014.2564] [Medline: 24664278]
23. 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]
24. Beck J, Greenwood DA, Blanton L, Bollinger ST, Butcher MK, Condon JE, 2017 Standards Revision Task Force. 2017 National Standards for Diabetes Self-Management Education and Support. Diabetes Educ 2018 Feb;44(1):35-50. [doi: 10.1177/0145721718754797] [Medline: 29346744]
25. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011;13(3):e65 [FREE Full text] [doi: 10.2196/jmir.1874] [Medline: 21979293]
26. American Diabetes Association. Standards of medical care in diabetes--2014. Diabetes Care 2014 Jan;37(Suppl 1):S14-S80. [doi: 10.2337/dc14-S014] [Medline: 24357209]
27. Duke DC, Barry S, Wagner DV, Speight J, Choudhary P, Harris MA. Distal technologies and type 1 diabetes management. Lancet Diabetes Endocrinol 2018 Dec;6(2):143-156. [doi: 10.1016/S2213-8587(17)30260-7] [Medline: 28867311]
28. Peng W, Yuan S, Holtz BE. Exploring the challenges and opportunities of health mobile apps for individuals with type 2 diabetes living in rural communities. Telemed J E Health 2016 Sep;22(9):733-738. [doi: 10.1089/tmj.2015.0180] [Medline: 26982017]

29. Pulman A, Taylor J, Galvin K, Masding M. Ideas and enhancements related to mobile applications to support type 1 diabetes. JMIR Mhealth Uhealth 2013;1(2):e12 [FREE Full text] [doi: 10.2196/mhealth.2567] [Medline: 25100684]
30. Ritholz MD, Beverly EA, Weinger K. Digging deeper: the role of qualitative research in behavioral diabetes. Curr Diab Rep 2011 Dec;11(6):494-502. [doi: 10.1007/s11892-011-0226-7] [Medline: 21932032]
31. Conway N, Campbell I, Forbes P, Cunningham S, Wake D. mHealth applications for diabetes: user preference and implications for app development. Health Informatics J 2016 Dec;22(4):1111-1120. [doi: 10.1177/1460458215616265] [Medline: 26635324]
32. Nie L, Xie B, Yang Y, Shan YM. Characteristics of Chinese m-Health applications for diabetes self-management. Telemed J E Health 2016 Jul;22(7):614-619. [doi: 10.1089/tmj.2015.0184] [Medline: 27171016]
33. Drincic A, Prahalad P, Greenwood D, Klonoff DC. Evidence-based mobile medical applications in diabetes. Endocrinol Metab Clin North Am 2016 Dec;45(4):943-965 [FREE Full text] [doi: 10.1016/j.ecl.2016.06.001] [Medline: 27823614]
34. Zhang X, Wen D, Liang J, Lei J. How the public uses social media wechat to obtain health information in china: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
35. Kennedy A, Narendran P, Andrews RC, Daley A, Greenfield SM, EXTOD Group. Attitudes and barriers to exercise in adults with a recent diagnosis of type 1 diabetes: a qualitative study of participants in the Exercise for Type 1 Diabetes (EXTOD) study. Br Med J Open 2018 Dec 24;8(1):e017813 [FREE Full text] [doi: 10.1136/bmjopen-2017-017813] [Medline: 29371269]
36. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [Medline: 25326092]
37. Boyle L, Grainger R, Hall RM, Krebs JD. Use of and beliefs about mobile phone apps for diabetes self-management: surveys of people in a hospital diabetes clinic and diabetes health professionals in New Zealand. JMIR Mhealth Uhealth 2017 Jun 30;5(6):e85 [FREE Full text] [doi: 10.2196/mhealth.7263] [Medline: 28666975]
38. Hou C, Xu Q, Diao S, Hewitt J, Li J, Carter B. Mobile phone applications and self-management of diabetes: a systematic review with meta-analysis, meta-regression of 21 randomized trials and GRADE. Diabetes Obes Metab 2018 Dec;20(8):2009-2013. [doi: 10.1111/dom.13307] [Medline: 29582538]
39. Holmen H, Wahl AK, Cvancarova Småstuen M, Ribu L. Tailored communication within mobile apps for diabetes self-management: a systematic review. J Med Internet Res 2017 Jun 23;19(6):e227 [FREE Full text] [doi: 10.2196/jmir.7045] [Medline: 28645890]
40. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet hospitals in China: cross-sectional survey. J Med Internet Res 2017 Jul 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
41. Xu J, Wang W, Li Y, Zhang J, Pavlova M, Liu H, et al. Analysis of factors influencing the outpatient workload at Chinese health centres. BMC Health Serv Res 2010 Jun 05;10:151 [FREE Full text] [doi: 10.1186/1472-6963-10-151] [Medline: 20525381]
42. Li Q, Xie P. Outpatient workload in China. Lancet 2013 Jun 08;381(9882):1983-1984. [doi: 10.1016/S0140-6736(13)61198-8] [Medline: 23746895]
43. Xinhua News Agency. 2018. The General Office of the State Council issued the "Opinions on Promoting the Development of "Internet + Medical Health"" URL: http://www.gov.cn/xinwen/2018-04/28/content_5286707.htm [accessed 2018-07-27] [WebCite Cache ID 71DDXUpUM]
44. Pal K, Dack C, Ross J, Michie S, May C, Stevenson F, et al. Digital health interventions for adults with type 2 diabetes: qualitative study of patient perspectives on diabetes self-management education and support. J Med Internet Res 2018 Jan 29;20(2):e40 [FREE Full text] [doi: 10.2196/jmir.8439] [Medline: 29463488]
45. Doe E. An analysis of the relationships between peer support and diabetes outcomes in adolescents with type 1 diabetes. J Health Psychol 2018 Sep;23(10):1356-1366. [doi: 10.1177/1359105316656228] [Medline: 27402732]
46. Suh S, Jean C, Koo M, Lee SY, Cho MJ, Sim KH, et al. A randomized controlled trial of an internet-based mentoring program for type 1 diabetes patients with inadequate glycemic control. Diabetes Metab J 2014 Apr;38(2):134-142 [FREE Full text] [doi: 10.4093/dmj.2014.38.2.134] [Medline: 24851207]
47. Piatt GA, Rodgers EA, Xue L, Zgibor JC. Integration and utilization of peer leaders for diabetes self-management support: results from project SEED (Support, Education, and Evaluation in Diabetes). Diabetes Educ 2018 Dec;44(4):373-382. [doi: 10.1177/0145721718777855] [Medline: 29806788]
48. Tang TS, Funnell M, Sinco B, Piatt G, Palmisano G, Spencer MS, et al. Comparative effectiveness of peer leaders and community health workers in diabetes self-management support: results of a randomized controlled trial. Diabetes Care 2014 Jun;37(6):1525-1534 [FREE Full text] [doi: 10.2337/dc13-2161] [Medline: 24722495]
49. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
50. Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. BMC Med 2015;13:106 [FREE Full text] [doi: 10.1186/s12916-015-0314-7] [Medline: 25943590]

51. Buch V, Varughese G, Maruthappu M. Artificial intelligence in diabetes care. Diabet Med 2018 Apr;35(4):495-497. [doi: 10.1111/dme.13587] [Medline: 29368355]
52. Zhou W, Chen M, Yuan J, Sun Y. Welltang-a smart phone-based diabetes management application-improves blood glucose control in Chinese people with diabetes. Diabetes Res Clin Pract 2016 Jun;116:105-110. [doi: 10.1016/j.diabres.2016.03.018] [Medline: 27321324]
53. Floch J, Zettl A, Fricke L, Weisser T, Grut L, Vilarinho T, et al. User needs in the development of a health app ecosystem for self-management of cystic fibrosis: user-centered development approach. JMIR Mhealth Uhealth 2018 May 08;6(5):e113 [FREE Full text] [doi: 10.2196/mhealth.8236] [Medline: 29739742]
54. Holtz BE, Murray KM, Hershey DD, Dunneback JK, Cotten SR, Holmstrom AJ, et al. Developing a patient-centered mHealth app: a tool for adolescents with type 1 diabetes and their parents. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e53 [FREE Full text] [doi: 10.2196/mhealth.6654] [Medline: 28428167]
55. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 08;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
56. Castensøe-Seidenfaden P, Reventlov Husted G, Teilmann G, Hommel E, Olsen BS, Kensing F. Designing a self-management app for young people with type 1 diabetes: methodological challenges, experiences, and recommendations. JMIR Mhealth Uhealth 2017 Oct 23;5(10):e124 [FREE Full text] [doi: 10.2196/mhealth.8137] [Medline: 29061552]

Abbreviations

DSME: diabetes self-management education
EMR: electronic medical record
HbA_{1c}: average glycosylated hemoglobin
HCP: health care professional
IQR: interquartile range
RCT: randomized controlled trial
T1DM: type 1 diabetes mellitus
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 25.06.18; peer-reviewed by A Pulman; comments to author 19.07.18; revised version received 13.08.18; accepted 04.09.18; published 21.09.18.

Please cite as:
Zhang Y, Li X, Luo S, Liu C, Liu F, Zhou Z
Exploration of Users' Perspectives and Needs and Design of a Type 1 Diabetes Management Mobile App: Mixed-Methods Study
JMIR Mhealth Uhealth 2018;6(9):e11400
URL: http://mhealth.jmir.org/2018/9/e11400/
doi:10.2196/11400
PMID:30249580

Original Paper

A Smartphone App and Personalized Text Messaging Framework (InDEx) to Monitor and Reduce Alcohol Use in Ex-Serving Personnel: Development and Feasibility Study

Daniel Leightley[1], BSc (Hons), MSc, PhD; Jo-Anne Puddephatt[2], BSc (Hons), MSc; Norman Jones[3], PhD; Toktam Mahmoodi[4], BSc (Hons), PhD; Zoe Chui[5], BSc (Hons), MSc; Matt Field[2,6], PhD; Colin Drummond[7,8], FRCPsych, MD; Roberto J Rona[1], PhD, FFPH; Nicola T Fear[1,3], BSc (Hons), MSc, D Phil (Oxon); Laura Goodwin[2], BSc (Hons), MSc, PhD

[1]King's Centre for Military Health Research, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[2]Department of Psychological Sciences, University of Liverpool, Liverpool, United Kingdom

[3]Academic Department of Military Mental Health, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[4]Department of Informatics, King's College London, London, United Kingdom

[5]Department of Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[6]UK Centre for Tobacco and Alcohol Studies, Department of Psychological Sciences, University of Liverpool, Liverpool, United Kingdom

[7]Addictions Department, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[8]South London and Maudsley NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Daniel Leightley, BSc (Hons), MSc, PhD
King's Centre for Military Health Research
Institute of Psychiatry, Psychology & Neuroscience
King's College London
Cutcombe Road
London, SE5 9RJ
United Kingdom
Phone: 44 78485334
Fax: 44 78485408
Email: daniel.leightley@kcl.ac.uk

Abstract

Background: Self-reported alcohol misuse remains high in armed forces personnel even after they have left service. More than 50% of ex-serving personnel meet the criteria for hazardous alcohol use; however, many fail to acknowledge that they have a problem. Previous research indicates that interventions delivered via smartphone apps are suitable in promoting self-monitoring of alcohol use, have a broad reach, and may be more cost-effective than other types of brief interventions. There is currently no such intervention specifically designed for the armed forces.

Objective: This study sought to describe the development of a tailored smartphone app and personalized text messaging (short message service, SMS) framework and to test the usability and feasibility (measured and reported as user engagement) of this app in a hard-to-engage ex-serving population.

Methods: App development used Agile methodology (an incremental, iterative approach used in software development) and was informed by behavior change theory, participant feedback, and focus groups. Participants were recruited between May 2017 and June 2017 from an existing United Kingdom longitudinal military health and well-being cohort study, prescreened for eligibility, and directed to download either Android or iOS versions of the ”Information about Drinking for Ex-serving personnel” (InDEx) app. Through the app, participants were asked to record alcohol consumption, complete a range of self-report measures, and set goals using implementation intentions (if-then plans). Alongside the app, participants received daily automated personalized text messages (SMS) corresponding to specific behavior change techniques with content informed by the health action process approach with the intended purpose of promoting the use of the drinks diary, suggesting alternative behaviors, and providing feedback on goals setting.

Results: Invitations to take part in the study were sent to ex-serving personnel, 22.6% (31/137) of whom accepted and downloaded the app. Participants opened the InDEx app a median of 15.0 (interquartile range [IQR] 8.5-19.0) times during the 4 week period (28 days), received an average of 36.1 (SD 3.2) text messages (SMS), consumed alcohol on a median of 13.0 (IQR 11.0-15.0) days, and consumed a median of 5.6 (IQR 3.3-11.8) units per drinking day in the first week, which decreased to 4.7 (IQR 2.0-6.9) units by the last week and remained active for 4.0 (IQR 3.0-4.0) weeks.

Conclusions: Personnel engaged and used the app regularly as demonstrated by the number of initializations, interactions, and time spent using InDEx. Future research is needed to evaluate the engagement with and efficacy of InDEx for the reduction of alcohol consumption and binge drinking in an armed forces population.

(JMIR Mhealth Uhealth 2018;6(9):e10074) doi:10.2196/10074

KEYWORDS

behavior change techniques; smartphone; alcohol misuse; binge drinking; text messaging; ex-serving; armed forces; mobile phones

Introduction

Alcohol misuse is common in the United Kingdom (UK) armed forces and the prevalence is higher in the military than in the general population [1-3], with the trend continuing after they leave service [1]. More than 50% of those who have left military service meet the criteria for hazardous alcohol use, defined as scoring 8 or more on the alcohol use disorders identification test (AUDIT; [4]). This prevalence rate is almost double of that found in the general population [5]. Additionally, 47% of ex-serving personnel report binge drinking, defined as 6 or more units for females and 8 or more units (1 UK unit=8g ethanol) for males, per session at least once per week [3].

Most people in the general population underestimate their drinking and do not perceive it as problematic, even when the level of consumption is potentially harmful to health [6]; young men are at particular risk of underestimating their drinking [6]. This pattern is similar among armed forces personnel with less than half of hazardous drinkers recognizing that they have an alcohol problem and seeking medical help [7]. There is a culture of heavy alcohol use in the armed forces, which may be encouraged or maintained by social determinants [8]; therefore, leaving service could provide an opportunity to initiate behavioral change in settings with less peer pressure to conform to social norms.

In the last decade, computer and Web-based interventions (eg, Down Your Drink [9]) have been harnessed to increase reach, provide real-time monitoring, and offer personalized delivery [9-11]. More recently, the mode of intervention delivery has shifted from Web-based to mobile-based [12]. Mobile apps for use in health have proven to be an effective and successful method of providing patient-centric interventions that are based on real-time data and needs [13].

There are a large number of alcohol-related apps available to the general population with a recent content analysis identifying more than 600 apps, of which 91 were identified as focusing on alcohol reduction [14]. It has been reported that many apps lack an evidence base and make no reference to the scientific literature [14,15]. Recent research has found the use of mobile apps as brief alcohol interventions to be effective compared with traditional delivery methods (eg, face-to-face) [16,17]; however, the content of most existing alcohol smartphone interventions is based on public health guidelines regarding safe alcohol limits [14,18]. These alcohol limits may not be *perceived* as credible because they are viewed as state sponsored and are often at odds with individual beliefs, prevailing social context, and perceptions of consumption [18-21]. Many users do not maintain engagement with mobile health interventions [22]. Further, the majority of existing alcohol mobile apps emphasize long-term health consequences which are seen as remote risks, especially by young drinkers [15,17,23]. A recent meta-analysis suggests that it may be more effective to focus on short-term detrimental consequences to encourage individuals to reduce their alcohol consumption [24].

Most existing alcohol apps include self-monitoring (eg, Drink Less [23], Drink Aware [25], One You Drinks Tracker [26]), wherein users are encouraged to regularly record and monitor (via visual graphics) their alcohol consumption within an app [23,27]. Self-monitoring was found to be the most effective behavior change technique (BCT) for reducing alcohol use; a BCT is defined as a specific, irreducible component of an intervention designed to change behavior and a putative active ingredient in an intervention [28]. A recent review of computer and Web-based delivered alcohol interventions suggested that provision of normative feedback, goal review, and inclusion of the social norms approach in combination were associated with better outcomes [24]. There is also evidence that short message service (SMS) text message interventions can be successful in encouraging people to change their behavior [29,30], and further benefits may be gained by incorporating user input to tailor the SMS text messages. However, to the authors' knowledge, there is no published work that seeks to develop an alcohol reduction app for ex-serving personnel.

We are not aware of any mobile health app that seeks to customize a brief alcohol intervention using personalized SMS text messages. In this study, we describe the development of the "Information about Drinking for Ex-serving personnel" (InDEx) mobile phone app, a tailored 4 week (28 day) intervention specifically designed to target ex-serving personnel who meet the criteria for hazardous alcohol use, which is likely to impact on their functioning. The purpose of this study was to design an engaging, responsive, and usable smartphone app that delivers personalized SMS text messages and gathers alcohol usage data and to test the usability and feasibility, measured and reported as user engagement, of this app in a

hard-to-engage ex-serving population. Our primary outcome measure was adherence with InDEx, which was measured by the number of weeks participants engaged with the app. Our secondary outcome measures were how many times participants used the app (eg, utilization of the drinks diary) and the proportion of participants using InDEx at the end of the study period.

Methods

Participants

Potential participants were eligible for inclusion if they had served in the UK military, were aged 18-65 years, owned an iPhone or Android device released after 2012, were willing to receive daily SMS text messages, currently resided in UK, and were capable of providing informed consent. Those who had an AUDIT score lower than 8 or greater than 19 were excluded because InDEx is focused on intervening among those drinking hazardously or harmfully, who are likely to be experiencing short-term consequences of their drinking, yet unlikely to be seeking any treatment for this misuse. Those scoring above 20 on the AUDIT meet criteria for probable alcohol dependency and we felt that they may require more intensive treatment. Potential participants took part in the King's Centre for Military Health Research cohort study [2,31] and consented to receive further contact. Participants were asked to use the InDEx app for a period of 4 weeks (28 days) between May 2017 and June 2017. Providing informed consent, downloading the app, and registering an account constituted enrollment in the study. Participants were compensated £40 for their time.

App Design and Development

Design and development of the InDEx app was undertaken on an Apple MacBook Pro, 2.5 GHz i5 Intel processor and 8GB RAM. Drifty Co IONIC Framework version 1 [32] was used as the cross-platform framework to enable iOS and Android deployments using Atom [33] as the development environment (see Multimedia Appendix 1 for an infographic of the InDEx ecosystem).

A full description of the development process, including the InDEx app source code, is available in [34]. A summary is provided hereafter.

Specification and Development

The development of the InDEx app was academic-led and supported by experts in smartphone app development, epidemiology, addiction psychiatry, and military mental health. The content of the intervention incorporated effective components of previous electronic alcohol interventions (eg, [24]) with SMS text messages informed by the health action process approach (HAPA). HAPA theorizes that individuals work through a number of stages to change their behavior, emphasizing the motivational processes underpinning behavioral intentions and the various processes that bring about behavior change [35,36]. The delivery was split into 3 stages, based upon the HAPA model, with the content of the app and SMS text messages corresponding to each stage, for example, goal setting was only introduced at stage 2 (and available for use in stage 3). The stages were:

- Stage 1: Normative feedback (defined below), action self-efficacy, and self-monitoring
- Stage 2: Maintenance of self-efficacy and action planning
- Stage 3: Recovery of self-efficacy and coping planning

The features were grouped into the following modules:

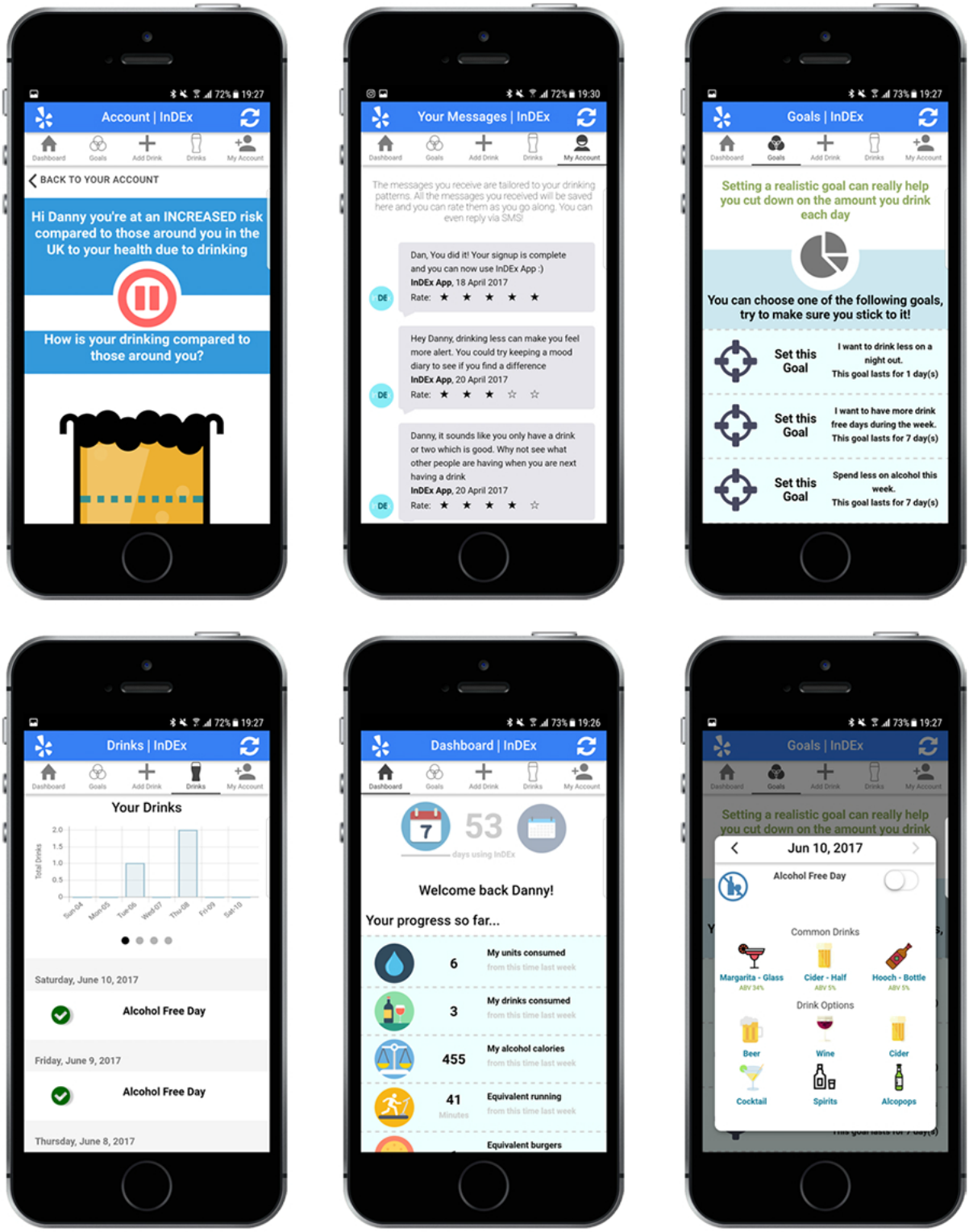
1. Account Management: Participants can modify personal information (eg, first name, last name, and mobile number), password, and app parameters (eg, automatic log-out and clear local storage).
2. Assessment and Normative Feedback: Captures the participant's response to a set of questions (defined by the research team) and aggregates responses to produce an infographic representing the participant's alcohol consumption in comparison to the general population.
3. Self-monitoring and Feedback: Records alcohol consumption by participants and provides a range of visual (eg, charts, figures, and text) metrics to allow for monitoring of consumption.
4. Goal (setting and review): Participants can set goal(s) based on the implementation intentions [37] methodology; visual feedback provides feedback on progress toward achieving goal(s) set.
5. SMS Text Messaging (review): Provides a facility to review SMS text messages sent to and from the InDEx central server system. Further, participants can rate automated SMS text messages (5 star Likert rating).

The app was developed using Agile development methodologies [38] in which an incremental design approach is employed and each increment builds upon the functionality of the previous. Each increment underwent rigorous testing by stakeholder and expert participants sourced from King's Centre for Military Health Research and University of Liverpool (n=17) to ensure software quality and usability. Stakeholders and expert participants were requested to provide feedback on usability, language, functionality, and errors at each increment point. The development cycle did not progress until functionality and source errors were addressed.

To create an account, a participant was required to provide their first name, last name, email address, mobile telephone number, username, password, and in-app informed consent. All sensitive information such as password was encrypted using Bcrypt hashing algorithms (salt factor 10).

InDEx app is presented in Figure 1. The app was designed with limited storage capabilities to avoid concerns regarding confidentiality and privacy of data. Only the username and a secure JSON Web Token denoting the user's time restricted session were stored on the local device with all other data being stored in temporary memory and accessible via application programming interface calls. The app was also available for limited offline use.

Figure 1. Example screenshots of interactions with the InDEx app (left to right, beginning at top): normative feedback, personalized text message history, set a goal, drink diary, dashboard and add a drink. Source: King's Centre for Military Health Research, King's College London.



Operating System Selection

In UK, 4 out of 5 adults own a smartphone; among 18-44-year-olds, adoption is higher at 91% [39] with the majority (over 90%) of smartphones operating either on Google Android or Apple iOS. Based on this information, InDEx was developed for use with both Google Android and Apple iOS enabled devices ensuring that a wide spread of participants could be included.

Personalized Text Messaging

The InDEx app was complemented by tailored SMS text messaging that provides prompts to use the drinks diary, suggests alternative behaviors, and provides feedback on goals. A bank of 180 tailored SMS text messages was developed in

line with delivery stages (defined earlier), which were informed by the HAPA framework and from discussion groups with ex-serving personnel to further refine the messages (Table 1). Each message had the following characteristics: what day it would be sent, message content, and a decision tree defining when it should be triggered. A participant would receive at least one SMS text message each day, up to a maximum of 2. The ultimate design and objective of each message was to prompt diary completion and to suggest alternative behavior related to their individual alcohol consumption.

InDEx uses baseline and contiguous measurements to inform the type of SMS text messages a participant receives to provide a participant-centric approach. Baseline measurements are used to identify suitable messages and as a participant engages with InDEx, continuous measurements are used to reflect current behavior and attitude; for example, if a participant reports feeling depressed or anxious (measured by the patient health questionnaire [40]), a message with suggestions for alternative behaviors to cope with these symptoms (eg, going for a walk) is sent. The messages covered a wide range of topics to target beliefs and motivations with the primary aim of increasing the participant's awareness of their drinking habits and behaviors. The messages were divided into 3 categories: (1) tailored: personalized to drinking habits, baseline, and weekly measurements; (2) tailored and triggered: tailored to baseline and contiguous measurements and a specific event occurring; and (3) targeted (generic): sent on specific days to highlight inactivity, a new feature, or to remind users about an issue. See Table 1 for examples of SMS text messages. The message bank and decision tree for sending SMS text messages are available upon request from the corresponding author.

SMS text messages and two-factor authentication codes (used to verify the participant's mobile phone number) were sent automatically using Twilio's Application Programming Interface via InDEx central command servers. No human involvement was required. All SMS text messages sent to participants were visible in the app ("My Messages" page). Participants could rate any message (rating scale 1-"poor" to 5-"excellent") and provide SMS text message responses, which were stored and displayed to the user but not monitored by the study team.

Submission and Testing

InDEx was submitted to the Google Play and Apple iTunes App stores via Google Play Developer Console and Apple iTunes Connect, respectively. For testing of InDEx, a private testing group was created; only those who had been given permission were able to access and download InDEx.

Measurement Reporting

All measurements were collected via the modules, as seen in Figure 2. The study team had no ability to modify or influence any measurement response.

Upon successful registration (referred to as "day 0"), participants completed several baseline questionnaires that collected the following information: (1) Age and sex; (2) Alcohol consumption and alcohol use disorders via alcohol use disorders identification test (AUDIT; [41]); (3) Symptoms of anxiety using the two item Generalized Anxiety Disorder Scale (GAD-2; [42]); (4) Symptoms of depression using the two item patient health questionnaire (PHQ-2; [40]); (5) Symptoms of Post-Traumatic Stress Disorder were assessed using the five item Diagnostic and Statistical Manual of Mental Disorders Post-Traumatic Stress Disorder Scale [43]; and (6) Readiness to Change and Self-efficacy Scales (score range: 0-10) [44].

Baseline measurement responses informed the type of SMS text message a participant would receive. Although this was optional, the baseline measures were asked again upon completion of the study (day 28).

Weekly Measurements

Participants were asked on days 8, 15, and 22 to complete GAD-2, PHQ-2, and Readiness to Change and Self-Efficacy Scales. Any response provided by the participant further informed the tailoring of the SMS text messages, for example, a participant who scores low on the Readiness to Change Scale is sent supportive messages to encourage a willingness to change.

Reporting Alcohol Consumption

Participants could *"record"* alcohol beverage(s) or an *"alcohol free day "* via the *"Add Drinks"* tab; Multimedia Appendix 2 illustrates the types of alcoholic drink a participant could record. Self-reported alcohol consumption is a standard method for assessing the efficacy of low-intensity interventions [14,17,23].

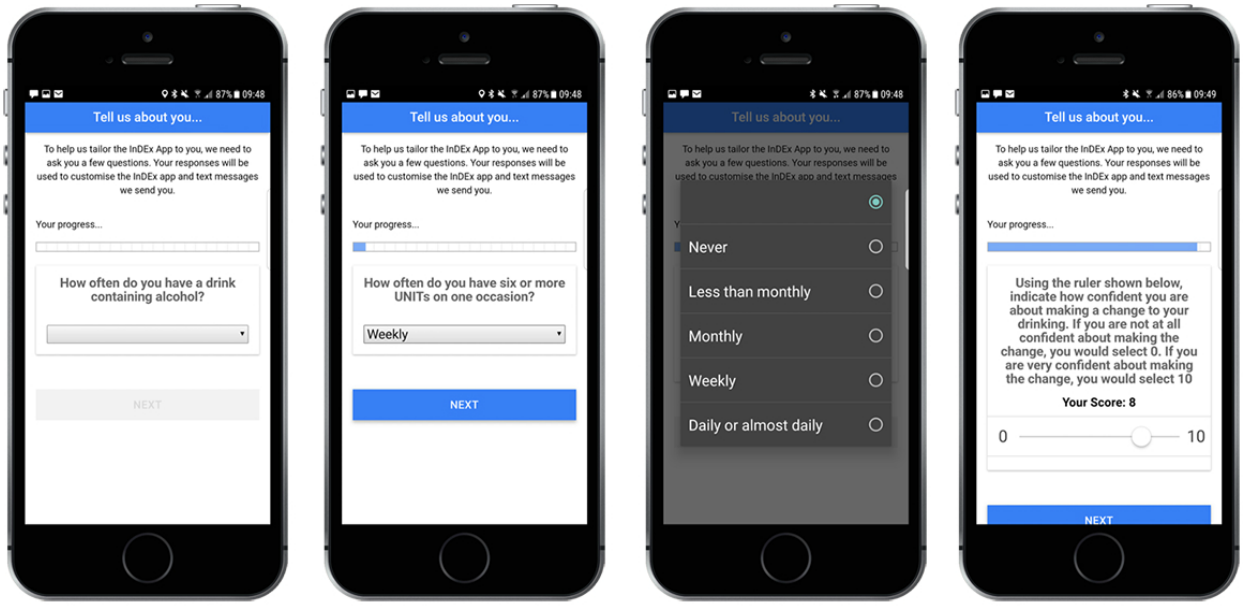
Table 1. An example of the type of *personalized* SMS text messages sent to an individual throughout their use of the app.

Day to be sent	Type	Related BCT[a]	Message
3	Tailored	Mental rehearsal of successful performance (BCT 15.2)[b]	Hi {name}, try thinking that if I am at the pub this week and feel like drinking then imagine how fresh I will feel the next day if I do not drink a lot.
8, 14, 21, 28	Tailored and triggered	Self-monitoring of behavior (BCT 2.3)[b]	Hi {name}, have you logged your drinks from last week? It's quick and easy to do, just go onto the "drinks" tab in the app.
8	Generic	Action planning (BCT 1.4)[b]	Hi {name}, why not set a goal to reduce the amount you drink? It has been found to really help reduce your drinking, you can start now by clicking on the "goals" tab in the app.

[a]BCT: behavior change technique.

[b]*Personalized* SMS text fields with reference to relevant *behavior change technique taxonomy.*

Figure 2. Example screenshots of the InDEx app measurement (questionnaire) module. Source: King's Centre for Military Health Research, King's College London.



Participants could optionally provide volume, strength, price, and calories; however, if no information was provided, UK standard data were used [45]. Further, participants could record who they were drinking with, where they were drinking, and, if consent was provided, their geographical position was recorded.

Engagement and Usability

We measured usability by frequency of engagement using a published procedure [46] that included the number of times the app was initialized (ie, started when not running in the background), the average session duration (ie, time spent using the app and overall and for each page), the number of times a participant performed an interaction (ie, synchronized data, added a drink, and added a SMS text message rating), and the number of weeks in which participants remained engaged with the app. User engagement was defined as having at least 3 client-server interactions in a 7 day period, other than receiving a SMS text message, and was used as a proxy for usability.

Participant engagement was tracked using Google Analytics for Mobile which recorded data when the participant was online or offline. It was not possible to confirm and track if a participant read the SMS text messages, except in cases where the participant provided a rating from within the app.

Clinical Monitoring and Risk Management

Prior to the study commencing, a risk protocol was developed and approved by the University of Liverpool Ethics Committee. Adverse health events were ascertained via automatic monitoring and reporting based on measurement responses and alcohol consumption. A clinician received all warning notifications, which were predefined by the research team for review. If the clinician felt that the event was clinically significant, participants were offered a call by a clinician (for those who declined, a reason was recorded) to discuss the adverse health event. All participants, irrespective of an adverse health event, were provided with a signposting and pathways to local support and assistance via a "Support" page within the app.

Data Analysis

We calculated descriptive statistics to estimate engagement and usability with the app, which were used as a proxy for the feasibility of the InDEx app (to address the primary outcome). Engagement statistics were reported as median and interquartile range (IQR) because the data were not normally distributed (evaluated using skewness and kurtosis values and visualizing the data). Popularity of pages was inferred from the summation of the total number of times each page was viewed by users, and pages were then ranked from highest to lowest number of views.

The average number of drinking days, drink free days, units consumed, units consumed per drinking day, and alcoholic drinks per drinking day were computed across participants and reported as median and IQR. In this study, the number of binge drinking days was computed per week based on the number of days participants reported consuming 6 or more alcoholic drinks (to address the secondary outcome). Self-reported baseline and weekly measurements were presented as median and IQR, except for Readiness to Change and Self-Efficacy Scales, which were presented as mean and SD. Analyses were undertaken using STATA SE 14.2.

Ethical Approval

Ethical approval was obtained from the local Research Ethics Committee at the University of Liverpool (reference: #0625).

Results

Recruitment, Study Enrollment, and Participant Demographics

As shown in Figure 3, 150 individuals were contacted via email to participate in this study; 13 emails bounced back as the email addresses were not valid. Overall, 22.6% (31/137) downloaded

and registered an account with InDEx, 87% (27/31) male and 13% (4/31) female. Of those who joined, 16% (5/31) were aged 25-39 years, 19% (6/31) were aged 40-44, 19% (6/31) were aged 45-49, 19% were aged 50-54 (6/31), and 26% (8/31) were aged 55-64. Finally, 84% (26/31) reported serving in the military for 12 years or more.

Engagement

Participants used the InDEx app for a median of 4.0 (IQR 3.0-4.0) weeks (primary outcome), initializing 15.0 (IQR 8.5-19.0) times over 4 weeks and engaging in 29.0 (IQR 20.0-40.5) sessions for a median of 48.8 seconds (IQR 35.1-73.1). Table 2 provides the engagement measures relating to the level of engagement and adherence; 74% (23/31) of participants used the app every week (maximum 4 weeks) with 87% (27/31) using the app in the final week. Table 3 describes the top 10 pages viewed by participants with the "Dashboard" (38.41%) page being the most popular.

Drinking Behaviors

Table 4 describes the frequency with which participants made a diary entry. Participants consumed alcohol a median of 13.0 (IQR 11.0-15.0) days, had 15.0 (IQR 13.0-17.0) drink free days, and recorded 2.0 (IQR 1.0-4.0) alcoholic drinks per drinking day with a median of 4.7 (IQR 2.3-9.1) units per day.

Table 5 illustrates the drinking behavior of participants over the study period. During week 1, participants reported a median of 2.0 (IQR 1.0-3.0) binge drinking days per week with a similar result in week 4.0 (2.0; IQR 1.0-2.5). However, reductions in units per drinking day from week 1 (5.6; IQR 3.3-11.8) to week 4 (4.7; IQR 2.0-6.9) and units consumed (week 1: 22.9; IQR 14.3-32.4 and week 4: 15.9; 11.6-26.9) was observed.

Measurement Responses

Table 6 summarizes participants' baseline and weekly self-reported measurement responses. Participants had a baseline median AUDIT score of 11 (IQR 10-12), indicating hazardous alcohol use, with an average Readiness to Change Scale score of 4.4 (SD 3.2), indicating some willingness to change. A small change in AUDIT score was observed for participants who self-reported for Day 0 (registration) and Day 28 (final day) based on median score; however, they would still be classified as hazardous drinkers. Most participants did not report anxiety or depression symptoms (measured via GAD-2 or PHQ-2) throughout the study.

Text Messaging

In total, 1083 (mean 36.1, SD 3.2) SMS text messages were sent. Participants were able to reply to messages but were informed that responses would not be monitored. There were 18 replies and 42 SMS text message ratings. The mean rating of content suitability was 2.5 (SD 1.3), indicating a neutral rating for the content of those messages. One participant withdrew consent for receiving SMS text messages on day 16 of the study.

Figure 3. Participant flow through the study.

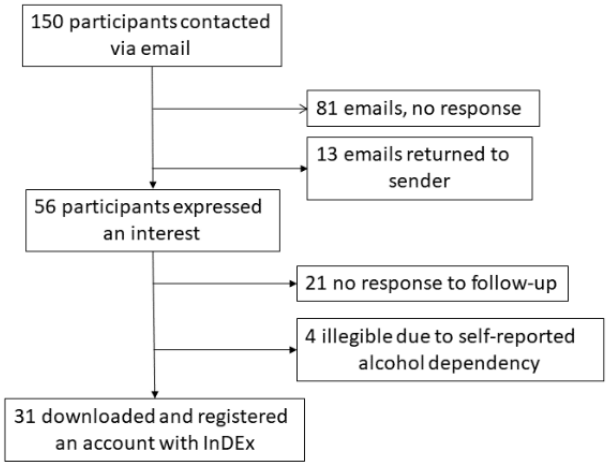


Table 2. Engagement measures over the study period per participant.

Engagement Measure	Median (IQR[a])
Initializations[b]	15.0 (8.5-19.0)
Session count	29.0 (20.0-40.5)
Session duration (s)	48.8 (35.1-73.1)
Interactions[c]	223.0 (182.3-303.5)
Weeks active	4.0 (3.0-4.0)

[a]IQR: interquartile range.

[b]App initialization reflects the app being opened without a background session existing.

[c]Defined as a participant performing a click event (eg, add drink, log-out, change page, change drinks diary chart).

Table 3. Top 10 viewed pages within the InDEx app visited by participants within the study period.

Page	n (%)
Dashboard	4045 (38.41)
Drinks diary	3031 (28.78)
Add drink	1160 (11.01)
Account	390 (3.70)
Goals	379 (3.59)
Normative feedback	244 (2.31)
Weekly assessment	166 (1.57)
Login	148 (1.40)
Support	102 (0.96)
Your messages	98 (0.93)
Other pages	766 (7.27)

Table 4. Number of drinking days, drink free days, units consumed, and alcoholic drinks per drinking day across the study period (4 weeks, n=31).

Reported alcohol consumption	Median (IQR[a])
Drinking days	13.0 (11.0-15.0)
Drink free days	15.0 (13.0-17.0)
Units per drinking day	4.7 (2.3-9.1)
Units consumed	79.4 (58.4-117.3)
Alcoholic drinks per drinking day	2.0 (1.0-4.0)

[a]IQR: interquartile range.

Table 5. Drinking behavior of participants over the study period; n denotes number of participants who recorded an alcohol event during the period.

Reported alcohol consumption	Week 1 (n=31), median (IQR[a])	Week 2 (n=30), median (IQR)	Week 3 (n=29), median (IQR)	Week 4 (n=31), median (IQR)
Drinking days	4.0 (3.0-5.0)	3.0 (3.0-4.0)	3.0 (3.0-4.0)	3.0 (2.0-3.0)
Drink free days	3.0 (2.0-4.0)	4.0 (3.0-4.0)	4.0 (3.0-4.0)	4.0 (4.0-5.0)
Units per drinking day	5.6 (3.3-11.8)	6.5 (2.3-9.1)	4.54 (2.3-8.9)	4.7 (2.0-6.9)
Units consumed	22.9 (14.3-32.4)	20.4 (14.6-25.0)	18.1 (12.7-26.3)	15.9 (11.6-26.9)
Alcoholic drinks per drinking day	2.0 (2.0-4.0)	3.0 (1.0-4.0)	2.0 (1.0-4.0)	2.0 (1.0-4.0)
Binge drinking days per week[b]	2.0 (1.0-3.0)	2.0 (1.0-2.0)	1.0 (0.0-2.0)	2.0 (1.0-2.5)

[a]IQR: interquartile range.

[b]Defined as having 6 or more alcoholic drinks in a session.

Table 6. Self-reported baseline and weekly measurement responses.

Variable	Day 0 (n=31)	Day 8 (n=25)	Day 15 (n=25)	Day 22 (n=21)	Day 28 (n=22)
Two item Generalized Anxiety Disorder Scale, median (IQR[a])	0 (0-1)	0 (0-0)	0 (0-1)	0 (0-0)	0 (0-0)
Two item patient health questionnaire, median (IQR)	0 (0-2)	0 (0-0)	0 (0-1)	0 (0-0)	0 (0-0)
Alcohol use disorders identification test, median (IQR)	11 (10-12)	N/A[b]	N/A	N/A	10 (8-12)
Self-efficacy, mean (SD)	6.7 (2.7)	5.9 (3)	4.9 (3.2)	6.3 (2.5)	4.5 (3.1)
Readiness to change, mean (SD)	4.4 (3.2)	4 (3.3)	3.4 (2.8)	4.9 (3.2)	3.7 (2.7)

[a]IQR: interquartile range.

[b]N/A: not applicable.

Discussion

Principal Findings

The aim of this paper was to design an engaging, responsive, and usable smartphone app that delivered personalized SMS text messages and gathered alcohol usage data. We tested the usability and feasibility, measured and reported as user engagement, of this app in a hard-to-engage ex-serving population. The InDEx app was codesigned by stakeholders and ex-serving personnel, with the results indicating successful user engagement and adherence. Based on the primary and secondary outcome measures, the participants used the app for the length of the study period, with two-thirds of participants using the app every week and the majority still using it in the final week (27/31, 87%). These engagement measures suggest that participants were highly active in using InDEx during the study period and that it is feasible to collect alcohol consumption data from this population. On average, most participants reported drinking on just under half of the days in the study period with participants reporting binge drinking on average 2 times a week. Reductions in units per drinking day and units consumed per week were observed across this 4 week study (yet the average number of drinks remained consistent); however, it is not possible to determine whether this may be due to participants changing the size and alcohol content of their drinks in this small feasibility study.

In this study, the most frequently opened page was the "Dashboard," the "Drinks Diary" page was the second most frequently accessed, and the "Add Drinks" page was third. The top 3 most viewed pages accounted for 78.20% (8236/10529) of all app views, indicating that most participants used the InDEx app primarily for monitoring drinks and the other features were not used as frequently. InDEx offered the ability to set a goal using an if-then format; however, participants used this feature rarely even after encouragement to set a goal via SMS text message and in-app prompts. This may be due to the sample not believing that they have a problem or being unable to navigate to and set a goal, which will be explored further in future work.

We applied behavior change theory [28] to create a smartphone app that incorporated a tailored SMS text messaging framework in an attempt to engage with users who are usually hard to reach [47-49]. It is difficult to ascertain if, and to what extent, SMS text messages encouraged alcohol reduction or app engagement. Future work is needed to assess the relationship between receiving a SMS text message and engagement with the app. The InDEx app takes advantage of a delivery method that circumvents practical and psychological barriers by utilizing digital technology. Participants were compensated for registering but had no financial incentive to use the app for the study period; nevertheless, they spent a median of 4 weeks engaging with the app.

InDEx has features not offered in other currently available alcohol apps [17,18,23]. First, it offers a user-centered and personalized design; features (ie, normative feedback) of the app were generated through codesign discussions with stakeholders and ex-serving personnel and developed using an iterative development framework to ensure that they were properly focused. The second major facet of the app was the use of BCTs in conjunction with data collected via the app to personalize the SMS text messages sent to participants. These features exploit contemporary technology which, as our feasibility study suggests, has the potential to promote the acceptability of InDEx and encourages users to engage with the app to record and thereby self-monitor their alcohol consumption. Third, InDEx is focused on reducing alcohol use among those meeting criteria for hazardous to harmful alcohol use (who may not recognize that they have a problem with alcohol), unlike other studies which have sought to support recovery for alcohol dependency (alcoholism) [21].

To the authors' knowledge, this was the first study to use SMS text messages embedded in an app to specifically focus on improving engagement and alternative behavior related to the individual alcohol consumption of ex-serving personnel. Although several studies have sought to investigate the impact that SMS text messages and tailoring can have on adherence, the combined use of the 2 strategies within the framework of a mobile app has never been attempted before.

Limitations

Notwithstanding the study strengths, our findings have some limitations. First, the baseline weekly alcohol consumption data were self-reported, albeit using reliable, consistent, and "gold standard" measurements. As with all self-report measures, recall and social desirability biases may have impacted responses to be more favorable than if collected using objective methods, such as transdermal alcohol monitoring [50,51]. Second, participants were asked to use the InDEx app for 4 weeks. Although the app appears feasible and acceptable to users based on engagement measurements during the study period, this study was not designed to ascertain the long-term benefits. Third, the sample size and design were appropriate for feasibility testing but not for assessing the efficacy of the app. Fourth, participants were recruited via the King's Centre for Military Health Research and offered an incentive to take part, resulting in a possible selection bias because participants had consented to participate in a research study previously. Finally, we studied InDEx in isolation and did not directly compare it with other app-based interventions.

Conclusions

In summary, the results of this study suggest that the InDEx app was feasible to implement and acceptable to participants, who typically engaged with the app for most of the study duration. It was feasible that participants reduced alcohol consumption during the study period, but this needs to be specifically addressed in a randomized controlled trial. Future research is needed to evaluate the engagement with and efficacy of InDEx for the reduction of alcohol consumption and binge drinking in an armed forces population.

Acknowledgments

This study was funded by the Medical Research Council (MR/N028244/2). We thank Margaret Jones of the King's Centre for Military Health Research, King's College London, for her assistance during the development of the InDEx app. The views expressed are those of the authors and do not necessarily reflect the views of the Department of Health, the National Institute for Health Research, Ministry of Defence or the Medical Research Council.

Conflicts of Interest

NTF is a member of the Independent Group Advising on the Release of Data for NHS (National Health Service) Digital and a trustee of The Warrior Programme. NJ is a serving member of the UK armed forces, but was not directed in any way by the Ministry of Defence. CD is partly funded by the National Institute for Health Research (NIHR) Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and King's College London and partly funded by the NIHR Collaborations for Leadership in Applied Health Research and Care South London at King's College Hospital NHS Foundation Trust.

Multimedia Appendix 1

Online supplement: Infographic representing the InDEx ecosystem.

[PNG File, 933KB - mhealth_v6i9e10074_app1.png]

Multimedia Appendix 2

List of alcohol types and categories included in the InDEx App.

[PDF File (Adobe PDF File), 19KB - mhealth_v6i9e10074_app2.pdf]

References

1. Fear N, Iversen A, Meltzer H, Workman L, Hull L, Greenberg N, et al. Patterns of drinking in the UK Armed Forces. Addiction 2007 Nov;102(11):1749-1759. [doi: 10.1111/j.1360-0443.2007.01978.x] [Medline: 17935583]
2. Fear N, Jones M, Murphy D, Hull L, Iversen A, Coker B, et al. What are the consequences of deployment to Iraq and Afghanistan on the mental health of the UK armed forces? A cohort study. The Lancet 2010 May 22;375(9728):1783-1797. [doi: 10.1016/S0140-6736(10)60672-1] [Medline: 20471076]
3. Rona RJ, Jones M, Fear NT, Hull L, Hotopf M, Wessely S. Alcohol misuse and functional impairment in the UK Armed Forces: a population-based study. Drug Alcohol Depend 2010 Apr 01;108(1-2):37-42. [doi: 10.1016/j.drugalcdep.2009.11.014] [Medline: 20047802]
4. Babor T, Higgins-Biddle J, Saunders J, Monteiro M. The alcohol use disorders identification test. In: Guidelines for use in primary care Internet. Geneva, Switzerland: World Health Organization; 2001.
5. McManus S, Bebbington P, Jenkins R, Brugha T, editors. Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014. In: NHS Digital. Leeds, England: NHS Digital; 2014.
6. Garnett C, Crane D, West R, Michie S, Brown J, Winstock A. Normative misperceptions about alcohol use in the general population of drinkers: a cross-sectional survey. Addict Behav 2015 Mar;42:203-206 [FREE Full text] [doi: 10.1016/j.addbeh.2014.11.010] [Medline: 25482365]
7. Hines LA, Goodwin L, Jones M, Hull L, Wessely S, Fear NT, et al. Factors affecting help seeking for mental health problems after deployment to Iraq and Afghanistan. Psychiatr Serv 2014 Jan 01;65(1):98-105. [doi: 10.1176/appi.ps.004972012] [Medline: 24037454]
8. Jones E, Fear NT. Alcohol use and misuse within the military: a review. Int Rev Psychiatry 2011 Apr;23(2):166-172. [doi: 10.3109/09540261.2010.550868] [Medline: 21521086]
9. Linke S, Brown A, Wallace P. Down your drink: a web-based intervention for people with excessive alcohol consumption. Alcohol Alcohol 2004;39(1):29-32. [Medline: 14691071]
10. Kypri K, Langley JD, Saunders JB, Cashell-Smith ML, Herbison P. Randomized controlled trial of web-based alcohol screening and brief intervention in primary care. Arch Intern Med 2008 Mar 10;168(5):530-536. [doi: 10.1001/archinternmed.2007.109] [Medline: 18332300]
11. Leightley D, McPhee J, Yap M. Automated Analysis and Quantification of Human Mobility Using a Depth Sensor. IEEE J Biomed Health Inform 2017 Dec;21(4):939-948. [doi: 10.1109/JBHI.2016.2558540] [Medline: 27254874]
12. Fowler LA, Holt SL, Joshi D. Mobile technology-based interventions for adult users of alcohol: A systematic review of the literature. Addict Behav 2016 Dec;62:25-34. [doi: 10.1016/j.addbeh.2016.06.008] [Medline: 27310031]
13. Bowles K, Dykes P, Demiris G. The use of health information technology to improve care and outcomes for older adults. Res Gerontol Nurs 2015;8(1):5-10 [FREE Full text] [doi: 10.3928/19404921-20121222-01] [Medline: 25625621]

14. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015 May 14;17(5):e118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
15. Weaver ER, Horyniak DR, Jenkinson R, Dietze P, Lim MS. "Let's get Wasted!" and Other Apps: Characteristics, Acceptability, and Use of Alcohol-Related Smartphone Applications. JMIR Mhealth Uhealth 2013 Jun 25;1(1):e9 [FREE Full text] [doi: 10.2196/mhealth.2709] [Medline: 25100681]
16. 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]
17. Attwood S, Parke H, Larsen J, Morton K. Using a mobile health application to reduce alcohol consumption: a mixed-methods evaluation of the drinkaware track & calculate units application. BMC Public Health 2017 Dec 17;17(1):394 [FREE Full text] [doi: 10.1186/s12889-017-4358-9] [Medline: 28511698]
18. Milward J, Deluca P, Drummond C, Watson R, Dunne J, Kimergård A. Usability Testing of the BRANCH Smartphone App Designed to Reduce Harmful Drinking in Young Adults. JMIR Mhealth Uhealth 2017 Aug 08;5(8):e109 [FREE Full text] [doi: 10.2196/mhealth.7836] [Medline: 28790022]
19. Kuntsche E, Knibbe R, Gmel G, Engels R. Who drinks and why? A review of socio-demographic, personality, and contextual issues behind the drinking motives in young people. Addict Behav 2006 Oct;31(10):1844-1857. [doi: 10.1016/j.addbeh.2005.12.028] [Medline: 16460883]
20. Winstock A. Global Drug Survey. 2017. DrinksMeter App URL: http://www.drinksmeter.com/ [accessed 2018-01-02] [WebCite Cache ID 6wAWecD3f]
21. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
22. Garnett C, Crane D, West R, Brown J, Michie S. Identification of Behavior Change Techniques and Engagement Strategies to Design a Smartphone App to Reduce Alcohol Consumption Using a Formal Consensus Method. JMIR Mhealth Uhealth 2015 Jun 29;3(2):e73 [FREE Full text] [doi: 10.2196/mhealth.3895] [Medline: 26123578]
23. Garnett C, Crane D, Michie S, West R, Brown J. Evaluating the effectiveness of a smartphone app to reduce excessive alcohol consumption: protocol for a factorial randomised control trial. BMC Public Health 2016 Dec 08;16:536 [FREE Full text] [doi: 10.1186/s12889-016-3140-8] [Medline: 27392430]
24. Black N, Mullan B, Sharpe L. Computer-delivered interventions for reducing alcohol consumption: meta-analysis and meta-regression using behaviour change techniques and theory. Health Psychol Rev 2016 Sep;10(3):341-357. [doi: 10.1080/17437199.2016.1168268] [Medline: 26999311]
25. Drinkaware. 2018. Drink Aware - Drink less alcohol URL: https://www.drinkaware.co.uk/tools/app/ [accessed 2018-05-21] [WebCite Cache ID 6zZtHhyXH]
26. NHS England. 2017. One You Drinks Tracker URL: https://www.nhs.uk/pages/home.aspx [accessed 2018-05-21] [WebCite Cache ID 6zZtobwPF]
27. Quanbeck A, Chih M, Isham A, Gustafson D. Mobile Delivery of Treatment for Alcohol Use Disorders: A Review of the Literature. Alcohol Res 2014;36(1):111-122 [FREE Full text] [Medline: 26259005]
28. Michie S, Whittington C, Hamoudi Z, Zarnani F, Tober G, West R. Identification of behaviour change techniques to reduce excessive alcohol consumption. Addiction 2012 Aug;107(8):1431-1440. [doi: 10.1111/j.1360-0443.2012.03845.x] [Medline: 22340523]
29. Redfern J, Thiagalingam A, Jan S, Whittaker R, Hackett ML, Mooney J, et al. Development of a set of mobile phone text messages designed for prevention of recurrent cardiovascular events. Eur J Prev Cardiol 2014 Apr;21(4):492-499. [doi: 10.1177/2047487312449416] [Medline: 22605787]
30. 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]
31. Hotopf M, Hull L, Fear N, Browne T, Horn O, Iversen A, et al. The health of UK military personnel who deployed to the 2003 Iraq war: a cohort study. The Lancet 2006 May;367(9524):1731-1741. [doi: 10.1016/S0140-6736(06)68662-5]
32. Ionic Framework. 2013. URL: https://ionicframework.com/ [accessed 2018-01-02] [WebCite Cache ID 6wAW1gmfr]
33. Github Inv. Atom Software Package URL: https://atom.io/ [accessed 2017-07-25] [WebCite Cache ID 6sDl3Gi3X]
34. Leightley D, Puddephatt J, Goodwin L, Rona R, Fear N. InDEx: Open Source iOS and Android Software for Self-Reporting and Monitoring of Alcohol Consumption. J Open Res Softw 2018 Mar 23;6:13 [FREE Full text] [doi: 10.5334/jors.207] [Medline: 29795769]
35. Schwarzer R. Self-efficacy in the adoptionmaintenance of health behaviors: Theoretical approachesa new model. In: Schwarzer R, editor. Self-efficacy: Thought control of action. Washington, DC, US: Hemisphere Publishing Corp; 1992:217-243.
36. Schwarzer R, Luszczynska A. How to Overcome Health-Compromising Behaviors. European Psychologist 2008 Jan;13(2):141-151. [doi: 10.1027/1016-9040.13.2.141]
37. Gollwitzer PM, Sheeran P. Implementation intentions and goal achievement: a meta‐analysis of effects and processes. Advances in Experimental Social Psychology 2006 May 01;38:69-119. [doi: 10.1016/S0065-2601(06)38002-1]

38. Keith C. GDCONF. 2008. An Agile Retrospective URL: https://tinyurl.com/yatuwjqp [accessed 2018-01-02] [WebCite Cache ID 6wAVvYvtA]
39. There's no place like phone - Global Mobile Consumer Survey 2016. Deloitte LLP. 2016. URL: http://www.deloitte.co.uk/mobileuk2016/assets/pdf/Deloitte-Mobile-Consumer-2016-There-is-no-place-like-phone.pdf
40. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010;8(4):348-353 [FREE Full text] [doi: 10.1370/afm.1139] [Medline: 20644190]
41. Saunders J, Aasland O, Babor T, de LFJR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804. [Medline: 8329970]
42. Kroenke K, Spitzer R, Williams J, Monahan P, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 06;146(5):317-325. [Medline: 17339617]
43. Prins A, Bovin M, Smolenski D, Marx B, Kimerling R, Jenkins-Guarnieri M, et al. The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5): Development and Evaluation Within a Veteran Primary Care Sample. J Gen Intern Med 2016 Oct;31(10):1206-1211 [FREE Full text] [doi: 10.1007/s11606-016-3703-5] [Medline: 27170304]
44. Readiness Ruler. In: Cleveland, Ohio. Cleveland, Ohio: Center for Evidence-Based Practices, Case Western Reserve University; 2010.
45. Health and Social Care Information Centre. In: Statistics on Alcohol, England 2016. Leeds, England: Health and Social Care Information Centre; 2016.
46. Iacoviello BM, Steinerman JR, Klein DB, Silver TL, Berger AG, Luo SX, et al. Clickotine, A Personalized Smartphone App for Smoking Cessation: Initial Evaluation. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e56 [FREE Full text] [doi: 10.2196/mhealth.7226] [Medline: 28442453]
47. Brief D, Rubin A, Enggasser J, Roy M, Keane T. Web-Based Intervention for Returning Veterans with Symptoms of Posttraumatic Stress Disorder and Risky Alcohol Use. J Contemp Psychother 2011 Dec 01;41(4):237-246 [FREE Full text] [doi: 10.1007/s10879-011-9173-5] [Medline: 25378713]
48. Wooten NR, Tavakoli AS, Al-Barwani MB, Thomas NA, Chakraborty H, Scheyett AM, et al. Comparing behavioral health models for reducing risky drinking among older male veterans. Am J Drug Alcohol Abuse 2017 Dec;43(5):545-555 [FREE Full text] [doi: 10.1080/00952990.2017.1286499] [Medline: 28410002]
49. Lapham GT, Hawkins EJ, Chavez LJ, Achtmeyer CE, Williams EC, Thomas RM, et al. Feedback from recently returned veterans on an anonymous web-based brief alcohol intervention. Addict Sci Clin Pract 2012 Aug 28;7:17 [FREE Full text] [doi: 10.1186/1940-0640-7-17] [Medline: 23186354]
50. Davis CG, Thake J, Vilhena N. Social desirability biases in self-reported alcohol consumption and harms. Addict Behav 2010 Apr;35(4):302-311. [doi: 10.1016/j.addbeh.2009.11.001] [Medline: 19932936]
51. Leffingwell T, Cooney N, Murphy J, Luczak S, Rosen G, Dougherty D, et al. Continuous objective monitoring of alcohol use: twenty-first century measurement using transdermal sensors. Alcohol Clin Exp Res 2013 Jan;37(1):16-22 [FREE Full text] [doi: 10.1111/j.1530-0277.2012.01869.x] [Medline: 22823467]

Abbreviations

AUDIT: alcohol use disorders identification test
BCT: behavior change technique
GAD: generalized anxiety disorder
HAPA: Health Action Process Approach
InDEx: Information about Drinking for Ex-serving
IQR: interquartile range
PHQ: patient health questionnaire
SMS: short message service

Edited by C Dias; submitted 08.02.18; peer-reviewed by C Garnett, D Fung, J Borodovsky, W Boendermaker; comments to author 28.02.18; revised version received 25.04.18; accepted 11.06.18; published 11.09.18.

Please cite as:
Leightley D, Puddephatt JA, Jones N, Mahmoodi T, Chui Z, Field M, Drummond C, Rona RJ, Fear NT, Goodwin L
A Smartphone App and Personalized Text Messaging Framework (InDEx) to Monitor and Reduce Alcohol Use in Ex-Serving Personnel: Development and Feasibility Study
JMIR Mhealth Uhealth 2018;6(9):e10074
URL: http://mhealth.jmir.org/2018/9/e10074/
doi:10.2196/10074
PMID:30206054

Original Paper

Evaluating the Carrot Rewards App, a Population-Level Incentive-Based Intervention Promoting Step Counts Across Two Canadian Provinces: Quasi-Experimental Study

Marc Mitchell[1], PhD; Lauren White[2], MSc; Erica Lau[3], PhD; Tricia Leahey[4], PhD; Marc A Adams[5], PhD; Guy Faulkner[3], PhD

[1]Western University, London, ON, Canada
[2]Carrot Insights Inc, Toronto, ON, Canada
[3]University of British Columbia, Vancouver, BC, Canada
[4]University of Connecticut, Storrs, CT, United States
[5]Arizona State University, Phoenix, AZ, United States

Corresponding Author:
Marc Mitchell, PhD
Western University
Somerville House, Room 2360C
London, ON, N6A 5B9
Canada
Phone: 1 519 661 2111 ext 87936
Email: marc.mitchell@uwo.ca

Abstract

Background: The Carrot Rewards app was developed as part of an innovative public-private partnership to reward Canadians with loyalty points, exchangeable for retail goods, travel rewards, and groceries for engaging in healthy behaviors such as walking.

Objective: This study examined whether a multicomponent intervention including goal setting, graded tasks, biofeedback, and very small incentives tied to daily step goal achievement (assessed by built-in smartphone accelerometers) could increase physical activity in two Canadian provinces, British Columbia (BC) and Newfoundland and Labrador (NL).

Methods: This 12-week, quasi-experimental (single group pre-post) study included 78,882 participants; 44.39% (35,014/78,882) enrolled in the Carrot Rewards "Steps" walking program during the recruitment period (June 13–July 10, 2016). During the 2-week baseline (or "run-in") period, we calculated participants' mean steps per day. Thereafter, participants earned incentives in the form of loyalty points (worth Can $0.04) every day they reached their personalized daily step goal (ie, baseline mean+1000 steps=first daily step goal level). Participants earned additional points (Can $0.40) for meeting their step goal 10+ nonconsecutive times in a 14-day period (called a "Step Up Challenge"). Participants could earn up to Can $5.00 during the 12-week evaluation period. Upon meeting the 10-day contingency, participants could increase their daily goal by 500 steps, aiming to gradually increase the daily step number by 3000. Only participants with ≥5 valid days (days with step counts: 1000-40,000) during the baseline period were included in the analysis (n=32,229).The primary study outcome was mean steps per day (by week), analyzed using linear mixed-effects models.

Results: The mean age of 32,229 participants with valid baseline data was 33.7 (SD 11.6) years; 66.11% (21,306/32,229) were female. The mean daily step count at baseline was 6511.22. Over half of users (16,336/32,229, 50.69%) were categorized as "physically inactive," accumulating <5000 daily steps at baseline. Results from mixed-effects models revealed statistically significant increases in mean daily step counts when comparing baseline with each study week (P<.001). Compared with baseline, participants walked 115.70 more steps (95% CI 74.59 to 156.81; P<.001) at study week 12. BC and NL users classified as "high engagers" (app engagement above sample median; 15,511/32,229, 48.13%) walked 738.70 (95% CI 673.81 to 803.54; P<.001) and 346.00 (95% CI 239.26 to 452.74; P<.001) more steps, respectively. Physically inactive, high engagers (7022/32,229, 21.08%) averaged an increase of 1224.66 steps per day (95% CI 1160.69 to 1288.63; P<.001). Effect sizes were modest.

Conclusions: Providing very small but immediate rewards for personalized daily step goal achievement as part of a multicomponent intervention increased daily step counts on a population scale, especially for physically inactive individuals and individuals who engaged more with the walking program. Positive effects in both BC and NL provide evidence of replicability.

(JMIR Mhealth Uhealth 2018;6(9):e178) doi:10.2196/mhealth.9912

KEYWORDS

behavioral economics; financial health incentives; mHealth; mobile phone; physical activity; public health

Introduction

The health benefits of regular physical activity are unquestionable. Regular moderate-intensity physical activity, brisk walking, for example, reduces the risk of several noncommunicable diseases, such as type 2 diabetes [1,2]. Regular physical activity has also been shown to improve cognition [3], prevent and manage depression [4], and prevent or delay the onset of dementia [5]. Furthermore, a recent analysis of objectively measured physical activity (n=5562 American adults) determined that participation in moderate-intensity physical activity was associated with substantial reduction in mortality risk [6]. For women, even modest participation in *low-intensity* physical activity, for example, slower walking without "huffing and puffing," was linked with lower mortality risk [6]. Unfortunately, physical inactivity remains a global pandemic [7,8]. Conservative estimates suggest that this pandemic cost the global economy US $53.8 billion in direct health care expenses in 2013 [9]. In Canada, as in most higher-income countries, the public sector bears the largest proportion of health care expenditures attributable to physical inactivity [9].

Behavioral economics, a branch of economics complimented by insights from psychology [10], has stimulated interest in using financial health incentives to promote physical activity [11]. Financial health incentives are defined as rewards with monetary value contingent on achievement of prespecified health behaviors or outcomes [12], such as rewarding people to walk more [13] or to lose weight [14]. One way timely financial incentives might work, according to behavioral economics, is by leveraging people's predictable tendency to act in favor of their immediate self-interest, a principal referred to as "present bias" [10]. In the case of physical activity, the likelihood that someone will be more physically active should increase if a financial incentive is at stake—and the more immediate the incentive, the stronger the nudge, according to this theoretical perspective [15].

Evidence supporting the use of financial health incentives is growing, with 2 systematic reviews [13,16] and 1 meta-analysis [17] finding that incentives generally increase physical activity in the short-term (≤3 months) and while they are still in place (ie, *before* they are withdrawn). However, evidence regarding *sustained* physical activity increases (ie, *after* incentives are removed) is more mixed, with some randomized controlled trials (RCTs) reporting postintervention benefits [18-21] and others not [22-24]. Finkelstein et al (2016) conducted the largest (N=800) of these trials and found that physical activity was higher among incentive group participants at 6 months, but this effect was not sustained 6 months after incentive removal [24]. The authors suggest that study design (eg, intervention duration), sample characteristics (eg, baseline physical activity), and incentive features (eg, generic, not tailored, physical activity goals) may have moderated postintervention responses. Discrepant findings and a still limited number of studies suggest that more research is needed to elucidate conditions under which incentives are more likely to drive postintervention changes.

In some cases, however, offering incentives for longer periods may be suitable, as Finkelsetin et al (2016) suggest—until a time when physical activity motives are internalized ("I walk because it makes me feel good") or until clinically meaningful health outcomes are achieved [24]. While acknowledging that more research is needed [25], the 3 RCTs that have tested physical activity incentives for ≥6 months have reported significant, positive effects [22,24,26]. However, the cost of longer term incentive programs may be prohibitive, especially if offered on a population scale. Therefore, at the same time research continues to examine conditions under which incentives drive sustained, long-term changes, efforts to increase efficiency, and thus scalability, of incentive interventions are also needed. The incentive magnitude typically used to promote physical activity in RCT settings (ie, US $1-US $2 per day) [15,19,22,23,27,28] may be simply too high for third-party payers and real-world implementation.

To reduce the cost of incentives and realistically operate within fixed government or insurer budgets, several incentive program features or reinforcement properties can be manipulated (eg, size, immediacy, probability, timing, type of incentive) [11,12,29,30]. For example, by shortening the time between behavior and reward so that rewards are delivered immediately after desired responses, the reward size needed to stimulate physical activity may decrease [11]. Smartphone technology presents an opportunity to provide incentives immediately upon physical activity goal completion (eg, steps per day). Built-in smartphone accelerometers now make it easier to track physical activity (ie, since the Apple Inc. iOS Health Kit app launched in 2014) [31]; furthermore, previously unavailable moment-by-moment physical activity data can now be used to set and personalize physical activity goals and provide immediate feedback in the form of rewards (eg, rewards automatically transmitted to Web-based accounts). Also, loyalty points (ie, points given by retailers to promote customer loyalty) have emerged as a promising new incentive type (vs cash, vouchers, or charity donations) [32-34]. Research shows that consumers tend to overvalue the points they collect (eg, although US $1 cash may have stimulated physical activity in the past, US $0.50 in loyalty points may produce the same effect) [35], possibly lowering the reward size needed to stimulate physical activity. These intervention features (using smartphones to track and reward physical activity with loyalty points) may appeal to governments and insurers looking to deploy financial health incentives more efficiently.

In Canada, such features are now available via the Carrot Rewards app, a new mHealth initiative that rewards Canadians with loyalty points (eg, retail goods, travel, groceries) to engage in healthy behaviors (eg, visiting flu shot clinic, walking) [34,36,37]. This study's purpose was to examine whether the

Carrot Rewards "Steps" walking program, which utilizes very small incentives (Can $ 0.04 in loyalty points) tied to daily step goal achievements could stimulate physical activity in two Canadian provinces.

Methods

Background

Carrot Insights Inc. is a private company that developed the free Carrot Rewards app with support from the Public Health Agency of Canada. The British Columbia (BC) Ministry of Health was the company's founding provincial Ministry partner. Newfoundland and Labrador (NL) was the second Canadian province to offer the app to its residents. Carrot Rewards was made available for BC and NL residents on the Apple iTunes and Google Play app stores on March 3 and June 13, 2016, respectively, in both English and French (Canada's official languages). Upon downloading the app, the users were asked to enter their age, gender, postal code, and loyalty program card number to complete registration (users without loyalty cards were directed to an easy sign-up page). To register successfully, users must have entered a valid BC or NL postal code and have been ≥13 years (age cutoff of participating loyalty programs). The walking program was not initially available in BC, but was introduced the day the app launched in NL. Carrot Insights Inc. partnered with 4 major Canadian loyalty programs to offer a variety of popular incentives (ie, points could be redeemed for groceries, travel, movies, or gas). While BC users could earn points via any of the 4 participating loyalty programs, NL users could earn points only for the 2 loyalty programs with a regional presence (ie, movies and travel). In addition to the 4 participating loyalty programs, Carrot Insights Inc. also partnered with 4 Canadian health charities (ie, Heart and Stroke Foundation of Canada, Diabetes Canada, Young Men's Christian Association Canada, and the BC Healthy Living Alliance), primarily for the purpose of reviewing and approving health education content offered in the app. The Behavioural Research Ethics Board of the University of British Columbia approved this study (UBC BREB Number H17-02814).

Recruitment

The marketing assets of the 4 loyalty programs and 1 charity partner were leveraged so that in the first few weeks, partners could heavily promote the app in both provinces (ie, in BC, partners sent 1.64 million emails to their loyalty members; in NL, the number of emails is unknown). The users were not automatically enrolled in the walking program, but were rather asked to opt-in. Study recruitment was open for approximately 1 month from June 13 to July 10, 2016. To participate, users had to agree to allow the app to access step data tracked and stored in their smartphones and were rewarded Can $0.60 in loyalty points for doing so.

Study Participants and Design

Registered users from BC (n=65,414) and NL (n=13,468) were eligible to participate in the walking program. However, only those with iPhone version 5S or higher could participate (ie, the Health Kit app, step data aggregator, is supported and preinstalled on these devices). Android smartphone users could also participate, but they were required to download the Health Kit equivalent (ie, Google Fit app) first. Only those who enabled the walking program on their smartphones (ie, allowed the app to access their data) received the intervention. From June 13 to July 10, 2016, 78,882 users from two Canadian provinces (BC and NL) were eligible to participate in the walking program, and 44.39% (35,014/78,882) ultimately activated it on their smartphones during the recruitment period. To examine the effect of this multicomponent intervention on objectively measured daily step counts, a 12-week quasi-experimental (single group pre-post) study design was employed. Testing the walking program simultaneously in 2 provinces provided a direct replication condition.

Theoretical Underpinnings

This intervention was theoretically based on principles from behavioral economics and self-determination theory. While behavioral economics describes how incentives exploit "present bias" to *stimulate* behaviors [10], self-determination theory focuses on the extent to which behaviors are controlled by external agents (eg, physicians) or contingencies (eg, incentives) and can be *sustained* [38]. A more thorough review of how these theories complement each other in a financial health incentive context is presented elsewhere [39]. Briefly, timely in-app notifications ("Congrats! You have achieved your 6600 daily step goal!"), very small incentives (not to be overly controlling and to protect autonomy), and a personalized approach to goal setting (realistic daily step goals, so users experience success early) were deployed to maintain fidelity to both behavioral economics and self-determination theory. As well, a range of behavior change techniques [40] are embedded in the app, including goal setting, self-monitoring, and biofeedback (ie, feedback using an external monitoring device), and graded tasks (ie, set at "easy" and then their difficulty increased).

Baseline Period

For a personalized walking goal to be generated (ie, steps per day), users must have accumulated at least 5 valid days during the initial 14-day baseline or "run-in" period. A valid day was defined as any day with step counts from 1000 to 40,000, as these numbers were considered reasonable, not outliers [41]. Days with step counts <1000 were considered days smartphones were not worn, and days with step counts above 40,000 were deemed suspiciously high (eg, technology bug) and were excluded. For users with at least 5 valid days, a daily step count average was calculated for the baseline period, and 1000 steps were added to set the first daily step goal (rounded to the nearest 100 steps). If users did not have a sufficient number of valid days (ie, ≤4 days) during the baseline period, a generic 5000 daily step goal was provided and they were excluded from analysis. The approximate the number of steps taken daily by the average Canadian adult is 5000, as measured by a popular smartphone-based activity tracking app [42].

Program

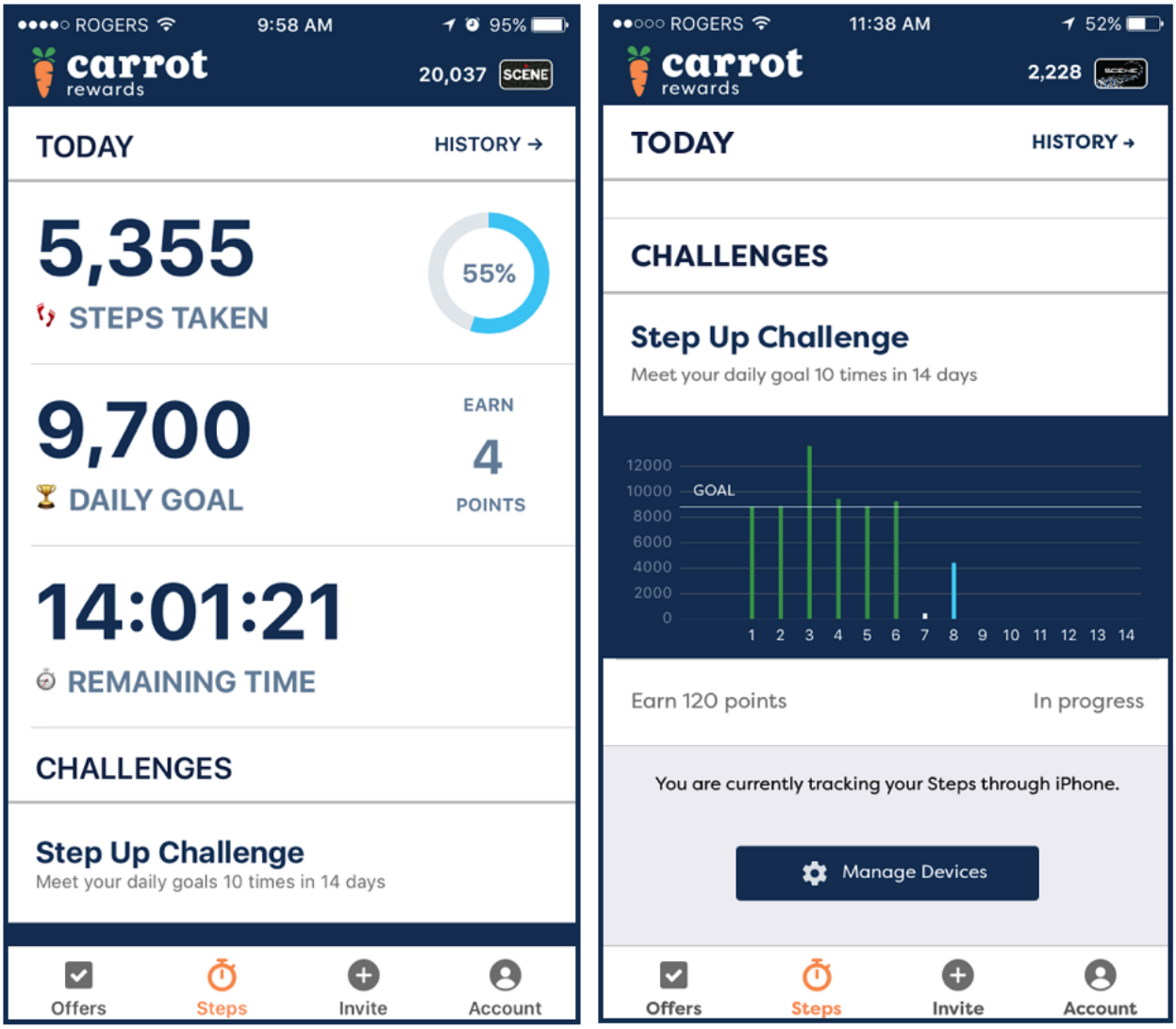
After the 14-day baseline period, users could begin to earn incentives for reaching or exceeding their individualized daily step goals; a progress wheel illustrated progress for the day (see Figure 1 for walking program screenshots). Incentives for daily

achievements were worth Can $0.04 in loyalty points. After 2 weeks of earning daily rewards in the form of points, users could then begin to earn bonus rewards worth Can $0.40 in points for reaching their daily goal ≥10 nonconsecutive times within a 14-day period, called a "Step Up Challenge." Incentives for longer term (eg, biweekly) physical activity goals, in addition to daily goals only, have worked well in past studies [24]. Users were automatically enrolled in the first "Step Up Challenge," but thereafter always had to accept the challenge when it became available. A bar graph to illustrate "Step Up Challenge" progress was also made available upon tapping "Accept" in the app (see Figure 1). For users who successfully completed the "Step Up Challenge," a new higher daily step goal was provided (ie, 500 steps more than the previous goal). For unsuccessful users, the previous goal persisted. Over the 3-month evaluation period, participants could earn a total of Can $5.00 in points (Can $0.60 for activating the walking program, Can $2.80 for daily step goal achievements, and Can $1.60 for successfully completing 4 "Step Up Challenges").

Outcome Measures

The primary outcome variable was mean daily step counts as measured by either built-in smartphone accelerometers, for example, iPhone 5S or higher for 53.63% (42,304/78,882) of users, Android devices for 37.48% (29,565/78,882) of users, or any Fitbit device for 7.18% (5664/78,882) of users. Recent validation studies found that the iPhone step counting feature (version 6 or newer), as well as those for Android smartphones (eg, HTC, Motorola) and Fitbit trackers (eg, hip-worn Zip, wrist-worn Flex) were accurate in laboratory and field conditions [43-45]. However, Duncan et al (2018) did determine that steps were underestimated by the iPhone step counting feature in their free-living condition by approximately 1340 steps per day [43]. According to the study authors, this likely reflects not carrying the iPhone continually throughout the day rather than inaccuracy in the step counting feature; they suggest that if adherence can be optimized, smartphones may be suitable for physical activity evaluations.

Figure 1. Carrot Rewards app's "Steps" walking program screenshots.



Covariates

The majority of demographic variables used to describe the study sample were self-reported (eg, age, gender, province). Median personal income was inferred by linking user postal codes with census data (ie, 2011 National Household Survey) at the local health area level (89) in BC and regional health authority level (4) in NL.

Data Analyses

Three different analytical approaches were used to account for missing data and to test the sensitivity of our assumptions with the analytical sample: (1) The "any" data approach included participants with valid baseline data (≥5 days in acceptable range during the 14-day baseline period) and at least 1 other valid week (ie, at least 4 valid days in a 7-day week) from study week 1 to 12 (32,229/35,019, 92.03% of those enabling the walking program met these criteria); (2) the "completer" approach included just participants with valid data at baseline and study week 12 (19,964/32,229, 61.94%); and (3) the "imputed" approach included participants with valid baseline data, but no valid data at study week 12 (29,261/32,229, 90.79%). Then, we imputed participants' "Pseudo study week 12" by carrying forward their baseline values. Therefore, among those included in the analysis (n=32,229), 61.94% (19,964/32,229) had complete datasets (completers). No differences were observed in demographic characteristics between completers and noncompleters (see Table 1). Since the 3 different analytic approaches yielded very similar results, given the public health nature of the intervention and that completers did not differ from noncompleters on key demographic characteristics, analyses using the "any" data approach are presented.

Statistical analysis was performed using R 3.3.0.68 Mavericks build (7202) Rstudio Version 1.0.136 (RStudio, Boston, MA, USA). Study week was treated as a categorical variable (baseline=0, study week 1=1, ..., study week 12=12) to allow for the nonlinear trajectory of daily step counts. Also, the estimate for each study week helped refine the program to maintain user engagement. Mixed-effects models were performed to examine whether there were significant changes in mean daily step counts between baseline and study week 12. We fitted a simple linear mixed-effects model that included study week as the independent variable (baseline data were used as the reference), followed by an adjusted model with random intercepts to account for measurements nesting within individuals and by controlling for age, gender, median personal income, and province as covariates. Analyses were performed on the entire sample, and participants were stratified by physical activity status as defined by Tudor-Locke et al [46] (ie, physically inactive: baseline mean steps per day<5000; physically active: baseline mean steps per day≥5000) and by province (ie, BC and NL).

As suggested by previous studies [47], we examined whether participants' engagement levels had a moderating effect on intervention outcome. Two additional variables, engagement and study week × engagement, were tested in all models. Engagement was a variable dichotomizing all participants into 2 categories, "high" or "low" engagers, based on the median percentage of days when a "Step Up Challenge" was accepted. The interaction term allows the difference between high and low engagers to differ at baseline and study week 12, while controlling for their baseline values and other covariates. Cohen f^2 for local effect sizes of mean daily step counts within mixed-effects models were calculated, with $f^2 \geq 0.02$, $f^2 \geq 0.15$, and $f^2 \geq 0.35$ representing small, medium, and large effect sizes, respectively [48]. Least-square means along with *P* values were obtained from mixed-effects models for comparing mean daily step counts between subgroups. All data were expressed in least-square means with 95% CIs. Statistical significance levels were set at $P<.05$.

Table 1. Baseline characteristics of Carrot Rewards users, by completion status, and for the general Canadian population.

Characteristics	Completers[a] (n=19,964)	Noncompleters[b] (n=12,265)	Canadian population (N=35,151,728)
Age in years, mean (SD)	33.8 (11.4)	33.5 (11.9)	40.6 (median)
Gender (% female)	66.1	66.1	50.4
Province (% British Columbia)	72.1	70.3	13.2
Median personal income (Can $1000/year), mean (SD)	29.7 (4.1)	29.6 (4.0)	33.9
Steps per day, baseline mean (SD)	6665.6 (4220.7)	6157.5 (4388.9)	N/A[c]
Engagement[d] (% high)	59.4	19.8	N/A

[a]Participants with valid data at baseline and study week 12.

[b]Participants with valid data at baseline, but not at study week 12.

[c]N/A: not applicable.

[d]A variable dichotomizing participants into 2 categories, "high" or "low" engagers, based on the median percentage of days when a "Step Up Challenge" was accepted.

Results

Baseline Characteristics

The mean age of the 32,229 participants with valid baseline data was 33.7 (SD 11.6) years; 66.11% (21,306/32,229) were female (Table 1). Participants from BC made up 71.41% (23,016/32,229) of the study sample owing to the province's larger population and to the app launching 3 months prior to its launch in NL. The mean personal median income was Can $29,650, slightly lower than that of 2014 BC and NL means of Can $31,610 and Can $30,450, respectively [49]. The mean daily step count at baseline was 6511.22 steps per day. Just over half of users 50.69% (16,336/32,229) were categorized as "physically inactive," having accumulated <5000 daily steps at baseline. Assuming age, income, and province were held constant, male participants walked 2297.50 steps more steps per day at baseline compared with females (P<.001), and participants from NL walked 992.95 fewer steps per day than those from BC (P<.001).

Weekly Means

The trends of daily step counts for the total group and the physically inactive subgroup over the 12-week intervention period are illustrated in Figure 2. The difference between baseline and the 12-week evaluation period average for the total group (5.01%) and physically inactive participants (21.14%) are also illustrated. Error bars show 95% CIs. For the total, some behavioral decay was observed in later weeks as the weekly steps per day average dropped below the 12-week intervention mean (6864.77 steps) in study weeks 9 (6772.68 steps) through 12 (6626.92 steps). The average increase in daily step counts over the 12-week intervention period was 353.56 steps, which represents a 5.01% difference from baseline. Among physically inactive users, an average increase of 861.12 steps per day was observed, representing a 21.14% difference from baseline. There was no evidence of behavioral decay in this subgroup as weekly steps per day persisted at or above the intervention mean (4621.76 steps) in study weeks 9 (4622.22 steps) to 12 (4634.83 steps).

Figure 2. Least-square means for daily steps at baseline and for each study week during the 12-week evaluation period for the total sample and physically inactive participants.

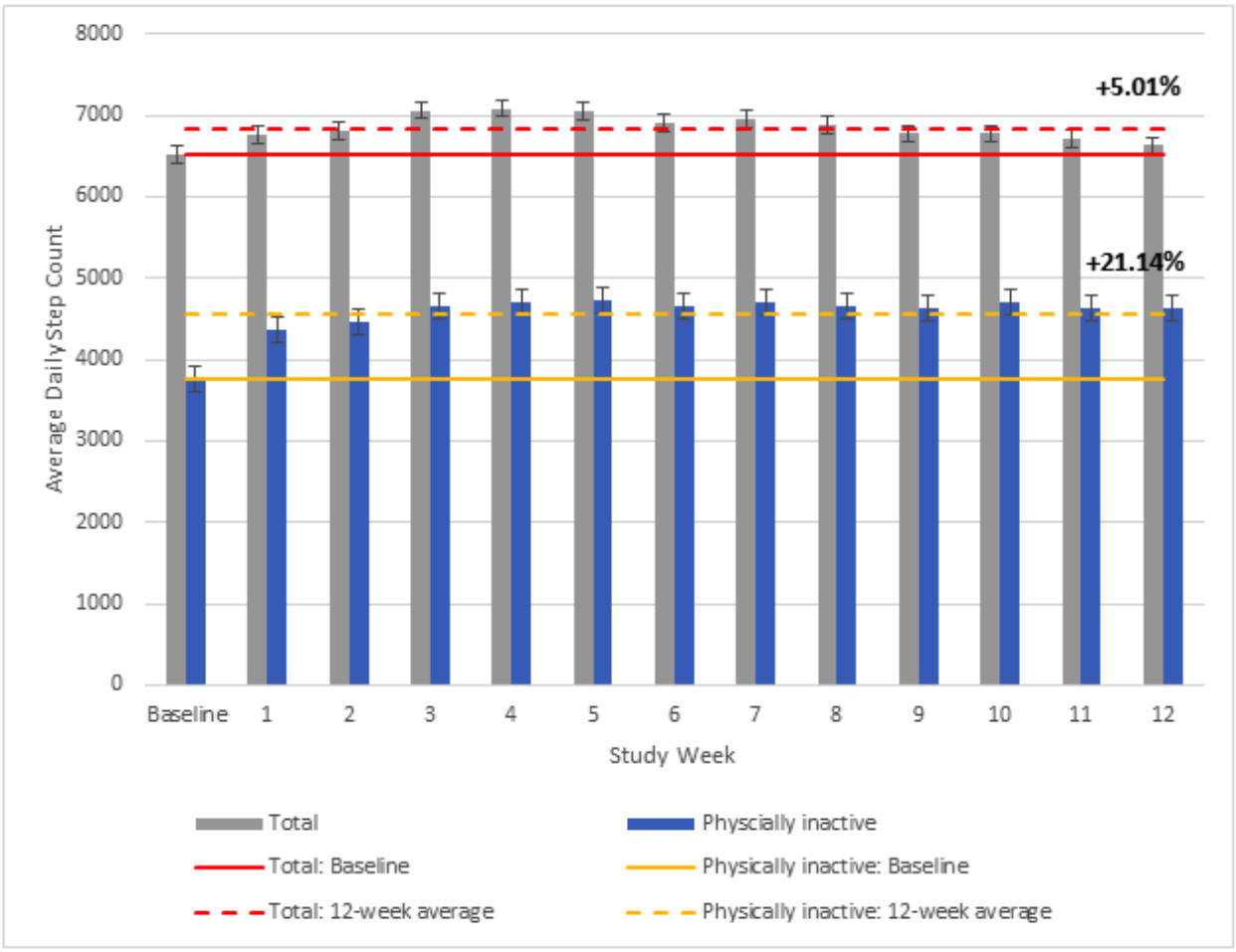


Table 2. Changes in mean daily step counts between baseline and study week 12.

Analysis	Baseline least-square means[a] (95% CIs)	Week 12 least-square means[a] (95% CIs)	Differences (Week 12 – baseline) least-square means[a] (95% CIs)	Cohen f^{2b}
Total sample analysis	6511.22 (6242.24 to 6780.19)	6626.92 (6357.34 to 6896.50)	115.70 (74.59 to 156.81)	0.0059
Subgroup analyses				
Physical activity status				
Physically inactive	3760.64 (3543.31 to 3977.96)	4634.83 (4416.56 to 4853.09)	874.19 (827.98 to 920.40)	0.0234
Physically active	8778.01 (8392.20 to 9163.81)	8297.19 (7910.38 to 8684.00)	−480.82 (−545.17 to −416.46)	0.0073
Province				
British Columbia	7064.82 (6796.12 to 7333.52)	7282.83 (7013.40 to 7552.26)	218.01 (169.56 to 266.46)	0.0061
Newfoundland and Labrador	6071.87 (5790.32 to 6353.43)	5938.22 (5654.07 to 6222.37)	−133.66[c] (−155.98 to −3.37)	0.0087
Engagement				
Low engager	6229.27 (5958.75 to 6497.80)	5738.52 (5466.67 to 6010.37)	−490.75 (−551.21 to −428.29	0.0073
High engager	6780.37 (6509.93 to 7050.81)	7411.27 (7140.55 to 7681.99)	630.90 (575.43 to 686.36)	N/A[d]
Physically inactive, low engager	3650.63 (3432.54 to 3868.72)	4132.55 (3 911.07 to 4354.04)	481.92 (414.62 to 549.22)	0.0055
Physically inactive, high engager	3838.79 (3628.93 to 4068.65)	5073.45 (4853.31 to 5293.59)	1224.66 (1160.69 to 1288.63)	N/A
Physically active, low engager	8588.00 (8200.39 to 8975.61)	7258.26 (6866.58 to 7649.93)	−1329.74 (−1427.93 to −1231.56)	0.0096
Physically active, high engager	8928.63 (8540.95 to 9316.31)	9131.09 (8742.81 to 9519.36)	202.26 (117.20 to 287.72)	N/A
British Columbia, low engager	6771.26 (6500.00 to 7042.52)	6353.11 (6078.918 to 6627.31)	−418.15 (−491.12 to −345.17)	0.0071
British Columbia, high engager	7316.68 (7044.75 to 7588.61)	8055.38 (7783.15 to 8327.61)	738.70 (673.81 to 803.54)	N/A
Newfoundland and Labrador, low engager	5769.62 (5480.24 to 6059.00)	5120.22 (4822.23 to 5418.21)	−649.40 (−763.50 to −535.30)	0.0074
Newfoundland and Labrador, high engager	6369.92 (6072.63 to 6667.21)	6715.92 (6417.20 to 7014.64)	346.00 (239.26 to 452.74)	N/A

[a]Least-square means adjusted for age, median personal income, gender, and province.

[b]Cohen f^2≥0.02, ≥0.15, and ≥0.35 representing small, medium, and large effect sizes, respectively. For the engagement subgroup analysis only, Cohen f^2 was calculated for the pre-post difference in steps between the low and high engagement groups (high engagement as the referent group).

[c]The difference between baseline and week 12 were statistically significant at P<.001 for total sample and all subgroup analyses, except for Province Newfoundland and Labrador (P<.001).

[d]N/A: not applicable.

Total Sample Analysis

The results from mixed-effects models revealed statistically significant increases in mean daily step counts when comparing baseline with each study week (P<.001). Changes in mean daily step count from baseline to study week 12 expressed in least-square means are presented in Table 2. Overall, compared with baseline, participants walked 115.70 more steps (95% CI 74.59 to 156.81; P<.001) at study week 12. The Cohen f^2 value was 0.0059 (P<.001), indicating the effect was modest. Adjusting for demographic variables (ie, age, gender, province, and median personal income) had little effect on the estimated difference between study week 12 and baseline.

Subgroup Analysis

The intervention effect was more pronounced in physically inactive users than in physically active users. As with the total sample analysis, the mean daily steps were significantly higher for physically inactive users at each study week than at baseline (P<.001), with an observed increase of 874.19 steps per day at

study week 12 (Table 2; 95% CI 827.98 to 920.40, P<.001). Cohen f^2 statistic indicated that the effect was small (0.0234, P<.001). At study week 12, compared with baseline, a highly significant decrease of 480.82 steps per day was observed among physically active participants (Table 2; 95% CI −545.17 to −416.46, P<.001, Cohen f^2=0.0073, P<.001). Participants from NL did not respond as well as participants from BC. At study week 12, compared with baseline, a highly significant increase of 218.01 was observed in BC (Table 2; 95% CI 169.56 to 266.46, P<.001, Cohen f^2=0.0061, P<.001), while a highly significant decrease of 133.66 steps per day was observed in NL (Table 2; 95% CI −155.98 to −3.37, P<.001, Cohen f^2=0.0087, P<.001).

Moderation Analysis

Participant engagement showed a significant moderating effect on the intervention outcome in all models (P<.001). Therefore, we also conducted subgroup analysis by participants' engagement levels. As shown in Table 2, all subgroups except physically active low engagers showed significant increase in step counts from baseline to study week 12. The difference from baseline to study week 12 for high (15,511/32,229; 48.13%) and low engagers (16,718/32,229; 51.87%) was +630.90 and −489.75 steps per day, respectively (P<.001). As well, users classified as high engagers in BC and NL walked 738.70 and 346.00 more steps per day, respectively (P<.001). Among users classified as high engagers and physically inactive (7,022/32,229; 21.08%), an average increase of 1224.66 steps per day was observed (P<.001).

Discussion

Principal Findings

In this large quasi-experimental study examining the impact of a multicomponent intervention on objectively measured daily step count, a small but significant effect overall was observed (5% average daily step count increase over 12 weeks vs baseline) with a more pronounced effect (21% increase) among physically inactive users (representing over half of the total sample). Notably, this effect was evident irrespective of age, gender, or median personal income. While the overall effect was small (ie, 116 steps per day), these results underscore the potential public health impact of using modest incentives (Can $ 0.04 per day) to stimulate physical activity, particularly among higher risk, physically inactive populations. When considering the clinical significance of this study's results, it is likely that health benefits (eg, better glucose control) [1] might be reserved for 51% of the analytic sample that increased their daily step counts by 874 steps per day (the physically inactive). Health economic implications of initiatives like this may be important, especially considering that a mere 1% reduction in the number of Canadians classified as physically inactive would yield annual health care savings of Can $2.1 billion [50]. The combination of immediate rewards in the form of loyalty points tied to smartphone-assessed physical activity outcomes may prove an efficient way of delivering financial health incentives while still producing a measurable effect.

Other reinforcement-based methods of increasing health behaviors have included using deposit contracts (ie, participants wager their own money) [51], chance-based designs (ie, 1 in 3 chance of earning Can $3 vs just Can $1 per day) [52] and loss-framing (ie, incentive given up front and then *taken away* if goal unmet) [23]. While deposit contract, chance-based and loss-framed designs may be effective, they may also limit enrollment (in the case of deposits) and may be less palatable to governments or insurers looking to deploy such programs (eg, raising concerns about gambling or punishing citizens or employees for not meeting health goals) [53]. This study provides evidence that even very small incentives, as modest as Can $0.04 per day, can be implemented as part of a multicomponent intervention and on a population scale to increase walking and other ambulatory behaviors effectively.

Attrition

Behavioral decay (ie, steps per day decline) was noted as time passed, with weekly steps per day averages dropping below the intervention mean in later weeks. While this was observed in the total sample (driven by the 480.82 daily step count *reduction* among physically active users), step counts persisted throughout the 12-week evaluation period in the physically inactive subgroup. At study week 12, for example, physically inactive participants were walking 874.19 more steps per day on average (vs baseline). This is consistent with incentives for physical activity literature that suggests that physically inactive adults are more sensitive to incentive interventions and more likely to sustain the behavior for longer periods [17]. Similarly, larger intervention effect sizes are observed among insufficiently active individuals in Web-based physical activity interventions [54]. Why daily step counts *decreased* among physically active participants remains unclear. Seasonal effects may partly explain the drop (the evaluation period began in warmer spring and summer seasons and ended in the colder fall). Smartphone (ie, accelerometer) wear time may also explain the decrease. Physically active users, being generally less sensitive to physical activity incentives, may have carried their smartphones less and less (and recorded fewer and fewer steps) as the intervention progressed.

Provincial Differences

Regarding provincial differences, NL users did not respond as well as BC users (−133.66 steps per day vs +218.01 steps per day at study week 12, respectively). This could be due to a number of factors. The most important factor may have to do with the walking program's availability to *all* NL participants right away (upon downloading the app), while BC users who were still engaging with the app 3 months after it launched could activate the walking program (self-selection bias). Additionally, these provinces are on opposite Canadian coasts, with distinct climates and chronic disease risk profiles. Regarding climate, in the final 3 weeks of the evaluation period (when the provincial step count disparity was greatest, ending on October 17, 2016), residents of St. John's, NL, experienced more "cold days" (ie, below our operational 13.0°C threshold) than their Vancouver, BC, counterparts; 43% (10/23) versus 13% (3/23) of days were "cold"; St. John's and Vancouver are the largest cities in NL and BC, respectively). Regarding chronic disease risk, while

BC has the lowest self-reported adult overweight and obesity rate in Canada (48.0%), NL has the highest (67.5%). Notably, while NL users in general experienced a 133.66 step per day decrease (with low engagers experiencing an even greater 649.40 steps per day drop), a 346.00 step per day increase was observed at study week 12 (vs baseline) among high engagers (3846/9209, 37.85% of the provincial sample). App engagement therefore appears to have boosted intervention effectiveness, regardless of province, suggesting potential effect replication in other jurisdictions. This aligns with broader evidence that greater engagement with a physical activity app or website is associated with increased intervention efficacy [55]. Developing innovative strategies to increase and maintain engagement is a priority (eg, machine learning informed push notifications when "Step Up Challenge" was not accepted within 3 days, rewards for just accepting challenges, small team-based challenges).

Limitations

The results of this population-level study should be interpreted with caution because there are a number of limitations to consider. First, neither the randomization of participants into intervention and control groups was logistically feasible within this quasi-experimental design nor was the identification of a nonequivalent control group (ie, a group not randomly assigned to receive or not receive the intervention) [56]. For this reason, internal validity (ie, the extent to which causality can be established) may be limited. To improve internal validity as much as possible in this real-world setting, we sought to define a time period that reflected the counter-factual (ie, outcome if the intervention had not been implemented) [56]. To do this, a preintervention time period clearly differentiated from the intervention was introduced. An immediate increase in daily step count compared with baseline was expected, and this is what was observed. This increase, however, may have occurred because participants simply started carrying their smartphones more (the most likely alternative explanation or rival hypothesis) to get credit for the steps they were taking. Disentangling "wear time" from increased actual daily step count is difficult, however, a limitation cited in more carefully controlled RCTs [24]. Additionally, more smartphone accelerometer validation studies are likely required in free-living conditions and with different demographic groups to increase confidence in results. Analysis-phase strategies were employed to improve internal validity as well, including (a) testing the sensitivity of assumptions made with 3 different analytic samples to handle missing data and (b) fitting an adjusted mixed-effects model to account for measurements nesting within individuals and controlling for key demographics. As well, an increase in steps in high, but not low, engagers provides further support for the main conclusion that this multicomponent intervention, when utilized above a threshold level, appears to have yielded daily step count improvements. That behavioral decay was noted in weeks 9-12 for the total sample, but not for the physically inactive subgroup (the group more likely to respond to an incentive-based intervention with realistic and personalized goals) also suggests that the intervention achieved its intended effect of stimulating physical activity among the least active. While traditional RCTs strongly prioritize internal validity, this quasi-experimental design seeks to achieve greater balance between internal and external validity in real-world conditions to facilitate real-world implementation. A second limitation was that participants were followed for only 12 weeks, so longitudinal work is required to elucidate longer term effects. Third, this analysis addressed only the earliest Carrot Rewards app adopters and includes just Canadian provinces, so results may not be generalizable to newer users or other countries. Next, only 44.39% (35,014/78,882) of eligible users who could enable the walking program and earn additional incentives did so during the 4-week recruitment period. How those who activated the program during the recruitment period compare with those who did not remains unknown. While on a population scale this recruitment rate is impressive, there is room to improve. The less than ideal recruitment rate may be because health app users in general discontinue use within days or weeks of first download [47] or a too-short recruitment period. Lastly, at what intensity any extra walking may have occurred is unknown. The association between physical activity and key health outcomes (eg, cardiovascular disease risk factor reduction) is stronger with higher intensity physical activities [6].

Future Research

To increase internal validity in this quasi-experimental environment, future studies might incorporate interrupted time series, stepped-wedge, intervention removal, or designs with a nonequivalent control group [56]. Future work might also compare different ways of setting and graduating daily step goals (eg, static vs adaptive goal setting) and include longitudinal analyses examining longer term (at least 6 months) impacts, as well as associated cost-effectiveness studies. For example, an adaptive goal setting feature was introduced in the app in February 2017 (after the study period), when step goals began to be recalculated every 2-4 weeks to encourage engagement (as opposed to the "set it and forget it" approach initially adopted). Examining alternative methods to promote sustained physical activity should continue to be a priority for researchers and others in this field (eg, moving from small, regularly scheduled incentives, to large, more irregular, and less predictable ones). To increase the chances of behavior maintenance, exploring opportunities for enhanced engagement that also promote social interaction and support could be a particular focus of future work (eg, encouraging social networking).

Conclusions

Until recently, financial health incentive programs have shown promise, but little potential for scalability given rewards' cost. This study adds to the understanding of how incentives can be delivered in ways that are not prohibitively costly. Providing immediate rewards for personalized daily step goal achievement as part of a multicomponent intervention appears to have increased daily step counts on a population scale, especially for higher risk, physically inactive individuals. Positive effects in both BC and NL provide evidence of replicability.

Acknowledgments

The Carrot Rewards initiative has been made possible in part through funding from the Public Health Agency of Canada. The views expressed herein do not necessarily represent the views of the Public Health Agency of Canada, the BC Ministry of Health, or the Ministry of Children, Seniors, and Social Development in NL. The authors also thank staff from the BC Ministry of Heath, the Ministry of Children, Seniors, and Social Development in NL, the BC Healthy Living Alliance, the Heart and Stroke Foundation of Canada, Diabetes Canada, and Young Men's Christian Association Canada for providing expert health content advice over the course of this project.

Conflicts of Interest

MM reports grant support from the Canadian Institutes of Health Research, the University Health Network, Green Shield Canada Inc, as well as in-kind research support from Cookson James Loyalty Inc Furthermore, he reports consulting income from Carrot Insights Inc and stock options in Carrot Insights Inc. LW is a Carrot Insights Inc employee and also reports stock options in Carrot Insights Inc. GF is supported by a Canadian Institutes of Health Research-Public Health Agency of Canada Chair in Applied Public Health. The authors with no financial relation to Carrot Insights Inc conducted the analyses (GF, EL). The other coauthors report no conflicts of interest.

References

1. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, et al. 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]
2. Warburton DE, Charlesworth S, Ivey A, Nettlefold L, Bredin SS. A systematic review of the evidence for Canada's Physical Activity Guidelines for Adults. Int J Behav Nutr Phys Act 2010 May 11;7:39 [FREE Full text] [doi: 10.1186/1479-5868-7-39] [Medline: 20459783]
3. Khan NA, Hillman CH. The relation of childhood physical activity and aerobic fitness to brain function and cognition: a review. Pediatr Exerc Sci 2014 May;26(2):138-146. [doi: 10.1123/pes.2013-0125] [Medline: 24722921]
4. Mammen G, Faulkner G. Physical activity and the prevention of depression: a systematic review of prospective studies. Am J Prev Med 2013 Nov;45(5):649-657. [doi: 10.1016/j.amepre.2013.08.001] [Medline: 24139780]
5. Blondell SJ, Hammersley-Mather R, Veerman JL. Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies. BMC Public Health 2014;14:510 [FREE Full text] [doi: 10.1186/1471-2458-14-510] [Medline: 24885250]
6. Borgundvaag E, Janssen I. Objectively Measured Physical Activity and Mortality Risk Among American Adults. Am J Prev Med 2017 Jan;52(1):e25-e31. [doi: 10.1016/j.amepre.2016.09.017] [Medline: 27816381]
7. Kohl HW, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, et al. The pandemic of physical inactivity: global action for public health. Lancet 2012 Jul 21;380(9838):294-305. [doi: 10.1016/S0140-6736(12)60898-8] [Medline: 22818941]
8. Sallis JF, Bull F, Guthold R, Heath GW, Inoue S, Kelly P, et al. Progress in physical activity over the Olympic quadrennium. Lancet 2016 Sep 24;388(10051):1325-1336. [doi: 10.1016/S0140-6736(16)30581-5] [Medline: 27475270]
9. Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van MW, et al. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet 2016 Sep 24;388(10051):1311-1324. [doi: 10.1016/S0140-6736(16)30383-X] [Medline: 27475266]
10. Camerer C, Loewenstein G. Behavioral Economics: Past, Present, Future. Advances in Behavioral Economics. Princeton: Princeton University Press; 2003.
11. Loewenstein G, Asch DA, Volpp KG. Behavioral economics holds potential to deliver better results for patients, insurers, and employers. Health Aff (Millwood) 2013 Jul;32(7):1244-1250 [FREE Full text] [doi: 10.1377/hlthaff.2012.1163] [Medline: 23836740]
12. Adams J, Giles EL, McColl E, Sniehotta FF. Carrots, sticks and health behaviours: a framework for documenting the complexity of financial incentive interventions to change health behaviours. Health Psychol Rev 2014;8(3):286-295. [doi: 10.1080/17437199.2013.848410] [Medline: 25053215]
13. Strohacker K, Galarraga O, Williams DM. The impact of incentives on exercise behavior: a systematic review of randomized controlled trials. Ann Behav Med 2014 Aug;48(1):92-99 [FREE Full text] [doi: 10.1007/s12160-013-9577-4] [Medline: 24307474]
14. Burns RJ, Donovan AS, Ackermann RT, Finch EA, Rothman AJ, Jeffery RW. A theoretically grounded systematic review of material incentives for weight loss: implications for interventions. Ann Behav Med 2012 Dec;44(3):375-388. [doi: 10.1007/s12160-012-9403-4] [Medline: 22907712]
15. Adams MA, Hurley JC, Todd M, Bhuiyan N, Jarrett CL, Tucker WJ, et al. Erratum to: Adaptive goal setting and financial incentives: a 2 × 2 factorial randomized controlled trial to increase adults' physical activity. BMC Public Health 2017 Dec 06;17(1):303 [FREE Full text] [doi: 10.1186/s12889-017-4231-x] [Medline: 28385148]
16. Barte JCM, Wendel-Vos GCW. A Systematic Review of Financial Incentives for Physical Activity: The Effects on Physical Activity and Related Outcomes. Behav Med 2017;43(2):79-90. [doi: 10.1080/08964289.2015.1074880] [Medline: 26431076]

17. Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am J Prev Med 2013 Nov;45(5):658-667. [doi: 10.1016/j.amepre.2013.06.017] [Medline: 24139781]
18. Charness G. Incentives to exercise. Econometrica 2009;77(3):909-931.
19. Kullgren JT, Harkins KA, Bellamy SL, Gonzales A, Tao Y, Zhu J, et al. A mixed-methods randomized controlled trial of financial incentives and peer networks to promote walking among older adults. Health Educ Behav 2014 Oct;41(1 Suppl):43S-50S [FREE Full text] [doi: 10.1177/1090198114540464] [Medline: 25274710]
20. Patel MS, Asch DA, Rosin R, Small DS, Bellamy SL, Eberbach K, et al. Individual Versus Team-Based Financial Incentives to Increase Physical Activity: A Randomized, Controlled Trial. J Gen Intern Med 2016 Mar 14. [doi: 10.1007/s11606-016-3627-0] [Medline: 26976287]
21. Andrade LF, Barry D, Litt MD, Petry NM. Maintaining high activity levels in sedentary adults with a reinforcement-thinning schedule. J Appl Behav Anal 2014;47(3):523-536 [FREE Full text] [doi: 10.1002/jaba.147] [Medline: 25041789]
22. Pope L, Harvey J. The efficacy of incentives to motivate continued fitness-center attendance in college first-year students: a randomized controlled trial. J Am Coll Health 2014;62(2):81-90. [doi: 10.1080/07448481.2013.847840] [Medline: 24456510]
23. Patel MS, Asch DA, Volpp KG. Framing Financial Incentives to Increase Physical Activity Among Overweight and Obese Adults. Ann Intern Med 2016 Dec 18;165(8):600. [doi: 10.7326/L16-0280] [Medline: 27750316]
24. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EEK, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
25. Mantzari E, Vogt F, Shemilt I, Wei Y, Higgins JPT, Marteau TM. Personal financial incentives for changing habitual health-related behaviors: A systematic review and meta-analysis. Prev Med 2015 Jun;75:75-85 [FREE Full text] [doi: 10.1016/j.ypmed.2015.03.001] [Medline: 25843244]
26. Jeffery RW, French SA. Preventing weight gain in adults: the pound of prevention study. Am J Public Health 1999 May;89(5):747-751. [Medline: 10224988]
27. Shin DW, Yun JM, Shin J, Kwon H, Min HY, Joh H, et al. Enhancing physical activity and reducing obesity through smartcare and financial incentives: A pilot randomized trial. Obesity (Silver Spring) 2017 Feb;25(2):302-310. [doi: 10.1002/oby.21731] [Medline: 28063226]
28. Kranker K. The Efficacy of Using Financial Incentives to Change Unhealthy Behaviors Among a Rural Chronically Ill and Uninsured Population. Am J Health Promot 2018 Feb;32(2):301-311. [doi: 10.1177/0890117117696621] [Medline: 28279086]
29. Mitchell MS, Goodman JM, Alter DA, Oh PI, Faulkner GEJ. Development of the Health Incentive Program Questionnaire (HIP-Q) in a cardiac rehabilitation population. Transl Behav Med 2015 Dec;5(4):443-459 [FREE Full text] [doi: 10.1007/s13142-015-0330-3] [Medline: 26622917]
30. Haff N, Patel MS, Lim R, Zhu J, Troxel AB, Asch DA, et al. The role of behavioral economic incentive design and demographic characteristics in financial incentive-based approaches to changing health behaviors: a meta-analysis. Am J Health Promot 2015;29(5):314-323. [doi: 10.4278/ajhp.140714-LIT-333] [Medline: 25928816]
31. Apple. www.apple.com. 2014. Apple Unveils iOS 8, the Biggest Release Since the Launch of the App Store URL: https://www.apple.com/ca/newsroom/2014/06/02Apple-Unveils-iOS-8-the-Biggest-Release-Since-the-Launch-of-the-App-Store/ [accessed 2018-08-29] [WebCite Cache ID 720c5yQfD]
32. Liu S, Hodgson C, Zbib AM, Payne AYM, Nolan RP. The effectiveness of loyalty rewards to promote the use of an Internet-based heart health program. J Med Internet Res 2014 Jul 02;16(7):e163 [FREE Full text] [doi: 10.2196/jmir.3458] [Medline: 24989982]
33. Goyal S, Morita PP, Picton P, Seto E, Zbib A, Cafazzo JA. Uptake of a Consumer-Focused mHealth Application for the Assessment and Prevention of Heart Disease: The <30 Days Study. JMIR Mhealth Uhealth 2016;4(1):e32 [FREE Full text] [doi: 10.2196/mhealth.4730] [Medline: 27012937]
34. Mitchell M, White L, Oh P, Alter D, Leahey T, Kwan M, et al. Uptake of an Incentive-Based mHealth App: Process Evaluation of the Carrot Rewards App. JMIR Mhealth Uhealth 2017 May 30;5(5):e70 [FREE Full text] [doi: 10.2196/mhealth.7323] [Medline: 28559224]
35. Van Osselaer SM, Alba J, Manchanda P. Irrelevant information and mediated intertemporal choice. Journal of Consumer Psychology 2004;14(3):257-270.
36. Government of Canada. www.canada.ca. 2015. National Healthy Living Platform: Carrot Rewards targets lifestyle improvements URL: https://www.canada.ca/en/news/archive/2015/07/national-healthy-living-platform-carrot-rewards-targets-lifestyle-improvements.html [accessed 2018-08-29] [WebCite Cache ID 720cRoNby]
37. Dale LP, White L, Mitchell M, Faulkner G. Smartphone app uses loyalty point incentives and push notifications to encourage influenza vaccine uptake. Vaccine 2018 Apr 23. [doi: 10.1016/j.vaccine.2018.04.018] [Medline: 29699784]
38. Deci E, Ryan R. Handbook of Self-Determination Research. Rochester, NY: University of Rochester Press; 2002.
39. Mitchell M, Faulkner G. A Nudge at All? The Jury Is Still Out on Financial Health Incentives. Healthcare Papers 2013;12(4).

40. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
41. Bassett DR, Wyatt HR, Thompson H, Peters JC, Hill JO. Pedometer-measured physical activity and health behaviors in U.S. adults. Med Sci Sports Exerc 2010 Oct;42(10):1819-1825 [FREE Full text] [doi: 10.1249/MSS.0b013e3181dc2e54] [Medline: 20305579]
42. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Dec 20;547(7663):336-339 [FREE Full text] [doi: 10.1038/nature23018] [Medline: 28693034]
43. Duncan MJ, Wunderlich K, Zhao Y, Faulkner G. Walk this way: validity evidence of iphone health application step count in laboratory and free-living conditions. J Sports Sci 2017 Nov 28:1-10. [doi: 10.1080/02640414.2017.1409855] [Medline: 29179653]
44. 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]
45. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
46. Tudor-Locke C, Craig CL, Thyfault JP, Spence JC. A step-defined sedentary lifestyle index: <5000 steps/day. Appl Physiol Nutr Metab 2013 Feb;38(2):100-114 [FREE Full text] [doi: 10.1139/apnm-2012-0235] [Medline: 23438219]
47. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, Present, and Future of eHealth and mHealth Research to Improve Physical Activity and Dietary Behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
48. Selya AS, Rose JS, Dierker LC, Hedeker D, Mermelstein RJ. A Practical Guide to Calculating Cohen's f(2), a Measure of Local Effect Size, from PROC MIXED. Front Psychol 2012;3:111 [FREE Full text] [doi: 10.3389/fpsyg.2012.00111] [Medline: 22529829]
49. Table17-10-0005-01 Population estimates on July 1st, by age and sex and Table 11-10-0008-01 Tax filers and dependants with income by total income, sex and age database on the Internet. 2017 URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1110000801&pickMembers%5B0%5D=1.35 [accessed 2018-06-29] [WebCite Cache ID 720ceEOsb]
50. Krueger H, Turner D, Krueger J, Ready AE. The economic benefits of risk factor reduction in Canada: tobacco smoking, excess weight and physical inactivity. Can J Public Health 2014;105(1):e69-e78. [Medline: 24735700]
51. Washington W, McMullen D, Devoto A. A Matched Deposit Contract Intervention to Increase Physical Activity in Underactive and Sedentary Adults. Translational Issues is Psychology Science 2016;2(2).
52. Leahey TM, Subak LL, Fava J, Schembri M, Thomas G, Xu X, et al. Benefits of adding small financial incentives or optional group meetings to a web-based statewide obesity initiative. Obesity (Silver Spring) 2015 Jan;23(1):70-76 [FREE Full text] [doi: 10.1002/oby.20937] [Medline: 25384463]
53. Promberger M, Brown RCH, Ashcroft RE, Marteau TM. Acceptability of financial incentives to improve health outcomes in UK and US samples. J Med Ethics 2011 Nov;37(11):682-687 [FREE Full text] [doi: 10.1136/jme.2010.039347] [Medline: 21670321]
54. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
55. Schoeppe S, Alley S, Van LW, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
56. Handley MA, Lyles CR, McCulloch C, Cattamanchi A. Selecting and Improving Quasi-Experimental Designs in Effectiveness and Implementation Research. Annu Rev Public Health 2018 Apr 01;39:5-25. [doi: 10.1146/annurev-publhealth-040617-014128] [Medline: 29328873]

Abbreviations

BC: British Columbia
NL: Newfoundland and Labrador
RCT: randomized controlled trials

Edited by G Eysenbach; submitted 25.01.18; peer-reviewed by E Mantzari, K De Cocker; comments to author 08.03.18; revised version received 03.05.18; accepted 29.06.18; published 20.09.18.

Please cite as:
Mitchell M, White L, Lau E, Leahey T, Adams MA, Faulkner G
Evaluating the Carrot Rewards App, a Population-Level Incentive-Based Intervention Promoting Step Counts Across Two Canadian Provinces: Quasi-Experimental Study
JMIR Mhealth Uhealth 2018;6(9):e178
URL: http://mhealth.jmir.org/2018/9/e178/
doi:10.2196/mhealth.9912
PMID:30148712

Original Paper

Longitudinal Validity and Reliability of Brief Smartphone Self-Monitoring of Diet, Stress, and Physical Activity in a Diverse Sample of Mothers

Dallas Swendeman[1*], PhD, MPH; Warren Scott Comulada[1*], Dr PH; Maryann Koussa[1*], MPH; Carol M Worthman[2*], PhD; Deborah Estrin[3*], PhD, MS; Mary Jane Rotheram-Borus[1*], PhD; Nithya Ramanathan[4*], PhD

[1]Department of Psychiatry and Biobehavioral Sciences, David Geffon School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[2]Department of Anthropology, Emory University, Atlanta, GA, United States

[3]Cornell Tech, Cornell University, New York, NY, United States

[4]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, United States

[*]all authors contributed equally

Corresponding Author:
Nithya Ramanathan, PhD
Department of Computer Science
University of California, Los Angeles
10920 Wilshire Boulevard
Los Angeles, CA, 90024
United States
Phone: 1 213 915 6729
Fax: 1 213 302 5247
Email: nithyaar@gmail.com

Abstract

Background: Multiple strategies can be used when self-monitoring diet, physical activity, and perceived stress, but no gold standards are available. Although self-monitoring is a core element of self-management and behavior change, the success of mHealth behavioral tools depends on their validity and reliability, which lack evidence. African American and Latina mothers in the United States are high-priority populations for apps that can be used for self-monitoring of diet, physical activity, and stress because the body mass index (BMI) of mothers typically increases for several years after childbirth and the risks of obesity and its' sequelae diseases are elevated among minority populations.

Objective: To examine the intermethod reliability and concurrent validity of smartphone-based self-monitoring via ecological momentary assessments (EMAs) and use of daily diaries for diet, stress, and physical activity compared with brief recall measures, anthropometric biomeasures, and bloodspot biomarkers.

Methods: A purposive sample (n=42) of primarily African American (16/42, 39%) and Latina (18/42, 44%) mothers was assigned Android smartphones for using Ohmage apps to self-monitor diet, perceived stress, and physical activity over 6 months. Participants were assessed at 3- and 6-month follow-ups. Recall measures included brief food frequency screeners, physical activity assessments adapted from the National Health and Nutrition Examination Survey, and the nine-item psychological stress measure. Anthropometric biomeasures included BMI, body fat, waist circumference, and blood pressure. Bloodspot assays for Epstein–Barr virus and C-reactive protein were used as systemic load and stress biomarkers. EMAs and daily diary questions assessed perceived quality and quantity of meals, perceived stress levels, and moderate, vigorous, and light physical activity. Units of analysis were follow-up assessments (n=29 to n=45 depending on the domain) of the participants (n=29 with sufficient data for analyses). Correlations, R^2 statistics, and multivariate linear regressions were used to assess the strength of associations between variables.

Results: Almost all participants (39/42, 93%) completed the study. Intermethod reliability between smartphone-based EMAs and diary reports and their corresponding recall reports was highest for stress and diet; correlations ranged from .27 to .52 (P<.05). However, it was unexpectedly low for physical activity; no significant associations were observed. Concurrent validity was demonstrated for diet EMAs and diary reports on systolic blood pressure (r=−.32), C-reactive protein level (r=−.34), and moderate and vigorous physical activity recalls (r=.35 to.48), suggesting a covariation between healthy diet and physical activity behaviors.

EMAs and diary reports on stress were not associated with Epstein–Barr virus and C-reactive protein level. Diary reports on moderate and vigorous physical activity were negatively associated with BMI and body fat (r=−.35 to −.44, *P*<.05).

Conclusions: Brief smartphone-based EMA use may be valid and reliable for long-term self-monitoring of diet, stress, and physical activity. Lack of intermethod reliability for physical activity measures is consistent with prior research, warranting more research on the efficacy of smartphone-based self-monitoring of self-management and behavior change support.

(JMIR Mhealth Uhealth 2018;6(9):e176) doi:10.2196/mhealth.9378

KEYWORDS

self-monitoring; mHealth; diet; physical activity; stress; multi-method; mobile phones; C-reactive protein

Introduction

Background

Smartphones are increasingly used and integrated into daily routines, creating opportunities for continuous, real-time data streams of health behaviors and states [1,2]. Such data streams include self-reports, such as ecological momentary assessments (EMAs) and daily diaries [3]. The reliability and validity of smartphone apps for self-monitoring health behaviors is not yet fully understood but is critical for generating an evidence on the growing field of mobile health or "mHealth" [1] and for its broad adoption by consumers [4].

Diet, stress, and physical activity are the key lifestyle factors associated with a broad range of physical and mental health issues, such as obesity, diabetes, cardiovascular disease, depression, and anxiety [5]; for example, cardiovascular disease is a significant health problem that is still largely ignored by women, especially young women [6], although it accounts for a higher mortality rate than all forms of cancer in women [7]. Mothers are included in a high-priority target population because their body mass index (BMI) typically increases by 5 kg for several years after giving birth [8].

Smartphones are well suited for real-time self-monitoring using daily diaries and more frequent EMAs of diet, perceived stress, and physical activity because these behaviors and states can be difficult to recall precisely over longer periods of time and can vary significantly within and across days [3]. Smartphone-based EMA and diaries of target behaviors may be more specific and feasible than biomarkers and biomeasures (eg, BMI and blood pressure), which typically reflect the accumulated impact of several factors on physiological systems over time. Active self-monitoring (ie, via self-report) is also an important behavior change technique [9,10], particularly for the self-management of diet and physical activity [11]. Self-monitoring is nearly a universal behavior change element in smartphone apps for diet and physical activity self-management [12-14].

Smartphones lighten the burden of one aspect of EMAs by allowing data entry on a readily available device that is close at hand; cumbersome paper diaries or personal digital assistants of yesteryears are no longer needed. The intensity of EMA that requires daily reporting at various time-points throughout the day remains. Not surprisingly, decreases in adherence to mobile phone-reported EMAs over time have been noted for disparate outcomes, including nutrition, mood, and use of substance measures [15,16]. Therefore, it is important to determine which measures that need to be captured via EMAs and those that can be captured less frequently. Such an assessment has been a challenge due to limited validity and reliability studies in mobile phone-reported diet, stress, and physical activity measures, which is the focus of this study because there are no accepted gold standards for in situ assessments of these behaviors that can be readily used for objective comparison [17-19]. Studies have demonstrated discrepancies between retrospective self-reports and EMAs as well as their benefits and limitations [20-22]. Some studies have compared self-reported health behaviors to more objective measures, such as self-reported physical activity and that obtained using a pedometer [23-26]. In this study, we observed an intermethod reliability between smartphone EMAs and diary reports and their corresponding recall reports in addition to EMAs and recall between health measures, such as diet and exercise, that we anticipate to be correlated to each other. Such studies on ethnic minorities and women are also limited [11,27]. Toward this goal, we examined data from a feasibility study that pilot-tested a health-behavior self-monitoring mobile app in a sample that mostly included ethnic minority mothers; a prior study has examined the predictors of self-monitoring adherence (BLINDED) [28]. This paper examined the validity and reliability of brief use of smartphone-based EMAs and daily diaries for diet, stress, and physical activity compared with those of the brief recall self-reports, simple anthropometric biomeasures (eg, weight and BMI), and laboratory biomarkers (ie, C-reactive protein and Epstein–Barr Virus) collected at 3-month intervals over 6 months. We evaluated the intermethod reliability and concurrent validity of the app, which are high priorities for mobile health, "mHealth," evidence development [1]. Different fields employ divergent conceptualizations of validity and reliability. However, in this study, intermethod reliability is broadly conceptualized as concurrency between different methods assessing the same domain, whereas concurrent validity is conceptualized as concurrency between the assessments of different but linked domains (ie, diet, stress, and physical activity), and it uses multiple assessment methods that provide information on the concurrent validity of methods.

Hypotheses

This study assessed several sets of interrelated hypotheses to evaluate the reliability and validity of the brief use of EMAs and diary questions designed for smartphone apps. First, we hypothesized that the brief use of smartphone-based EMAs and daily diary questions would demonstrate intermethod reliability through associations with their corresponding recall self-reports. Second, we hypothesized that the EMAs and daily diary measures would demonstrate concurrent validity through

associations with anthropometric biomeasures, bloodspot biomarkers (for stress), and recall and EMAs and diary reports on other domains. Third, given that EMAs are designed to be independent and captured close to or in the moment, we hypothesized that EMAs would be more reliable than daily diaries for diet and perceived stress, which may be difficult to recall due to a high variability throughout the day. A study indicating that daily diary reports are comparable with EMA would suggest that a less burdensome diary method as preferable for future applications. However, we hypothesized that the daily diaries for physical activity would be sufficient to achieve minimal recall biases. Thus, EMAs for physical activity were not assessed to minimize burden.

Methods

Participatory Sensing

Development and study procedures were based on a participatory sensing approach used in mobile phone sensing projects developed by computer scientists [29]. User-centered design principles prioritize participant autonomy and choice over which features of the app to use (eg, responding to surveys) and recognize varying ability and motivation to adopt and sustain the activities. Similar to pragmatic designs in implementation research [30], participatory sensing's emphasis on naturalistic use prioritizes the external validity or generalizability of the sensing tool used across diverse user preferences, participation options, and motivations for participation. In this study, user preferences were collected through focus groups (BLINDED) and iterative trials with participants. The participatory sensing approach was used as a basis for designing the questions for EMA and diary questions which are brief, engaging, and meaningful for self-monitoring using the app for self-management rather than being granular and precise as a gold standard that is more typical in basic behavioral EMA studies.

Ethics Statement

The institutional review board of the University of California Los Angeles reviewed and approved the study. All participants signed informed consent. The study was conducted in accordance with the Declaration of Helsinki.

Recruitment and Participants

Mothers residing in an urban area who had at least one child living in the household were recruited to participate in the study from January 2012 to September 2012. Recruitment flyers framed the study as seeking support to develop and pilot-test a smartphone app that can help in the self-monitoring and self-management of diet, stress, and physical activity. Recruitment included weekly visits to local farmer markets; classes and groups at a community center; outreach at local grocery stores, churches, and targeted community organizations; and posting on local online groups regarding parenting and children. Women with a child below 18 years who is living at home, those who were not pregnant or breastfeeding, and those with a BMI ≤18.5 (ie, dangerously underweight) were included in the study. The recruitment plan aimed to include a sample of mothers with diverse BMIs (about one-third were normal in terms of weight, overweight, or obese) and those who were primarily African American and Latina. However, other race or ethnic groups were not excluded.

Measures

Smartphone Ecological Momentary Assessment and Daily Diary

Participants were assigned Samsung Vibrant smartphones and were instructed to complete the smartphone EMAs and daily diary surveys by responding to time-based (ie, alarm) prompts and event-based reports (ie, self-initiated). The participants selected the start and end times for three 3-hour windows for time-based EMA prompts (morning, midday, and late afternoon) and one end-of-day daily diary survey. The first three surveys asked about diet "since the last survey" and stressful events "over the last 2 hours." The end-of-day "daily diary" survey asked about diet, stress, and physical activity for the entire day. The following variables were used in the analyses:

Diet Ecological Momentary Assessment

Three times a day, the participants were asked the following question: "Rate the nutritional quality of this meal?" The responses were as follows: (1) low; (2) medium; and (3) high.

Diet Diary

At the end of the day, the participants were asked the following question: "How healthy would you rate your eating today, in terms of both quality and quantity, on a scale of 1 to 5 (with 5 being very healthy)?"

Stress Ecological Momentary Assessment

Three times a day, the participants were asked the following question: "Have you felt stressed in the last two hours?" The responses were as follows: (1) not at all, (2) slightly, (3) moderately, and (4) very.

Stress Diary:

At the end of the day, the participants were asked the following question: "How stressful was your day overall on a scale of 1-5 (with 5 being very stressful)?"

Physical Activity Diary

At the end of the day only, three questions were asked: "How many minutes of activity did you do today." The responses were as follows: for three intensities of physical activity, light (ie, "no increase in breathing or heart rate, eg, stretching"), moderate (ie, "small increase in breathing or hearth rate, eg, fast walking"), and vigorous ("significant increase in breathing or heart rate, eg, running"). To compare the recall reports for each activity type, the variables were calculated as number of days (over 30 days), minutes per day on average, and total minutes (minutes × days).

Baseline and Follow-up Recall Self-Reports

Measures were brief and reflective of the scope and scale of measures that are likely to be used in clinical practice or large-scale survey research. Recall periods were retained from the original measures, which were designed to minimize recall biases and capture general habits rather than detailed, gold-standard assessments.

Demographics Characteristics of the Participants

The background factors assessed included age, race or ethnicity, highest level of education, work hours per week (paid, volunteer, or student), and number of children living in the home.

Dietary Behaviors

Food frequency questionnaires from the California Health Interview Survey (CHIS) were used to assess diet; CHIS 2009 Adult questionnaire ver 3.4 (Public) March 1, 2011, Section C, p 32. These brief screening measures have a validity sufficient to discriminate higher or lower intakes among individuals, particularly for examining the relationships between diet and other variables, and they were used by the Applied Research Program in the Division of Cancer Control and Population Sciences at the National Cancer Institute for diet screening. Ten questions assessed the number of times a participant ate or drank various foods over the past 30 days, including prompts to estimate times per day, week, or month. Food types were categorized into three variables in the analyses: intake of fruits and vegetables (three questions on fruits, green vegetables or salad or beans, and nonfried potatoes), intake of food with high-sugar content (four questions on sugar-rich drinks or soda, sweetened fruit drinks, cookies, cake, and ice cream), and intake of fast food (one question over the past 7 days).

Perceived Stress

The brief, nine-item psychological stress measure (PSM-9), which is designed to assess for stress in primary care settings, was used [31]. The PSM-9 was developed from an original 49-item version and then two 25-item versions, which showed a high internal consistency (Cronbach alpha coefficients of 0.92 and 0.93) and test-retest reliability (0.68-0.80). Convergent, divergent, concomitant, and predictive validity were established by comparing the measures of various constructs, such as depression and anxiety [31]. PSM-9 questions were established according to both related and redundant contents from the longer item versions, covering domains, such as feeling calm, stressed, rushed, worried, confused, and energetic, physical symptoms, and difficulty controlling reactions or emotions.

Physical Activity

Physical activity was assessed using the questions from the CHIS 2009 survey that ask about the number of times per week (within the last 7 days) and average minutes per day for walking (for transport and recreation or physical activity), moderate activity (ie, breathe somewhat harder than normal), and vigorous activity (eg, aerobic sports, breathe significantly harder than normal). These questions are similar to those used in the National Health and Nutrition Examination Survey and other studies [17]. The two walking domains were combined, and the following variables were used for the analyses: walking (minutes per day, number of days, and total minutes), moderate physical activity (minutes per day, number of days, and total minutes), vigorous physical activity (minutes per day, number of days, and total minutes).

Biomeasures and Biomarkers

The biomeasures and biomarkers used in this study were selected for meeting the criteria on feasibility, acceptability, minimal invasiveness, and the comprehensive indicators of diet, stress, and physical activity. Simple anthropometric measures (height and weight to determine BMI, body fat measurements, and waist circumference) were used as indicators of physical activity [17]. Blood pressure was used as an indicator of overall health, which includes factors correlated to diet, physical activity, and stress and stressors [32,33]. Bloodspot C-reactive protein and Epstein–Barr virus antibodies were used as proxy measures of stress that respectively identified inflammation and cardiovascular risk and allostatic load correlated to a variety of stressors and stress-related impairment of innate immune capacity; this result is similar to that of prior research on maternal stress [34]. Although low-grade inflammation and weakened innate immunity have been linked to lifestyle and psychosocial factors, such as overnutrition, depression, and obesity [35-38], in the literature, stress is identified as a common cause of inflammation and alterations in innate immunity [39-41]. C-reactive protein and Epstein–Barr virus antibodies have been positively associated with stress, and they are often used as biomarkers of stress in research, as well as in this study [42-44]. Body fat was measured with the Tanita body composition analyzer (Model Tbf-300a) that uses bioelectrical impedance analysis, a commonly used method for estimating body composition [45,46]. Because fluctuations in hydration may affect body composition results, the participants were asked to avoid diuretics for 7 days, alcohol for 48 hours prior, intense physical activity for 12 hours prior, eating or drinking for 4 hours prior, emptying of bladder 30 minutes prior to assessment, and rescheduling if they became ill.

Body Mass Index

Height was measured in meters using a portable stadiometer, and the Tanita body composition analyzer scale was also used to measure weight in kilograms. Company and model information for the stadiometer is unknown. BMI was calculated as weight in kilograms divided by height in meters squared.

Waist Circumference

Waist measurements were taken using standardized procedures recommended in the Anthropometric Standardized Reference Manual [47]. The measuring tape was placed around the waist area (midway between the rib cage and hip bone). It was ensured that the tape was snug without compressing the skin and was parallel to the floor. Measurements were taken after the participant had exhaled, and the average of the three separate measurements was obtained.

Blood Pressure

Systolic and diastolic blood pressure were measured at the midpoint of the upper arm using an Omron automatic blood pressure monitor (HEM-705CP). The participants were asked to be seated with their back supported and legs uncrossed and were instructed to support their arm on a table so that the midpoint of their upper arm was at the level of the heart. Three readings were obtained. The first reading was discarded, and the average was taken using the final two readings [48].

Bloodspot Tests

A single finger stick from a microlancet (Becton-Dickson Microtainer contact-activated lancet, high blood flow 366594) was used to collect 5 drops of blood and place them onto the

preprinted filter paper card (250 µL), which is commonly used for neonatal screening and does not require immediate freezing [49]. Samples were labeled, dried for 4 hours, and stored in a plastic container in a locked refrigerator. Bloodspot samples were shipped each week to the laboratory and stored at −28°C until analysis for C-reactive protein and Epstein–Barr virus.

C-Reactive Protein

We used a validated biotin-streptavidin immunofluorometric assay for bloodspot C-reactive protein level, as reported elsewhere [50]. Based on the methods outlined herein [50], we developed an algorithm for serum equivalent values: serum (C-reactive protein value)=1.7*(bloodspot C-reactive protein value). The values indicate the extent of chronic low-grade systemic inflammation associated with cardiovascular and metabolic risk [51]. The C-reactive protein values were classified as low (<1 mg/L), medium (1-3 mg/L), and high risk (>3 mg/L; [52,53]). Values >10 mg/L indicate acute inflammation, and they were not used in the statistical analyses.

Epstein–Barr Virus Antibodies

Epstein–Barr virus antibody titers reflect the degree of immune response impairment. The enzyme-linked immunosorbent assay (ELISA) assay used for Epstein–Barr virus antibodies in blood spots is a modification of a commercially available kit (Number P001606A; DiaSorin Corporation, Stillwater, MN); method and validity have been reported elsewhere [49,54]. Antibody titers were presented as ELISA units. Values <20 indicate undetectable antibody levels [54], and they were not used in the statistical analyses. In the absence of standard health-risk categories for Epstein–Barr virus, we compared the mean levels of our sample to those from other samples obtained mostly from ethnic minority women in Illinois (n=183; mean Epstein–Barr virus level=136.8; [34])

Design

The participants (N=56) were randomly categorized into two groups. The experimental group (n=44) was asked to self-monitor their condition using Android smartphones with the self-monitoring app. The control group (n=12) was not provided with smartphones. The primary aim of the study was to assess the validity and reliability of using smartphones in measuring diet, stress, and physical activity. The control group was designed for secondary, preliminary efficacy aims not reported in this paper, and it did not show statistically significance in the preliminary analyses. Retrospective assessments and biomarker collection were conducted at baseline as well as 3 and 6 months after enrollment to estimate the time between clinical visits for moderately acute ill patients. The participants assigned to the smartphone group were instructed to use the app over the 6-month study period. Two of the 44 participants in the self-monitoring group got pregnant within 2 months after enrollment, dropped out of the study, and were excluded from the sample. The units of analysis were 3-month follow-up during the 6-month study period for 42 participants (n=84 possible units, two periods for each participant).

Procedure

Ohmage: Smartphone App. Ohmage is a mobile to web platform that supports the collection, storage, analysis, and visualization of self-report and passive sensor data streams. Ohmage has been released as an open source, and it can be downloaded for free online. The platform is feature-rich and extensible, and facilitates the collection of multidimensional, heterogeneous, and complex personal data streams. Ohmage adds a time and location stamp to each data point. Web interfaces are available for researchers to access and view data. The Ohmage user interface was designed based on expert feedback from behavioral science collaborators and nonexpert pilot users through focus group (BLINDED) and one-on-one interviews during a pilot phase. Notably, most of the discussions with the participants focused on reducing the user burden when self-reporting data.

Incentives

At the end of the study, upon returning the assigned study equipment and completion of the final assessment, the participants received reimbursement for parking and transportation and cash or a gift certificate up to US $355 for completing the following: US $40 each for baseline, 3-month, and 6-month follow-up assessments with biomarkers and biomeasures and US $1.30/day for completing at least one EMA or diary survey per day for 6 months (approximately 180 days).

Analyses

Prior to analysis, EMA and diary measures were averaged during a 30-day period that ended on and included the date of recall assessment and biomarker measurement. Thirty-day periods were chosen to broadly capture the time frame imposed by the recall measures. Analyses exclude 30-day periods with less than 14 days of EMAs or diary reports (ie, less than 50% reporting) for each specific comparison with a recall or biomarker measure.

The primary analytic goal was to examine the relationship between pairs of measurements, where the first measurement is an EMA or diary measure, aggregated over 30-day periods, and the second measurement was a recall or biomarker or biomeasure. To fully explore meaningful relationships between measurement pairs, several metrics of association that address both clinical significance (eg, correlations) and statistical significance (ie, *P* values) were presented.

Pearson product-moment correlations were calculated using measurement pairs as a basic measure of association. To calculate *P* values for the statistical significance of the correlation between measurement pairs, random effects linear regression models were utilized for correlations between measurements in the same individual, and they were expressed as follows:

$$Y_{ij} = \beta_0 + X_{ij}\beta_1 + \lambda_i + \varepsilon_{ij}$$

Where Y_{ij} is an EMA measurement for participant i at time point j, β_0 is an intercept term, and β_1 is a regression coefficient for X_{ij}, either a recall or biomarker measurement and represents the association within measurement pairs, which is similar to the correlation. Participant-level random effect λ_i identifies the correlation between repeated measurements in the same individual, and ε_{ij} are the residual error terms. *P* values were presented for the significance of β_1.

Table 1. Linear relationships between the pairs of ecological momentary assessments or daily diary responses.

Ecological momentary assessments and diary response predictor	n	r[a]	*P* value[b]	R^2_{Ed}[c,d]	R^2_{Total}[e]
Diet ecological momentary assessments (3 times per day)					
Recall: Intake of fruits and vegetables	36	0.34[f]	0.006	0.23	0.05
Recall: Intake of food with high-sugar content	34	−0.52[f]	0.004	0.27	0.24
Recall: Intake of fast food	33	−0.42[g]	.02	0.17	0.18
Recall: Moderate physical activity (min/day)	35	0.48[g]	.03	0.17	0.19
Biomarker variable: Body fat	38	−0.33	.07	0.13	0.07
Biomarker variable: Systolic blood pressure	38	−0.32[g]	.02	0.14	0.08
Biomarker variable: Diastolic blood pressure	38	−0.26	.10	0.08	0.05
Biomarker variable: C-reactive protein level	35	−0.34[g]	.02	0.16	0.11
Diet diary					
Recall: Intake of fruits and vegetables	30	0.27	.09	0.10	0.07
Recall: Intake of food with high-sugar content	29	−0.40[g]	.04	0.16	0.14
Recall: Moderate physical activity (min/day)	31	0.45	.09	0.11	0.14
Recall: Vigorous physical activity (min/day)	31	0.47[g]	.02	0.34	0.17
Recall: Vigorous physical activity (total min)	30	0.35[f]	.01	0.45	0.11
Stress ecological momentary assessments (3 times per day)					
Recall: PSM-9[h] (stress inventory)	45	0.30[g]	.01	0.27	0.08
Recall: Walking (min/day)	42	0.27[g]	.05	0.18	0.06
Stress diary					
Recall: PSM-9 (stress inventory)	31	0.50[g]	.02	0.28	0.24
Recall: Walking (min/day)	30	0.38[g]	.05	0.14	0.18
Recall: Walking (total min)	30	0.36	.06	0.14	0.15
Biomarker variable: C-reactive protein level	29	−0.28	.09	0.12	0.04
Light physical activity diary (days)					
Recall: Walking (days)	31	0.16	.26	0.05	−0.01
Recall: Moderate physical activity (days)	31	0.43[9]	.04	0.15	0.16
Moderate physical activity diary (days)					
Recall: Moderate physical activity (days)	31	0.33	.07	0.11	0.08
Biomarker variable: Body mass index	32	−0.40[g]	.02	0.16	0.13
Biomarker variable: Body fat	32	−0.35	.06	0.18	0.09
Biomarker variable: Systolic blood pressure	32	−0.24	.20	0.06	0.03
B: Diastolic blood pressure	32	−0.34	.07	0.13	0.09
Vigorous physical activity diary (days)					
Recall: Vigorous physical activity (days)	30	0.34	.07	0.11	0.09
Biomarker variable: Body mass index	32	−0.44[g]	.02	0.26	0.16
Biomarker variable: Body fat	32	0.44[g]	.02	0.25	0.16
Biomarker variable: Systolic blood pressure	32	−0.25	.22	0.05	0.03
Biomarker variable: Diastolic blood pressure	32	−0.35	.07	0.12	0.09
Vigorous physical activity diary (min/day)					

Ecological momentary assessments and diary response predictor	n	r[a]	*P* value[b]	R^2_{Ed}[c,d]	R^2_{Total}[e]
Recall: Vigorous physical activity (min/day)	31	0.27	.30	0.04	0.03
Biomarker variable: Body mass index	32	−0.38	.06	0.17	0.10
Biomarker variable: Body fat	32	−0.37	.07	0.15	0.09
Vigorous physical activity diary (total min)					
Recall: Vigorous physical activity (total min)	30	0.35	.11	0.10	0.09
Biomarker variable: Body mass index	32	−0.37	.07	0.15	0.09
Biomarker variable: Body fat	32	−0.35	.09	0.13	0.07

[a]r: Pearson product-moment correlations.

[b]*P* values for regression coefficients.

[c]R^2 statistics based on random effects linear regression models.

[d]Based on the calculation of Edwards et al (2008).

[e]Based on the ratio of unexplained-to-total variance.

[f]*P* value≤.01.

[g]*P* value≤.05.

[h]PSM-9: nine-item psychological stress measure.

Lastly, R^2 statistics that describe the amount of variation explained by the model were presented. There is no standard R^2 calculation for the random effects models. Two reasonable formulations were presented. Edwards et al [55] have calculated R^2 for regression coefficients as follows:

$$R^2_{Ed} = (q-1)v^{-1}F(\beta^*,\Sigma^*)/[1 + (q-1)v^{-1}F(\beta^*,\Sigma^*)]$$

Where q-1 is the number of regression coefficients minus the intercept term, v is the residual degrees of freedom based on the Kenward-Rogers approximation, and $F(\beta^*,\Sigma^*)$ is the F statistic that is used to test the null hypothesis stating that q-1 regression coefficients are equal to 0. Models were fit using the PROC MIXED procedure in SAS software version 9.3 (SAS Institute Inc, Cary, NC). The second R^2 statistics was based on the total variation explained and was expressed as follows:

$$R^2_{Total} = (Var_{int} – Var_{full}) / Var_{int}$$

Where Var_{int} is the total variation from a model that contains an intercept term, random effects, and a residual error term and Var_{full} is the total variation from the full model with the intercept term and all covariates.

Multivariate Analyses

As a secondary analytic goal, multiple-predictor linear regression models that regress EMAs and diary measurements on recall and biomarker measurements were established to examine the significant associations, and they further provided insights about the reliability and validity of the EMAs and diary measures. The models were summarized as follows: recall and biomarker→EMA. The candidate predictors for the multiple-predictor regression models include all the variables shown in Table 1, and they were selected based on the four statistics that were presented for the primary analyses: Pearson correlation coefficients, *P* values, β1, and two R^2 statistics. We aimed to strike a balance between parsimony and prediction in our model. Therefore, *P* values ≤.10 were roughly considered for the entry of predictors into the multiple-predictor models as well as the direct corresponding recall measure regardless of *P* value. Moreover, we wanted to achieve reasonable levels of explained variation based on the R^2 statistics and strengths of associations based on the Pearson correlation coefficient. Once the multiple-predictor models were built, a backward stepwise selection procedure was used to select predictors for the final models. Predictors were retained at a .05 alpha level.

Results

Demographic Characteristics of the Participants

Table 2 and Table 3 present the demographic characteristics of the participants (n=42) assigned with smartphones and those (n=29) with sufficient data for the analysis based on the participation rates (described below). No significant differences were found in terms of the characteristics of the participants who were included (n=29) and excluded (n=13) from the analyses owing to missing data. Most participants self-identified as ethnic minority (85%, 35/41), of which 39% (16/41) identified as African American, 44% (18/41) as Latina, and 2% (1/41) as Other. The age of the participants ranged from 20 to 43 years with an average age of 31.2 years. Approximately one-third of the participants were working full-time or part-time (between 4 and 20 hours a week) or not working. A little over half of the participants were obese or very obese (57%, 24/42) in line with an average body fat percentage of 40%. On average, the participants had blood pressure readings within the normal range (mean systolic blood pressure=122.3 and diastolic blood pressure=79.8). Average C-reactive protein levels of 3.2 mg/L indicated intermediate risk for cardiovascular disease. Average Epstein–Barr virus levels of 140.3 ELISA units were comparable to Epstein–Barr virus levels discussed earlier [34].

Table 2. Demographic and baseline biomarker characteristics of mothers who were included and those excluded from the analyses due to low ecological momentary assessments completion rates.

Characteristics	Analysis data (n=29), n (%)	Excluded mothers (n=13), n (%)	Total sample (n=42), n (%)
Ethnicity[a]			
African American	11 (39)	5 (39)	16 (39)
Asian	2 (7)	0 (0)	2 (5)
Latina	13 (46)	5 (39)	18 (44)
White	2 (7)	2 (15)	4 (10)
Others	0 (0)	1 (8)	1 (2)
Education[a]			
High school or lower	10 (35)	3 (25)	13 (32)
College	10 (35)	2 (17)	12 (27)
College degree or higher	9 (31)	9 (58)	16 (39)
Number of children			
1	12 (41)	7 (54)	19 (45)
2-4	17 (59)	6 (46)	23 (55)
Body mass index category			
Normal weight	5 (17)	2 (15)	7 (17)
Overweight	8 (28)	3 (23)	11 (26)
Obese	6 (21)	4 (31)	10 (24)
Extremely obese	10 (35)	4 (31)	14 (33)

[a]n=41, number of participants excluded due to missing responses.

Table 3. Demographic and baseline biomarker characteristics of mothers who were included and those excluded from the analyses due to low ecological momentary assessments completion rates.

Characteristics	Analysis data (n=29), mean (SD)	Excluded mothers (n=13), mean (SD)	Total sample (n=42), mean (SD)
Age	30.9 (6.3)	32.0 (6.0)	31.2 (6.2)
Work hours	22.0 (17.8)	19.5 (20.3)	21.2 (18.4)
Body fat (%)	39.1 (7.6)	40.9 (6.8)	39.6 (7.4)
Systolic blood pressure (mm Hg)	123.3 (15.7)	120.1 (11.0)	122.3 (14.4)
Diastolic blood pressure (mm Hg)	81.3 (9.5)	76.4 (10.0)	79.8 (9.8)
C-reactive protein level (mg/L)[a]	3.2 (2.4)	3.0 (2.6)	3.2 (2.4)
Epstein–Barr virus (enzyme-linked immunosorbent assay units)[b]	135.0 (56.8)	151.8 (67.8)	140.3 (60.2)

[a]n=37, number of participants excluded due to five C-reactice protein values >10.

[b]n=41, number of participants excluded due to one Epstein–Barr virus value <20.

Participation Rates

Almost all the 42 participants, who used a smartphone, completed the 6-month study (93%; n=39). The analyses excluded 13 participants, including two who were lost to follow-up before the 3-month follow-up, one who moved out of the state before the 6-month follow-up, and 10 participants who had reporting rates <50%.

In total,15,103 time-prompted EMA and diary surveys were completed by 29 participants included in analyses. The responses were distributed uniformly across the four surveys with 4043 (26.8%) morning EMAs, 3855 (25.5%) midday EMAs, 3643 (24.1%) late afternoon EMAs, and 3562 (23.6%) end-of-day diary surveys. Table 4 shows EMA and diary survey participation rates in terms of the mean days of reporting. The participants completed at least two surveys a day for 151.8 days (84%) on average, and the goal was 180 days.

Table 4. Ecological momentary assessments and diary reports of the participants, including participants with delayed 6-month follow-up assessments who continued the ecological momentary assessments and diary.

Characteristics	Mean (range)
Number of days it took to complete at least 1 survey	184.2 (126-242)
Number of days it took to complete at least 2 surveys	149.3 (18-230)
Number of days it took to complete at least 4 surveys	47.8 (10-169)

Table 5. Participation rates for ecological momentary assessments and diary surveys, recall self-reports, and biomarker assessments (n=29).

Characteristics	Participants, n (%)
Recall self-reports	
Number of follow-up assessment meetings	58[a]
Physical activity	48 (82.8)
Intake of fruits and vegetables	50 (86.2)
Intake of food with high-sugar content	47 (81.0)
Intake of fast food	47 (81.0)
Biomarker assessments	
Body fat and body mass index and blood pressure	58 (100.0)
Epstein–Barr virus and C-reactive protein level	54 (93.1)

[a]Number of follow-ups that were completed by participants (29 participants who completed two follow-ups each).

As shown in Table 5, the participants completed 58 follow-ups (29 participants × 2 follow-ups). The completion rates for the individual recall and biomarker measures were fairly high on average (range 81.0%-100% of the recall or biomarker measures).

Associations Between Ecological Momentary Assessment and Diary Reports and Recall and Biomarker Data

Table 1 presents the bivariate associations across domains (diet, stress, and physical activity) between the pairs of EMA and diary measures and either a recall or biomarker measure. Considering the large number of possible pairs, Table 1 presents only the associations between EMA and diary measures and alternative measures of the same domain (eg, recalls) or the different domains if the *P* value was ≤.10 for the association. Small-to-moderate correlations were observed.

Diet quality EMA and daily diary ratings were both negatively associated with the recall of the intake of food with high-sugar content (r=−.52, *P*<.01 and r=−.40, *P*=.04, respectively). Diet quality EMA ratings, but not those of daily diary, were positively associated with the intake of fruits and vegetables (r=.34, *P*<.01) and moderate physical activity in terms of minutes per day (r=48, *P*=.03) and negatively associated with the intake of fast food (r=−.42, *P*=.02), systolic blood pressure (r=−.32, *P*=.02), and C-reactive protein levels (r=−.34, *P*=.02). Ratings of the diet diary were also associated with vigorous physical activity in terms of minutes per day (r=.47, *P*=.02) and total minutes (r=.35, *P*<.01). Based on both R^2 statistics, roughly a quarter of the variance in the relationship between diet EMA reports and high-sugar counts was explained (R^2_{Ed}=.25 and R^2_{Total}=.28). Based on the R^2_{Ed} statistics, a similar level of variability was found for the relationship between reports on diet diary and vigorous physical activity in terms of minutes per day (R^2=.34) and total minutes (R^2=.45). A low level of variability was observed for the remaining relationships between diet measures with recall reports and biomarkers.

Daily stress diary reports correlated with the PSM-9 recall or global measure had the highest correlation among the results (r=.50, *P*=.02) and had trends in correlations with walking activity recalls. Stress EMA was also correlated with PSM-9 (r=.30, *P*=.01) and recall for walking minutes per day (r=.27, *P*=.05). Similar to the dietary report, the R^2_{Ed} statistics for stress EMA reports (R^2_{Ed}=.27) and both R^2statistics for stress daily reports (R^2_{Ed}=.28 and R^2_{Total}=.24) indicate that approximately a quarter of the variation was caused by the relationship with PSM-9 recall.

Daily diary physical activity reports were not significantly associated with their corresponding recall measures, whether counted by days, minutes per day, or total minutes. Light physical activity (counted as days) was associated with moderate activity recall days (r=.43, *P*=.04). Moderate physical activity days were negatively associated with BMI (r=−.40, *P*=.02). Vigorous physical activity days were negatively associated with BMI and body fat (r=−.44, *P*=.02 for both measures). Based on the R^2_{Ed} statistics, a quarter of the variance for physical activity days was explained by the relationships with BMI (R^2_{Ed}=.26) and body fat (R^2_{Ed}=.25). The variance was fairly low for the remaining relationships between physical activity reports and both recall reports and biomarkers.

Table 6. Final random effects linear regression models showing the associations between ecological momentary assessment and diary and recall-based predictor variables.

Ecological momentary assessment or diary response predictor	n	β[a]	SE	R^2_{Ed}[b]	R^2_{Total}[c]
Diet ecological momentary assessment (3 times per day): Intake of high-sugar food[d]	34	−0.0055	0.0018	0.27	0.24
Diet diary: Model 1	29	—[e]	—	0.39	0.29
Intake of high-sugar food[f]	—	−0.011	0.0051	—	—
Vigorous physical activity (min/day)[f]	—	0.0071	0.0026	—	—
Diet diary: Model 2	28	—	—	0.45	0.23
Intake of high-sugar food[f]	—	−0.011	0.0054	—	—
Vigorous physical activity (total min)[d]	—	0.0021	0.00066	—	—
Stress ecological momentary assessment (3 times per day): PSM-9[g] (stress inventory)[f]	45	0.014	0.0050	0.27	0.08
Stress diary: PSM-9 (stress inventory)[f]	31	0.024	0.0092	0.28	0.24

[a]β: regression coefficients.

[b]Based on the calculation of Edwards et al (2008).

[c]Based on ratio of unexplained-to-total variance.

[d]*P* value<.01.

[e]Not applicable.

[f]*P* value<.05.

[h]PSM-9: nine-item psychological stress measure.

Multivariate Regression Analyses

Table 6 shows the multivariate regression results for EMA and diary measures that were significantly associated with a recall or biomarker measure as well as the direct corresponding recall measures (ie, variables shown in Table 1). Results indicated that diet EMA is most strongly associated with high-sugar food recalls, indicating the particular sensitivity for low-quality foods. The associations between stress EMA and diary reports as well as their corresponding PSM-9 recall were confirmed in the multivariate models and washing out associations with low physical activity indicated in the bivariate models. The only multivariate model with multiple significant predictors was the diet diary rating in which both high-sugar food and vigorous physical activity (total minutes and minutes per day) recalls retained their significant associations and together explained almost half of the variability diet diary rating based on the R^2 measures.

Discussion

Principal Findings

The results in this analysis present several sets of inferences on the reliability and validity of the brief EMA and daily diary reports on diet, stress, and physical activity. Intermethod reliability between EMA and diary reports and their corresponding recall reports is moderate for stress and diet, as hypothesized. However, it was unexpectedly low for physical activity. In contrast to intermethod reliability, concurrent validity with other measures was demonstrated for diet and physical activity EMA and diary reports. However, it was not observed for stress.

Diet EMA as simple and subjective as “high or medium or low quality” ratings showed reliability with brief food frequency questionnaire recall methods and concurrent validity with physical activity. This is remarkable considering the subjective quality of the question and response options. As hypothesized, EMA appears to be more reliable and valid compared with the end-of-day diaries, particularly for food quality reports, and to some key diet- or systemic-related biomarkers. By contrast, daily diary diet ratings are not associated with the biomarkers and are more significantly associated with vigorous activity recalls than food count recalls. Daily diary diet reports are likely associated with a combination of recall and rounding errors that reflect the linked lifestyle habits of healthy eating and regular physical activity or poor diet and sedentary behaviors.

The simple stress self-monitoring questions (both EMA and daily diary) used in this study also indicated good intermethod reliability with the PSM-9 brief recall measure. The trends in the associations between stress EMA and diary as well as walking recall report data demand further explanation and analysis. Qualitative reports from formative work (BLINDED) have identified that mothers have lower levels of physical activity due to stress, specifically, time pressures associated with stress result in the lack of time to engage in moderate or vigorous physical activity. Walking as physical activity in daily life (eg, for transport and walking pets) may be more salient for reportings when stressed and unable to engage in intentional moderate or vigorous physical activity.

The moderate negative associations between vigorous physical activity diary reports and BMI and body fat and similar trends in the associations with moderate physical activity indicate the concurrent validity of the physical activity diary reports. Vigorous physical activity reports likely reflect the classes of

highly active (or inactive) participants who are consistent enough in their activity levels to observe the associations with biomarkers. Notably, the crudest calculations of physical activity (ie, days vs minutes for vigorous and moderate physical activity) show potential trends for intermethod reliability correlations with their corresponding recall reports, whereas light physical activity diary reports showed significant correlations with moderate activity.

Concurrent validity with different biomarkers and biomeasures for different activity measurement methods also suggests that the underlying aspects of physical activity measured using different methods may be somewhat independent [17]. Similar challenges were observed for food frequency questionnaires [18] and are reported to vary in terms of individuals and population groups [56-58].

Limitations

The primary limitation of this study is its relatively small sample size. A larger sample size would increase statistical power, which would likely result in trends showing statistical significance (ie, P=.10) and more elaboration between variables in the multivariate analyses. Another limitation is the participant-centered design that prioritized brief recall measures and salient EMA and diary question design over highly detailed measures that might have better reliability and validity but higher user burden and lower engagement, sustainability, and completion rates. There are additional limitations in evaluating the magnitude of the associations between measures. Given that some study measures, such as diet quality, measure of the self-perceptions of health, the degree of association between self-reported measures likely includes method bias; for example, a high perception of health may lead to positive reports on both diet quality and physical activity. Finally, the 6-month study period is in accordance with the time periods for weight loss interventions and other lifestyle modification programs; 6 months is a long period for an EMA study and calls for additional research to determine optimal EMA schedules that compliment lifestyle modification program schedules.

Conclusions

The results of this study suggest that simple and brief EMA measures for diet and daily diary measures for stress, may be good enough tools for long periods of self-monitoring. This is especially important in the study population that mostly includes ethnic minority mothers, several of whom expressed that time is a barrier in monitoring and engaging in healthy behaviors. The inconsistencies between self-report and objective measurement methods and the lack of gold standards have resulted in recommendations to use a combination of methods, particularly when examining impacts on health status [17,59,60]. Future studies with larger sample sizes must be conducted while examining active and passive self-monitoring strategies and ecological momentary interventions [61] that trigger microinterventions based on the self-monitoring and contextual data (eg, global positioning system location) of smartphone apps [62,63]. There are significant challenges in addressing this in future studies. The intersecting issues of burden, participation (compliance), and timelines for changing and then sustaining daily health routines must be carefully considered. Initially, intensive EMA and diary self-reporting could support changes in behaviors. Once improvement in health status is achieved (eg, weight loss and improved cardiovascular fitness), the next hurdle is to maintain positive outcomes by continuously performing behavioral routines. Apps must be adaptive to the stages of change, participation burnout, and varying patterns of setbacks that individuals experience in adopting and maintaining healthier routines. However, smartphone apps are well suited for the task.

Acknowledgments

This work was sponsored by NHLBI Grant 5RC1HL099556 to the last author (PI) and the co-authors as co-investigators. Comulada's time was also supported by NIMH Grant K01MH089270. Swendeman's time was supported by a scholar award from the William T Grant Foundation (#180039). Other support was provided by: the Center for HIV Identification, Prevention, and Treatment (CHIPTS) NIMH Grant MH58107; the UCLA Center for AIDS Research (CFAR) Grant 5P30AI028697; and the National Center for Advancing Translational Sciences through UCLA CSTI Grant UL1TR000124. The content is solely the responsibility of the authors and does not necessarily represent the views of the NIH.

Conflicts of Interest

None declared.

References

1. 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]
2. Miller G. The Smartphone Psychology Manifesto. Perspect Psychol Sci 2012 May;7(3):221-237. [doi: 10.1177/1745691612441215] [Medline: 26168460]
3. Shiffman S, Stone A. Introduction to the special section: Ecological momentary assessment in health psychology. Health Psychology 1998;17(1):3-5. [doi: 10.1037/h0092706]

4. Mottl J. Fierce Healthcare. 2015. Consumers, 'validation' will be key to mHealth growth in 2015 URL: https://www.fiercehealthcare.com/mobile/consumers-validation-will-be-key-to-mhealth-growth-2015 [accessed 2017-11-07] [WebCite Cache ID 6unmrRfac]
5. Goldstein MG, Whitlock EP, DePue J, Planning Committee of the Addressing Multiple Behavioral Risk Factors in Primary Care Project. Multiple behavioral risk factor interventions in primary care. Summary of research evidence. Am J Prev Med 2004 Aug;27(2 Suppl):61-79. [doi: 10.1016/j.amepre.2004.04.023] [Medline: 15275675]
6. Liu K, Daviglus ML, Loria CM, Colangelo LA, Spring B, Moller AC, et al. Healthy lifestyle through young adulthood and the presence of low cardiovascular disease risk profile in middle age: the Coronary Artery Risk Development in (Young) Adults (CARDIA) study. Circulation 2012 Feb 28;125(8):996-1004 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.060681] [Medline: 22291127]
7. Lloyd-Jones, Adams R, Carnethon M, De SG, Ferguson TB, Flegal K, American Heart Association Statistics CommitteeStroke Statistics Subcommittee. Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 2009 Jan 27;119(3):e21-181 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.108.191261] [Medline: 19075105]
8. Gunderson EP, Murtaugh MA, Lewis CE, Quesenberry CP, West DS, Sidney S. Excess gains in weight and waist circumference associated with childbearing: The Coronary Artery Risk Development in Young Adults Study (CARDIA). Int J Obes Relat Metab Disord 2004 Apr;28(4):525-535 [FREE Full text] [doi: 10.1038/sj.ijo.0802551] [Medline: 14770188]
9. Webb T. L. ., Using the Internet to Promote Health Behavior Change: A Systematic Review and Meta-analysis of the Impact of Theoretical Basis, Use of Behavior Change Techniques, and Mode of Delivery on Efficacy. Journal of Medical Internet Research 2010;12(1):A.
10. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
11. Burke L. E., J. Wang, and M. A. Sevick, Self-Monitoring in Weight Loss: A Systematic Review of the Literature. Journal of the American Dietetic Association 2011;111(1):A-102.
12. Hebden L, Cook A, van der Ploeg HP, Allman-Farinelli M. Development of smartphone applications for nutrition and physical activity behavior change. JMIR Res Protoc 2012 Aug 22;1(2):e9 [FREE Full text] [doi: 10.2196/resprot.2205] [Medline: 23611892]
13. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
14. Hutchesson MJ, Rollo ME, Callister R, Collins CE. Self-monitoring of dietary intake by young women: online food records completed on computer or smartphone are as accurate as paper-based food records but more acceptable. J Acad Nutr Diet 2015 Jan;115(1):87-94. [doi: 10.1016/j.jand.2014.07.036] [Medline: 25262244]
15. Courvoisier DS, Eid M, Lischetzke T. Compliance to a cell phone-based ecological momentary assessment study: the effect of time and personality characteristics. Psychol Assess 2012 Sep;24(3):713-720. [doi: 10.1037/a0026733] [Medline: 22250597]
16. Comulada WS, Lightfoot M, Swendeman D, Grella C, Wu N. Compliance to Cell Phone-Based EMA Among Latino Youth in Outpatient Treatment. J Ethn Subst Abuse 2015;14(3):232-250 [FREE Full text] [doi: 10.1080/15332640.2014.986354] [Medline: 26114764]
17. Atienza AA, Moser RP, Perna F, Dodd K, Ballard-Barbash R, Troiano RP, et al. Self-reported and objectively measured activity related to biomarkers using NHANES. Med Sci Sports Exerc 2011 May;43(5):815-821. [doi: 10.1249/MSS.0b013e3181fdfc32] [Medline: 20962693]
18. Cade J, Burley VJ, Warm DL, Thompson RL, Margetts BM. Food-frequency questionnaires: a review of their design, validation and utilisation. Nutr Res Rev 2004 Jun;17(1):5-22. [doi: 10.1079/NRR200370] [Medline: 19079912]
19. Prince S, Adamo KB, Hamel ME, Hardt J, Connor GS, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
20. Solhan MB, Trull TJ, Jahng S, Wood PK. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. Psychol Assess 2009 Sep;21(3):425-436 [FREE Full text] [doi: 10.1037/a0016869] [Medline: 19719353]
21. Stone A, Schwartz JE, Broderick JE, Shiffman SS. Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. Pers Soc Psychol Bull 2005 Oct;31(10):1340-1346. [doi: 10.1177/0146167205275615] [Medline: 16143666]
22. Gorin A, Stone A. Recall biases and cognitive errors in retrospective self-reports: A call for momentary assessments. Handbook of health psychology 2001;23:405-413.
23. Vallance JKH, Courneya KS, Plotnikoff RC, Yasui Y, Mackey JR. Randomized controlled trial of the effects of print materials and step pedometers on physical activity and quality of life in breast cancer survivors. J Clin Oncol 2007 Jun 10;25(17):2352-2359. [doi: 10.1200/JCO.2006.07.9988] [Medline: 17557948]

24. Biddle S, Gorely T, Marshall SJ, Cameron N. The prevalence of sedentary behavior and physical activity in leisure time: A study of Scottish adolescents using ecological momentary assessment. Prev Med 2009 Feb;48(2):151-155. [doi: 10.1016/j.ypmed.2008.10.025] [Medline: 19046984]
25. Treiber FA, Baranowski T, Braden DS, Strong WB, Levy M, Knox W. Social support for exercise: relationship to physical activity in young adults. Prev Med 1991 Nov;20(6):737-750. [Medline: 1766945]
26. Tudor-Locke CE, Bell RC, Myers AM, Harris SB, Lauzon N, Rodger NW. Pedometer-determined ambulatory activity in individuals with type 2 diabetes. Diabetes Res Clin Pract 2002 Mar;55(3):191-199. [Medline: 11850095]
27. Arab L, Tseng C, Ang A, Jardack P. Validity of a multipass, web-based, 24-hour self-administered recall for assessment of total energy intake in blacks and whites. Am J Epidemiol 2011 Dec 01;174(11):1256-1265 [FREE Full text] [doi: 10.1093/aje/kwr224] [Medline: 22021561]
28. Comulada WS, Swendeman D, Koussa MK, Mindry D, Medich M, Estrin D, et al. Adherence to self-monitoring healthy lifestyle behaviours through mobile phone-based ecological momentary assessments and photographic food records over 6 months in mostly ethnic minority mothers. Public Health Nutr 2018 Mar;21(4):679-688. [doi: 10.1017/S1368980017003044] [Medline: 29199630]
29. Reddy S. Evaluating participation and performance in participatory sensing. Urban Sense 2008;08(4).
30. Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. Trials 2009:10.
31. Lemyre L, Lalande-Markon M. Psychological Stress Measure (PSM-9): integration of an evidence-based approach to assessment, monitoring, and evaluation of stress in physical therapy practice. Physiother Theory Pract 2009 Jul;25(5-6):453-462. [Medline: 19842868]
32. Fagard RH. Effects of exercise, diet and their combination on blood pressure. J Hum Hypertens 2005 Dec;19 Suppl 3:S20-S24. [doi: 10.1038/sj.jhh.1001956] [Medline: 16302006]
33. Kaplan MS, Nunes A. The psychosocial determinants of hypertension. Nutr Metab Cardiovasc Dis 2003 Feb;13(1):52-59. [Medline: 12772438]
34. Borders A, Grobman WA, Amsden LB, McDade TW, Sharp LK, Holl JL. The relationship between self-report and biomarkers of stress in low-income reproductive-age women. Am J Obstet Gynecol 2010 Dec;203(6):577.e1-577.e8 [FREE Full text] [doi: 10.1016/j.ajog.2010.08.002] [Medline: 20870203]
35. Kolb H, Mandrup-Poulsen T. The global diabetes epidemic as a consequence of lifestyle-induced low-grade inflammation. Diabetologia 2010 Jan;53(1):10-20. [doi: 10.1007/s00125-009-1573-7] [Medline: 19890624]
36. Lionetti L, Mollica MP, Lombardi A, Cavaliere G, Gifuni G, Barletta A. From chronic overnutrition to insulin resistance: the role of fat-storing capacity and inflammation. Nutr Metab Cardiovasc Dis 2009 Feb;19(2):146-152. [doi: 10.1016/j.numecd.2008.10.010] [Medline: 19171470]
37. Kiecolt-Glaser JK, Derry HM, Fagundes CP. Inflammation: depression fans the flames and feasts on the heat. Am J Psychiatry 2015 Nov 01;172(11):1075-1091. [doi: 10.1176/appi.ajp.2015.15020152] [Medline: 26357876]
38. Herder C, Schneitler S, Rathmann W, Haastert B, Schneitler H, Winkler H, et al. Low-grade inflammation, obesity, and insulin resistance in adolescents. J Clin Endocrinol Metab 2007 Dec;92(12):4569-4574. [doi: 10.1210/jc.2007-0955] [Medline: 17911172]
39. Glaser R, Kiecolt-Glaser JK. Stress-induced immune dysfunction: implications for health. Nat Rev Immunol 2005 Dec;5(3):243-251. [doi: 10.1038/nri1571] [Medline: 15738954]
40. Black PH. The inflammatory response is an integral part of the stress response: Implications for atherosclerosis, insulin resistance, type II diabetes and metabolic syndrome X. Brain Behav Immun 2003 Oct;17(5):350-364. [Medline: 12946657]
41. Padgett DA, Glaser R. How stress influences the immune response. Trends Immunol 2003 Aug;24(8):444-448. [Medline: 12909458]
42. McDade TW, Stallings JF, Worthman CM. Culture change and stress in Western Samoan youth: Methodological issues in the cross-cultural study of stress and immune function. Am J Hum Biol 2000 Nov 01;12(6):792-802. [doi: 10.1002/1520-6300(200011/12)12:6<792::AID-AJHB7>3.0.CO;2-F] [Medline: 11534070]
43. Ernst L, Grobman WA, Wolfe K, Huang MH, McDade TW, Holl JL, et al. Biological markers of stress in pregnancy: associations with chronic placental inflammation at delivery. Am J Perinatol 2013 Aug;30(7):557-564. [doi: 10.1055/s-0032-1329187] [Medline: 23271381]
44. Weaver L, Worthman CM, DeCaro JA, Madhu SV. The signs of stress: embodiments of biosocial stress among type 2 diabetic women in New Delhi, India. Soc Sci Med 2015 Apr;131:122-130. [doi: 10.1016/j.socscimed.2015.03.002] [Medline: 25771481]
45. Kyle UG, Bosaeus I, De LAD, Deurenberg P, Elia M, Gómez JM, Composition of the ESPEN Working Group. Bioelectrical impedance analysis--part I: review of principles and methods. Clin Nutr 2004 Oct;23(5):1226-1243. [doi: 10.1016/j.clnu.2004.06.004] [Medline: 15380917]
46. Kyle UG, Bosaeus I, De LAD, Deurenberg P, Elia M, Manuel GJ, ESPEN. Bioelectrical impedance analysis-part II: utilization in clinical practice. Clin Nutr 2004 Dec;23(6):1430-1453. [doi: 10.1016/j.clnu.2004.09.012] [Medline: 15556267]
47. Lohman T, Roche A, Martorell R. Anthropometric Standardization Reference Manual. Human Kinetics 1988.

48. Parker ED, Schmitz KH, Jacobs DR, Dengel DR, Schreiner PJ. Physical activity in young adults and incident hypertension over 15 years of follow-up: the CARDIA study. Am J Public Health 2007 Apr;97(4):703-709. [doi: 10.2105/AJPH.2004.055889] [Medline: 17329668]
49. Worthman CM, Stallings JF. Hormone measures in finger-prick blood spot samples: new field methods for reproductive endocrinology. Am J Phys Anthropol 1997 Sep;104(1):1-21. [doi: 10.1002/(SICI)1096-8644(199709)104:1<1::AID-AJPA1>3.0.CO;2-V] [Medline: 9331450]
50. Copeland W. E.., Cumulative depression exposure predicts later c-reactive protein levels: a prospective analysis. Biol Psychiatry 2012;71(1):a-21.
51. Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon RO, Criqui M, Centers for Disease ControlPrevention, American Heart Association. Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. Circulation 2003 Jan 28;107(3):499-511 [FREE Full text] [Medline: 12551878]
52. Ansar W, Ghosh S. C-reactive protein and the biology of disease. Immunol Res 2013 May;56(1):131-142. [doi: 10.1007/s12026-013-8384-0] [Medline: 23371836]
53. Yeh ETH, Willerson JT. Coming of age of C-reactive protein: using inflammation markers in cardiology. Circulation 2003 Jan 28;107(3):370-371 [FREE Full text] [Medline: 12551854]
54. McDade T, Stallings JF, Angold A, Costello EJ, Burleson M, Cacioppo JT, et al. Epstein-Barr virus antibodies in whole blood spots: a minimally invasive method for assessing an aspect of cell-mediated immunity. Psychosom Med 2000;62(4):560-567. [Medline: 10949102]
55. Edwards LJ, Muller KE, Wolfinger RD, Qaqish BF, Schabenberger O. An R2 statistic for fixed effects in the linear mixed model. Stat Med 2008 Dec 20;27(29):6137-6157 [FREE Full text] [doi: 10.1002/sim.3429] [Medline: 18816511]
56. Kristal AR, Feng Z, Coates RJ, Oberman A, George V. Associations of race/ethnicity, education, and dietary intervention with the validity and reliability of a food frequency questionnaire: the Women's Health Trial Feasibility Study in Minority Populations. Am J Epidemiol 1997 Nov 15;146(10):856-869. [Medline: 9384206]
57. Liu K, Slattery M, Jacobs D, Cutter G, McDonald A, Van Horn L, et al. A study of the reliability and comparative validity of the cardia dietary history. Ethn Dis 1994;4(1):15-27. [Medline: 7742729]
58. Mayer-Davis EJ, Vitolins MZ, Carmichael SL, Hemphill S, Tsaroucha G, Rushing J, et al. Validity and reproducibility of a food frequency interview in a Multi-Cultural Epidemiology Study. Ann Epidemiol 1999 Jul;9(5):314-324. [Medline: 10976858]
59. Ferrari P, Friedenreich C, Matthews CE. The role of measurement error in estimating levels of physical activity. Am J Epidemiol 2007 Oct 01;166(7):832-840. [doi: 10.1093/aje/kwm148] [Medline: 17670910]
60. Harris TJ, Owen CG, Victor CR, Adams R, Ekelund U, Cook DG. A comparison of questionnaire, accelerometer, and pedometer: measures in older people. Med Sci Sports Exerc 2009 Jul;41(7):1392-1402. [doi: 10.1249/MSS.0b013e31819b3533] [Medline: 19516162]
61. 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]
62. Goldstein S, Evans BC, Flack D, Juarascio A, Manasse S, Zhang F, et al. Return of the JITAI: Applying a Just-in-Time Adaptive Intervention Framework to the Development of m-Health Solutions for Addictive Behaviors. Int J Behav Med 2017 Dec;24(5):673-682 [FREE Full text] [doi: 10.1007/s12529-016-9627-y] [Medline: 28083725]
63. Dunton GF. Ecological Momentary Assessment in Physical Activity Research. Exerc Sport Sci Rev 2017 Dec;45(1):48-54 [FREE Full text] [doi: 10.1249/JES.0000000000000092] [Medline: 27741022]

Abbreviations

BMI: body mass index
CHIS: California Health Interview Survey
ELISA: enzyme-linked immunosorbent assay
EMA: ecological momentary assessment

Edited by G Eysenbach; submitted 09.11.17; peer-reviewed by J Smyth, JH Lee; comments to author 15.03.18; revised version received 10.05.18; accepted 21.06.18; published 21.09.18.

Please cite as:
Swendeman D, Comulada WS, Koussa M, Worthman CM, Estrin D, Rotheram-Borus MJ, Ramanathan N
Longitudinal Validity and Reliability of Brief Smartphone Self-Monitoring of Diet, Stress, and Physical Activity in a Diverse Sample of Mothers
JMIR Mhealth Uhealth 2018;6(9):e176
URL: http://mhealth.jmir.org/2018/9/e176/
doi:10.2196/mhealth.9378
PMID:30249576

Original Paper

Mobile Apps for Oral Health Promotion: Content Review and Heuristic Usability Analysis

Brooks Tiffany[1], MS; Paula Blasi[1], MPH; Sheryl L Catz[2], PhD; Jennifer B McClure[1], PhD

[1]Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States

[2]Betty Irene Moore School of Nursing, University of California, Davis, Sacramento, CA, United States

Corresponding Author:
Jennifer B McClure, PhD
Kaiser Permanente Washington Health Research Institute
1730 Minor Avenue, Suite 1600
Seattle, WA,
United States
Phone: 1 206 287 2737
Email: Jennifer.B.McClure@kp.org

Abstract

Background: There has been an increase in consumer-facing mobile health (mHealth) apps in recent years. Prior reviews have characterized the availability, usability, or quality of popular mHealth apps targeting a range of health behaviors, but none has examined apps that promote better oral health care. Oral disease affects billions of people worldwide and mobile phone use is on the rise, so the market for well-designed and effective oral health apps is substantial.

Objective: We examined the content and usability of popular oral health promotion apps to better understand the current state of these self-help interventions and inform the need and opportunity for future app development.

Methods: Between February and March 2018, we identified oral health-focused apps that were designed for Android or iOS, available in English, and targeted adult consumers (as opposed to children or dental health professionals). The sample was limited to the most popular and highly rated apps on each platform. For each app reviewed, we assessed its basic descriptive characteristics (eg, platform, cost), evidence of a theoretical basis or empirical validation, key program functionality, and the extent to which the app addressed diet and tobacco and alcohol use as risk factors for oral disease. We characterized the framing (ie, gain vs loss) of all persuasive messaging and conducted a heuristic analysis to assess each app's usability as a persuasive health technology.

Results: Thirty-three apps were eligible for review based on the selection criteria. Two-thirds (22/33, 67%) were geared toward the general public as opposed to dental clinic patients, insurance plan members, or owners of specific electric toothbrushes. Most (31/33, 94%) were free to download, and a majority (19/33, 58%) were sponsored by software developers as opposed to oral health experts. None offered any theoretical basis for the content or had been empirically validated. Common program features included tools for tracking or reminding one to brush their teeth and assistance scheduling dental appointments. Nineteen apps (58%) included educational or persuasive content intended to influence oral health behavior. Only 32% (6/19) of these included a larger proportion of gain-framed than loss-framed messaging. Most of the apps did not mention diet, alcohol or tobacco—important risk factors for oral disease. Overall, the apps performed poorly on standard usability heuristics recommended for persuasive health technologies.

Conclusions: The quality of the reviewed apps was generally poor. Important opportunities exist to develop oral health promotion apps that have theoretically grounded content, are empirically validated, and adhere to good design principles for persuasive health technologies.

(JMIR Mhealth Uhealth 2018;6(9):e11432) doi:10.2196/11432

KEYWORDS

oral health; oral hygiene; dental care; mobile health; eHealth; review

Introduction

Untreated oral conditions, including dental caries, severe periodontitis, and edentulism, affect about 3.5 billion people worldwide [1]. Oral health is integral to overall health [2]; oral disease contributes to unnecessary pain and suffering and is the fourth most costly disease to treat in most industrialized countries [3]. Routine oral hygiene, including daily brushing and flossing, is important for preventing oral disease and maintaining good oral health [4,5]. Routine dental visits are also important for maintaining good oral health because dentists can check for early signs of oral disease, provide teeth cleaning, and offer counseling about oral health behaviors [6]. Thus, promoting better oral health care is an important public health goal.

We believe mobile phones offer a promising strategy for reaching the public to deliver low-cost oral health promotion apps. According to a 2018 survey by the Pew Research Center, 77% of US residents now own a mobile phone [7]. Across 40 countries, a global median of 43% of residents owned a mobile phone in 2015, and mobile phone ownership is growing rapidly in countries with emerging and developing economies [8], making this technology a viable platform for direct-to-consumer public health interventions.

Along with the growth in mobile phone usage, there has been rapid growth in consumer-facing health promotion apps [9]. About 29% of those who have downloaded an app to a mobile phone or tablet report downloading a health-related app [10], and in 2017, there were 325,000 health-related apps available for download [11]. Mobile phone apps have been used to promote a variety of healthy behaviors including tobacco cessation [12], diabetes self-management [13], diet and nutrition [14], and physical activity [14].

Prior reviews have examined the efficacy of mobile health (mHealth) apps for behaviors such as weight loss and physical activity [15], self-management of chronic conditions [16], and identifying skin cancers [17]. Other reviews have sought to characterize the availability, usability, or quality of popular mHealth apps for common health issues such as supporting care during pregnancy [18], promoting physical activity [19] and encouraging smoking cessation [12]. But very little is known about the state of publicly available mHealth apps for adult oral health promotion. In a review of the literature, we only identified 1 cross-sectional survey of people who had used an app (Brush DJ) to improve their oral health self-care [20] and no prior reviews of the overall content, usability, or quality of existing oral health apps for adults. However, research has shown that text messaging interventions are associated with improvements in tooth brushing frequency [21], reducing plaque [22], and changing one's oral hygiene index and gingival index [23], so it is plausible that well-designed oral health apps could be effective at improving knowledge, self-care behaviors, or facilitating receipt of needed professional dental care among adults.

To address the gap in the literature and inform future oral health intervention development, we conducted a systematic review of the most popular oral health apps available for Android and iOS phones. To do this, we sought to review the content and assess the usability of each app. Our content review included an examination of the basic features and functionality, inclusion of key oral health discussion topics, analysis of message framing, and assessment of whether the developers cited a theoretical or empirical basis for the content. The findings from this comprehensive review inform the overall quality of popular oral health promotion apps and speak to both the need and opportunity for additional mHealth intervention development.

Methods

Sample

We identified oral health-focused apps in the App Store and Google Play between February and March 2018. Apps were identified using the following search phrases: *oral health*, *dental health*, *teeth health*, *tooth health*, *mouth health*, *dental care*, *teeth care*, *and oral care*. The initial search revealed 1975 Android and 1005 iOS apps. To narrow the field, we excluded apps that were not in English and those designed specifically for dental health professionals, dental students, patients of specific dental practices, or children. We then restricted the sample to the most popular available for each platform based on the assumption that these were likely the highest quality and most frequently used apps. As such, we believe these represent a reasonable sample for evaluating the content and quality of current adult-focused oral health apps. A similar strategy has been used by others to limit the number of apps reviewed when it was not feasible to review all available apps [24]. We did not eliminate any apps that met our inclusion criteria based on their perceived quality.

To limit the search to those most popular, we followed the example of Abroms et al [24] and selected Android apps that had been installed at least 1000 times (n=24). Of these, 12 were only available for Android and 10 were also available for iOS. Similar download information was not available from the App Store, but according to Apple Inc, presentation order is correlated with user ratings, reviews, and user downloads [25], even though ratings are not always published, so we selected the first 12 iOS apps (excluding duplicates also found for Android), allowing a balance between Android-only and iOS-only apps. However, 1 iOS app was not compatible with iOS 11, so we were unable to include it in our final review. Thus, the final sample included 33 apps (12 for Android, 11 for iOS, and 10 for both Android and iOS). Those available on both platforms were reviewed through Android since the Google Play store provided more complete information about each app than was available at the App Store. Several were available to the public but included content that was restricted to consumers who had purchased electric toothbrushes sold by the app sponsor. For these, we only reviewed the free-to-access content. All apps were reviewed on tablet devices.

Content Review: Key Features, Functionality, Content Basis, and Messaging

For each app we documented the developer/sponsor, platform, cost, number of installations, and user ratings. We summarized the basic functionality and noted whether the app or its

descriptive summary at the app store indicated that it had been empirically validated. We also conducted a literature review to identify empirical evaluations that might be included in the list. Next, we documented whether each app or its descriptive summary included any reference to a theoretical or empirical basis for the educational or persuasive content.

In addition, we characterized the framing of all persuasive health messages used. According to prospect theory, potential losses are more motivating than potential gains when risky actions are being considered, but gains are more motivating than losses for low-risk behaviors [26]. This theory is commonly applied when writing persuasive health messages. In fact, prior research has shown that gain-framed messages are particularly effective for promoting oral hygiene behaviors [27-31]. Gain-framing highlights the benefits of action, such as healthier teeth and gums or fresh breath. In contrast, loss-framed messages highlight the risks of inaction, such as increased cavities or oral disease. Thus, we were interested in understanding the extent to which this strategy was being used in each app. To assess framing, we captured screenshots of all content and then systematically reviewed and coded all educational or persuasive written messages. Messages were deemed gain-framed if they emphasized the benefits of positive oral health behaviors like brushing and flossing. Messages were deemed loss-framed if they emphasized the risks of not engaging in positive health behaviors. Messages that were neither gain- nor loss-framed were coded as neutral.

Because diet and tobacco and alcohol use are important risk factors for oral disease [32], we noted whether each topic was discussed (yes/no) and characterized the level of discussion as either brief (ie, up to a few sentences), partial (ie, a subsection dedicated to this topic, such as a short paragraph), or full (ie, 1 or more full sections dedicated to the topic). We then summarized the key talking points of this discussion to assess the quality and relevance of the content. In particular, we were interested in whether each discussion identified diet or alcohol or tobacco use as a risk factor for oral disease.

All content was reviewed and characterized using the methods outlined above by the first author (BT). A second reviewer (JM) reviewed the summarized findings and validated the first author's characterization of key content elements including message framing, level of discussion focused on key topics such as tobacco and alcohol, and whether the content linked diet, alcohol, or tobacco to oral disease risk.

Heuristic Analysis

Each of the apps reviewed is an example of a persuasive health technology [33], intended to change users' attitudes and/or behavior. The effectiveness of a persuasive technology is not limited to the quality of its written content; it is also affected by the app's user interface and functionality. If the app is difficult to operate, consumers will stop using it, and as a result the design will undermine the app's intended effects. Thus, it is essential that persuasive technologies comply with basic best practice design principles.

To further assess the quality of the selected apps, we reviewed and rated each based on 10 heuristics recommended for improving the usability of persuasive health technologies [34]. Kientz et al [34] argue that these heuristics are the most appropriate for technologies designed to persuade the end user to be healthier, and her team has found that they identify severe usability problems in persuasive technologies more frequently than the general heuristics proposed by Nielsen [35]. The heuristics recommended by Kientz et al [34] are as follows:

- Appropriate functionality: technology should meet usability, mobility, visibility, and durability needs according to the settings in which it might be used. The technology should function effectively in the user's environment by being easy to use and integrate into one's daily life and routine.
- Not irritating or embarrassing: technology should not irritate or embarrass the user, even after using the product repeatedly and regularly over a long period of time. This relates to aspects such as the presence of the product itself in the user's environment; the degree to which the technology intrudes upon the user's daily life; the timing, type, accuracy, and amount of feedback given; and the capability for customized settings and privacy controls.
- Protects user's privacy: system allows users to keep personal information private. Users can control what, when, to whom, and how much information is made public. Any public information is kept abstract.
- Use of positive motivation strategies: technology recognizes when target behaviors have been performed or goals have been met and uses positive reinforcement strategies to promote continued progress. App avoids use of punishment for failure to perform target behaviors or meet goals.
- Usable and aesthetically appealing design: visual design of the technology is attractive and appealing and adheres to basic usability standards. Design captures and sustains the user's interest, enhances user engagement with the technology, and adds to the credibility and usability of the product.
- Accuracy of user information: technology should not inaccurately record or misrepresent the user's behavior (for instance, due to limitations in automatic sensing capabilities or the inability to use the device in certain environments). If necessary—to obtain an accurate, comprehensive account of behavior—the technology should allow users to edit data records and/or manually input additional data that the device is incapable of detecting automatically.
- Appropriate time and place: information, feedback, and assistance are provided at an opportune time and place (ie, when and where it is needed, at the most appropriate time, and in the most effective manner).
- Visibility of user's status: technology should always keep the user informed about progress toward goals through appropriate feedback within a reasonable time frame. Feedback is accurate and easily understood (eg, through use of abstract displays, summary data).
- Customizability: users should be able to customize aspects of the technology, for example, creating personalized goals and customizing product settings (public/private data, interface, etc). However, customizability should not interfere with persuasive aspects.
- Educate users: users should understand why their actions promote positive behaviors and how their goals are being

met. This includes which specific behaviors lead to the accomplishment of a larger goal. The technology should engage users in an active process whereby they learn information and gain skills relevant to their goals, particularly skills that would enable them to continue to progress toward goals even in the absence of the technology.

Table 1. Best and worst case examples of usability by heuristic domain.

Heuristic	Best case	Worst case
Appropriate functionality	• Has a comprehensive set of persuasive health features such as a toothbrush timer, goal tracking, and reminders that run smoothly • Is intuitive and easy to use • Free of errors and crashes	• Primary function is a toothbrush timer, which doesn't work • Lacks basic features such as reminders, sound controls, etc. • Has frequent errors and crashes repeatedly
Not irritating or embarrassing	• Free of ads and other unnecessary interruptions • Functions smoothly, navigation is intuitive, and screens and features load quickly • Gives user the ability to customize settings and controls such as sound, timer length, and privacy	• Frequently interrupts user with full-screen ads, flashing banners, and solicitations • Slow loading times; disorganized, inconsistent navigational controls • No user guides, assistance, or help of any kind provided
Protects user's privacy	• Log-in/password protected • Informs/reassures user about privacy protections and rights • Allows user to control their data permissions/usage	• No log-in/password option • No mention of privacy policy or what happens to data • Shares data publicly or with third parties such as advertisers
Use of positive motivation strategies	• Primarily uses gain-framed messaging and positive imagery throughout • Positively reinforces achievement of target behavior: "Great job! Keep it up!" • Rewards user with badges, points, or other incentives when achieving target behaviors/goals	• No positive motivation strategies are used at all; says nothing when you complete a tooth brushing session • Primarily uses loss-framed messaging and negative imagery (eg, cartoon of person suffering from bad breath) • Warns potential mates will run from you if you have bad breath
Usable and aesthetically appealing design	• Interface is well designed and professional looking • High-quality, attractive images and graphics are used • Navigational elements and controls are intuitive and easy to use, keep the user well oriented	• Unappealing color choices, such as a puke-green background • Poor image quality; pixelated, unattractive imagery and graphics; reuses the same image for different sections • Disorganized, disorienting, and inconsistent navigation controls
Accuracy of user information	• Option to track multiple aspects of user information and behavior • Tracking features are accurate, editable, and function appropriately • Allows user to manually input, edit, or delete their information	• Has no ability to track or store user information/behavior • Tracking features do not function properly or are highly inaccurate • Does not provide a way to input, edit, or delete existing user information
Appropriate time and place	• Information is coherent and well organized for consumption • Provides user with mini-tutorials when accessing new features • Simply shaking the device allows the user to easily report an issue	• Textual information is plagued by spelling, grammatical, or syntax errors; difficult to consume • No guides, assistance, or help of any kind is provided for the user • App randomly asks user to solve math problems
Visibility of user's status	• Provides summary data of user's stats on home screen • Provides more in-depth explanation of user progression data with historical graphs/charts • Shows progression bars as user works toward goals	• No user data is tracked, so there is no feedback on status or progress towards goals • Does not let the user know which content they've reviewed; doesn't provide option to do so • Toothbrush timer is tiny; difficult to see progress
Customizability	• Provides ability to customize colors, sounds, and notifications • Provides ability to choose from among several oral hygiene goals • Provides ability to create multiple user profiles so a family can share the device, for instance	• No ability to customize colors, sounds, or notifications • No ability to set personalized goals or tailor oral hygiene strategies, such as timer length • No ability to customize privacy settings or control user data
Educates users	• Provides educational material on the benefits of oral hygiene practices • Provides videos that explain how to properly brush, floss, and perform other types of oral care • Positively reinforces target behaviors by reminding user of end-goal benefits	• No educational, instructional, or persuasive content whatsoever • No clear goals for the user to achieve • Does not attempt to engage the user in any process or explain why user should perform certain types of behaviors

For each of the heuristics above, we applied a standard severity scoring system recommended by Nielsen [36] and adopted by Kientz [34]; however, we modified our application of the scoring system for the purposes of this paper. Rather than scoring each identified usability problem, we assigned a single severity score for each heuristic domain for each app and then calculated an overall severity score for each app. This methodology allowed us to better compare the usability of the selected apps to one another and comment on the overall user-centeredness and quality of the apps. The scoring criteria applied were as follows: 0=not a usability issue, 1=cosmetic problem only, 2=minor usability problem, 3=major usability issue, and 4=usability catastrophe. We took the perspective that an ideal persuasive technology should adhere to the objectives of all 10 heuristics. If a domain was not addressed at all, it was coded as a usability catastrophe. Due to resource limitations, a single reviewer (BT) coded each app, but the application of the scoring system was discussed and agreed upon by all coauthors and individual items were randomly selected and reviewed by the team to ensure a consistent application of criteria.

Table 1 includes specific examples of how the scoring criteria were applied, illustrating the type of characteristics observed in the reviewed apps that would receive a low severity score (best case) and a high severity score (worst case).

Results

Basic Features and Functionality

A total of 33 apps met the selection criteria and were reviewed (Table 2). Most were sponsored by a software developer (19/33, 58%) or dental care provider (6/33, 17%) and were targeted to the general public (22/33, 67%). However, 12% (4/33) were offered by companies as an adjunct for their electric toothbrushes. All but 2 apps (31/33, 94%) were free to download.

Common design features included the ability to provide feedback on the app (12/33, 36%), customize aspects of appearance or sounds (12/33, 36%), set up a log-in account (9/33, 27%), access customer support (9/33, 27%), customize one's oral health goals (8/33, 24%), and share progress with others (5/33, 15%). Four apps (12%) specifically promoted oral hygiene products for purchase; 6 included gamification features such as awarding badges for behavioral milestones and accomplishments.

Key functions are presented in Table 3. Most (20/33, 61%) included a timer for brushing teeth, and for 45% (15/33) of the apps reviewed, this was the primary function of the app. Other common functions included tips for better managing oral hygiene, reminders and alerts for improving oral hygiene (eg, notification to floss before bed or replace an old toothbrush), and general oral health-related educational content. Five apps (15%) allowed people to communicate with a dental health professional to either ask questions, request a video-based oral health assessment, or inquire about dental insurance. Five (15%) were designed to help users look for a dental care provider.

Theoretical or Empirical Basis

None of the apps cited a theoretical foundation for design or content. Only 1 mentioned any evaluation of the app's impact on changing users' knowledge, attitudes, or behavior. The cited study, however, was limited to a cross-sectional survey of users' perceptions and did not assess the actual effectiveness of the intervention in a randomized or longitudinal study [20]. None of the apps appeared to have been empirically validated.

Key Messaging: Content and Framing

More than half (19/33, 57%) of the apps included written content intended to influence oral health behavior as opposed to simply providing instructional directions for how to use the features or functionality (eg, toothbrush timer). The majority of these apps (15/19, 79%) included a mix of gain- and loss-framed messages (Table 4). A third (6/19, 32%) included a larger proportion of gain-framed messages compared to loss-framed messages, and 1 app included only neutral messaging with neither a gain nor loss frame.

Examples of gain-framed messages included:

> *To keep your teeth and gums healthy you should clean between teeth daily with floss or an interdental cleaner.*
>
> *Brush your way to a fresh smile.*

Examples of loss-framed messages included:

> *In general, the higher the frequency and quantity of sugary foods and drinks you intake per day, the more at risk you are of developing tooth decay.*
>
> *If you don't floss and brush your teeth regularly, any food trapped between your teeth will be broken down by the bacteria and may be responsible for bad breath.*

Twelve apps (36%) included some discussion of diet, but the discussion was either brief (5/12, 42%) or partial (5/12, 42%). Only 2 provided more comprehensive dietary information. Most of the apps that mentioned diet (7/12, 58%) recommended limiting or avoiding sugary foods that cause tooth decay or bad breath, and 33% (4/12) mentioned that acidic food or drinks can damage teeth. Several apps (3/12, 25%) endorsed eating a diet rich in fruits and vegetables, but 1 suggested eating fruits and vegetables could stain one's teeth.

Eleven apps (33%) briefly or partially discussed tobacco use. None included a full discussion of the effects of tobacco on oral health outcomes. Five of the 11 apps that included some tobacco discussion (45%) advised people not to use or to quit tobacco, 36% (4/11) emphasized tobacco's role in causing bad breath, 36% (4/11) emphasized that tobacco causes tooth stains, and 36% (4/11) linked tobacco use to oral disease. None referred people to stop-smoking treatment services.

Nine apps (27%) briefly addressed alcohol use. Seven of these (7/9, 78%) suggested drinking be limited or avoided to reduce the risk of unwanted oral health issues such as bruxism, bad breath, dry mouth, tooth decay, or stained teeth. Only 1 app noted that alcohol use is a risk factor for oral cancer. None referred people to treatment services to reduce their drinking.

Table 2. Reviewed apps.

Name	Platform	Sponsor	Cost	Number of installs[a]	Number of user ratings[a]	Average user rating[a]
24/7 Live Dentist Response	iOS	Digital health company	Free	Unknown	Unknown	Unknown
Bad Breath	Android	App developer	Free	1000	3	3.7
Bad Breath Guide	iOS	Dental provider	Free	Unknown	Unknown	Unknown
Brush DJ	Android & iOS	App developer	Free	100,000	1207	4.1
Brushing and Whitening Teeth	Android	App developer	Free	5000	106	4.4
Brush'n'Save	Android	Dental provider	Free	10,000	74	4
Brushy	Android	App developer	Free	10,000	115	3.3
Colgate Connect	iOS	Oral hygiene product company	Free	Unknown	81	4.3
DDS Anywhere	Android & iOS	App developer	Free	5000	23	4
Delta Dental	Android & iOS	Insurance provider	Free	100,000	793	3
Dentacare - Health Training	Android	App developer	Free	10,000	Unknown[b]	Unknown[b]
DentAdvisor: Oral Care Expert	iOS	App developer	$1.99	Unknown	Unknown	Unknown
Dental Care	Android	App developer	Free	1000	9	4.6
Dental Care—Target Smile	Android	App developer	Free	1000	40	4.9
Dental Desk	Android	Dental provider	Free	10,000	66	4.6
Do I Grind	Android & iOS	Digital health company	Free	1000	40	4.1
FoodForTeeth[c]	iOS	Dental provider	Free	Unknown	Unknown	Unknown
Healthy Teeth[d]	iOS	App developer	Free	Unknown	Unknown	Unknown
Kolibree	Android & iOS	Oral hygiene product company	Free	5000	136	3.3
Let’s Brush Free	iOS	App developer	Free	Unknown	Unknown	Unknown
Moment of Tooth	iOS	App developer	Free	Unknown	8	4.9
My Dental Care	Android & iOS	Dental provider	Free	500	14	5
MySmile	Android & iOS	App developer	Free	10,000	126	4.1
Oral-B App	Android & iOS	Oral hygiene product company	Free	1,000,000	12,489	3.3
Philips Sonicare	Android & iOS	Oral hygiene product company	Free	50,000	564	2.7
QuickBrush Toothbrush Timer	Android	App developer	Free	10,000	176	4.2
Smile—Dental Hygiene Analysis	iOS	Dental provider	Free	Unknown	14	3.6
Tooth Notes	iOS	App developer	$0.99	Unknown	Unknown	Unknown
Toothbrush Pacer	Android	App developer	Free	10,000	98	4
Toothbrush timer	Android	App developer	Free	100,000	1046	3.5
Toothbrush Timer	Android	App developer	Free	50,000	776	3.8
Toothy: Brush Floss Rinse!	iOS	App developer	Free	Unknown	401	4.6
United Concordia Dental Mobile	Android & iOS	Insurance provider	Free	50,000	232	3.3

[a]Details on the number of installations, number of user ratings, and average rating are only available in the App Store for some iOS-based apps. Data presented on apps available for both Android and iOS list data from Google Play only. User ratings scored on a 1 (worst) to 5 (best) scale. All data were current as of the time of this review.

[b]This app was released as a beta version, so rating information was not yet being tracked.

[c]Full app name: FoodForTeeth—Food Database and Diet Diary.

[d]Full app name: Healthy Teeth—Tooth Brushing reminder with timer.

Table 3. Summary of key app functions.

Key function	Total, n (%)
Tooth brushing timer	20 (61)
Tips for better oral hygiene	16 (48)
Oral health educational content	13 (39)
Oral hygiene alerts or reminders	13 (39)
Tracking of oral health behaviors (eg, brushing)	9 (27)
Tracking of dental appointments	7 (21)
Ability to communicate with a dental professional	5 (15)
Ability to search for dentists	5 (15)

Table 4. Gain and loss framing in apps with persuasive health messages. Table does not include apps that provided instructional content only (n=14).

App name	Total messages, N	Gain framed, n (%)	Loss framed, n (%)	Neutral framed, n (%)
MySmile	22	0 (0)	9 (41)	13 (59)
Do I Grind	21	3 (14)	11 (52)	7 (33)
Delta Dental	81	8 (10)	30 (37)	43 (53)
Bad Breath	159	31 (20)	46 (29)	82 (52)
United Concordia Dental Mobile	113	7 (6)	18 (16)	89 (79)
Bad Breath Guide	352	48 (14)	74 (21)	220 (63)
Brushing and Whitening Teeth	424	9 (2)	16 (4)	98 (23)
Dental Desk	172	6 (4)	16 (9)	154 (90)
FoodForTeeth[a]	499	50 (10)	73 (15)	377 (76)
My Dental Care	691	45 (7)	70 (10)	576 (83)
DentAdvisor: Oral Care Expert	909	42 (5)	77 (9)	777 (86)
QuickBrush Toothbrush Timer	13	1 (8)	1 (8)	11 (85)
Brush DJ	19	0 (0)	0 (0)	19 (100)
Dental Care	430	69 (16)	65 (15)	295 (69)
Dental Care—Target Smile	48	1 (2)	0 (0)	47 (98)
Moment of Tooth	21	8 (38)	5 (24)	8 (38)
Let's Brush Free	21	5 (24)	0 (0)	16 (76)
Philips Sonicare	11	8 (73)	1 (9)	2 (18)
Oral-B App	23	18 (78)	2 (9)	3 (13)

[a]Full app name: FoodForTeeth—Food Database and Diet Diary.

Heuristic Review

Although not all the reviewed apps included explicit health persuasion messages, all were intended to influence users' oral health attitudes and behavior. Thus, we reviewed all 33 apps using the heuristics recommended for persuasive health technologies.

Table 5 summarizes the number of apps that received each severity score by heuristic. Most apps received a severity rating of 3 or 4 across each heuristic. In general, apps scored poorly on protecting users' privacy but had fewer issues with usability and aesthetic appeal. We also calculated an overall severity score for each app by summing the scores across all domains. Across all apps, the average overall severity score was 30 out of 40 (range 16 to 39). This score reflects major, systemic usability problems.

Table 5. Number of apps receiving usability severity scores for each persuasive health technology heuristic.

Heuristic	No issues (score=0)	Cosmetic issues (score=1)	Minor issues (score=2)	Major issues (score=3)	Usability catastrophe (score=4)
Appropriate functionality	0	0	5	26	2
Not irritating or embarrassing	0	1	9	20	3
Protects user privacy	2	0	4	5	22
Use of positive motivation strategies	0	3	1	17	12
Usable and aesthetically appealing design	0	6	13	12	2
Accuracy of user information	0	0	3	14	16
Appropriate time and place	0	4	2	25	2
Visibility of user status	0	4	1	20	8
Customizability	0	0	5	10	18
Educates users	0	0	4	20	9

Discussion

Principal Findings

We selected 33 of the most popular and highly rated oral health apps available for Android and iOS and reviewed the design, content, and usability of each. As a group, apps were of poor quality. All were intended to influence adults' oral health attitudes and behaviors, but none were empirically validated to demonstrate their effectiveness. None cited any theoretical basis for the content, which is not unusual for mHealth apps, but since the majority were created by developers who did not appear to be affiliated with oral health or behavioral science experts, it is likely that the content and design of most were not driven by sound behavioral theory. As a case in point, prospect theory suggests that gain-framed messages are more effective at promoting preventive health behaviors than loss-framed messages [26,37]. This theory has some empirical basis, as prior research has shown gain-framed messages to be effective at promoting oral hygiene behaviors [27-31]. As such, one might expect to see gain-framing (as opposed to loss-framing) as a central feature in apps designed to improve oral hygiene behavior. However, of the 19 apps that included any explicit persuasive messaging, only 6 included content balanced in favor of gain-framed messaging and most of these still included loss-framed messages. Only 2 apps included gain-framed messaging without loss-framed messaging, but both included only a few gain-framed messages—far less than might be considered ideal for a persuasive intervention.

The oral health educational content had other issues, as well. For example, diet, tobacco, and alcohol are significant risk factors for oral disease [32], yet one-third or less of the apps reviewed discussed each of these important topics. When included, the content was typically brief and sometimes included information that might be considered counter-productive to the goal of improving users' health behavior. A prime example is the app that advised users that eating fruits and vegetables can stain one's teeth. While there is some evidence that eating certain fruits, vegetables, dairy, or soy products may contribute to the development of dark spots on the teeth called "black stain," which are microflora deposits, the prevention or management of this condition is more complex than avoiding fruits and vegetables [38,39] and, in fact, the American Dental Association recommends that eating fruits and vegetables can benefit oral health [40]. In other examples, the developers' attempts to use humor seemed to undermine the intended message. For instance, 1 app advised users that whether or when they clean their tongue should depend on whether they will be meeting "the crush of their life" and what they have eaten recently. In contrast, according to the American Dental Association, there is no evidence that brushing or scraping one's tongue prevents bad breath [41].

It is also notable that of the apps that included any discussion of tobacco and alcohol, very few linked use of these substances to oral cancer risk. Most highlighted the cosmetic (eg, stained teeth) or social implications (eg, bad breath) of their use instead. Given the independent and synergistic effects that these substances have on oral disease [32], future oral health apps should more clearly articulate this risk and consider providing treatment referral information for those interesting in quitting or cutting back.

Finally, the reviewed apps all had significant usability issues based on our heuristic review. Only 2 took measures to protect users' privacy, and none received a perfect severity score (ie, 0) on the other 9 heuristics. Significant usability issues (severity score of 3 or 4) were noted for most apps across the heuristic domains. Three apps sponsored by companies promoting oral health products (Oral B, Philips Sonicare, and Colgate Connect) were generally well designed and received higher scores but were narrowly intended to promote use of their products, resulting in lower overall scores as persuasive health technologies.

Strengths and Limitations

This study has a number of notable strengths, including its novelty. To our knowledge, this is the first review of oral health promotion apps. As such, this paper addresses an important gap in the literature. It also establishes the need and opportunity to create high-quality oral health promotion apps targeting adults. Oral disease affects billions of people worldwide [1] and mobile phone use is rapidly expanding around the world [8], so the

impact of effective oral health interventions delivered via mobile phone could be substantial.

Another strength of this review is its comprehensive nature, in which both the quality of persuasive health content and user-centered design of the apps were evaluated. Our application of Kientz's persuasive health technology heuristics is also novel. Typically, these heuristics are used to identify a range of user issues and each issue is scored based on its severity rather than assigning a score to each heuristic domain or an overall severity score to the entire app. However, our approach was reasonable for the purposes of this review since it allowed us to compare the relative quality of the user design across apps using a common evaluation and scoring scheme.

Study limitations should also be noted. First, it is possible that there are existing oral health apps which are of higher quality than that of the most popular apps we reviewed. Resource restraints prevented us from reviewing all of the available apps (nearly 3000 were identified based on our keywords), but we believe our focus on the most popular apps is reasonable because these are the apps which are most frequently being used by the public. Next, the downside of our comparative heuristic review is that it treated each heuristic as equally important, which may not be the case. For this reason, we presented both overall app severity scores and domain-specific scores so readers can get a better sense of where usability issues were observed. But we acknowledge that depending on the nature of the app, performance in some heuristic categories may be more important than others so in future reviews it might be more appropriate to differentially weight the domains. Another limitation is that our heuristic review was performed by a single coder trained in user-centered design using standardized scoring criteria. More typically, if the goal had been to delineate all of the observed issues, a group of coders might be used. This was prohibited due to limited project resources but is also unlikely to have changed our conclusion about the quality of the design of these apps as a group based on the pervasiveness and severity of issues observed. Finally, we note that our review was limited to content that was available through installation of each app. Additional content only available to certain audiences such as health plan members or owners of purchased electric toothbrushes and protected via account log-in could not be viewed.

Conclusions

Many oral health apps are available to consumers, but based on this review of the most popular and highly rated ones, the quality of these apps is generally poor. Important opportunities exist to develop oral health promotion apps whose content is theoretically grounded and evidence-based and that adhere to good design principles for persuasive health technologies.

Acknowledgments

This research was supported by the National Institute for Dental and Craniofacial Research (U01DE024462; principal investigator: JBM). We are grateful to Ella Thompson for her management of this project, Tooka Zokaie for her participation in an initial review of available oral health apps, and Annie Shaffer for her assistance with the final manuscript preparation. A portion of the findings in this paper were presented at the 2018 annual meeting of the Society of Behavioral Medicine in New Orleans, Louisiana.

Conflicts of Interest

None declared.

References

1. Kassebaum NJ, Smith AGC, Bernabé E, Fleming TD, Reynolds AE, Vos T, GBD 2015 Oral Health Collaborators. Global, regional, and national prevalence, incidence, and disability-adjusted life years for oral conditions for 195 countries, 1990-2015: a systematic analysis for the global burden of diseases, injuries, and risk factors. J Dent Res 2017 Apr;96(4):380-387 [FREE Full text] [doi: 10.1177/0022034517693566] [Medline: 28792274]
2. Oral health in America: a report of the Surgeon General. Rockville: US Department of Health and Human Services, National Institute of Dental and Craniofacial Research, National Institutes of Health; 2000. URL: https://www.nidcr.nih.gov/sites/default/files/2017-10/hck1ocv.%40www.surgeon.fullrpt.pdf [accessed 2018-08-20] [WebCite Cache ID 71p5ANVo6]
3. Petersen PE, Bourgeois D, Ogawa H, Estupinan-Day S, Ndiaye C. The global burden of oral diseases and risks to oral health. Bull World Health Organ 2005 Sep;83(9):661-669 [FREE Full text] [Medline: 16211157]
4. Kumar S, Tadakamadla J, Johnson NW. Effect of toothbrushing frequency on incidence and increment of dental caries: a systematic review and meta-analysis. J Dent Res 2016 Oct;95(11):1230-1236. [Medline: 27334438]
5. Sambunjak D, Nickerson JW, Poklepovic T, Johnson TM, Imai P, Tugwell P, et al. Flossing for the management of periodontal diseases and dental caries in adults. Cochrane Database Syst Rev 2011 Dec 07(12):CD008829. [Medline: 22161438]
6. Institute of Medicine. Advancing Oral Health in America. Washington: The National Academies Press; 2011.
7. Mobile fact sheet. Washington: Pew Internet and American Life Project; 2018 Feb 05. URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-08-20] [WebCite Cache ID 70LYMn3N0]
8. Poushter J. Smartphone ownership rates skyrocket in many emerging economies, but digital divide remains. Washington: Pew Internet and American Life Project URL: http://assets.pewresearch.org/wp-content/uploads/sites/2/2016/02/pew_research_center_global_technology_report_final_february_22__2016.pdf [accessed 2018-06-21] [WebCite Cache ID 70LZtqBw8]

9. 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]
10. Purcell K. What types of apps are adults downloading?. Washington: Pew Internet and American Life Project; 2011. URL: http://www.pewinternet.org/2011/11/02/part-iv-what-types-of-apps-are-adults-downloading/ [accessed 2018-06-21] [WebCite Cache ID 70LaMeba6]
11. Research2Guidance. 2018. mHealth Economics 2017 – Current Status and Future Trends in Mobile Health URL: https://research2guidance.com/wp-content/uploads/2017/10/1-mHealth-Status-And-Trends-Reports-Cover3.jpg [WebCite Cache ID 70Lay0CKE]
12. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
13. Holmen H, Wahl AK, Cvancarova SM, Ribu L. Tailored communication within mobile apps for diabetes self-management: a systematic review. J Med Internet Res 2017 Jun 23;19(6):e227 [FREE Full text] [doi: 10.2196/jmir.7045] [Medline: 28645890]
14. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
15. 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]
16. 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;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
17. Rat C, Hild S, Rault SJ, Gaultier A, Quereux G, Dreno B, et al. Use of smartphones for early detection of melanoma: systematic review. J Med Internet Res 2018 Apr 13;20(4):e135 [FREE Full text] [doi: 10.2196/jmir.9392] [Medline: 29653918]
18. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The usability and effectiveness of mobile health technology-based lifestyle and medical intervention apps supporting health care during pregnancy: systematic review. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e109 [FREE Full text] [doi: 10.2196/mhealth.8834] [Medline: 29691216]
19. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
20. Underwood B, Birdsall J, Kay E. The use of a mobile app to motivate evidence-based oral hygiene behaviour. Br Dent J 2015 Aug 28;219(4):E2. [doi: 10.1038/sj.bdj.2015.660] [Medline: 26315196]
21. Schluter P, Lee M, Hamilton G, Coe G, Messer-Perkins H, Smith B. Keep on brushing: a longitudinal study of motivational text messaging in young adults aged 18-24 years receiving Work and Income Support. J Public Health Dent 2015;75(2):118-125. [doi: 10.1111/jphd.12079] [Medline: 25409935]
22. Bowen TB, Rinchuse DJ, Zullo T, DeMaria ME. The influence of text messaging on oral hygiene effectiveness. Angle Orthod 2015 Jul;85(4):543-548. [doi: 10.2319/071514-495.1] [Medline: 25343689]
23. Jadhav HC, Dodamani AS, Karibasappa GN, Naik RG, Khairnar MR, Deshmukh MA, et al. Effect of reinforcement of oral health education message through short messaging service in mobile phones: a quasi-experimental trial. Int J Telemed Appl 2016;2016:7293516 [FREE Full text] [doi: 10.1155/2016/7293516] [Medline: 26941793]
24. Abroms LC, Lee WJ, 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]
25. Apple Inc. 2018. Optimizing for app store search URL: https://developer.apple.com/app-store/search/ [WebCite Cache ID 70LbLuynh]
26. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979 Mar;47(2):263. [doi: 10.2307/1914185]
27. Rothman AJ, Salovey P. Shaping perceptions to motivate healthy behavior: the role of message framing. Psychol Bull 1997 Jan;121(1):3-19. [Medline: 9000890]
28. Rothman AJ, Martino SC, Bedell BT, Detweiler JB, Salovey P. The systematic influence of gain-and loss-framed messages on interest in and use of different types of health behavior. Pers Soc Psychol Bull 2016 Jul 02;25(11):1355-1369. [doi: 10.1177/0146167299259003]
29. Sherman DK, Updegraff JA, Mann T. Improving oral health behavior: a social psychological approach. J Am Dent Assoc 2008 Oct;139(10):1382-1387. [Medline: 18832274]

30. Brick C, McCully SN, Updegraff JA, Ehret PJ, Areguin MA, Sherman DK. Impact of cultural exposure and message framing on oral health behavior: exploring the role of message memory. Med Decis Making 2016 Dec;36(7):834-843 [FREE Full text] [doi: 10.1177/0272989X15570114] [Medline: 25654986]
31. O'Keefe DJ, Jensen JD. The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease prevention behaviors: a meta-analytic review. J Health Commun 2007;12(7):623-644. [doi: 10.1080/10810730701615198] [Medline: 17934940]
32. US Department of Health and Human Services. Oral health in America: a report of the Surgeon General. J Calif Dent Assoc 2000 Sep;28(9):685-695. [Medline: 11324049]
33. Fogg B. Persuasive Technology: Using Computers to Change What We Think and Do. San Francisco: Morgan Kaufman Publishers; 2003.
34. Kientz J, Chloe E, Birch B, Maharaj R, Fonville A, Glasson C, et al. Heuristic evaluation of persuasive health technologies. 2010 Presented at: Proceedings of the 1st ACM International Health Informatics Symposium; 2010; Arlington.
35. Nielsen J. Ten usability heuristics. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2018-06-21] [WebCite Cache ID 70LbgnxBi]
36. Nielsen J. Severity ratings for usability problems. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2018-06-21] [WebCite Cache ID 70LbgnxBi]
37. Rothman A, Bartels R, Wlaschin J, Salovey P. The strategic use of gain- and loss-framed messages to promote healthy behavior: how theory can inform practice. J Communication 2006;56:S202-S220. [doi: 10.1111/j.1460-2466.2006.00290.x]
38. Chen X, Zhan J, Lu H, Ye W, Zhang W, Yang W, et al. Factors associated with black tooth stain in Chinese preschool children. Clin Oral Investig 2014 Dec;18(9):2059-2066. [doi: 10.1007/s00784-013-1184-z] [Medline: 24430339]
39. Żyła T, Kawala B, Antoszewska-Smith J, Kawala M. Black stain and dental caries: a review of the literature. Biomed Res Int 2015;2015:469392 [FREE Full text] [doi: 10.1155/2015/469392] [Medline: 25802850]
40. Mouth Healthy. American Dental Association; 2018. What you eat affects your health URL: https://www.mouthhealthy.org/en/nutrition/food-tips [WebCite Cache ID 70LbqwlBE]
41. Mouth Healthy. American Dental Association; 2018. Tongue scrapers and cleaners URL: https://www.mouthhealthy.org/en/az-topics/t/tongue-scrapers [WebCite Cache ID 70LbxazRn]

Edited by C Dias; submitted 28.06.18; peer-reviewed by P Krebs, J Owens; comments to author 16.07.18; revised version received 02.08.18; accepted 14.08.18; published 04.09.18.

Please cite as:
Tiffany B, Blasi P, Catz SL, McClure JB
Mobile Apps for Oral Health Promotion: Content Review and Heuristic Usability Analysis
JMIR Mhealth Uhealth 2018;6(9):e11432
URL: http://mhealth.jmir.org/2018/9/e11432/
doi:10.2196/11432
PMID:30181114

Original Paper

Health and Fitness Apps for Hands-Free Voice-Activated Assistants: Content Analysis

Arlene E Chung[1,2,3,4*], MD, MHA, MMCi; Ashley C Griffin[3,4*], MSPH; Dasha Selezneva[3]; David Gotz[4,5], PhD

[1]Division of General Medicine & Clinical Epidemiology, Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC, United States

[2]Division of General Pediatrics and Adolescent Medicine, Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC, United States

[3]Program on Health and Clinical Informatics, University of North Carolina School of Medicine, Chapel Hill, NC, United States

[4]Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[5]School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[*]these authors contributed equally

Corresponding Author:
Arlene E Chung, MD, MHA, MMCi
Division of General Medicine & Clinical Epidemiology
Department of Medicine
University of North Carolina School of Medicine
5034 Old Clinic Building, CB 7110
Chapel Hill, NC, 27599-7110
United States
Phone: 1 9199662276
Fax: 1 9199662274
Email: arlene_chung@med.unc.edu

Abstract

Background: Hands-free voice-activated assistants and their associated devices have recently gained popularity with the release of commercial products, including Amazon Alexa and Google Assistant. Voice-activated assistants have many potential use cases in healthcare including education, health tracking and monitoring, and assistance with locating health providers. However, little is known about the types of health and fitness apps available for voice-activated assistants as it is an emerging market.

Objective: This review aimed to examine the characteristics of health and fitness apps for commercially available, hands-free voice-activated assistants, including Amazon Alexa and Google Assistant.

Methods: Amazon Alexa Skills Store and Google Assistant app were searched to find voice-activated assistant apps designated by vendors as health and fitness apps. Information was extracted for each app including name, description, vendor, vendor rating, user reviews and ratings, cost, developer and security policies, and the ability to pair with a smartphone app and website and device. Using a codebook, two reviewers independently coded each app using the vendor's descriptions and the app name into one or more health and fitness, intended age group, and target audience categories. A third reviewer adjudicated coding disagreements until consensus was reached. Descriptive statistics were used to summarize app characteristics.

Results: Overall, 309 apps were reviewed; health education apps (87) were the most commonly occurring, followed by fitness and training (72), nutrition (33), brain training and games (31), and health monitoring (25). Diet and calorie tracking apps were infrequent. Apps were mostly targeted towards adults and general audiences with few specifically geared towards patients, caregivers, or medical professionals. Most apps were free to enable or use and 18.1% (56/309) could be paired with a smartphone app and website and device; 30.7% (95/309) of vendors provided privacy policies; and 22.3% (69/309) provided terms of use. The majority (36/42, 85.7%) of Amazon Alexa apps were rated by the vendor as mature or guidance suggested, which were geared towards adults only. When there was a user rating available, apps had a wide range of ratings from 1 to 5 stars with a mean of 2.97. Google Assistant apps did not have user reviews available, whereas most of Amazon Alexa apps had at least 1-9 reviews available.

Conclusions: The emerging market of health and fitness apps for voice-activated assistants is still nascent and mainly focused on health education and fitness. Voice-activated assistant apps had a wide range of content areas but many published in the health and fitness categories did not actually have a clear health or fitness focus. This may, in part, be due to Amazon and Google

policies, which place restrictions on the delivery of care or direct recording of health data. As in the mobile app market, the content and functionalities may evolve to meet growing demands for self-monitoring and disease management.

(JMIR Mhealth Uhealth 2018;6(9):e174) doi:10.2196/mhealth.9705

KEYWORDS

voice-activated assistant; intelligent personal assistant; virtual personal assistant; Amazon Alexa; Google Assistant; artificial intelligence; voice-activated technology; voice assistant

Introduction

Hands-free voice-activated assistants (VAAs) have recently gained popularity with the release of commercial products, including Amazon Alexa and Google Assistant and their associated speaker devices. VAAs are also referred to as "intelligent personal assistants," "voice assistants," or "virtual personal assistants" [1]. A voice-activated assistant is a software agent that can perform tasks or services for an individual and uses voice activation for interaction through a smart speaker device. Apple's Siri is an example of a VAA, but requires the user to press a button before one can use voice for interaction. However, hands-free VAAs allow the user to command the speaker device without having to touch the device and by using only their voice. Historically, VAA technologies have been able to perform a range of rudimentary tasks delegated by the user with the primary functions being to organize and manage information [2], such as provide facts or play music. With advancements in machine learning, artificial intelligence, and natural language processing, VAAs can now handle more complex interactions, such as commanding smart home devices and placing orders for merchandize [3-8].

Released in the United States in November 2014, the Amazon Echo was the first commercially available hands-free device controlled by voice interaction. Alexa is the cloud-based, personal voice assistant integrated into Amazon's VAA devices, which include the Echo, Dot, Tap, Look, Spot, and Show. Amazon has a Skills Store that houses "skills," which are apps that "…add new capabilities that create a more personalized experience with your Alexa-enabled devices…" [9]. The Skills Store is similar to the iOS and Android app stores and allows a variety of skills to be "enabled" for use on Amazon Alexa devices. Although many of them are free to "enable," some have associated fees or require accounts to use the skills. Google Assistant is the voice assistant powering Google Home, which was released in November 2016 in the United States. In addition to Google Home, Google Home Mini and Google Home Max devices were released in late 2017. Similar to the Amazon Alexa-powered devices, Google Assistant has in-house and third-party apps called "actions." As with Amazon, some apps require that the user link a mobile phone app account to their Google account before using the service with the Google Assistant. Most Google Assistant apps are already enabled by default. Both Amazon Alexa and Google Assistant have platforms for developers to create apps.

There are a number of potential health-related use cases for VAAs because they could be used in a variety of settings (eg, patients' homes or in clinics and hospitals) for many different functions, such as home monitoring of symptoms or health education. Evidence suggests that VAAs could increase accessibility to information for those with physical, sensory, and cognitive impairments and facilitate self-management or education [6,8,10-13]. Prior research concerning use cases for VAAs have featured health tracking and monitoring, assistance with locating health providers, and collecting data to aid in decision making [3-5], but VAAs studied were not commercially available products. At this writing, there are only a few studies that have utilized a commercially available VAA for a health use case. These studies have used Amazon Alexa to assess deaf speech [14], provide task support for individuals with cognitive disabilities [15], and receive voice input from patients to determine "unexpected changes in mood" [5]. Additionally, a research study using Amazon Alexa was recently launched to increase physical activity among overweight or obese cancer survivors [16]. There is also limited research on the attitudes of patients or health care providers regarding the use of VAAs for health or fitness. Two studies based on customer reviews indicated that users are interested in potentially utilizing these devices for self-management, as a memory aid, or overcoming accessibility issues [12,13].

Although there is rich literature regarding the characteristics and use of health and fitness mobile phone apps [17-25], no prior studies have described the characteristics of health and fitness apps for commercially available hands-free VAAs. Previous VAA studies have focused only on research-grade VAAs that were not commercially available, were not hands-free, and were primarily focused on usability and design of the devices rather than the apps that could be used with VAAs [1,2,6,7,10,26,27]. Thus, this study is the first to examine the features and characteristics of hands-free commercially available VAA apps for health and fitness based on information available from app marketplaces.

Methods

Selection Criteria and Methodology

VAA apps are uniquely different from mobile phone apps, in that the full scope of the types of interactions are not clearly delineated by interacting with the voice interface because it does not have the same transparency as interacting with a physical user interface (eg, the screen of a mobile phone). VAA vendors generally provide only a few examples of invocation commands, and there is no menu of features or functionalities such that the user or evaluator could understand the full spectrum of the types of commands one could ask the VAA app or what types of information could be provided by the app. Thus, these aspects of VAAs and voice-based interfaces do not allow for the direct application of the traditional review methods used for mobile health apps. Because there were no review

methodologies specific to VAA apps, we used existing methods for evaluating the content of mobile health apps for guidance [17-21,28]. This study specifically focused on conducting a descriptive content analysis based on the information provided by vendors to determine the types of apps released in the health and fitness categories for commercially available VAAs because this is the information consumers use to select apps from VAA app marketplaces.

As of the review date April 19, 2017, Amazon and Google were the only companies with commercially available hands-free VAAs (Amazon Alexa and Google Assistant). Thus, the Amazon Alexa Skills Store website and mobile phone app (iOS and Android) and Google Assistant mobile phone app (iOS and Android) were searched to determine the availability of eligible voice apps. The full list of categorized voice apps for the Google Assistant was only available through the Google Assistant mobile phone app (iOS and Android). There were 23 types of VAA app categories listed on the Alexa Skills Store and 17 categories listed on the Google Assistant mobile phone app.

Inclusion criteria included any VAA apps that were categorized by vendors in the "health and fitness" categories. Both Amazon Alexa and Google Assistant have a "health and fitness" category for apps, and apps could be cross-listed in multiple categories; for example, an Amazon Alexa nutrition app could be listed by the app vendor in both the health and fitness category and the food and drink category. For the apps meeting the inclusion criteria, information provided by vendors was extracted into an evaluation form (Textbox 1). The Amazon Alexa skill release date was retrieved from a third-party website [9,29] but was manually verified using the dates available on the Amazon Skills Store. The Google Assistant app release dates were not publicly available except for a press release, which listed all the apps released as of April 19, 2017. Thus, this was the date selected for data extraction. Figure 1 presents a flowchart for identification, screening, and review of apps.

Analysis

A codebook of categories for the type of health and fitness app, intended age group, and the target audience was created to evaluate each app (Table 1). Codebook development was guided by definitions from mobile phone app reviews of app content [18,22] because there were no VAA app reviews to use for guidance. Intended age was used to classify the age group that the app was designed for, and the target audience was used to classify the population most likely to use the app or be the end user [23]; for example, a baby monitoring app would be coded with an intended age category of children, and the target audience would be parents or families. If a vendor provided a rating of "guidance suggested" or "mature," these were coded as for adults because Amazon states these types of apps contain nudity, violence, references to substance use, profanity, or sexuality and that these apps are for adults only [23]. Google apps did not provide any vendor ratings.

A single app could be coded into multiple health and fitness, intended age, and target audience categories. The health and fitness categories for baby naming, beauty tips, baby monitoring and tracking, dog monitoring and tracking, brain training and games, and time and task management were added during the iterative coding process.

Textbox 1. Extracted vendor information.

General information

- App name
- Release date, if available
- Cost to enable or link app to voice-activated assistant device
- Vendor rating (mature audience or guidance suggested), if available

Vendor information

- Name of the developer
- Has a developer policy
- Has a privacy policy

User ratings

- Number of user reviews, if available
- User rating from 1-5 stars

Features

- Ability to pair with a mobile phone app, website, or device
- Description of the app
- Example voice interaction or invocation word(s)

Figure 1. Health app identification, screening, and review assessment flowchart.

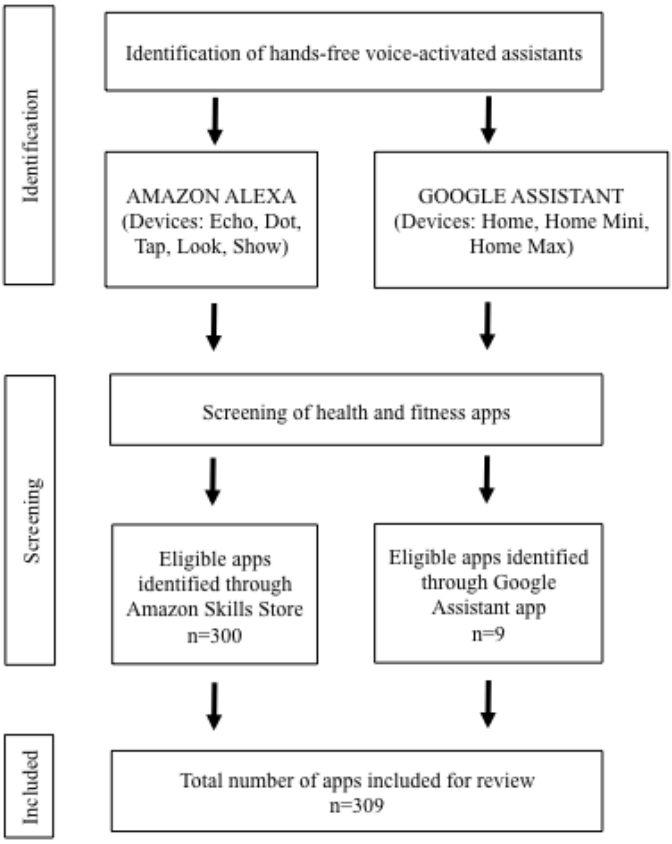


Table 1. Definitions of health and fitness, target audience, and intended age group categories.

Name of Category	Definition
Health and fitness category	
Air quality and monitoring	Inform users about air condition, air quality, or pollutants
Baby naming	Provide suggestions for baby names
Baby monitoring and tracking	Track feeding, diaper changes, naps, or other baby and toddler events
Beauty tips	Provide tips for beauty topics, such as hair, skin, and foot care
Brain training and games	Contain games or tests for memory or mental awareness
Care giving	Provide information for caregivers
Diet and caloric intake	Track calories, food, or beverage intake
Dog monitoring and tracking	Track feeding, walks, sleep, or other dog events like stools
Fitness and training	Provide workout plans, improve physical fitness, or assist with training
Global positioning system or geographic information system	Track location or assist with navigation to locations
Health education	Provide information or tips about health and wellness topics
Health location	Locate medical resources, fitness centers, or wait times at medical facilities
Health monitoring	Review vital signs or labs (blood pressure, heart rate, height, weight, body mass index, blood sugars, etc) and track medications or symptoms
Immunization tracking	Track immunizations
Meditation	Provide practices or techniques to promote relaxation, build internal energy, or mindfulness
Mental health	Provide information on mental health issues, such as depression or anxiety
Motivational	Provide guidance on the desire to perform a specific action or behavior
Nutrition	Provide advice or information on dietary topics, meal or snack planning, etc
Pain management	Inform users about pain management relief strategies
Pregnancy tracking	Provide guidance for users who are planning or expecting to have a baby
Stress management	Inform users about ways to manage stress
Sleep	Inform users on ways to improve sleep or to track sleep
Smoking cessation	Provide information about smoking cessation strategies
Time or task management	Provide strategies for tracking or managing time and tasks
Other	Does not fall into any of the above categories
Target audience (end user)	
Family	Family relations (parents, child, extended family)
Medical professional	Person who provides medical care (doctor, nurse, etc)
Parent	Person with one or more children
Patient	Person seeking medical care or has a medical condition
Pregnant women	Woman expecting a baby
Women	For women
Men	For men
Pet owner	Person with one or more pets
Student	Person who attends an educational institution
General or not specified	Does not specify an intended user
Other	Does not fall into any of the above categories
Intended age group	
Adult[a]	≥18 years

Name of Category	Definition
Children	Newborn to <18 years
Older adults	≥65 years
Not specified	Does not specify a target age

[a]The vendor rating guidelines from Amazon include the following definitions: 1) guidance suggested: may have nudity or suggestive content or require supervision due to account linking, location detection, etc and 2) mature: content is for adults only. Any app that had a vendor rating of guidance suggested or mature was coded as for adults. Additionally, apps that specifically mention adults in the app description were coded for adults.

Two reviewers (DS and ACG) independently coded each app based upon information provided by vendors into one or more health and fitness categories, intended age groups, and target audience groups. Any discordant codes were reconciled through discussion with a third reviewer (AEC) until consensus was reached. Descriptive statistics were used to summarize the number of apps in each of the health and fitness, intended age, and target audience categories. Additionally, the interrater agreement between primary coders was calculated.

Results

There were 309 apps in total that met inclusion criteria for being listed as "health and fitness" apps by vendors, as seen in Figure 1. We identified 300 apps from the Amazon Skills Store website and Amazon Alexa mobile phone app, and 9 apps were identified from the Google Assistant mobile phone app. Of the 9, 5 Google apps were also available in the Amazon Alexa Skills Store. The percent agreement for coding between raters was 91%.

Apps reviewed had a release date between November 6, 2015 and April 19, 2017. The Google Assistant app store did not have any user reviews available. With respect to Amazon Alexa apps, 174 apps had user reviews with a total of 1862 reviews ranging from 1 to 447 reviews per app (Table 2). On average, health and fitness apps were rated 2.97 out of 5 stars. All apps were free to enable, though some required an associated account, which may or may not charge a subscription fee. Some of the apps could be paired with a mobile phone app, website, or a device (56/309, 18.1%).

Table 2. Health and fitness app characteristics[a].

App characteristics	Google (n=9)	Amazon (n=300)	Total (N=309)
Vendor rating[b], n (%)			
Guidance suggested	0 (0)	33 (11.0)	33 (10.7)
Mature	0 (0)	3 (1.0)	3 (1.0)
Not available	9 (100.0)	264 (88.0)	273 (88.3)
User rating (1-5 stars), n (%)			
1-1.9	0 (0)	41 (13.7)	41 (13.2)
2-2.9	2 (22.2)	41 (13.7)	43 (13.9)
3-3.9	6 (66.7)	30 (10.0)	36 (11.7)
4-5	0 (0)	62 (20.6)	62 (20.1)
Not available	1 (11.1)	126 (42.0)	127 (41.1)
User reviews[b], n (%)			
1-9	0 (0)	152 (50.7)	152 (49.2)
10-99	0 (0)	18 (6.0)	18 (5.8)
≥100	0 (0)	4 (1.3)	4 (1.3)
Not available	9 (100.0)	126 (42.0)	135 (43.7)
Cost: free to enable	9 (100.0)	300 (100.0)	309 (100.0)
Has a developer policy	5 (55.6)	64 (21.3)	69 (22.3)
Has a privacy policy	9 (100.0)	86 (28.7)	95 (30.7)
Ability to pair with a mobile phone app, website, or device	3 (33.3)	53 (17.7)	56 (18.1)

[a]There were 309 apps evaluated. Apps could be included in multiple categories and were not mutually exclusive.

[b]Google Assistant does not provide vendor ratings or user reviews. The vendor rating guidelines from Amazon include: 1) guidance suggested: may have nudity or suggestive content or require supervision due to account linking, location detection, etc, and 2) mature: content is for adults only.

Table 3. Number of apps by health category, target audience, and age group (N=309). Apps could be included in multiple categories and were not mutually exclusive.

Categories	Google	Amazon	Total
Health category			
Health education	3	84	87
Fitness and training	3	69	72
Nutrition	0	33	33
Brain training and games	1	30	31
Health monitoring	0	25	25
Motivational	0	22	22
Meditation	0	16	16
Other	0	15	15
Health location	0	15	15
Stress management	0	12	12
Global positioning system or geographic information system	0	9	9
Diet and caloric tracking	0	9	9
Sleep	0	8	8
Mental health	0	7	7
Air quality monitoring	1	6	7
Baby monitoring and tracking	0	5	5
Smoking cessation	1	5	6
Baby naming	2	3	5
Care giving	0	4	4
Time and task management	1	2	3
Pregnancy tracking	0	3	3
Dog Monitoring and tracking	0	2	2
Beauty tips	0	2	2
Immunization tracking	0	1	1
Target audience			
General or not specified	5	261	266
Patients	1	32	33
Parents	2	14	16
Family	1	13	14
Medical professionals	0	6	6
Pregnant women	0	4	4
Pet owners	1	2	3
Women	0	2	2
Other	0	1	1
Men	0	0	0
Intended age			
Not specified	8	249	257
Adults	0	42	42
Children	1	9	10
Older adults	0	1	1

Collectively, 10.7% (33/309) of apps were rated by the vendors as "guidance suggested" and 1.0% (3/309) were rated as "mature." Furthermore, 30.7% (95/309) of apps had a privacy policy, and 22.3% (69/309) had a developer's policy for terms of use.

The most frequently occurring types of health and fitness VAA apps were health education (87), fitness and training (72), nutrition (33), brain training and games (31), health monitoring (25), motivational (22), and meditation (16), as seen in Table 3. The most common target audience (population that would use the app) was general or not specified (266), followed by patients (33), parents (16), and families (14). In terms of the intended age group, 257 apps did not have age specified, and there were 42 apps focused on adults, 10 focused on children, and 1 focused on older adults (Table 3). The majority of Amazon's apps coded for adults were rated by the vendor as mature or guidance suggested (36/42, 85.7%).

Discussion

Principal Findings

In this study, which is the first to report an analysis of VAA apps for health and fitness, we found that although the marketplace appeared to have many apps in these topical areas, there were mainly apps that focused on either health education or fitness and training and many did not seem to actually have a clear health or fitness focus at all (eg, baby naming). These apps were mostly targeted toward adults and general audiences with only a few apps specifically geared toward older adults or those with disabilities. However, these are the populations that may potentially benefit the most from VAA technologies [8,10,15]. Strikingly, very few apps were categorized as care giving or targeted specifically toward patients, caregivers, or medical professionals, yet these populations could be supported by VAAs; for example, the use of non-commercial VAAs in assisted living facilities has been associated with higher quality of living and improved recovery from illness [6,10]. Additionally, VAA apps that focus on social interaction and support, communication, care coordination, reminders, remote monitoring, locating providers, and scheduling appointments and transportation could be potentially impactful for both patients and their caregivers, but there are currently limited VAA apps available for these purposes. The main potential barriers to advancing the use of VAA apps for health are the restrictions and limitations for publishing VAA apps in the marketplace, security and privacy issues, and the credibility of these apps.

Health Monitoring

Our analysis revealed there was only a limited number of health monitoring apps (25). This may be, in part, due to the restrictions within Amazon and Google policies for publishing health-related apps. Amazon does not allow apps to be certified for release in the app store if the app "collects information relating to any person's physical or mental health or condition, the provision of health care to a person, or payment for the same;" "does not include a disclaimer in the skill description stating that the skill is not a substitute for professional advice;" and "claims to provide life-saving assistance through the skill or in the skill name, invocation name, or skill description" [30]. Thus, associated mobile phone apps could be used to collect health data to be reviewed within the VAA app, but the VAA app does not ask the user to directly document health data, such as blood pressure readings, using the voice interface. Google advises that "health care providers, health plans, or health care clearinghouses wishing to develop an action should be aware that Google is not able to commit that the actions on Google platform meet the requirements of the Health Insurance Portability and Accountability Act (HIPAA) or other relevant legal provisions [31]." There is no specific information regarding HIPAA on the Amazon website. More clarity and guidance in interpreting policy restrictions would help developers better understand what is possible and allowed for release to the market; for example, it is unclear whether biometric data collected outside of the VAA via a wearable device or mobile phone app but not directly captured by the VAA is allowable under Amazon's restrictions. Ultimately, HIPAA will need to be addressed if health data are collected via VAAs and integrated into electronic health records similar to other patient-generated health data from devices. These restrictions may be barriers to actualizing the potential benefits of leveraging VAAs for health care because VAAs could be used to collect health data for remote monitoring and to deliver care.

Security and Privacy

Only a small proportion of vendors provided privacy (95/309, 30.7%) or terms of use (69/309, 22.3%) policies (Table 2). Although voice recognition technologies could potentially restrict access to a specific person, an unauthorized user could still gain access to account information. Google Assistant was the first to distinguish between various users' voices and provides different levels of security access to the Google Home device through multiple user accounts. Amazon recently followed with a similar feature.

There are a number of potential security and privacy issues when using VAAs. However, customers may not recognize or realize the security or privacy implications of using VAA devices. Both the Amazon and Google voice-activated devices remain in a passive listening state for a specific keyword or "wake" word to activate the device to begin recording and transmitting audio; for example, Burger King revealed a vulnerability of VAA devices when their television advertisement stated the wake word for the Google Assistant ("OK, Google") and asked a question about the Whopper [32]. Additionally, because voice is a unique identifier, users should be concerned about how VAA companies collect, store, analyze, or share this information. These privacy issues were highlighted when prosecutors issued a warrant to obtain audio recordings from a murder suspect's Amazon Echo device [33]. Although Amazon did not provide the recordings in this case and cited first amendment protection over the information gathered and sent by the device, this highlighted the fact that Amazon VAA devices record and store data. Amazon and Google Home devices have a button that can be pressed to "mute" the device from listening, but instruction manuals do not have explicit language describing that the devices are always listening and that they are recording and storing audio. Both Amazon and Google VAA devices permit the user to delete search histories

and audio recordings but discourage the user from doing so as it will limit personalization. The feature to delete recordings may also not be apparent to users, and it is also unclear how long recordings are stored for. Moreover, these devices can access other accounts, which could contain private information; for example, Google Home can be enabled to access your calendar, email account, and shopping accounts, which raises additional security and privacy issues. Therefore, it is vital to have vendor transparency about when a user's voice data are stored and transmitted and to whom and what audio and other data are recorded and stored either on the device or in the cloud and for how long.

Developer Platforms and Application Programming Interfaces

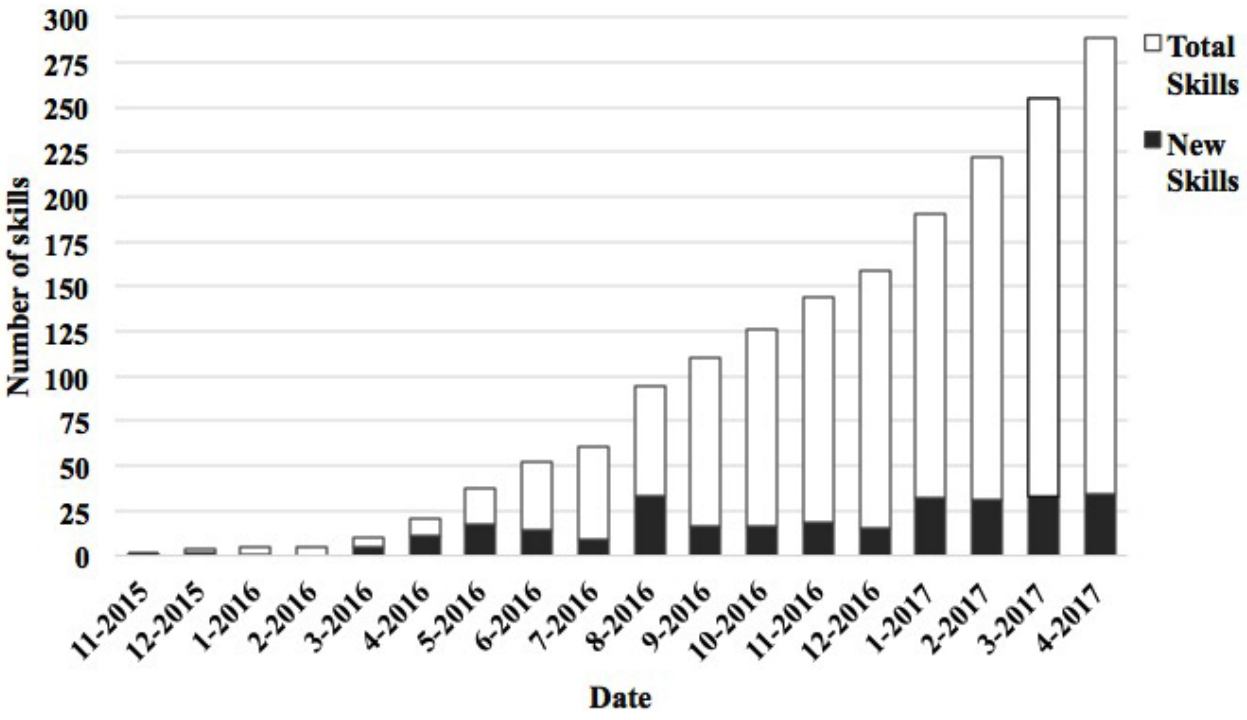
Figure 2 shows that there has been a steady increase in health and fitness apps released by Amazon since 2015 [9,29]. The addition of the Alexa Skills Kit (June 2015), a self-service API that contains a collection of tools and sample code, likely contributed to this increase because it took some time for a developer community to become established. Google Assistant released its API in December 2016. Although our search returned only 9 Google Assistant apps, there may be an uptick in release of apps over time as their developer platform and community matures. Only 18.1% (56/309) of apps examined in this study could be paired with a website, app, or device, but this is likely to expand as smart health and home devices continue to emerge (Table 2). Providing an open ecosystem with developer platforms and APIs accelerates the adoption and use of devices and the development of apps, thereby expanding the customer and market base for VAAs. Thus, future VAA devices should consider providing these self-service tools.

Credibility

Both Amazon and Google provide only limited information about each VAA app or its features and functionalities. Similar to mobile phone apps, it is difficult to determine the credibility and value of the content of apps based solely on vendor descriptions; for example, it would be difficult for a customer to determine whether health care systems, hospitals, or providers endorsed or developed the apps unless it was explicitly noted in the app name, such as the Boston Children's Hospital's KidsMD app. It is also unclear whether apps were developed with user-centered design principles or evidence-based guidelines or materials, particularly for health education materials. For health-focused apps that "…meet the regulatory definition of a device but pose minimal risk to patients or consumers, the Food and Drug Administration (FDA) exercises enforcement discretions…" and does not expect manufacturers to register and list their apps with the FDA [34]. Additionally, FDA does not deem entities that distribute mobile phone apps to be medical device manufacturers. Although there are no explicit comments from FDA regarding VAAs, it is likely that the regulations related to mobile phone apps would be applied similarly to VAA apps.

Traditionally, the development of health mobile phone apps has lacked stakeholder involvement [17,18,26], which has contributed to high rates of app abandonment due to lack of usability and poor user experience. Design and usability principles should also guide VAA app development and be focused on the distinct challenges and benefits of interactive voice-based user interfaces.

Figure 2. Total number of Alexa skills released over time (Alexa Skills Kit application programming interface released June 2015).



To integrate VAAs into health care, the design of voice-enabled user interfaces must consider the various needs of the end user, the setting (eg, background noise and multiple users of a single device in a home, such as the patient, their caregiver, home health nurse, etc), and the unique privacy and security issues that come with using these devices that are always listening in the background. With advancements in speech recognition and algorithms that enhance accuracy, VAA apps can become more sophisticated to recognize and respond to a diverse realm of users [3].

The translation of voice to text through speech recognition provides an opportunity to track and understand patient interactions and behaviors, though the integration and interoperability of the data, particularly with the electronic health record and patient portal, have not yet come to fruition but have much promise. Within health care, VAAs must take into account the context of the person to offer timely, appropriate, and valuable feedback, including the ability to understand speech, provide meaningful feedback, and generate accurate results. VAAs are also limited in their ability to complete complex tasks because users must rely on working memory instead of being able to visually browse through a Web, tablet, or mobile phone interface for cues and assistance [35]. Thus, a combination of user interactions with both VAA and mobile phone or tablet could help end users navigate more complex tasks.

As we learn more about how people use and interact with VAAs and the shortcomings of VAAs in terms of unmet needs or expectations, app designers, developers, and researchers can start to customize user experiences that align more closely to user needs and enhance usability. An exploratory ethnographic analysis of user reviews from Amazon Echo and Dot revealed a number of concepts around user experience, such as health care-related workarounds, quality of life improvement and physical disability, companionship, and benefits to health care [12]. A deeper understanding of the way patients potentially utilize VAAs for health and fitness could also help provide potential use cases for future app development and refinement of app functionality. The aforementioned challenges are important opportunities for future research.

Limitations

There are a few limitations to our study. First, the authors only examined apps published in the “health and fitness” category, though there may potentially be other health-related apps that exist in the “smart home,” “food and drink,” or other categories in the Amazon Skills Store and Google Assistant websites or apps. Apps are released, modified, updated, and discontinued on a regular basis. As a result, there may be apps that were reviewed in this study that are no longer available, and there may have been modifications in app descriptions since data extraction. Authors also relied solely on information published by vendors on the Amazon Skills Store and Google Assistant websites and apps. Thus, it is possible that the features listed may not be present in the actual app, which is not a unique issue to our study but app stores in general. Moreover, user reviews and star ratings may change over time as additional users make submissions. As with other customer reviews and ratings, these VAA reviews and ratings may not be representative of the quality of the apps because they may not reflect the actual content or usability of these VAA apps. Despite these limitations, our study is the first to review health and fitness apps for VAAs and contributes toward an understanding of the characteristics of health and fitness apps available for commercially available, hands-free VAA.

Implications

There are a number of key implications of this research. Because the VAA app marketplace is still evolving, a cursory look through the names of available apps may suggest that there is, in fact, a growing number of VAA apps focused on health and fitness. However, this impression is misleading, and our study findings provide a clearer picture of the number and scope of VAA apps available, which are predominantly focused on fitness, training, and health education. Understanding what is available in the marketplace also helps to illuminate where there may be a health or fitness use case or need but no apps currently available.

Additionally, we found that most apps are focused on general audiences than on specific health use cases. In general, the VAA app market does not contain as much health-focused content in comparison to the mobile health app market, where chronic disease, monitoring, and self-management apps have proliferated. In particular, VAAs offer advantages when compared with the physical user interfaces of computers and mobile devices (tablets and mobile phones) because voice is used for interactions with the app. This could improve accessibility for those with limited sight, physical limitations, limited literacy, and limited computer proficiency. VAA apps could also be used as a vehicle to deliver clinical and behavioral interventions, as a data collection tool for research, and to deliver health care, but these are also currently lacking in the marketplace.

To evolve the current market, privacy, security, and HIPAA compliance need to be addressed along with lessening the stringent requirements from publishers such that health care and direct health monitoring could potentially be enabled or delivered via hands-free VAAs. It is also likely that support for integration with apps that run on other platforms (phones, tablets, Web, medical devices, smart home devices, etc.) will be important to overcome some of the limitations of VAA technologies highlighted in this paper, to enhance the user experience, and to leverage the opportunities that stem from voice-based user interfaces. Additionally, because there is usually a proliferation of apps when there are APIs and developer platforms available, their availability will also be critical to encourage innovation and a strong user base and to enable researchers to develop VAA apps for interventions and for facilitating data collection.

Conclusions

The emerging market of health and fitness apps for hands-free VAAs is still nascent and mainly focused in the areas of health education and fitness. As with other health technologies, the usability and credibility of health apps are critical to ensuring adoption and long-term use. Further work is necessary to

evaluate the usefulness, usability, user experience, quality, privacy, and security implications of VAA apps. Future research should also consider the development of an evaluation method for VAA apps given the unique nature of the voice interface, such that the content of the apps can be assessed for quality and usability. It will also be imperative to understand workflow barriers and facilitators required to optimally integrate VAAs into clinical care contexts and within patients' homes and lives and to determine the acceptability and feasibility of deploying VAAs for health care use cases.

Acknowledgments

ACG is supported by the National Institutes of Health National Library of Medicine grant (1T15LM01250001). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Authors' Contributions

AEC conceived the study concept, study design, review method, search strategy, and development of the review and led the study. ACG contributed to the review method and search strategy. ACG and DS extracted information about apps and performed qualitative coding of the apps. AEC performed adjudication. AEC, ACG, and DS discussed until consensus was reached. AEC, ACG, and DG analyzed and interpreted the data. ACG and AEC developed an initial draft of the manuscript. All authors revised the manuscript critically for important intellectual content and approved the manuscript prior to its submission.

Conflicts of Interest

None declared.

References

1. Jiang J, Awadallah A, Jones R, Ozertem U, Zitouni I, Kulkarni R, et al. Automatic online evaluation of intelligent assistants. 2015 May Presented at: Proceedings of the 24th International Conference on World Wide Web; 2015; Florence, Italy p. 18-22. [doi: 10.1145/2736277.2741669]
2. Chaudhri V, Cheyer A, Guili R, Jarrold B, Myers K, Niekrasz J. A case study in engineering knowledge base for an intelligent personal assistant. In: CEUR Workshop Proceedings. 2006 Nov 06 Presented at: Proceedings of the 5th International Conference on Semantic Desktop and Social Semantic Collaboration; 2006 Nov 6; Athens, Georgia URL: http://ceur-ws.org/Vol-202/SEMDESK2006_0033.pdf
3. Neustein A, Markowitz J. Mobile speech and advanced natural language solutions. New York, NY: Springer; 2013:320-332.
4. Eriksson H, Timpka T. The potential of smart homes for injury prevention among the elderly. Inj Control Saf Promot 2002 Jun;9(2):127-131. [doi: 10.1076/icsp.9.2.127.8694] [Medline: 12461840]
5. Bafhtiar G, Bodinier V, Despotou G, Elliott M, Bryant N, Arvantis T. Providing patient home clinical decision support using off-the-shelf cloud-based smart voice recognition. In: West Midlands Health Informatics Network. 2017 Presented at: WIN Conference CDS; 2017 Jan 4; Coventry, England URL: https://wmhin.org/wp-content/uploads/2017/01/bafhtiar_g_win2017.pdf
6. Wolters MK, Kelly F, Kilgour J. Designing a spoken dialogue interface to an intelligent cognitive assistant for people with dementia. Health Informatics J 2016 Dec;22(4):854-866. [doi: 10.1177/1460458215593329] [Medline: 26276794]
7. Robinson JL, Braimah AV, Chun R, Pusateri G, Jay WM. Usage of Accessibility Options for the iPhone and iPad in a Visually Impaired Population. Semin Ophthalmol 2017 Jul;32(2):163-171. [doi: 10.3109/08820538.2015.1045151] [Medline: 26154395]
8. Gentry T. Smart homes for people with neurological disability: state of the art. NeuroRehabilitation 2009;25(3):209-217. [doi: 10.3233/NRE-2009-0517] [Medline: 19893189]
9. Amazon. 2017. Alexa Skills Guide URL: https://www.amazon.com/Alexa-Skills-Guide/b/ref=sv_a2s_4?ie=UTF8&node=15144553011 [WebCite Cache ID 6vXQzUsVd]
10. Pollack M. Intelligent technology for an aging population: the use of AI to assist elders with cognitive impairment. AI Mag 2005;26(2):9. [doi: 10.1609/aimag.v26i2.1810]
11. Beaudin JS, Intille SS, Morris ME. To Track or Not to Track: User Reactions to Concepts in Longitudinal Health Monitoring. J Med Internet Res 2006;8(4):e29 [FREE Full text] [doi: 10.2196/jmir.8.4.e29]
12. Coyne M, Thomas C, Collimore A, Franzese C, Hwang C. Early User Centered Insights on Voice Integrated Technologies Through Retrospective Analysis. JMIR 2017 Sep 22;3(1):e49 [FREE Full text] [doi: 10.2196/iproc.8576]
13. Pradhan, Alisha, Mehta, Kanika, Findlater L. Accessibility came by accident: use of voice-controlled assistants by people with disabilities. 2018 Presented at: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; April 21 - 26, 2018; Montreal QC, Canada. [doi: 10.1145/3173574.3174033]
14. Bigham JP, Kushalnagar R, Huang TK, Flores JP, Savage S. On How Deaf People Might Use Speech to Control Devices. 2017 Presented at: Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility; Oct 20-Nov 1, 2017; Baltimore, Maryland. [doi: 10.1145/3132525.3134821]

15. Carroll C, Chiodo C, Lin A, Nidever M, Prathipati J. Robin: enabling independence for individuals with cognitive disabilities using voice assistive technology. 2017 Presented at: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems; May 6-11,2017; Denver, Colorado. [doi: 10.1145/3027063.3049266]
16. Hassoon A, Schrack J, Naiman D, Lansey D, Baig Y, Stearns V, et al. Increasing Physical Activity Amongst Overweight and Obese Cancer Survivors Using an Alexa-Based Intelligent Agent for Patient Coaching: Protocol for the Physical Activity by Technology Help (PATH) Trial. JMIR Res Protoc 2018 Feb 12;7(2):e27 [FREE Full text] [doi: 10.2196/resprot.9096] [Medline: 29434016]
17. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a Pain App for That": Review of Patient-targeted Smartphone Applications for Pain Management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
18. Rivera J, McPherson A, Hamilton J, Birken C, Coons M, Iyer S, et al. Mobile Apps for Weight Management: A Scoping Review. JMIR Mhealth Uhealth 2016 Jul 26;4(3):e87 [FREE Full text] [doi: 10.2196/mhealth.5115] [Medline: 27460502]
19. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
20. Azar KMJ, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
21. Knight E, Stuckey MI, Prapavessis H, Petrella RJ. Public health guidelines for physical activity: is there an app for that? A review of android and apple app stores. JMIR Mhealth Uhealth 2015;3(2):e43 [FREE Full text] [doi: 10.2196/mhealth.4003] [Medline: 25998158]
22. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014 May;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
23. Radovic A, Vona PL, Santostefano AM, Ciaravino S, Miller E, Stein BD. Smartphone Applications for Mental Health. Cyberpsychol Behav Soc Netw 2016 Jul;19(7):465-470 [FREE Full text] [doi: 10.1089/cyber.2015.0619] [Medline: 27428034]
24. Higgins JP. Smartphone Applications for Patients' Health and Fitness. Am J Med 2015 Jun 17. [doi: 10.1016/j.amjmed.2015.05.038] [Medline: 26091764]
25. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014 Apr;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
26. Wallace LS, Dhingra LK. A systematic review of smartphone applications for chronic pain available for download in the United States. J Opioid Manag 2014 Jan;10(1):63-68. [doi: 10.5055/jom.2014.0193] [Medline: 24604571]
27. Moussalli S, Cardoso W. Are commercial "personal robots" ready for language learning? Focus on second language speech. CALL communities and culture - short papers from EUROCALL 2016. [doi: 10.14705/rpnet.2016.EUROCALL2016.9781908416445]
28. 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;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
29. Amazon. 2017. Alexa Skills Store URL: https://www.alexaskillstore.com [accessed 2017-12-12] [WebCite Cache ID 6ve3MLuPl]
30. Amazon. 2017. Alexa Skills Kit Policy Testing URL: https://developer.amazon.com/docs/custom-skills/policy-testing-for-an-alexa-skill.html [accessed 2017-12-12] [WebCite Cache ID 6ve3Z2CAe]
31. Google. 2017. Policies for Actions on Google URL: https://developers.google.com/actions/policies/general-policies [accessed 2017-12-12] [WebCite Cache ID 6ve3hEK7u]
32. Anderson M. Phys.org. 2017. How Burger King revealed the hackability of voice assistants URL: https://phys.org/news/2017-05-burger-king-revealed-hackability-voice.html [WebCite Cache ID 6ve3s77jq]
33. CNN. 2016. Amazon pushes back on prosecutor request for Alexa smart speaker info URL: http://www.cnn.com/2016/12/28/tech/amazon-echo-alexa-bentonville-arkansas-murder-case-trnd/ [WebCite Cache ID 6ve424F1C]
34. FDA. Center for Devices and Radiological Health; 2015. Mobile Medical Applications URL: https://www.fda.gov/MedicalDevices/DigitalHealth/MobileMedicalApplications/ucm255978.htm [WebCite Cache ID 6ve3mQ9Vo]
35. Cohen M, Giangola J, Balogh J. Voice user interface design. Boston, MA: Addison-Wesley Professional; 2004.

Abbreviations

API: application programming interface
FDA: Food and Drug Administration
HIPAA: Health Insurance Portability and Accountability Act
VAA: voice-activated assistant

Edited by C Dias; submitted 27.12.17; peer-reviewed by R Randell, N Keshav, JH Lee; comments to author 20.01.18; revised version received 07.05.18; accepted 08.06.18; published 24.09.18.

Please cite as:
Chung AE, Griffin AC, Selezneva D, Gotz D
Health and Fitness Apps for Hands-Free Voice-Activated Assistants: Content Analysis
JMIR Mhealth Uhealth 2018;6(9):e174
URL: http://mhealth.jmir.org/2018/9/e174/
doi:10.2196/mhealth.9705
PMID:30249581

Original Paper

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

Yan Guo[1,2,3*], PhD; Zhimeng Xu[1*], MS; Jiaying Qiao[1*], MS; Y Alicia Hong[4*], PhD; Hanxi Zhang[1], MS; Chengbo Zeng[1], MS; Weiping Cai[5], MD; Linghua Li[5], MD; Cong Liu[5], MSN

[1]Department of Biostatistics and Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]Sun Yat-sen Center for Migrant Health Policy, Sun Yat-sen University, Guangzhou, China

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

[4]Department of Health Promotion and Community Health Sciences, School of Public Health, Texas A&M University, College Station, TX, United States

[5]Department of Infectious Disease, Eighth People's Hospital, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Y Alicia Hong, PhD
Department of Health Promotion and Community Health Sciences
School of Public Health
Texas A&M University
212 Adriance Lab Road
MS 1266
College Station, TX, 77843
United States
Phone: 1 979 436 9343
Email: yhong@sph.tamhsc.edu

Abstract

Background: Most people living with HIV (PLWH) reside in middle- and low-income countries with limited access to health services. Thus, cost-effective interventions that can reach a large number of PLWH are urgently needed.

Objective: The objective of our study was to assess the feasibility and acceptability of an mHealth intervention among PLWH in China.

Methods: Based on previous formative research, we designed an mHealth intervention program that included sending weekly reminders to participants via text messages (short message service, SMS) and articles on HIV self-management three times a week via a popular social media app WeChat. A total of 62 PLWH recruited from an HIV outpatient clinic were randomly assigned to intervention or control group. The intervention lasted for 3 months, and all participants were assessed for their medication adherence, presence of depression, quality of life (QoL), and CD4 (cluster of differentiation 4) counts. Upon completing the intervention, we interviewed 31 participants to further assess the feasibility and acceptability of the study.

Results: At baseline, the intervention and control groups did not differ in terms of demographic characteristics or any of the major outcome measures. About 85% (53/62) of the participants completed the intervention, and they provided valuable feedback on the design and content of the intervention. Participants preferred WeChat as the platform for receiving information and interactive communication for ease of access. Furthermore, they made specific recommendations about building trust, interactive features, and personalized feedback. In the follow-up assessment, the intervention and control groups did not differ in terms of major outcome measures.

Conclusions: This pilot study represents one of the first efforts to develop a text messaging (SMS)- and WeChat-based intervention that focused on improving the medication adherence and QoL of PLWH in China. Our data indicates that an mHealth intervention is feasible and acceptable to this population. The data collected through this pilot study will inform the future designs and

implementations of mHealth interventions in this vulnerable population. We recommend more innovative mHealth interventions with rigorous designs for the PLWH in middle- and low-income countries.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1800017987; http://www.chictr.org.cn/showprojen.aspx?proj=30448 (Archived by WebCite at http://www.webcitation.org/71zC7Pdzs)

Registered Report Ientifier: RR1-10.2196/

(JMIR Mhealth Uhealth 2018;6(9):e10274) doi:10.2196/10274

KEYWORDS

mHealth; social media; medication adherence; people living with HIV; randomized controlled trial

Introduction

Most people living with HIV (PLWH) reside in middle- and low-income countries [1]. Delivering effective interventions to this vulnerable and stigmatized population remains a critical public health challenge. In recent years, mobile-based interventions, or mHealth interventions, have emerged as a promising solution to deliver health services to PLWH. For example, the WelTel program in Kenya sent short message service (SMS) text messages to PLWH, and this resulted in improved antiretroviral therapy (ART) adherence [2]. A recent Cochrane review has identified 17 SMS text message interventions to promote medication adherence, most of which have shown initial potential [3]. However, of the existing mHealth interventions that promote ART adherence among PLWH in middle- or low-income countries, only few have employed a rigorous design (eg, randomized controlled trial, RCT) or assessed the quality of life (QoL) measures and clinical outcomes such as CD4 (cluster of differentiation 4) counts of these people.

More than 64% of the world's population owns a mobile phone, and in several high- and middle-income countries, a majority of the population owns a smartphone [4]. Because of this, there has been a growing interest in delivering mHealth interventions via social media such as Facebook and mobile apps [5-8]. In a recent review of social media interventions that deliver HIV services, out of the 26 studies identified, 18 were conducted in high-income countries, 8 in middle-income countries, and none in low-income countries. Furthermore, of the 26 studies, only 1 was designed to improve ART adherence and none to promote the QoL of PLWH [8].

As the most populous country in the world, China has 1.3 billion mobile phone users, with a 95% penetration rate [9]. More than 740,000 PLWH live in China; they face a high level of stigma, and the prevalence of depression in this group is high [10]. Although Facebook and Twitter are not accessible in China, Chinese people are active on other social media platforms. With more than 570 million users, WeChat is the most popular social media platform in China; 93% of the residents in the major cities of China log into WeChat every day [11]. Recently, WeChat-based behavioral interventions have shown feasibility and acceptability [12,13]. The high ownership rates of mobile phones and wide popularity of WeChat suggest a promising platform to deliver low-cost interventions to the stigmatized population of PLWH. Although HIV-related mHealth interventions have amassed growing interests and have shown initial potential globally [5,6,8], the use of such programs has been limited in China despite the high rates of mobile phone use and access to social media. Recent reports on mHealth interventions for HIV-affected populations in China were either study protocols [14,15] or the delivery of SMS text messages only [16]. There are few mHealth interventions for PLWH in China that have been tested with rigorous design. Accordingly, we developed one of the first mHealth (WeChat+SMS text message) interventions to improve the ART adherence and QoL of PLWH in China and pilot-tested its feasibility and acceptability via an RCT. We hypothesized that the mHealth intervention would be feasible for and acceptable to PLWH in China.

Methods

Study Setting

This study was conducted in a hospital that has been offering services to PLWH in a large metropolitan area in South China from October 2016 to March 2017. The hospital serves more than 14,000 patients with HIV in the region.

Intervention Program

The development of the intervention program was guided by the information-motivation-behavioral skills model [17], the literature of mHealth interventions, and our formative research. The initial intervention protocol was developed based on prior mHealth interventions to improve medication adherence in PLWH [2,18,19] and our earlier studies in this population [20,21].

The final intervention program consisted of two major components. The first component was weekly SMS text message greetings and reminders regarding medication adherence and regular exercise. The second component consisted of short articles on side effect management, medication self-management, stress management, and healthy lifestyle, which were sent via WeChat three times a week. Detailed information on the contents of the articles is shown in Textbox 1. To track patient engagement, the participants would receive 3 multiple choice questions on the information about the articles every other week on WeChat; for example, "what does good medication adherence mean?" with 4 options (>80%, >85%, >90%, and >95% adherence to the prescribed medication).

Textbox 1. Times and titles of articles sent to the participants in the intervention group.

Week 1

- Common health problems after infection I
- Common health problems after infection II
- Treatment adherence—key to living a healthy life

Week 2

- How to exercise scientifically I
- How to exercise scientifically II
- Take your medicine on time—key to medication adherence

Week 3

- How to have healthy babies for HIV-seropositive men?
- Knowledge about DTG, a new drug
- Something important for men who have sex with men

Week 4

- When to begin HIV treatment, sooner or later?
- To those who are depressed
- Tips on how to quit smoking

Week 5

- Health issues on *Pneumocystis carinii* pneumonia and lactic acidosis
- My life, my choice
- Health issues on bones

Week 6

- Things you need to know about taking medicine I
- Things you need to know about taking medicine II
- Common side effects of medication

Week 7

- I am HIV positive—can I drink alcohol?
- What is drug resistance?
- Consequences of poor medication adherence

Week 8

- Tips for HIV-positive patients on physical checkup I
- Tips for HIV-positive patients on physical checkup II
- His healthy life: a story of an HIV-positive man

Week 9

- Tips about taking medicines
- What to do when you are upset
- Love yourself, love your family

Week 10

- Tips for psychological adjustment
- A brief introduction to opportunistic infection I

- Why you need to take medicine on time every day?

Week 11

- Disclosure of HIV status I
- Disclosure of HIV status II
- Disclosure of HIV status III

Week 12

- Improve your mood, live healthier
- Tips for pregnant women and lactating women
- A brief introduction to opportunistic infection II

Comparatively, the control group received articles on nutrition sent via WeChat three times a week. Each article typically had 1200 Chinese characters and took 3-5 minutes to read through for both intervention and control groups. All the articles were adopted from the authoritative websites of the World Health Organization (WHO) and China's Centers for Disease Control and Prevention and adapted for PLWH in China.

Participant Recruitment

The following were the eligibility criteria. The participant should be (1) at least 18 years old, (2) HIV seropositive, (3) on HIV treatment for at least 1 month, and (4) able to read and write. Patients with severe mental illnesses that prohibited them from participating in such intervention were excluded. We recruited participants from the outpatient clinic of the hospital described above. Our research staff approached patients in the waiting areas and invited them to participate in our research project. Those who were interested were taken to a private space for further explanation of the project. Those who met the inclusion criteria and were willing to participate signed an informed consent form before completing a baseline survey. All eligible participants were provided with free breakfast (milk and bread) upon completion of the baseline survey. The study protocol was approved by the Human Subjects Review Board of the School of Public Health, Sun Yat-sen University.

Randomized Controlled Trial Design

The intervention was delivered as a single-blinded RCT. A total of 62 eligible participants completed the baseline survey and were randomly assigned to the intervention or control group. To ensure that the two groups were balanced in terms of confounding factors, block randomization using SAS statistical software version 9.4 (SAS Institute, Inc., Cary, NC, United States) was conducted [22]. During the 12-week program, participants in the intervention group received a total of 12 SMS text message reminders and 36 WeChat articles. Meanwhile, participants in the control group received a total of 36 WeChat articles; however, they did not receive a SMS text message reminder. Following the HIV/AIDS treatment standard in China, all patients visited their primary health care providers in the designated hospital every 3 months for medication refilling and CD4 testing; thus, we conducted a follow-up survey when the participants returned to the hospital for their medical visit 3 months after the baseline.

Outcome Measures

All the participants completed the baseline and follow-up surveys using tablets while waiting for their appointments in the outpatient clinic. The survey covered the following domains: participants' demographic characteristics, medication adherence, mental health, and QoL. Along with patients' consent, we also obtained data on their CD4 counts from their medical records as a biomarker. Medication adherence was the primary outcome; CD4 count, depression, and QoL were secondary outcomes.

The demographic characteristics of the participants included age, gender, educational level, marital status, sexual orientation, income, and residence (rural or urban). Medication adherence was assessed using the question "In the last 30 days, have you ever missed taking any dose of your HIV medication?" We categorized adherence as a binary variable named "ever missed medication in previous 30 days." Participants' depression was measured using the Center for Epidemiological Studies Depression Scale, Chinese version [23,24]. The Cronbach alpha of the scale was 0.9. The total score ranged from 0 to 60 (2-36 in this study); patients with a score of ≥16 were considered to have depressive symptoms. The QoL was measured using the 31-item WHO Quality of Life HIV short form [25]. The total score ranged from 24 to 120 (55-120 in this study); the Cronbach alpha was 0.88.

Evaluation of the Feasibility and Acceptability of the Study

We conducted semistructured interviews to collect data on the feasibility and acceptability of the intervention and interviewed a total of 31 participants from both the intervention and control groups upon completion of the intervention. The participants were chosen to represent different demographic groups. They were asked about their experience with the intervention, including the design and implementation, and their recommendations on how to improve the program. All interviews were audiotaped and transcribed for content analysis.

Data Analysis

For the outcome measures of the pilot study, we used the SAS statistical software to perform the analysis. First, we used descriptive statistics to analyze the participants' characteristics and primary outcomes. Second, we used *t* test (for normally distributed continuous variables), Mann-Whitney U test (for nonnormally distributed continuous variables), and chi-square

test (for categorical variables) to compare the demographic characteristics and primary outcomes between the intervention and control groups at baseline and follow-up. Third, we used *t* test to compare the pre-post changes between the intervention and control groups. A two-sided *P* value < 0.05 was considered statistically significant. In total, 85% (53/62) participants finished both the baseline and 3-month follow-ups. For postintervention and pre-post analyses, only 53 participants were included.

For the qualitative data collected from postintervention interviews, we used Nvivo version 10.0 (QSR International Pty Ltd. Doncaster, Victoria, Australia). All audiotaped interviews were transcribed verbatim. Data analysis was started with reading and rereading the transcripts, followed by open-coding the transcripts. Detailed summaries with substantial retention of original quotes were prepared to facilitate further discussion and elaboration among team members. Coding themes and domains were developed by constant comparisons of codes across transcripts and consensus among team members. Coding themes were further analyzed in the original transcripts for consistency and accuracy. Quote excerpts and summaries were then categorized according to participants' characteristics and coding domains; they were further compared and reviewed for interrelationships and correspondence. A summary report was generated from the qualitative data analysis. This report covered the themes developed in the interview guides as well as new themes identified during the coding process. Each theme was explained using detailed excerpts and summaries of participants' characteristics.

Results

Baseline Characteristics of the Participants

A total of 62 participants completed the baseline survey; of them, 90% (56/62) were male. The mean age of the participants was 28.3 (SD 6.1) years. Among the participants, 76% (47/62) were gay or bisexual, 81% (50/62) had attained high school education or higher, only 8% (5/62) were married, 47% (29/62) were living in urban residences, and only 36% (22/62) had a monthly income >7000 yuan (the average monthly income in Guangzhou being 7425 yuan) [26]. The mean duration since HIV diagnosis was 2.7 years. As shown in Table 1, at baseline, there were no significant differences between the intervention and control groups in terms of demographic characteristics as well as outcome variables of medication adherence, mental health, QoL, and CD4 counts.

Primary and Secondary Outcomes

Table 2 shows the primary and secondary outcome measures at follow-up and the pre-post changes in these outcomes between the intervention and control groups. There were no significant differences in terms of the primary and secondary outcomes between the intervention and control groups at follow-up. In addition, none of the changes in the primary outcomes were statistically significant between the two groups.

Feasibility and Acceptability

Figure 1 shows the flowchart of the pilot study. Of all the participants, 85% (53/62) completed the intervention and finished the follow-up survey. Upon completion, 50% (31/62) of the participants were selected for postintervention interviews, with 54% (17/62) being from the intervention group and 46% (14/62) from the control group. They shared the following feedback: (1) The participants were, in general, satisfied with the program and appreciated being cared for. (2) They liked the articles sent to their WeChat account more than the SMS text message reminders. The participants preferred information that was more tailored for the different groups of PLWH, was more personalized, and provided more social support. The biweekly questions to check their reading of the articles were additional burdens for them. (3) The participants were more willing to follow an advice only after having built a trusting relationship with the research staff. They preferred having more interactive communications with the research staff. (4) As most participants had maintained good medication adherence, they were more interested in information to improve their QoL, especially the strategies to reduce anxiety and depression. (5) The participants made specific suggestions regarding the design and content of the WeChat-based program, for example, WeChat-based appointment system and testing result notification and multimedia functions to deliver audio- or video-based interactive programs. Table 3 shows the list of some sample responses from the interviews about feasibility and acceptability.

Table 1. Participants' characteristics and primary outcome measures at baseline.

Characteristics	Total (n=62)	Intervention group (n=31)	Control group (n=31)	*P* value
Age (years), mean (SD)	28.3 (6.1)	29.2 (6.5)	27.4 (5.7)	.26
Gender, n (%)				.20
Male	56 (90)	26 (84)	30 (97)	
Female	6 (10)	5 (16)	1 (3)	
Education, n (%)				>.99
<High school	12 (19)	6 (19)	6 (19)	
≥High school	50 (81)	25 (81)	25 (81)	
Sexual orientation, n (%)				>.99
Heterosexual	15 (24)	7 (23)	8 (26)	
Gay or bisexual	47 (76)	24 (77)	23 (74)	
Marital status, n (%)				.35
Married	5 (8)	4 (13)	1 (3)	
Unmarried	57 (92)	27 (87)	30 (97)	
Residence, n (%)				.20
Urban residence	29 (47)	17 (55)	12 (39)	
Rural residence	33 (53)	14 (45)	19 (61)	
Monthly income (yuan), n (%)				.30
＜3000	15 (24)	5 (16)	10 (32)	
3000–7000	25 (40)	14 (45)	11 (35)	
＞7000	22 (36)	12 (39)	10 (32)	
Duration since HIV diagnosis (years), mean (SD)	2.7 (2.4)	3.1 (2.2)	2.3 (2.5)	.24
CD4[a] cell counts (cells/μL), median (interquartile range)	392 (277-517)	380 (283-542)	414 (260-513)	.89
Missed medication within the last 30 days, n (%)	5 (8)	3 (10)	2 (6)	>.99
Depression, mean (SD)	16.9 (9.4)	15.8 (9.4)	18.0 (9.3)	.36
Quality of Life (total scores), mean (SD)	83.4 (12.7)	84.3 (14.2)	82.6 (11.2)	.60

[a]CD4: cluster of differentiation 4.

Table 2. Postintervention endpoint analyses and pre-post analyses of the primary outcomes.

Characteristics	Total (n=53)	Intervention group (n=26)	Control group (n=27)	*P* value
CD4[a] cell counts (cells/μL), median (interquartile range)	399 (270-564)	379 (254-570)	401 (272-524)	.89
CD4 change (cells/μL), mean (SD)	5 (111)	11 (122)	0 (101)	.71
Missed medication within the last 30 days, n (%)	3 (6)	2 (8)	1 (4)	.39
Depression, mean (SD)	16.7 (10.1)	15.5 (9.1)	17.9 (11.1)	.38
Changes in depression change, mean (SD)	−0.28 (8.15)	−0.42 (7.04)	−0.15 (9.22)	.90
QoL[b], mean (SD)	82.2 (13.7)	85.0 (13.2)	79.5 (13.8)	.15
Changes in QoL, mean (SD)	−2.1 (9.9)	−0.7 (10.6)	−3.5 (9.3)	.32

[a]CD4: cluster of differentiation 4.

[b]QoL: Quality of Life.

Figure 1. Flowchart of the pilot study.

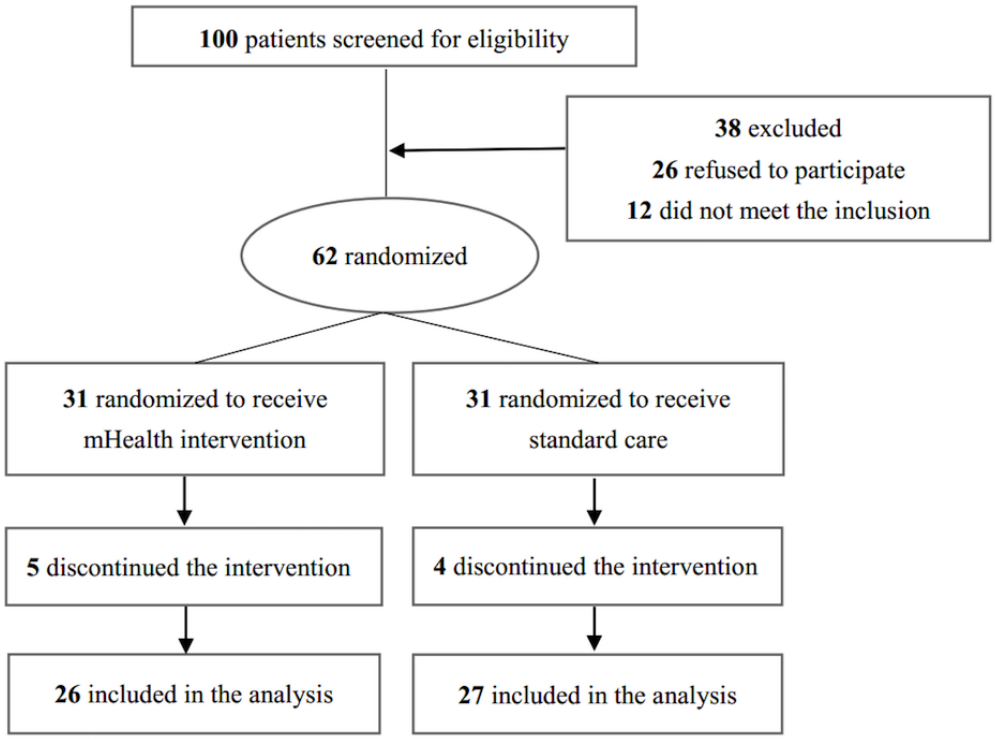


Table 3. Summary of the qualitative analysis on the feasibility and acceptability evaluation.

Domains	Feasibility and acceptability assessment questions	Typical answers
WeChat versus short message service (SMS) text messages	How do you like the articles sent via WeChat and reminders via SMS text messages?	• "I think they are helpful, I learned a lot about the disease that I had not known before." • "All of my friends are using WeChat instead of short message now. So I prefer the articles sent via WeChat."
Content of the articles	What topics are you most interested in?	• "I follow the latest progress of HIV treatment." • "I have been feeling awful since I realized I was infected. I feel my life is ruined. Anything to help me feel better would be helpful." • "In my opinion, the articles are not very targeted. If you can tell us something specific for us gay patients, it will be useful."
Format or style of the articles	What kind of format or style of the articles do you like?	• "There is too much information on the internet. Sometimes it can be very confusing. I hope the articles can be professional and authoritative." • "Well, reading words is kind of boring, it'd be better if I can listen to it, and video would be perfect."
Intervention adherence	To what extent did you read the articles and messages?	• "I use WeChat every day; reading articles on WeChat is convenient for me, so I read most of them. But I often missed messages via SMS as I do not use SMS much. Questions you asked are difficult for me. If I do not know the answer immediately, I do not answer them." • "I have subscribed many Subscription Accounts; they send me articles every day, so I do sometimes miss some of your articles."
Satisfaction	Are you satisfied with our intervention?	• "Of course, I am so pleased that someone like you cares about people like me."
Suggestions	Do you have any suggestions for our mHealth intervention?	• "If I can make an appointment and get my testing results via the system, it will save me a lot of time." • "I didn't have any motivation to read your articles. Maybe you can have more appealing titles for your articles." • "Repeated messages became burdens. I do not trust machine. What I really need are communication and interaction with real people."

Discussion

This pilot study represents one of the first efforts to develop an mHealth intervention for improving medication adherence in and QoL of PLWH in China. The outcome analyses indicated nonsignificant results in this study. Studies on prior SMS text message-based interventions to improve medication adherence in PLWH in middle- and low-income countries have mostly reported nonsignificant results, with few studies reporting significant viral suppression [8,27]. The following reasons might explain the nonsignificant results observed in this pilot study. First, a ceiling effect existed in the primary outcome of ART adherence; 92% of the participants had good adherence at baseline. Such high adherence rate might be due to the fact that 64.5% (40/62) of the participants were men who have sex with men (MSM) aged under 30 years, with a mean age of 28.3 years, and 38.7% (24/62) participants had attained college education or higher. The characteristics of our participants, being young and highly educated, were similar to those of the participants in earlier studies on MSM in China, a group with high rates of medication adherence [28]. Second, in the RCT, the control group received HIV-related nutrition information at the same frequency as the intervention group. Their feedback after the project has suggested that they were very interested in such information, and they became more cautious of their HIV self-management including medication adherence. Third, we observed a limited interaction between health care providers and patients as well as among patients during the intervention. Thus, the intervention mostly took effect at an individual level, with minimal effects at health care and community levels. Fourth, our SMS text message+WeChat systems could not track whether the participants had actually opened or read the information sent by us. Thus, we could not measure patient engagement or intervention exposure. From the postintervention feedback, we learned that some participants did not read all the articles sent by us, suggesting the need for better content design and innovative strategies to track and engage participants. Last, the small sample size of the pilot study might limit the power to detect significant differences hypothesized by us.

Despite these nonsignificant effects, this study showed feasibility and acceptability of the mHealth intervention in PLWH in China. Patients generally welcomed articles sent via WeChat and made specific recommendations to improve the intervention design and implementation. For example, they preferred receiving information via WeChat instead of SMS text messages; they welcomed more appealing design with multimedia functions. Furthermore, they expressed a strong need for programs that help combat depression and anxiety.

Based on the data collected from the pilot study, we proposed the following modifications for mHealth interventions for PLWH. First, the programs need to address the primary concern of the participants. For example, for PLWH in urban areas in China, poor mental health is the primary challenge that they face on a daily basis. It should be a priority in future health services provided to PLWH. Second, mHealth programs must go beyond traditional usability design and be more user centered. Because most mobile phone users are overloaded with information, the user experience is critical. For example, more tailored design with personalized feedback must be utilized. Third, theory-guided and evidence-based mHealth interventions should also incorporate tracking systems to measure user engagement and intervention exposure. Fourth, when testing the intervention, we need to recruit a diverse sample of PLWH of different age groups, educational levels, and transmission modes. Fifth, building a trust-based relationship between the participants and research staff of the program through personalized interactive communication is important for intervention adherence and participant retention. Finally, the efficacy trial must have a sufficient sample size and multiple follow-ups for the observation of intervention effects [29].

With the high penetration rates of mobile phones, mHealth interventions for PLWH have become more popular [3,26,30]. To date, the evidence on the effectiveness of mHealth interventions to promote medication adherence in PLWH has been preliminary [14,18,19,31,32] and the clinical evidence about viral suppression has been minimal [7,8,27]. The experiences from this study provided valuable inputs on the design and implementation of mHealth interventions for PLWH in middle- and low-income countries. Our team has revised the intervention protocol based on the experience from this pilot study, and a larger RCT is underway. We call for more evidence-based mHealth interventions with rigorous designs to serve the vulnerable population of PLWH in middle- and low-income countries.

Acknowledgments

This study was supported by a grant from the National Natural Science Foundation of China (Grant No. 71573290). We are grateful for the tremendous support from the patients and health care providers.

Authors' Contributions

YAH and YG conceptualized and designed the study. YG led the field implementation of the study with assistance from WC, LL, and CL. ZX, HZ, and JQ were involved in intervention delivery and data collection. YG and ZX led the statistical analysis. All authors critically reviewed and revised the article and approved the final version submitted for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 622KB - mhealth_v6i9e10274_app1.pdf]

References

1. UNAIDS. Fact sheet - latest statistics on the status of the AIDS epidemic URL: http://www.unaids.org/en/resources/fact-sheet [accessed 2018-03-01] [WebCite Cache ID 6xaH51Ye2]
2. 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]
3. Horvath T, Azman H, Kennedy GE, Rutherford GW. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. Cochrane Database Syst Rev 2012;3:CD009756. [doi: 10.1002/14651858.CD009756] [Medline: 22419345]
4. Statista. Number of mobile phone users worldwide from 2013 to 2019 URL: https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/ [accessed 2018-03-01] [WebCite Cache ID 6xaLwzq8d]
5. Cao B, Liu C, Durvasula M, Tang W, Pan S, Saffer AJ, et al. Social Media Engagement and HIV Testing Among Men Who Have Sex With Men in China: A Nationwide Cross-Sectional Survey. J Med Internet Res 2017 Jul 19;19(7):e251 [FREE Full text] [doi: 10.2196/jmir.7251] [Medline: 28724510]
6. Garett R, Smith J, Young SD. A Review of Social Media Technologies Across the Global HIV Care Continuum. Curr Opin Psychol 2016 Jun 01;9:56-66. [doi: 10.1016/j.copsyc.2015.10.024] [Medline: 26925455]
7. Simoni JM, Kutner BA, Horvath KJ. Opportunities and Challenges of Digital Technology for HIV Treatment and Prevention. Curr HIV/AIDS Rep 2015 Sep 28. [doi: 10.1007/s11904-015-0289-1] [Medline: 26412082]
8. Muessig KE, LeGrand S, Horvath KJ, Bauermeister JA, Hightow-Weidman LB. Recent mobile health interventions to support medication adherence among HIV-positive MSM. Curr Opin HIV AIDS 2017 Sep;12(5):432-441. [doi: 10.1097/COH.0000000000000401] [Medline: 28639990]
9. Statista. Number of mobile cell phone subscriptions in China from November 2016 to November 2017 URL: https://www.statista.com/statistics/278204/china-mobile-users-by-month/ [accessed 2018-03-01] [WebCite Cache ID 6xaO1olfD]
10. Niu L, Luo D, Liu Y, Silenzio VMB, Xiao S. The Mental Health of People Living with HIV in China, 1998-2014: A Systematic Review. PLoS One 2016;11(4):e0153489 [FREE Full text] [doi: 10.1371/journal.pone.0153489] [Medline: 27082749]
11. Zhang X, Wen D, Liang J, Lei J. How the public uses social media WeChat to obtain health information in china: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
12. He C, Wu S, Zhao Y, Li Z, Zhang Y, Le J, et al. Social Media-Promoted Weight Loss Among an Occupational Population: Cohort Study Using a WeChat Mobile Phone App-Based Campaign. J Med Internet Res 2017 Oct 23;19(10):e357 [FREE Full text] [doi: 10.2196/jmir.7861] [Medline: 29061555]
13. Lyu K, Zhao J, Wang B, Xiong G, Yang W, Liu Q, et al. Smartphone Application WeChat for Clinical Follow-up of Discharged Patients with Head and Neck Tumors: A Randomized Controlled Trial. Chin Med J (Engl) 2016 Dec 05;129(23):2816-2823 [FREE Full text] [doi: 10.4103/0366-6999.194635] [Medline: 27900995]
14. Zou H, Meng X, Grulich A, Huang S, Jia T, Zhang X, et al. A randomised controlled trial to evaluate the impact of sexual health clinic based automated text message reminders on testing of HIV and other sexually transmitted infections in men who have sex with men in China: protocol for the T2T Study. BMJ Open 2017 Dec 10;7(7):e015787 [FREE Full text] [doi: 10.1136/bmjopen-2016-015787] [Medline: 28698334]
15. Yan J, Zhang A, Zhou L, Huang Z, Zhang P, Yang G. Development and effectiveness of a mobile phone application conducting health behavioral intervention among men who have sex with men, a randomized controlled trial: study protocol. BMC Public Health 2017 Dec 24;17(1):355 [FREE Full text] [doi: 10.1186/s12889-017-4235-6] [Medline: 28438144]
16. Ruan Y, Xiao X, Chen J, Li X, Williams AB, Wang H. Acceptability and efficacy of interactive short message service intervention in improving HIV medication adherence in Chinese antiretroviral treatment-naïve individuals. Patient Prefer Adherence 2017;11:221-228 [FREE Full text] [doi: 10.2147/PPA.S120003] [Medline: 28228652]
17. Fisher JD, Fisher WA, Amico KR, Harman JJ. An information-motivation-behavioral skills model of adherence to antiretroviral therapy. Health Psychol 2006 Jul;25(4):462-473. [doi: 10.1037/0278-6133.25.4.462] [Medline: 16846321]
18. Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de WD, 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]
19. Mbuagbaw L, Thabane L, Ongolo-Zogo P, Lester RT, Mills EJ, Smieja M, et al. The Cameroon Mobile Phone SMS (CAMPS) trial: a randomized trial of text messaging versus usual care for adherence to antiretroviral therapy. PLoS One 2012;7(12):e46909 [FREE Full text] [doi: 10.1371/journal.pone.0046909] [Medline: 23236345]

20. Zhou Q, Wu Y, Hong YA, Yang C, Cai W, Zhu Y, et al. Association between perceived social norm and condom use among people living with HIV/AIDS in Guangzhou, China. AIDS Care 2017 Jan;29(1):91-97. [doi: 10.1080/09540121.2016.1198752] [Medline: 27327624]
21. Zeng C, Li L, Hong YA, Zhang H, Babbitt AW, Liu C, et al. A structural equation model of perceived and internalized stigma, depression, and suicidal status among people living with HIV/AIDS. BMC Public Health 2018 Dec 15;18(1):138 [FREE Full text] [doi: 10.1186/s12889-018-5053-1] [Medline: 29334959]
22. Rosenberger W, Lachin J. Randomization in Clinical Trials: Theory and Practice (Wiley Series in Probability and Statistics). Virginia: Wiley; 2015.
23. Zhang J, Norvilitis JM. Measuring Chinese psychological well-being with Western developed instruments. J Pers Assess 2002 Dec;79(3):492-511. [doi: 10.1207/S15327752JPA7903_06] [Medline: 12511017]
24. Lin N. Measuring depressive symptomatology in China. J Nerv Ment Dis 1989 Mar;177(3):121-131. [Medline: 2918295]
25. Hsiung P, Fang C, Wu C, Sheng W, Chen S, Wang J, et al. Validation of the WHOQOL-HIV BREF among HIV-infected patients in Taiwan. AIDS Care 2011 Aug;23(8):1035-1042. [doi: 10.1080/09540121.2010.543881] [Medline: 21500023]
26. Guangzhou Statistics Bureau. Guangzhou statistical yearbook URL: http://www.gzstats.gov.cn/ [WebCite Cache ID 6xaOwj7G4]
27. 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]
28. Tao J, Qian H, Kipp AM, Ruan Y, Shepherd BE, Amico KR, et al. Effects of depression and anxiety on antiretroviral therapy adherence among newly diagnosed HIV-infected Chinese MSM. AIDS 2017 Dec 28;31(3):401-406 [FREE Full text] [doi: 10.1097/QAD.0000000000001287] [Medline: 27677168]
29. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
30. 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 03;7(11):e017604 [FREE Full text] [doi: 10.1136/bmjopen-2017-017604] [Medline: 29101138]
31. Huang D, Sangthong R, McNeil E, Chongsuvivatwong V, Zheng W, Yang X. Effects of a Phone Call Intervention to Promote Adherence to Antiretroviral Therapy and Quality of Life of HIV/AIDS Patients in Baoshan, China: A Randomized Controlled Trial. AIDS Res Treat 2013;2013:580974 [FREE Full text] [doi: 10.1155/2013/580974] [Medline: 23401755]
32. Sabin LL, Bachman DM, Gill CJ, Zhong L, Vian T, Xie W, et al. Improving Adherence to Antiretroviral Therapy With Triggered Real-time Text Message Reminders: The China Adherence Through Technology Study. J Acquir Immune Defic Syndr 2015 Aug 15;69(5):551-559 [FREE Full text] [doi: 10.1097/QAI.0000000000000651] [Medline: 25886927]

Abbreviations

ART: antiretroviral therapy
CD4: cluster of differentiation 4
MSM: men who have sex with men
PLWH: people living with HIV
QoL: quality of life
RCT: randomized controlled trial
SMS: short message service
WHO: World Health Organization

Edited by G Eysenbach; submitted 03.03.18; peer-reviewed by R Amico, P Cook; comments to author 05.04.18; revised version received 31.05.18; accepted 18.06.18; published 04.09.18.

Please cite as:
Guo Y, Xu Z, Qiao J, Hong YA, Zhang H, Zeng C, Cai W, Li L, Liu C
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;6(9):e10274
URL: http://mhealth.jmir.org/2018/9/e10274/
doi:10.2196/10274
PMID:30181109

Original Paper

Smartphone-Based Contingency Management Intervention to Improve Pre-Exposure Prophylaxis Adherence: Pilot Trial

John T Mitchell[1,2], PhD; Sara LeGrand[3], PhD; Lisa B Hightow-Weidman[4], MD; Mehri S McKellar[5], MD; Angela DM Kashuba[6], PharmD; Mackenzie Cottrell[6], PharmD; Tony McLaurin[1], MSPH; Goutam Satapathy[7], PhD; F Joseph McClernon[1,2], PhD

[1]Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, United States

[2]Duke Center for Addiction Science and Technology, Durham, NC, United States

[3]Center for Health Policy and Inequalities Research at Duke University, Duke Global Health Institute, Durham, NC, United States

[4]Institute of Global Health and Infectious Diseases, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[5]Division of Infectious Diseases, Duke University Medical Center, Durham, NC, United States

[6]Division of Pharmacotherapy and Experimental Therapeutics, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[7]Intelligent Automation Incorporated, Rockville, NC, United States

Corresponding Author:
John T Mitchell, PhD
Department of Psychiatry and Behavioral Sciences
Duke University Medical Center
2608 Erwin Road
Pavilion East, Suite 300
Durham, NC, 27705
United States
Phone: 1 919 681 0012
Fax: 1 919 681 0016
Email: john.mitchell@duke.edu

Abstract

Background: Pre-exposure prophylaxis (PrEP) provides a strong preventative benefit to individuals at risk for HIV. While PrEP adherence is highly correlated with its efficacy, adherence rates are variable both across and within persons.

Objective: The objective of this study was to develop and pilot-test a smartphone-based intervention, known as mSMART, that targets PrEP adherence. mSMART provides contingency management in the form of monetary incentives for daily PrEP adherence based on a real-time adherence assessment using a camera-based medication event-monitoring tool as well as medication reminders, PrEP education, individualized behavioral strategies to address PrEP adherence barriers, and medication adherence feedback.

Methods: This was a 4-week open-label, phase I trial in a community sample of young men who have sex with men already on PrEP (N=10).

Results: Although adherence composite scores corresponding to PrEP biomarkers indicated that 90% (9/10) of the sample already had an acceptable baseline adherence in the protective range, by the end of the 4-week period, the scores improved for 30% (3/10) of the sample—adherence did not worsen for any participants. Participants reported mean PrEP adherence rates of 91% via daily entries in mSMART. At the end of the 4-week period, participants indicated acceptable ratings of satisfaction, usability, and willingness to recommend mSMART to others. There were no technical difficulties associated with smartphone compatibility, user misunderstandings about mSMART features that interfered with daily use, or study attrition.

Conclusions: This study is the first to apply contingency management to PrEP adherence. Findings indicated that mSMART is feasible and acceptable. Such an adherence intervention administered via a user-friendly smartphone app can allow for widespread dissemination. Future efficacy trials are needed.

Trial Registration: ClinicalTrials.gov NCT02895893; https://clinicaltrials.gov/ct2/show/NCT02895893 (Accessed by Webcite at http://www.webcitation.org/72JskjDJq)

(JMIR Mhealth Uhealth 2018;6(9):e10456) doi:10.2196/10456

KEYWORDS

HIV; preexposure prophylaxis; mobile health

Introduction

Pre-exposure prophylaxis (PrEP) in the form of tenofovir disoproxil fumarate and emtricitabine is a highly effective tool to prevent HIV infection [1-9]. However, adherence rates to this once daily medication are highly variable in clinical trials, ranging from 12% to 82% [4,10-15]. This is particularly significant for HIV prevention because the effectiveness of oral PrEP is strongly associated with sustained adherence [3,4,16]. Among those receiving PrEP in a 72-week open-label extension trial, HIV incidence significantly dropped from 4.7 infections per 100 person-years if the drug was not detected in blood to 2.3, 0.6, and 0.0 infections per year if blood levels correlated with participants' taking less than 2 tablets per week, 2-3 tablets per week, and ≥4 tablets per week, respectively [3]. Young men who have sex with men (MSM) are particularly at risk for HIV and could therefore benefit from PrEP. MSM represent just 2% of the US population but account for 67% of all new HIV diagnoses, which is driven in part by increased rates in young MSM [17]. In conjunction with trials indicating that younger participants, including MSM, are less likely to be adequately adherent to PrEP [3,18,19], interventions that target PrEP adherence are needed.

Despite the importance of PrEP adherence, there are few empirically supported interventions targeting adherence. One pilot trial indicated that a cognitive behavioral intervention, including 4-6 face-to-face sessions, improved PrEP adherence among MSM in comparison to a time-matched control intervention [20]. Although such interventions are promising, easily disseminated and wide-reaching interventions that maintain fidelity to rigorous intervention protocols may further enhance efforts to promote PrEP adherence. Smartphones offer such a platform for personalized and flexible interventions to improve health outcomes that can be administered in a uniform and user-friendly format [21]. Smartphones are used by an increasing segment of the US population (eg, 77% owned one in 2016 up from 35% in 2011) [22]—people who carry smartphones generally have them within reach and switched on at all times [23]. However, despite the fact that there are more than 800 medication adherence apps for a range of conditions, only a few have been widely studied [24]. Although some research is beginning to investigate the use of daily texting to support PrEP adherence [25], a smartphone app targeting HIV prevention that includes PrEP screening [26], and a smartphone app that incorporates PrEP adherence among MSM in an ongoing trial [27], to our knowledge, there are no published studies on medication adherence smartphone apps for PrEP.

Contingency management, administered via smartphones, may be a promising intervention approach for improving PrEP adherence. Contingency management is a behavioral intervention that uses systematic reinforcement dependent on the occurrence of a specific behavior and is effective in improving adherence to medications for a range of medical and psychiatric conditions [28]. Contingency management has been used to successfully improve adherence to antiretroviral medications among HIV-positive and HIV-exposed individuals [29,30], but it has yet to be examined as an intervention for PrEP adherence.

The aim of this study was to develop and pilot test a smartphone-based contingency management intervention, known as mSMART, that targets PrEP adherence. In addition to contingency management, mSMART provides medication reminders, PrEP education, individualized behavioral strategies to address PrEP adherence barriers, and medication adherence feedback. mSMART also assesses adherence in real time using a camera-based medication event-monitoring tool. This was a 4-week open-label, phase I trial. We examined the feasibility and acceptability of mSMART in a sample of young MSM prescribed PrEP in a community setting.

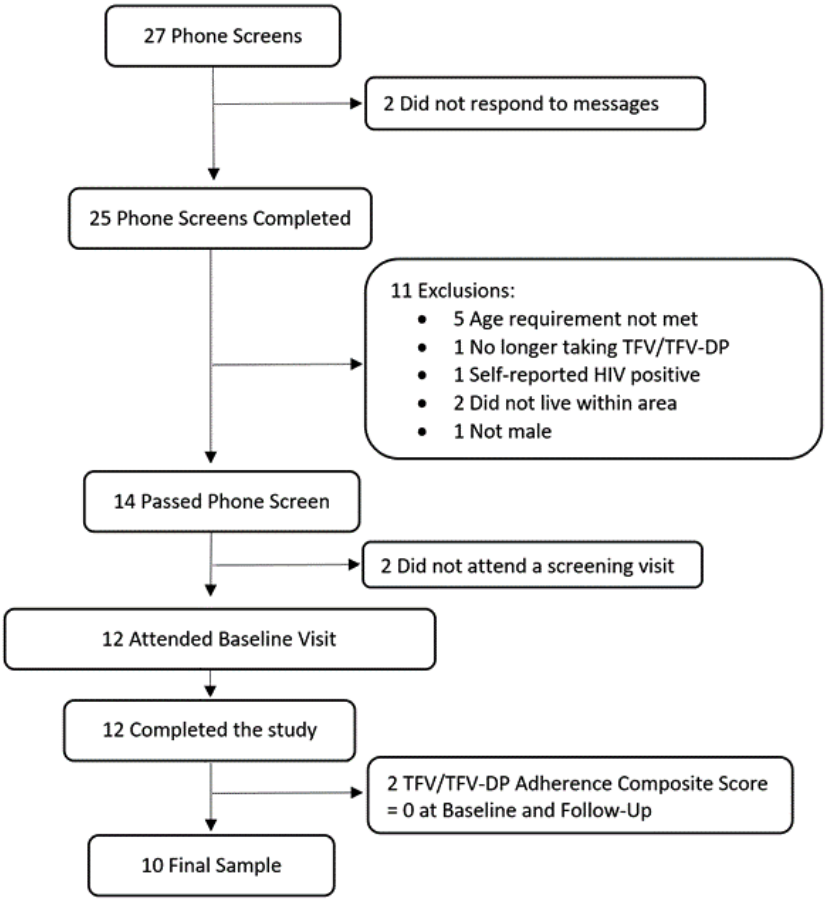
Methods

Participants

Inclusion criteria were male at birth, age 18-30 years, self-report having sex with men in the last 6 months, self-report being currently prescribed and taking PrEP for HIV prevention, English speaking, and own an Android or an iPhone compatible with the mSMART smartphone app. Exclusion criteria included significant medical or psychiatric conditions that may interfere with study participation (eg, suicidality) or being unable to attend both study visits. There were no inclusion or exclusion criteria pertaining to the amount of time participants were prescribed PrEP prior to enrollment. Participants were recruited via community advertisements and word of mouth.

A total of 27 screens conducted over the phone were held, and 14 individuals were invited for the baseline assessment. Individuals were not invited for a baseline visit for the following reasons: they did not respond to phone messages (n=2), they did not meet the age inclusion criterion (n=5), they were not currently prescribed PrEP (n=2), they self-reported HIV-positive status (n=1), they did not live close enough to attend laboratory visits (n=2), and they were not male (n=1). Among the 14 invited individuals for a baseline visit, 2 participants did not attend the visit. A total of 12 participants were consented, but 10 participants were included in this analysis; 2 participants had baseline tenofovir (TFV)/TFV-diphosphate (TFV-DP) levels, indicating that they were not taking PrEP, as seen in Figure 1. We decided to exclude these 2 participants from the analysis post hoc because, in contrast to the final sample of 10, we could not verify that these participants were ever prescribed PrEP using our biomarker analysis. Furthermore, there was concern about the validity of the self-report data these participants provided. For example, both participants stated that they were on PrEP for the past 8 and 9 months and self-reported only missing doses approximately 12 and 2 times over those time periods, respectively. This contrasted with the biomarker analysis that indicated that they had no detectable levels of TFV/TFV-DP.

Figure 1. Sample recruitment and participation flowchart. TFV: tenofovir, TFV-DP: tenofovir-diphosphate.



Procedures and Measures

In total, 2 laboratory visits were required: baseline and follow-up. Participants were provided the mSMART smartphone app on their smartphone and asked to use it daily during the 4-week period between visits.

Baseline Visit

After obtaining the informed consent, participant demographic, medication, and medical and psychiatric history information were collected in a paper-and-pencil format to characterize the sample. Additional questionnaires were administered via a Web-based survey tool during the visit to further characterize the sample. The 6-item Risk Behavior Assessment for MSM [31] recommended by the Centers for Disease Control and Prevention [32,33] was administered to assess HIV risk over the past 6 months. The 9-item Patient Health Questionnaire-9 [34] was used to assess depressed mood over the past 2 weeks. Substance use was assessed using the 10-item Drug Abuse Screening Test [35] and the 10-item Alcohol Use Disorders Identification Test [36]. The number of perceived barriers to PrEP adherence was assessed using the 20-item Adherence Starts with Knowledge questionnaire (ASK-20) [37]—the original version of this scale was modified to inquire about PrEP specifically. This measure was used to characterize perceived PrEP adherence barriers at baseline as well as an outcome measure to compare with follow-up visit ratings.

At the conclusion of the baseline visit, the study participants were registered with the mSMART app by the study team on a secure website [38] and the app was downloaded by participants from the appropriate app distribution platform for their operative system (eg, Apple Store, Google Play). Participants received brief instructions from an experimenter on the functions of mSMART. Overall, the baseline visit took approximately 90-120 minutes to complete.

Follow-Up Visit

The ASK-20 questionnaire modified for PrEP was readministered. Treatment acceptability ratings were provided by participants based on responses to individual items examining overall satisfaction with mSMART (ie, “What was your overall satisfaction with mSMART?”), mSMART usability on a daily basis (ie, “How usable was mSMART on a daily basis?”), difficulty learning how to use mSMART (ie, “How difficult was it to learn how to use mSMART?”), willingness to recommend mSMART to others (ie, “Would you recommend mSMART to a friend who is taking a medication?”), and overall user-friendliness of mSMART (ie, “How user-friendly was mSMART?”) on a Likert scale ranging from 1 (not at all) to 4 (extremely). These items were administered in an in-person interview format and were adapted from our past use of a similar scale [39]. The System Usability Scale (SUS) [40] was administered via a Web-based survey tool as a measure of treatment acceptability. SUS is a 10-item scale that assesses responses on a 5-point Likert scale with scores ranging from 0

to 100. Semistructured exit interviews were also conducted for qualitative analysis of participant experiences and perceptions of mSMART. Interview questions addressed topics such as mSMART design features, navigation, barriers to use, and features that facilitated regular use similar to other studies examining participant experience with smartphone-based interventions (eg, [41]).

Participants also completed predetermined tasks within the smartphone app during the follow-up study visit following guidelines from another smartphone app development study [41]. An experimenter sat next to the participant, provided instructions on 6 different tasks, and recorded the time to complete each task. These 6 tasks involved (a) taking a picture of their medication, (b) changing the reminder time for daily dosing, (c) checking how much money was earned using the smartphone app, (d) checking for any questions prompted by mSMART, (e) looking up a detail about medication side effects, and (f) looking up a second detail about medication side effects. The time recorded for each task was based on the first attempt to complete it.

Bioanalytical Adherence Assessment

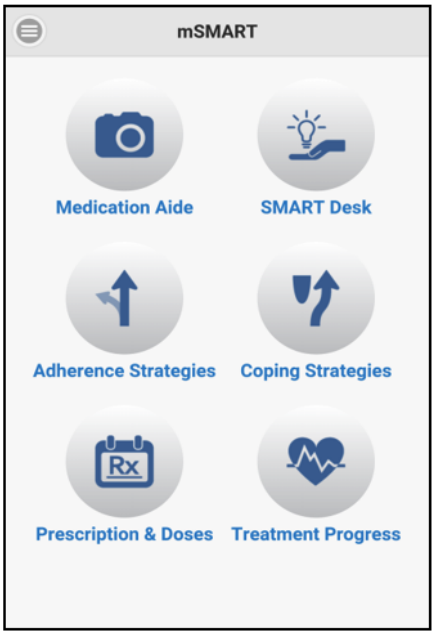
Blood samples were collected at both the baseline and follow-up visits to assess for biomarkers of PrEP adherence. The blood samples were used to assess the concentrations of TFV in plasma and intracellular TFV-DP in the upper layer packed cells both to characterize baseline levels and as a comparison with the follow-up assessment levels using methods previously described by Adams et al [42]. These levels were used to develop a semiordinal composite adherence score over the past 4 weeks, ranging from 0 (low or no doses of drug identified: no detectable TFV and <10,000 fmol/mL TFV-DP) to 5 (good adherence: >10 ng/mL TFV and >1,000,000 fmol/mL TFV-DP) [15]. A score of 4 (ie, 4-5 doses per week) or 5 (approximately daily dosing) is typically considered as the level of adherence at which PrEP is efficacious among MSM [43].

mSMART Intervention

mSMART was developed through a multistage process initially as a smartphone app for medication adherence for cigarette smokers during quit attempts [44]. It was adapted for PrEP by the research team for this study. The adaptation of mSMART for PrEP was informed by studies on adherence barriers in PrEP trials (eg, [45-48]) and feedback from experts working with and developing interventions for individuals at risk for HIV. The Information, Motivation, and Behavioral Skills (IMB) model [49], which conceptualizes health behavior change as a product of mediators, including information about the behavior, motivation to change, and behavioral skills, guided the refinement of mSMART for PrEP. For example, Information (the first IMB model component) was conveyed through an interactive daily question-and-answer format involving PrEP and HIV knowledge as well as self-assessments of general medication adherence difficulties (see the SMART Desk feature described below). Information was also provided about adherence with mSMART via visual feedback about logging doses each day (see the Treatment Progress feature described below). Motivation (the second IMB model component) to adhere to PrEP was provided in the form of contingent reinforcement when doses were logged daily and daily medication reminders (see the Medication Aide feature described below).

Behavioral Skills (the third IMB component) were taught by mSMART with behavioral skill instruction on how to improve adherence (eg, how to remember to take a daily dose if forgetfulness is a barrier to adherence) and how to cope with short-term side effects that may deter daily adherence (see the Adherence Strategies and Coping Strategies features described below). The IMB model has been used to guide the development of numerous HIV prevention interventions and encapsulates other theoretical HIV risk reduction models (eg, [50]).

Figure 2. mSMART home screen.



mSMART was used by participants over a 4-week period between the baseline and follow-up visits. mSMART is composed of 6 different components, as seen in Figure 2, that target medication adherence.

Table 1 summarizes the 6 different components of mSMART. The Medication Aide component of mSMART included the contingency management procedures. Upon receiving their daily PrEP reminder notification, participants touched the Medication Aide icon and were then directed to enter a dose of PrEP that they were either about to take or had already taken. In either case, as long as participants reported taking their daily dose of PrEP within 2 hours of their predetermined dosing time each day (dosing times were selected by participants), they received reinforcement. A fixed-ratio schedule of reinforcement was adopted. Participants received feedback that they earned US $2 every time they logged a dose (whether using the camera feature or manually) within a 2-hour window of their daily dosing. Feedback about money earned upon taking their daily dose was provided immediately by mSMART. Over the 4-week period, participants could earn up to US $56. Participants could opt to receive the money that they earned in accordance with the contingency management procedures weekly or at the end of the 4-week intervention period.

Data Analysis

Adherence Outcomes

Perceived and objective PrEP adherence outcomes were assessed via change scores on ASK-20 and medication adherence scores based on TFV/TFV-DP, respectively.

Table 1. Description of mSMART features.

mSMART feature	Description
Medication Aide	Participants used this feature to enter a dose of PrEP[a] that they were either about to take or had already taken. If a participant indicated that he was about to take his daily PrEP dose, the camera-based medication event-monitoring tool was activated. This involved taking the participant to another screen that prompted him to touch a pill icon that would open up the camera feature on his phone. The mSMART app would automatically take a picture and would take approximately 5-10 s to focus on the pill the participant was holding in his hand—a feedback bar indicating progress was provided over the top portion of the picture. For this study, these pictures were not examined by the study team or saved, although mSMART has that capability. If participants had already taken their daily dose of PrEP, they would manually enter when they took their daily dose of PrEP.
SMART Desk	This component was an interactive space where mSMART prompted brief daily surveys (ie, 1-4 questions per day pertaining to knowledge or concerns about PrEP, knowledge about HIV, and general medication use concerns or problems). These questions were phased out after any 7-day window only if participants were achieving 100% adherence by logging daily PrEP doses in that window, but were resumed if a dose was not logged. For participants who were not logging all daily PrEP doses, they continued to receive daily questions from the SMART Desk. Notifications informing participants of missing a PrEP dose were also provided through the SMART Desk.
Adherence Strategies[b]	This component described behavioral strategies to address PrEP adherence barriers identified in the literature [45-48]. These strategies were prioritized in list form based on responses from participants in the SMART Desk and could be accessed at any time by clicking on the Adherence Strategies icon. For example, if a participant indicated he did not have difficulty remembering to take daily PrEP doses but had a relatively poor understanding of how PrEP prevents HIV on previous SMART Desk questions, the Adherence Strategies component presented educational information about how PrEP prevents HIV before presenting behavioral strategies to help the participant remember to take his medication. Thus, adherence strategies were individualized based on participant responses in the SMART Desk.
Coping Strategies	This feature listed common PrEP side effects. Participants could access a list of side effects at any time and click on any to view strategies to mitigate them. The most common side effects reported in the literature (eg, upset stomach, headache, and vomiting [32,33]) were included.
Prescription and Doses	With this feature, the participants were able to set up their preferred time to receive medication reminders. Participants could change this setting at any time and therefore could modify it on days they anticipated taking PrEP at a different time.
Treatment Progress	This feature provided feedback about the participant's overall PrEP adherence in the form of percentage of days they logged a dose (within the 2-h window) within the Medication Aide feature. Participants could also click on this feature to see how much money they had earned based on the contingency management procedures.

[a]PrEP: Pre-exposure prophylaxis.

[b]Other examples of adherence strategies that this mSMART component addresses include strategies to organize materials to take medication daily, ways to remember to take medication, education about different aspects of PrEP (eg, explaining why daily adherence is important, describing a typical medical visit schedule once on PrEP, and addressing concerns regarding possible long-term health effects of PrEP), financial aspects related to being on PrEP, information about communicating with health care workers about PrEP and sexual behavior, and eliciting support from family and friends to support PrEP adherence. In addition to accessing adherence strategies by clicking on the Adherence Strategies icon, participants were automatically routed to specific Adherence Strategies from the SMART Desk after completing the questions in the SMART Desk. For example, if the SMART Desk asked about remembering to take medication, the participant would be routed to a strategy within Adherence Strategies to address medication forgetfulness. This routing occurred regardless of the response selected with the intent to increase exposure to a variety of adherence strategies, which was balanced against personalized presentation of strategies based on SMART Desk responses described above.

Treatment Feasibility

Feasibility was assessed in the following ways: study attrition rate and any smartphone-mSMART compatibility incidents, daily engagement with mSMART, time needed to complete the predetermined tasks on mSMART, and number of prompts (initiated by either the participant or the experimenter) to assist participants in completing the tasks.

Treatment Acceptability

Acceptability was assessed in multiple ways. First, responses to individual treatment acceptability questions about mSMART (ie, overall satisfaction, usability on a daily basis, difficulty learning mSMART, willingness to recommend mSMART to others, and overall user-friendliness) were descriptively analyzed. Second, SUS scores at or above 68 were considered as acceptable [51]. Third, we considered responses to semistructured exit interviews for qualitative analysis. Interviews were digitally recorded, transcribed, and qualitatively analyzed. Qualitative analysis involved identification of categories that emerged. These categories were identified through an iterative process following procedures similar to those involved in our past qualitative approaches [52,53]. That is, an initial list of anticipated categories based on the study team's experience with mSMART in other populations and separate experiences with young MSM. These categories were subsequently refined based on one of the authors' experience conducting the exit interviews and reading all interview transcripts. Another rater then read through the transcripts to comment on the category descriptions and identify any additional categories not previously considered. Next, both raters identified any discrepancies in category identification, reconciled these discrepancies, and finalized the categories. Following this process, the 2 raters separately read through the transcripts (n=5 per rater) in a Microsoft Word document and identified category endorsements for each participant. Each interview excerpt that was identified with a category endorsement was transferred to an Excel document so that frequency counts for particular categories could be summed across the full sample. We have adopted similar procedures in past studies [52]. Interrater reliability between raters was assessed on a subset of interview excerpts. Kappa coefficient between raters was .90 when determining whether a category should be endorsed.

Results

Sample Characteristics

The sample (N=10) contained predominantly white (n=7) and highly educated (n=7) participants, earning at least an undergraduate degree. The average number of months on PrEP was 8.3 with use ranging from 0.5 to 12 months. All but 1 participant reported being on PrEP for at least 5 months. Each participant exceeded the MSM Risk Index Score of 10 used to evaluate appropriateness for PrEP [32,33], indicating high risk for HIV. In addition, participants yielded low scores for depressed mood, drug use, and alcohol use (see Table 2 for a summary).

Adherence Outcomes

Objective Adherence

PrEP composite adherence scores based on TFV/TFV-DP values indicated that PrEP adherence increased for 30% (3/10) of the sample and did not change for 70% (7/10) of the sample. For participants who did not indicate any change, PrEP adherence scores were already at a level considered efficacious (ie, ≥4 doses per week) at baseline. Among the 3 participants whose PrEP adherence scores increased, 1 had a baseline score below what is considered efficacious. No PrEP composite adherence scores decreased. Table 3 provides the baseline and follow-up scores.

Perceived Adherence

The perceived number of barriers to PrEP adherence was measured using the modified ASK-20 at baseline and follow-up. A comparison of scores within participants indicated an increase in the number of perceived barriers for 1 participant. This participant indicated on an ASK-20 item that his belief that PrEP was helpful in reaching his overall health goals had decreased. However, 3 participants indicated that the number of barriers they perceived to PrEP adherence decreased, including barriers associated with the financial cost of PrEP. There was no change in modified ASK-20 scores for 50% (5/10) of the sample. One participant did not complete the modified ASK-20 at follow-up.

Treatment Feasibility

There was no study attrition. Furthermore, there were no smartphone-mSMART incompatibility events in which the mSMART app was not able to function on a study participant's phone. In terms of daily engagement with mSMART, participants logged a PrEP dose in mSMART (using either the camera-based medication event-monitoring tool or manual entry option) 91% of the time over the 4-week intervention period, as seen in Figure 3, with the mean amount earned per contingency management guidelines being US $53 per participant. Among these logged doses in mSMART, 88% involved the use of the camera-based medication event-monitoring tool, as seen in Figure 4. Overall, 40% (4/10) of the sample did not miss any days logging a PrEP dose in mSMART. An additional 40% (4/10) did not log a PrEP dose in mSMART between 1 and 5 days while in the study. Among the remaining participants, 1 did not log a PrEP dose for 6 days and the other did not log a PrEP dose for 12 days. Furthermore, 70% (7/10) of the sample responded to all of the mSMART daily surveys. During the follow-up study visit, all participants were able to complete each of the 6 predetermined tasks on mSMART without any prompts (initiated by either the participant or the experimenter). The amount of time it took to complete these tasks was 5.39 seconds (average across all tasks; see Multimedia Appendix 1).

Table 2. Sample characteristics (N=10).

Characteristic	Value
Age, mean (SD)	24.10 (2.38)
Race, n (%)	
Black	0 (0)
White	7 (70)
Asian	2 (20)
Multiracial	1 (10)
Ethnicity, n (%)	
Hispanic	0 (0)
Not Hispanic	9 (90)
Not reported	1 (10)
Education, n (%)	
High school graduate	1 (10)
Partial college	2 (20)
College graduate	5 (50)
Postgraduate studies	2 (20)
Employment status, n (%)	
Full-time	3 (30)
Part-time	2 (20)
Assistance	0 (0)
Unemployed	1 (10)
Dependent or student	3 (30)
Not reported	1 (10)
Salary range, n (%)	
US $0-$10,000	3 (30)
US $10,000-$25,000	3 (30)
US $25,000-$50,000	0 (0)
US $50,000-$75,000	2 (20)
>US $75,000	1 (10)
Not reported	1 (10)
Months prescribed PrEP[a], mean (SD)	8.30 (3.45)
MSM[b] Risk Index Score[c], mean (SD)	21.50 (5.48)
Smartphone, n (%)	
iPhone	9 (90)
Android	1 (10)
Patient Health Questionnaire-9, n (%)	
Minimal depression (scores=0-5)	9 (90)
Mild depression (score=6)	1 (10)
Drug Abuse Screening Test, n (%)	
None	7 (70)
Low	3 (30)
Alcohol Use Disorders Identification Test, n (%)	

Characteristic	Value
Low risk	10 (100)

[a]PrEP: Pre-exposure prophylaxis.

[b]MSM: Men who have sex with men.

[c]100% of the sample exceeded the cut-off score of 10 and therefore are recommended to evaluate for PrEP per Centers for Disease Control and Prevention guidelines [32,33].

Table 3. Frequency of pre-exposure prophylaxis (PrEP) composite adherence scores at study baseline and follow-up visits.

Composite score[a]	Baseline (%)	Follow-up (%)
0	0 (0)	0 (0)
1	1 (10)	0 (0)
2	0 (0)	0 (0)
3	0 (0)	0 (0)
4	8 (80)	7 (70)
5	1 (10)	3 (30)

[a]Composite scores were based on concentrations of tenofovir (TFV) in plasma and intracellular TFV-diphosphate (TFV-DP) in upper layer packed cells. Scores assess adherence in the past 4 weeks, ranging from 0 (low or no doses of drug identified) to 5 (good adherence). A score of 4 (ie, 4-5 doses per week) or 5 (approximately daily dosing) is typically considered as a good level of adherence in which PrEP is efficacious. Because 1 participant was on PrEP for only 2 weeks, the baseline visit adherence score for this participant could have been artificially lower as a result of taking PrEP for a shorter duration in comparison to other study participants (ie, all other participants reported being on it for at least 5 months). However, this participant yielded a baseline adherence score of 4, indicating an adequate level of protection since starting on PrEP and that his score was likely not artificially lower.

Figure 3. Percentage of time the individual participants logged a dose in mSMART using either the camera-based medication event-monitoring tool or manual entry option.

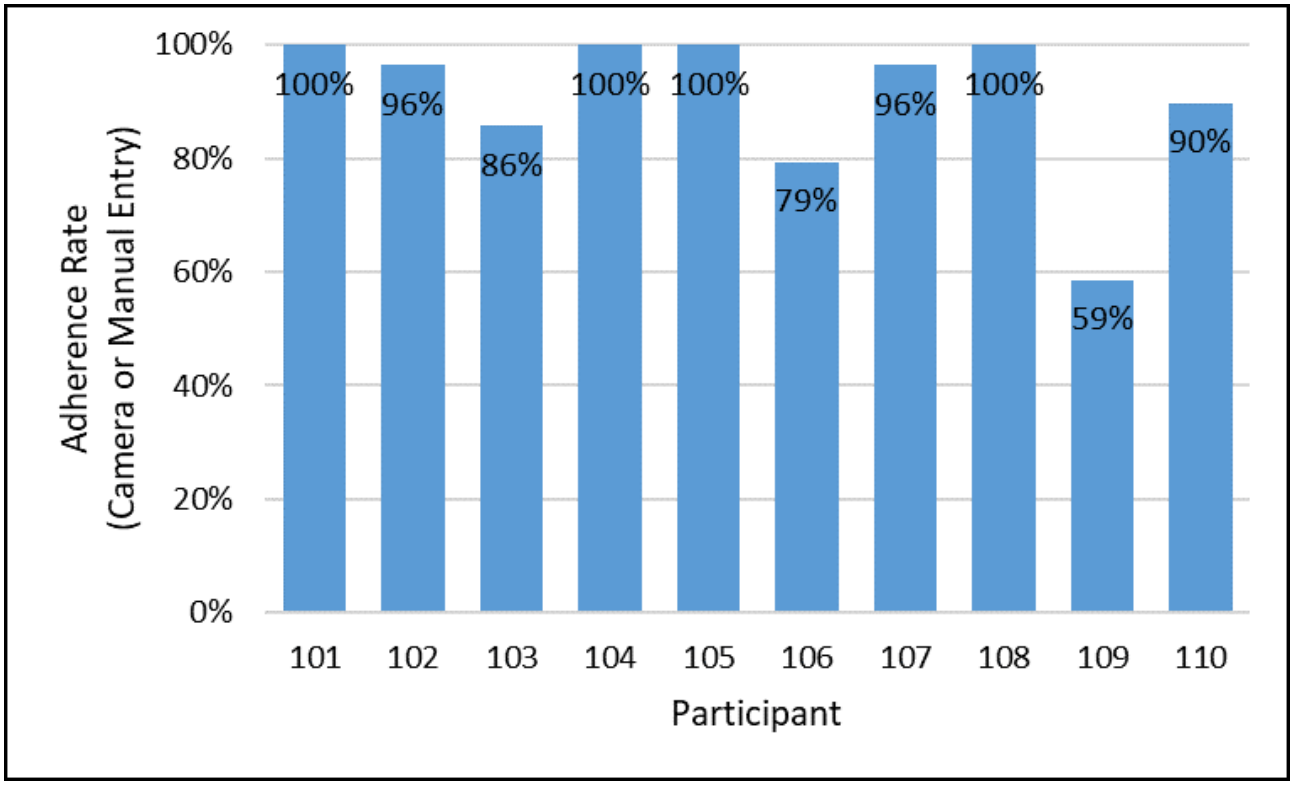
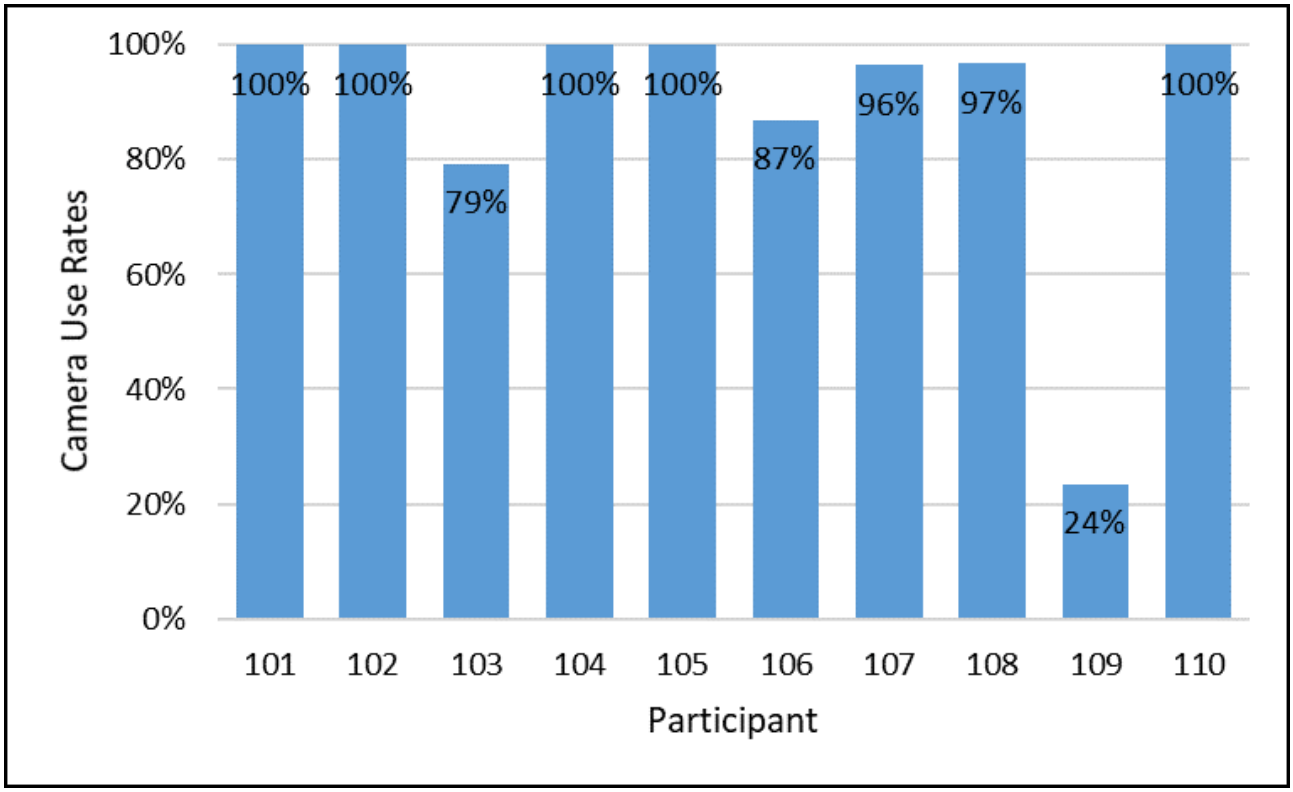


Figure 4. Percentage of time the camera-based medication event-monitoring tool was used among participants when they logged a dose in mSMART.



Treatment Acceptance: Quantitative Analysis

In-person interview items at the follow-up visit indicated that the mean rating on a scale of 1 (not at all) to 4 (extremely) for overall satisfaction with mSMART was 2.80 (SD 0.63), mSMART usability on a daily basis was 3.50 (SD 0.53), willingness to recommend mSMART to others prescribed PrEP was 2.70 (SD 0.82), and user-friendliness of mSMART was 2.80 (SD 0.79). Participants also indicated that difficulty learning how to use mSMART was 1.20 (SD 0.42).

On the SUS, with total scores ranging from 0 to 100, the mean score was 68.25 (SD 15.10). Using a score of ≥68 as an indicator of mSMART user-acceptability, 60% (6/10) of the sample met this cut-off.

Treatment Acceptance: Qualitative Analysis

We classified participant comments about mSMART into the following 6 different categories: (1) mSMART features that were liked or disliked, (2) daily mSMART use, (3) mSMART aesthetics, (4) learning how to use mSMART, (5) mSMART features that should be added, and (6) the likelihood of using mSMART. The first 4 domains were further subdivided into comments that were either positive or negative feedback about mSMART (see Multimedia Appendix 2 for a summary of endorsement rates across domains).

mSMART Features: Liked and Disliked

The majority of the sample, 8 participants, commented on features that they both liked and disliked, and the other 2 participants commented only on features that they liked and did not report any dislikes. Particular mSMART features that were most frequently mentioned included using the camera feature (5 participants liked, 2 participants disliked, 2 participants both liked and disliked, and 1 participant did not comment on this feature), receiving daily questions (5 participants liked, 0 participants disliked, 3 participants both liked and disliked, and 2 participants did not comment on this feature), and receiving medication reminders (5 participants liked, 1 participant disliked, 1 participant both liked and disliked, and 3 participants did not comment on this feature), for example, when asked about his overall impression of mSMART, 1 participant responded in the following way about the medication reminders:

> *I think [mSMART] is helpful, I mean, because it also offers some useful information on Truvada and, uhh, the most important thing is it offers reminders and I mean, at least for now, I still need the reminder to remind me to take my medication. Before I started using this app...it's really easy to forget every day.*

Another participant commented on how the daily questions from the SMART Desk were helpful as follows:

> *[mSMART] figured out what I struggled with by the questions.*

Although some participants indicated that they both liked and disliked the camera feature, most of the comments about disliking the camera feature actually involved initial difficulty learning how to use this feature, for example, 1 participant stated the following:

> *I think there were like, especially in the first week, there were four or five times I would try to take a picture of the pill and it went straight to uhhh, it wouldn't, it didn't read it. And mainly it was times when I was taking the picture and it was too dark, right? I was just like in my room in the morning and didn't have any lights on, or in our kitchen and it was*

> *just really dark. And that was frustrating. Umm, it wasn't so, I mean it was a minor inconvenience in the grand scheme of the world, right?*

After describing this initial experience with the camera feature, this participant went on to say,

> *Personally, I definitely think that taking a photo was good.*

Other participants indicated that the camera feature helped with consolidating memories to increase confidence that they took their medication, for example,

> *...using the camera like, really forced me to use it and kind of was a mental check guard for myself to make sure I took the medication...Like I was telling you that, I'm like, did I take the pill? Was it yesterday when I was going to work? Or was that today? And so [inaudible] taking the picture at 5:37, like I did do it today. So I'm not confused. I know when I took it, and that was today.*

The money earned in accordance with the contingency management procedures was not among the mSMART features that participants considered to find helpful, that is, only 4 participants indicated liking the contingency management payment feature of mSMART; 2 participants said that it was not a helpful feature in adhering to PrEP and another 4 participants indicated that they neither liked nor disliked this feature.

Daily Use: Facilitators and Barriers

Four participants commented on factors that facilitated successful daily use of mSMART, whereas 6 participants commented on factors that both facilitated mSMART use and barriers to mSMART use. The most frequently endorsed factor facilitating daily mSMART use (7 participants) was using it at the time they received a medication reminder. For example, 1 participant stated the following:

> *I pretty much only used it when I needed to log my, umm, medication, which was at night.*

Regarding barriers to daily use, the most frequently identified feature was that the speed of mSMART was too slow (ie, time to transition from one screen to the next or to perform a function) and was therefore a barrier to use (2 participants):

> *So, sometimes when you click on the medic—, like the camera function, it takes a second and then it'll go, umm, and then it'll take a minute to get to the next slide and the next screen or whatever you want to call it, which is fine, but I'm just saying like for someone who is going to use it every day and does not have the incentive of here's...money at the end of the trial, you know what I mean? It could be, people could, someone might get frustrated.*

mSMART Aesthetics: Liked and Disliked

An equal proportion of participants either commented that they disliked mSMART aesthetics (4 participants) or both liked and disliked the aesthetics of the app (4 participants); the other 2 participants commented either only on aesthetics that they liked (n=1) or did not comment about aesthetics at all (n=1). The most frequently identified aesthetic that was disliked was how the content was displayed in text format (7 participants), for example,

> *I would go to some of the coping strategies just to like look through them, and I would say, like, you open it, and there's kind of just a large block of text—that might be a little intimidating....on the one hand I thought it was helpful because it felt like a pretty clinical tone, umm, like from a healthcare provider, like in a good way. Like, if that's what you want, you know, if you want that...But, then sometimes I was thinking that maybe I would want a more just like a friend...*

The most frequently identified aesthetics that was liked was the overall design of mSMART (4 participants), for example,

> *I thought it was pretty well-designed. And I guess, yeah, I mean it was clearly laid out to me. Umm, the, like the functions, like everything opened when you tapped it. There was no like glitch, there was no, the camera, everything worked.*

Similarly, another participant stated the following about the design when asked:

> *I like the simple breakdown into the six different sections, I think that's what makes it user-friendly. Umm, I mean, it's easy to follow when you go into like the different coping strategies and whenever you try to highlight the hyperlinks are really easy to kind of delineate what you're looking for.*

Learning How to Use mSMART: Easy and Difficult

Although 9 participants indicated that learning how to use mSMART was easy, 1 participant indicated that it was difficult. Typical comments about learning how to use mSMART included

> *I think the app is actually pretty, uhh, like user-friendly. It doesn't, it doesn't take a whole lot of learning.*

Features of mSMART That Should Be Added

In total, 90% (9/10) of participants commented on features of mSMART that they thought should be modified. The most commonly endorsed feature that should be added was a snooze option for medication reminders (4 participants), for example, when discussing modifying the alarm feature of mSMART, 1 participant suggested the following change:

> *Almost like on your phone if you snoozed or something...If something happens, it will alert you again....to physically turn it off almost.*

Likelihood of Using mSMART

The majority of the sample, 9 participants, commented on how they thought mSMART would be most appropriate for individuals either just starting PrEP or those who have PrEP adherence problems. In particular, 6 participants—all of whom reported taking PrEP for at least 5 months—commented on how

it would have been helpful to use mSMART when they began taking PrEP. For example, 1 participant commented the following on how mSMART would have been helpful when starting on PrEP:

> *I do think, like the first two months probably would've been the time this app would've been the most helpful...Cause there were also times that I straight up, like I forgot if I had taken it or not; in that first month.*

Another participant spoke less about his own initial difficulties with PrEP adherence but spoke more broadly about individuals initiating PrEP as follows:

> *...as more and more people learn about it and find out about it that information might be less, so that there might be more questions about understanding side effects, especially in the first month where you're most likely to have side effects. So like, I think that [is] where it can be useful. So like figuring out how do I cope with these side effects? Are they going to go away? What's the duration?...Umm I think, one thing I can imagine is like, you know, suppose that when you're first starting your medication, you're less likely to have a routine, so you're more likely to miss a dose, and in some cases you might wonder, well like what, let's say I usually take my dose at 8 in the morning and its 3 in the afternoon, and I just realized I didn't take my pill, should I take my pill or not? That's a question that I think people might have, and your doctor may or may not have given you guidance on what to do in that situation...So, that's an area where I can imagine the role this app can fill.*

Discussion

The large-scale implementation of PrEP is an ongoing challenge that requires diverse models of delivery addressing multiple facets of the PrEP continuum of care [54,55]. This study was a 4-week pilot trial of mSMART as a mobile health PrEP adherence intervention. Adherence outcomes, treatment feasibility, and treatment acceptability were examined in 10 young MSM already prescribed PrEP in the community. Findings from this treatment development study are preliminary but yield promising results and indications for treatment refinement for future efficacy trials.

PrEP adherence outcomes were measured both objectively and subjectively. For the former, PrEP composite adherence scores based on TFV/TFV-DP were examined. Although baseline scores indicated a high level of adherence prior to using mSMART with 90% (9/10) of the sample was at or above a level of PrEP adherence considered as efficacious and therefore a ceiling effect would likely occur, 30% (3/10) of the sample's scores improved at follow-up. In addition, no composite adherence scores worsened over the course of the study. Our use of biomarkers as an adherence outcome is a strength given that there is substantial within-subject variability in adherence based on the measurement method selected among individuals on PrEP [56]. Alternative methods such as self-report [57] and electronic pill bottles [58] among young MSM have noted limitations. In terms of a future direction, because PrEP adherence scores showed an already high rate at baseline, future studies are needed that would address whether mSMART improves PrEP adherence among those with a poor medication adherence history and whether this impact on adherence is clinically significant.

The subjective measure of adherence was an adapted medication questionnaire measuring perceived PrEP adherence barriers administered at baseline and follow-up. Although 50% (5/10) of the sample did not report any change in barriers to PrEP adherence, 30% (3/10) reported a decrease in barriers and 10% (1/10) reported an increase in barriers. Participants who reported a decrease in barriers indicated that factors such as barriers associated with the financial cost of PrEP; the participant who reported an increase in barriers indicated that the belief that PrEP was helpful in reaching overall health goals had changed. Although these changes in perceived barriers (either increasing or decreasing) emerged, future studies that include a control condition are needed to address whether these changes occurred because of mSMART or other factors. Overall, across the methods of adherence examined in this study, the findings were relatively consistent.

mSMART feasibility was positive as evidenced by 0% (0/10) study attrition, the absence of any smartphone incompatibility events, and daily engagement with mSMART. Regarding daily engagement, we looked at the rates of logged PrEP doses and the proportion that responded to all of the mSMART daily surveys. Using either the camera-based medication event-monitoring tool or the manual entry option within a 2-hour window of when participants identified the time they should take their PrEP pill, the overall adherence rate was 91%. Although contingency management guidelines in this trial considered a medication event as valid either if there was a picture taken of the PrEP pill or if it was entered retrospectively, a more methodologically rigorous contingency management approach would require an objective assessment of behavior that does not rely on self-report (eg, use of the camera-based medication event-monitoring tool only). Given that the majority of times medication adherence was reported via mSMART involved camera-based entries (88%), a more rigorous contingency management intervention appears feasible in future mSMART studies. In terms of responses to daily surveys on mSMART, 70% (7/10) of participants responded to all of the questions.

Feasibility was also examined by measuring the time it took participants to complete different tasks on mSMART. Although there is no standardization of scores on these tasks (ie, the number of seconds to complete each task within mSMART), the performance on these tasks can inform treatment development efforts, such as determining whether basic procedures within the smartphone app are understood and can be executed independently [41]. In this sample, no prompts were requested by participants and the majority of tasks (92%, 55/60 tasks completed across the whole sample) were carried out in 10 seconds or less, which indicated that mSMART was a feasible tool for young MSM on PrEP.

Acceptability of mSMART was examined with mixed results. Ratings on a 4-point Likert scale indicated that participants on average "moderately" agreed that mSMART was usable on a daily basis and somewhat less than "moderately" agreed that overall, they were satisfied with mSMART as an intervention to improve PrEP adherence, they would be willing to recommend mSMART to others on PrEP, and it was user-friendly. Difficulty learning how to use mSMART was minimal. SUS indicated that 60% (6/10) of the sample found the mSMART intervention to be usable. Although a sample size of 10 is small, guidelines for SUS indicate that it measures perceived usability of a system with a small sample around this size [59,60].

Qualitative analyses of exit interviews were conducted to complement the quantitative analyses of acceptability and examined aspects of mSMART that could be maintained, discarded, or adapted in future iterations. Although some features of mSMART were generally perceived favorably (eg, use of reminders, the camera feature, and daily questions), participants indicated that even these features could be adapted in future trials. For example, some participants expressed that a "snooze" function or multiple reminder alarms should be added. One particularly notable feedback theme was that mSMART was too text heavy with suggestions to minimize wording and make the display of such wording more visually appealing (eg, in bulleted formatting, as opposed to paragraphs). Although mSMART is a multicomponent intervention (eg, contingency management, behavioral skills training, and use of reminders), one feature that we anticipated to emerge in our exploratory qualitative analysis was for participants to view contingency management favorably. However, other features of mSMART emerged that were more favored than contingency management. Therefore, as mSMART undergoes further development, comparative trials should consider whether mSMART is viewed just as favorably without contingency management. This aspect of our findings pertain to contingency management's acceptability and not contingency management's efficacy. It is unclear whether the magnitude of the reinforcer adopted in this study actually impacted PrEP adherence. In addition, it may be that reinforcer saliency (ie, US $2 for each logged dose) was too low for participants to find engaging and other contingency management approaches may be warranted to improve PrEP acceptability. Although this could come in the form of a higher dollar amount as a reinforcer, the cost of such contingency management approaches may be prohibitive. To reconcile this, studies should consider lower cost contingency management approaches that are engaging (eg, the "fishbowl" technique) [61-63] or a time-limited use of mSMART with higher reinforcer amounts (eg, during PrEP initiation, as was recommended in our qualitative analysis).

Future studies are needed to build on these pilot trial findings. In addition to the factors mentioned above, efficacy trials are needed to examine whether mSMART improves PrEP adherence in comparison to a control group. This would necessitate larger samples that are statistically powered to detect group differences as well as consideration of sample composition (eg, those initiating PrEP or who have struggled with PrEP adherence at baseline, as opposed to the sample in this study in which 90% (9/10) had protective levels at baseline and therefore may have already established adherence habits). Relatedly, although this study examined a group at risk for HIV infection—young MSM—young black MSM are a particularly at-risk group [17]. However, the sample for this study was 70% (7/10) white and did not contain any black MSM, which limits generalizability. Finally, because PrEP use is extending into adolescent MSM, adherence interventions are needed that address unique challenges that emerge in working with this younger age group than those included in this study [18].

In conclusion, this was a phase I trial of a mobile health intervention that aims to improve PrEP adherence. To our knowledge, mSMART is the first PrEP adherence intervention administered via smartphones to integrate contingency management. Given its mobile health format and the ubiquity of smartphone use among younger populations recommended for PrEP [22], this is a PrEP adherence intervention that would be scalable and likely easily disseminated into clinical care settings. In clinical practice, mSMART could be integrated with electronic health records and allow for real-time communication between health care providers and patients. However, although our findings indicate that mSMART is a promising intervention to improve adherence rates, the results are preliminary and future studies are needed to demonstrate efficacy. These studies should also consider our findings indicating areas in which mSMART can be adapted to more comprehensively meet the needs of young MSM prescribed PrEP.

Acknowledgments

This work was supported by the Duke Department of Psychiatry and Behavioral Sciences Research Incentive and Development Program. This publication also resulted in part from research supported by the Duke University Center for AIDS Research, a National Institutes of Health-funded program (5P30 AI064518). In addition, this publication resulted in part from research supported by the University of North Carolina at Chapel Hill Center for AIDS Research, a National Institutes of Health-funded program (P30 AI50410). We thank Rebecca Pratt for assisting with data collection.

Conflicts of Interest

JTM and FJM have served as site principal investigator and coinvestigator on a subcontract from Intelligent Automation Incorporated (IAI) funded by the National Institutes of Health (N44DA132236) to develop mSMART for a different indication (ie, smoking cessation). JTM also served as site principal investigator on a subcontract from Intelligent Automation Incorporated funded under N43DA130022. GS is employed by IAI and represents IAI as mSMART product owner.

Multimedia Appendix 1

Time it took all participants to complete each of the 6 predetermined tasks on mSMART without any prompts during the follow-up study visit.

[PDF File (Adobe PDF File), 45KB - mhealth_v6i9e10456_app1.pdf]

Multimedia Appendix 2

Summary of endorsement rates across domains.

[PDF File (Adobe PDF File), 23KB - mhealth_v6i9e10456_app2.pdf]

References

1. Grant R, Anderson P, McMahan V, Liu A, Amico K, Mehrotra M. Results of the iPrEx open-label extension (iPrEx OLE) in men and transgender women who have sex with men: PrEP uptake, sexual practices, and HIV incidence. Presented at: International AIDS Conference; 2014; Melbourne, Australia.
2. Molina J, Capitant C, Charreau I, Meyer L, Spire B, Pialoux G. On demand PrEP with oral TDF-FTC in MSM: results of the ANRS Ipergay Trial. Presented at: Conference on Retroviruses Opportunistic Infections; 2015; Seattle, WA.
3. Grant RM, Anderson PL, McMahan V, Liu A, Amico KR, Mehrotra M, et al. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. Lancet Infect Dis 2014 Sep;14(9):820-829. [doi: 10.1016/S1473-3099(14)70847-3] [Medline: 25065857]
4. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]
5. Baeten J, Heffron R, Kidoguchi L, Celum C. Near elimination of HIV transmission in a demonstration project of PrEP and ART. Presented at: Conference on Retroviruses Opportunistic Infections; 2015; Seattle, WA.
6. Martin M, Mock PA, Curlin ME, Vanichseni S. Preliminary follow-up of injecting drug users receiving preexposure prophylaxis. Presented at: Conference on Retroviruses Opportunistic Infections; 2015; Seattle, WA.
7. Peterson L, Taylor D, Roddy R, Belai G, Phillips P, Nanda K, et al. Tenofovir disoproxil fumarate for prevention of HIV infection in women: a phase 2, double-blind, randomized, placebo-controlled trial. PLoS Clin Trials 2007 May 25;2(5):e27 [FREE Full text] [doi: 10.1371/journal.pctr.0020027] [Medline: 17525796]
8. McCormack S, Dunn D. Pragmatic open-label randomised trial of preexposure prophylaxis: the PROUD Study. Presented at: Conference on Retroviruses Opportunistic Infections; 2015; Seattle, WA.
9. Grohskopf LA, Chillag KL, Gvetadze R, Liu AY, Thompson M, Mayer KH, et al. Randomized trial of clinical safety of daily oral tenofovir disoproxil fumarate among HIV-uninfected men who have sex with men in the United States. J Acquir Immune Defic Syndr 2013 Sep 1;64(1):79-86. [doi: 10.1097/QAI.0b013e31828ece33] [Medline: 23466649]
10. Baeten JM, Donnell D, Ndase P, Mugo NR, Campbell JD, Wangisi J, et al. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. N Engl J Med 2012 Aug 2;367(5):399-410 [FREE Full text] [doi: 10.1056/NEJMoa1108524] [Medline: 22784037]
11. Choopanya K, Martin M, Suntharasamai P, Sangkum U, Mock PA, Leethochawalit M, et al. Antiretroviral prophylaxis for HIV infection in injecting drug users in Bangkok, Thailand (the Bangkok Tenofovir Study): a randomised, double-blind, placebo-controlled phase 3 trial. Lancet 2013 Jun 15;381(9883):2083-2090. [doi: 10.1016/S0140-6736(13)61127-7] [Medline: 23769234]
12. Marrazzo JM, Ramjee G, Richardson BA, Gomez K, Mgodi N, Nair G, et al. Tenofovir-based preexposure prophylaxis for HIV infection among African women. N Engl J Med 2015 Feb 5;372(6):509-518 [FREE Full text] [doi: 10.1056/NEJMoa1402269] [Medline: 25651245]
13. Thigpen MC, Kebaabetswe PM, Paxton LA, Smith DK, Rose CE, Segolodi TM, et al. Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana. N Engl J Med 2012 Aug 2;367(5):423-434. [doi: 10.1056/NEJMoa1110711] [Medline: 22784038]
14. Van Damme L, Corneli A, Ahmed K, Agot K, Lombaard J, Kapiga S, et al. Preexposure prophylaxis for HIV infection among African women. N Engl J Med 2012 Aug 2;367(5):411-422 [FREE Full text] [doi: 10.1056/NEJMoa1202614] [Medline: 22784040]
15. Corneli AL, Deese J, Wang M, Taylor D, Ahmed K, Agot K, et al. FEM-PrEP: adherence patterns and factors associated with adherence to a daily oral study product for pre-exposure prophylaxis. J Acquir Immune Defic Syndr 2014 Jul 1;66(3):324-331 [FREE Full text] [doi: 10.1097/QAI.0000000000000158] [Medline: 25157647]
16. Amico KR, Stirratt MJ. Adherence to preexposure prophylaxis: current, emerging, and anticipated bases of evidence. Clin Infect Dis 2014 Jul;59 Suppl 1:S55-S60 [FREE Full text] [doi: 10.1093/cid/ciu266] [Medline: 24926036]
17. Centers for Disease Control and Prevention. https://stacks.cdc.gov/view/cdc/37333. Archived at: http://www.webcitation.org/6y6uQgvT4. 2015. Trends in U.S. HIV diagnoses, 2005-2014 URL: https://stacks.cdc.gov/view/cdc/37333 [WebCite Cache ID 6y6uQgvT4]

18. Koss CA, Hosek SG, Bacchetti P, Anderson PL, Liu AY, Horng H, et al. Comparison of measures of adherence to HIV preexposure prophylaxis among adolescent and young men who have sex with men in the United States. Clin Infect Dis 2018 Jan 06;66(2):213-219. [doi: 10.1093/cid/cix755] [Medline: 29020194]
19. Hosek SG, Landovitz RJ, Kapogiannis B, Siberry GK, Rudy B, Rutledge B, et al. Safety and feasibility of antiretroviral preexposure prophylaxis for adolescent men who have sex with men Aged 15 to 17 years in the United States. JAMA Pediatr 2017 Nov 01;171(11):1063-1071. [doi: 10.1001/jamapediatrics.2017.2007] [Medline: 28873128]
20. Mayer KH, Safren SA, Elsesser SA, Psaros C, Tinsley JP, Marzinke M, et al. Optimizing pre-exposure antiretroviral prophylaxis adherence in men who have sex with men: results of a pilot randomized controlled trial of "Life-Steps for PrEP". AIDS Behav 2017 May;21(5):1350-1360. [doi: 10.1007/s10461-016-1606-4] [Medline: 27848089]
21. Miller G. The smartphone psychology manifesto. Perspect Psychol Sci 2012 May;7(3):221-237 [FREE Full text] [doi: 10.1177/1745691612441215] [Medline: 26168460]
22. Pew Research Center. 2017. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-03-23] [WebCite Cache ID 6y8fB1HN3]
23. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
24. Ahmed I, Ahmad NS, Ali S, Ali S, George A, Saleem DH, et al. Medication adherence apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e62 [FREE Full text] [doi: 10.2196/mhealth.6432] [Medline: 29549075]
25. Moore DJ, Jain S, Dubé MP, Daar ES, Sun X, Young J, et al. Randomized controlled trial of daily text messages to support adherence to PrEP in at-risk for HIV individuals: the TAPIR Study. Clin Infect Dis 2017 Nov 24. [doi: 10.1093/cid/cix1055] [Medline: 29228144]
26. Sullivan PS, Driggers R, Stekler JD, Siegler A, Goldenberg T, McDougal SJ, et al. Usability and acceptability of a mobile comprehensive HIV prevention app for men who have sex with men: a pilot study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e26 [FREE Full text] [doi: 10.2196/mhealth.7199] [Medline: 28279949]
27. Available from: https://clinicaltrials.gov/ct2/show/NCT03320512?term=hightow-weidman&rank=3. Archived at: http://www.webcitation.org/6y73X0EQd. P3 (Prepared, Protected, emPowered) (P3) URL: https://clinicaltrials.gov/ct2/show/NCT03320512?term=hightow-weidman&rank=3 [WebCite Cache ID 6y73X0EQd]
28. Petry NM, Rash CJ, Byrne S, Ashraf S, White WB. Financial reinforcers for improving medication adherence: findings from a meta-analysis. Am J Med 2012 Sep;125(9):888-896 [FREE Full text] [doi: 10.1016/j.amjmed.2012.01.003] [Medline: 22800876]
29. Mbuagbaw L, Sivaramalingam B, Navarro T, Hobson N, Keepanasseril A, Wilczynski NJ, Patient Adherence Review (PAR) Team. Interventions for enhancing adherence to antiretroviral therapy (ART): a systematic review of high quality studies. AIDS Patient Care STDS 2015 May;29(5):248-266 [FREE Full text] [doi: 10.1089/apc.2014.0308] [Medline: 25825938]
30. Landovitz RJ, Fletcher JB, Shoptaw S, Reback CJ. Contingency management facilitates the use of postexposure prophylaxis among stimulant-using men who have sex with men. Open Forum Infect Dis 2015 Jan;2(1):ofu114 [FREE Full text] [doi: 10.1093/ofid/ofu114] [Medline: 25884003]
31. Menza TW, Hughes JP, Celum CL, Golden MR. Prediction of HIV acquisition among men who have sex with men. Sex Transm Dis 2009 Sep;36(9):547-555 [FREE Full text] [doi: 10.1097/OLQ.0b013e3181a9cc41] [Medline: 19707108]
32. United States Public Health Service. 2014. Preexposure prophylaxis for the prevention of HIV infection in the United States: clinical practice guideline URL: https://www.cdc.gov/hiv/pdf/guidelines/PrEPguidelines2014.pdf [WebCite Cache ID 6y7AEYDCm]
33. United States Public Health Service. Preexposure prophylaxis for the prevention of HIV infection in the United States: clinical providers supplement. 2014 URL: https/www.cdc.gov/hiv/pdf/guidelines/PrEPProviderSupplement2014.pdf [WebCite Cache ID 6y8IhPpAv]
34. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
35. Skinner HA. The drug abuse screening test. Addict Behav 1982;7(4):363-371. [Medline: 7183189]
36. Saunders JB, Aasland OG, Babor TF, de LFJR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [Medline: 8329970]
37. Hahn SR, Park J, Skinner EP, Yu-Isenberg KS, Weaver MB, Crawford B, et al. Development of the ASK-20 adherence barrier survey. Curr Med Res Opin 2008 Jul;24(7):2127-2138. [doi: 10.1185/03007990802174769] [Medline: 18554431]
38. Intelligent Automation, Inc. mSMART Medication Aide URL: https://msmart.i-a-i.com/ [accessed 2018-03-23] [WebCite Cache ID 6y8giOU9A]
39. Mitchell JT, McIntyre EM, English JS, Dennis MF, Beckham JC, Kollins SH. A pilot trial of mindfulness meditation training for ADHD in adulthood: impact on core symptoms, executive functioning, and emotion dysregulation. J Atten Disord 2017 Nov;21(13):1105-1120. [doi: 10.1177/1087054713513328] [Medline: 24305060]
40. Brooke J. SUS-A quick and dirty usability scale. Usability Eval Ind 1996;189:4-7.

41. Vilardaga R, Rizo J, Kientz JA, McDonell MG, Ries RK, Sobel K. User Experience evaluation of a smoking cessation app in people with serious mental illness. Nicotine Tob Res 2016 May;18(5):1032-1038. [doi: 10.1093/ntr/ntv256] [Medline: 26581430]
42. Adams JL, Sykes C, Menezes P, Prince HMA, Patterson KB, Fransen K, et al. Tenofovir diphosphate and emtricitabine triphosphate concentrations in blood cells compared with isolated peripheral blood mononuclear cells: a new measure of antiretroviral adherence? J Acquir Immune Defic Syndr 2013 Mar 1;62(3):260-266 [FREE Full text] [doi: 10.1097/QAI.0b013e3182794723] [Medline: 23111578]
43. Anderson PL, Glidden DV, Liu A, Buchbinder S, Lama JR, Guanira JV, et al. Emtricitabine-tenofovir concentrations and pre-exposure prophylaxis efficacy in men who have sex with men. Sci Transl Med 2012 Sep 12;4(151):151ra125 [FREE Full text] [doi: 10.1126/scitranslmed.3004006] [Medline: 22972843]
44. Nanda N, Satapathy G, Holt L, McClernon FJ, Mitchell JT. mSMART: personalized mobile app for improving medication adherence. Presented at: 3rd annual mHealth@Duke conference; 2015; Durham, NC.
45. Amico KR, Mansoor LE, Corneli A, Torjesen K, van der Straten A. Adherence support approaches in biomedical HIV prevention trials: experiences, insights and future directions from four multisite prevention trials. AIDS Behav 2013 Jul;17(6):2143-2155 [FREE Full text] [doi: 10.1007/s10461-013-0429-9] [Medline: 23435697]
46. Gilmore HJ, Liu A, Koester KA, Amico KR, McMahan V, Goicochea P, et al. Participant experiences and facilitators and barriers to pill use among men who have sex with men in the iPrEx pre-exposure prophylaxis trial in San Francisco. AIDS Patient Care STDS 2013 Oct;27(10):560-566 [FREE Full text] [doi: 10.1089/apc.2013.0116] [Medline: 24093809]
47. Haberer JE, Bangsberg DR, Baeten JM, Curran K, Koechlin F, Amico KR, et al. Defining success with HIV pre-exposure prophylaxis: a prevention-effective adherence paradigm. AIDS 2015 Jul 17;29(11):1277-1285 [FREE Full text] [doi: 10.1097/QAD.0000000000000647] [Medline: 26103095]
48. Liu AY, Cohen SE, Vittinghoff E, Anderson PL, Doblecki-Lewis S, Bacon O, et al. Preexposure prophylaxis for HIV infection integrated with municipal- and community-based sexual health services. JAMA Intern Med 2016 Jan;176(1):75-84 [FREE Full text] [doi: 10.1001/jamainternmed.2015.4683] [Medline: 26571482]
49. Fisher J, Fisher W. Theoretical approaches to indivdual-level change in HIV risk behavior. In: Peterson JL, DiClemente RJ. editors. Handbook of HIV prevention. 1st ed. New York: Kluwer Academic/Plenum Press; 2000.
50. Aliabadi N, Carballo-Dieguez A, Bakken S, Rojas M, Brown W, Carry M, et al. Using the information-motivation-behavioral skills model to guide the development of an HIV prevention smartphone application for high-risk MSM. AIDS Educ Prev 2015 Dec;27(6):522-537. [doi: 10.1521/aeap.2015.27.6.522] [Medline: 26595265]
51. Sauro J. A practical guide to the system usability scale: background, benchmarks,best practices. Measuring Usability LLC 2011.
52. Mitchell JT, Sweitzer MM, Tunno AM, Kollins SH, McClernon FJ. "I use weed for my ADHD": a qualitative analysis of online forum discussions on cannabis and ADHD. PLoS One 2016;11(5):e0156614 [FREE Full text] [doi: 10.1371/journal.pone.0156614] [Medline: 27227537]
53. Mitchell JT, Weisner TS, Jensen PS, Murray DW, Molina BSG, Arnold EL, et al. How substance users with ADHD perceive the relationship between substance use and emotional functioning. J Atten Disord 2017 Feb 01:1087054716685842. [doi: 10.1177/1087054716685842] [Medline: 28166690]
54. Mayer KH, Chan PA, Flash CA, Krakower DS. Evolving models and ongoing challenges for HIV pre-exposure prophylaxis implementation in the United States. J Acquir Immune Defic Syndr 2018 Feb 01;77(2):119-127. [doi: 10.1097/QAI.0000000000001579] [Medline: 29084044]
55. Kelley CF, Kahle E, Siegler A, Sanchez T, Del RC, Sullivan PS, et al. Applying a PrEP continuum of care for men who have sex with men in Atlanta, Georgia. Clin Infect Dis 2015 Nov 15;61(10):1590-1597. [doi: 10.1093/cid/civ664] [Medline: 26270691]
56. Abaasa A, Hendrix C, Gandhi M, Anderson P, Kamali A, Kibengo F, et al. Utility of different adherence measures for PrEP: patterns and incremental value. AIDS Behav 2018 Apr;22(4):1165-1173 [FREE Full text] [doi: 10.1007/s10461-017-1951-y] [Medline: 29090394]
57. Baker Z, Javanbakht M, Mierzwa S, Pavel C, Lally M, Zimet G, et al. Predictors of over-reporting HIV pre-exposure prophylaxis (PrEP) adherence among young men who have sex with men (YMSM) in self-reported versus biomarker data. AIDS Behav 2017 Oct 27. [doi: 10.1007/s10461-017-1958-4] [Medline: 29079950]
58. Koss CA, Hosek SG, Bacchetti P, Anderson PL, Liu AY, Horng H, et al. Comparison of measures of adherence to human immunodeficiency virus preexposure prophylaxis among adolescent and young men who have sex with men in the United States. Clin Infect Dis 2018 Jan 06;66(2):213-219. [doi: 10.1093/cid/cix755] [Medline: 29020194]
59. Albert W, Tullis T. Measuring the user experience: collecting, analyzing, and presenting usability metrics. 2nd ed. Amsterdam, Netherlands: Morgan Kaufmann; 2013.
60. Tullis T, Stetson J. A comparison of questionnaires for assessing website usability. Presented at: Usability Professionals' Association Conference; 2004; Minneapolis, MN.
61. Petry NM, Martin B. Low-cost contingency management for treating cocaine- and opioid-abusing methadone patients. J Consult Clin Psychol 2002 Apr;70(2):398-405. [Medline: 11952198]

62. Petry NM, Martin B, Cooney JL, Kranzler HR. Give them prizes, and they will come: contingency management for treatment of alcohol dependence. J Consult Clin Psychol 2000 Apr;68(2):250-257. [Medline: 10780125]
63. Petry NM, Martin B, Simcic F. Prize reinforcement contingency management for cocaine dependence: integration with group therapy in a methadone clinic. J Consult Clin Psychol 2005 Apr;73(2):354-359. [doi: 10.1037/0022-006X.73.2.354] [Medline: 15796645]

Abbreviations

ASK-20: 20-item Adherence Starts with Knowledge questionnaire
IMB: Information, Motivation, and Behavioral
MSM: Men who have sex with men
PrEP: Pre-exposure prophylaxis
SUS: System Usability Scale
TFV: tenofovir
TFV-DP: tenofovir-diphosphate

Edited by G Eysenbach; submitted 26.03.18; peer-reviewed by A Corneli, S Buchbinder; comments to author 02.05.18; revised version received 07.06.18; accepted 16.06.18; published 10.09.18.

Please cite as:
Mitchell JT, LeGrand S, Hightow-Weidman LB, McKellar MS, Kashuba ADM, Cottrell M, McLaurin T, Satapathy G, McClernon FJ Smartphone-Based Contingency Management Intervention to Improve Pre-Exposure Prophylaxis Adherence: Pilot Trial
JMIR Mhealth Uhealth 2018;6(9):e10456
URL: http://mhealth.jmir.org/2018/9/e10456/
doi:10.2196/10456
PMID:30201601

Original Paper

Social Media Users' Perception of Telemedicine and mHealth in China: Exploratory Study

Ricky Leung[1*], PhD; Huibin Guo[2*], PhD; Xuan Pan[3], PhD

[1]Department of Health Policy, Management and Behavior, School of Public Health, University at Albany, State University of New York, Rensselaer, NY, United States

[2]School of Economics and Management, Hebei University of Economics and Business, Shijiazhuang, China

[3]Department of Economics and Finance, School of Economics and Management, Tongji University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Ricky Leung, PhD
Department of Health Policy, Management and Behavior
School of Public Health
University at Albany, State University of New York
One University Place, Room 181
Rensselaer, NY, 12144
United States
Phone: 1 5184026512
Fax: 1 5184020414
Email: rleung@albany.edu

Abstract

Background: The use of telemedicine and mHealth has increased rapidly in the People's Republic of China. While telemedicine and mHealth have great potential, wide adoption of this technology depends on how patients, health care providers, and other stakeholders in the Chinese health sector perceive and accept the technology.

Objective: To explore this issue, we aimed to examine a social media platform with a dedicated focus on health information technology and informatics in China. Our goal is to utilize the findings to support further research.

Methods: In this exploratory study, we selected a social media platform—HC3i.cn—to examine the perception of telemedicine and mHealth in China. We performed keyword analysis and analyzed the prevalence and term frequency–inverse document frequency of keywords in the selected social media platform; furthermore, we performed qualitative analysis.

Results: We organized the most prominent 16 keywords from 571 threads into 8 themes: (1) Question versus Answer; (2) Hospital versus Clinic; (3) Market versus Company; (4) Doctor versus Nurse; (5) Family versus Patient; (6) iPad versus Tablet; (7) System versus App; and (8) Security versus Caregiving. Social media participants perceived not only significant opportunities associated with telemedicine and mHealth but also barriers to overcome to realize these opportunities.

Conclusions: We identified interesting issues in this paper by studying a social media platform in China. Among other things, participants in the selected platform raised concerns about quality and costs associated with the provision of telemedicine and mHealth, despite the new technology's great potential to address different issues in the Chinese health sector. The methods applied in this paper have some limitations, and the findings may not be generalizable. We have discussed directions for further research.

(JMIR Mhealth Uhealth 2018;6(9):e181) doi:10.2196/mhealth.7623

KEYWORDS

mHealth; telemedicine; China; social media; text mining; keyword analysis; mobile phone

Introduction

Background

In recent years, the use of telemedicine and mHealth has increased rapidly in the People's Republic of China (hereafter China) [1]. This new technology enables health care providers and patients to meet over the internet and save transportation costs. Research has estimated the Chinese mHealth market to be worth 1.86 billion Chinese Renminbi (RMB), equivalent to

US $271 million, during 2012-2013 [2]. By the end of 2017, the market could reach 10 billion RMB (US $1.46 billion) [2].

In this study, we examined a social media platform with a dedicated focus on health information technology and informatics in China. The platform has attracted subscribers from diverse backgrounds: physicians, patients, entrepreneurs, information technology professionals, and other social groups. Exploring the posts from these subscribers allows for understanding the perceptions regarding telemedicine and mHealth in a particular country. Our goal is to utilize the findings from this study to support further research.

Telemedicine and mHealth in China

First, we provide the definitions of two key terms in this paper. Telemedicine can be defined as provision of medical services from one site to another using electronic communication devices [3,4]. This technology has been developed for several decades. In China, telemedicine was used as early as in the mid-1980s [1], when Chinese physicians consulted patients via telegram; "mHealth" expands telemedicine. It can be defined as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [5]. mHealth devices expand telemedicine because they enable patients to receive care from health providers without using a desktop computer at a physically fixed location [6-8].

China's telecommunication infrastructure is quite mature; thus, it can support the delivery of telemedicine and mHealth services for health providers [2,4-5]. Several large telecommunication networks have been actively involved in providing telemedicine services, including International MedioNet of China and the Golden Health Network [5]. While there are more than 940 million mobile phone and app users in China [9], keen competitions among companies have made voice calls and data usage inexpensive for mobile phone subscribers. For instance, many Chinese mobile phone subscribers use the function of text messaging, either synchronously or asynchronously. Mobile phone carriers charge subscribers approximately US $0.015 for sending a text message [10].

Regarding health service provision, there are significant disparities between rural and urban China [11]. Young, highly qualified physicians prefer staying in urban communities [9-13]. To reduce disparities in health services, Chinese government agencies and hospitals are eager to utilize telemedicine and mHealth technology [14-16]. These organizations also believe that elderly patients can benefit from telemedicine and mHealth in various ways [17]. Nonetheless, many Chinese patients and physicians are still skeptical about the new technology [18]. Moreover, for hospitals and clinics, installation and maintenance costs are associated with the provision of telemedicine and mHealth services [17,19].

Methods

Social Media Platform Selection

In this exploratory study, we selected a social media platform—HC3i.cn—to examine the perception of telemedicine and mHealth in China. Research has shown that there are different relative advantages of examining "generic" (eg, Facebook or QQ) and "specialized" (eg, patientslikeme.com or Dingxiang Yuan) platforms [20]. Our selected platform is specialized, enabling us to focus on collecting relevant data. The major drawback is that it does not necessarily represent the broad concerns of all health stakeholders in China. We have paid attention to this drawback, however, when drawing conclusions from data collected for this paper.

HC3i.cn was founded in 2010. It is the first internet-based platform that focuses on health informatics, internet-based medicine, and mHealth in China. As of now, it has attracted 150,000 registered users, with a monthly average of 8 million page views. Within the larger HC3i platform, there are more than 30 million posts. They are organized by specific forums; we have examined one of them in this paper, and it is entitled "telemedicine or mHealth" (or "*yuan cheng yi liao or yi dong yi liao* " in Chinese). The selection is consistent with our research focus in this paper. As of May 16, 2014 (our research cutoff date), this specific forum had 571 threads with 2811 participation posts (ie, each new thread consisted of 2811/571=4.92 posts on average). The earliest thread was initiated on November 21, 2009, and new posts continued to appear on the research cutoff date (May 16, 2014).

Some researchers have pointed out that robots ("bots") can generate social media content [21]. There are specific algorithms to detect bot activities [21]. Nonetheless, data limitations disallow us to utilize these algorithms. Importantly, bot activities normally do not change the nature of the interactions. They might increase or decrease the magnitude of regular human users' activities to a certain extent. A full investigation of bot activities in the social media forum under our research will be the subject of a subsequent publication.

Keyword Analysis

As a text mining technique, we adopted a modified "bag of words" approach [22] to extract keywords in the selected forum. This approach is appropriate for exploratory research. Under this approach, we treat each post in the selected forum as a "document" that consists of individual words. All the posts that we included in the analysis constitute the corpus of documents. Within the corpus, every word is potentially important and can be regarded as a "keyword." Keyword denotes the importance of a word, defined by frequency or other criteria (eg, a term that leads to many discussions). The frequency criterion indicates the prevalence of a term, which is sometimes called "term frequency (TF) representation." This approach is straightforward and computationally efficient to operate [22].

We follow several data cleaning steps. If the corpus contains English words, we ignore grammar, word order, sentence structure, and punctuation. In English documents, term normalization steps include changing all words to lower case, so that "iPhone," "iphone," "IPHONE" are treated as the same term. Stemming removes suffixes and normalizes tenses and plurals of a word. That is, "play," "playing," and "played" are the same. "Stop-words" that are common but have no substantive meaning are also removed. For example, "and," "the," "of," and other prepositions are typically eliminated in frequency representation. For Chinese texts, additional data

cleaning steps are necessary [22]. Different from English, Chinese sentences are written in a continuous sequence of characters without the use of delimiters such as the blank space. Besides, individual Chinese characters may or may not constitute a word, so the segmentation of Chinese words needs to be predefined with a dictionary or processed with specific techniques that can distinguish syntactic and semantic units [23].

We applied the "dictionary approach" to identify keywords. Two coders on our research team examined the first post of each of the 571 threads using the word count function in Excel and identified a number of keywords (or categories of them) that have substantive meaning. Notably, not all words that appeared frequently were selected (eg, pronouns, prepositions); only words with a substantive meaning were selected for further analysis.

After analyzing the first posts, we increased the sophistication of keyword identification using an additional criterion. That is, while frequency is concerned with the prevalence of a term, we also considered the sparseness of a term as an additional criterion of importance. Following the text mining literature, a term that occurs in every single document within the entire corpus of documents does not necessarily have any distinguishing power, whereas a term that occurs only in a few documents could have strong distinguishability [24]. The sparseness of a term may be used as a factor to increase the weight of a term's frequency. The inverse document frequency (IDF) is used in the literature to capture this weight:

IDF(t)=1+log ((Total number of documents)/(Number of documents containing t))

Combining TF and IDF, we can evaluate each term's term frequency–inverse document frequency (TFIDF) using the following equation:

TFIDF (t, d) = TF(t, d) * IDF (t)

Where t denotes the term (or keyword) and d denotes the document.

Qualitative Analysis

To supplement quantitative keyword analysis, we organized the identified keywords into meaningful themes to facilitate interpretations. In the text mining literature, there are quantitative and qualitative approaches to organize keywords into meaningful groups. Clustering is a quantitative technique that considers the coexistence and correlation of multiple terms. Simply put, keywords that are highly correlated are often used together, and they may constitute a theme [25]. An alternative approach is akin to qualitative research [26] and is appropriate for exploratory research. This approach requires the researcher to "code" text data by categorization and interpretations [26]. It is a common technique in qualitative and ethnographic research, which involves elaborate procedures to ensure consistency and reliability [27,28].

Specifically, we modified the hierarchical procedure to code keywords and organize them into themes [26]. First, two coders identified meaningful keywords from the corpus. This step was facilitated by keyword counting, as reported above. Second, each coder independently selected meaningful pairs (or groups) of keywords to generate a theme. The selection required coders to include a note about the coder's rationale based on knowledge from the literature or experience with the Chinese health sector. Then, the coders were required to crosscheck with each other to reach a consensus and ensure consistency. Based on consensus and practical concerns, the research team determined that 16 prominent keywords with a high frequency of appearance in the selected forum were worthy of greater in-depth analysis and interpretations. These themes facilitated comparisons in the keyword analysis. In the qualitative analysis, the research team further collapsed the themes into 2 higher-level categories in relation to the major research question(s).

Results

Keyword Analysis

The number of participation entries to each of these 571 threads—including the first post and all subsequent "replies"—varies, ranging from 0 to 59 as of our research cutoff date. The most popular thread has attracted 59 participation entries. The first post was titled "Education needed: Right now how many level-3A hospitals in China have used telemedicine?" ("Level-3A" indicates the highest rank in an order developed to represent the level of a Chinese hospital's infrastructure development) [29,30].

Table 1 groups these keywords into 8 themes according to the coding procedures reported above. Thus, we organized the most prominent 16 keywords from 571 threads into 8 themes: (1) Question versus Answer; (2) Hospital versus Clinic; (3) Market versus Company; (4) Doctor versus Nurse; (5) Family versus Patient; (6) iPad versus Tablet; (7) System versus App; and (8) Security versus Caregiving.

Of all first posts, 16.3% (93/571) included a question mark, indicating that many HC3i participants were seeking answers or solutions to some problems; in comparison, 4.9% (28/571) of first posts mentioned "case study," which were typically discussions on a technical or company solution to problems in implementing telemedicine or mHealth technology. Another intriguing pair was "hospital" and "clinic." Results suggested that Chinese stakeholders might still see telemedicine and mHealth technology as used primarily within hospitals. While 10.2% (58/571) of first posts mentioned "hospital or inpatient," only 5.4% (31/571) of first posts mentioned "clinic or outpatient."

Regarding business opportunities, 3.7% (21/571) of first posts mentioned "market" as opposed to 1.6% (9/571) that mentioned "company." In terms of health professionals, 3.9% (22/571) of all first posts mentioned "doctor," but only 1.2% (7/571) mentioned "nurse." In terms of health service recipients, 1.2% (7/571) of all first posts mentioned "patient" and 0.7% (4/571) mentioned "family."

Table 1. Keywords identified in threads.

Number and themes	Frequency, n (%)[a]
1	
Question	93 (16.3)
Answer	28 (4.9)
2	
Hospital	58 (10.2)
Clinic	31 (5.4)
3	
Market	21 (3.7)
Company	9 (1.6)
4	
Doctor	22 (3.9)
Nurse	7 (1.2)
5	
Family	4 (0.7)
Patient	7 (1.2)
6	
iPad	13 (2.3)
Tablet	1 (1.2)
7	
System	64 (11.2)
App	18 (3.2)
8	
Security	7 (1.2)
Caregiving	17 (3.0)

[a]Percentage: Frequency/Number of threads. Number of threads=571.

In terms of technological devices, "iPad" appeared in 2.3% (13/571) of all first posts, whereas "tablet" (participants typically used the word "tablet" to mean an Android-based tablet) appeared in only 1.2% (1/571) of all first posts. A greater percentage of first posts were about how telemedicine and mHealth could work with a health organization's larger technical "system," occupying 11.2% (64/571) of all first posts; in contrast, "app" was mentioned only in 3.2% (18/571) of all first posts. Finally, the comparison between "caregiving" and "security" concerns was 3.0% (17/571) versus 1.2% (7/571).

Term Frequency–Inverse Document Frequency of Keywords From Follow-up Posts

We then examined follow-up posts in the threads generated by the first posts, focusing on threads that had generated at least 20 replies. These replies become the corpus of documents in the subsequent analysis. Using the formula for TFIDF mentioned above, we computed the TFIDF of the same keywords that we used in the previous step to examine the first post. We omitted keywords that either had a lot of missing information or made little sense in the interpretation. Table 2 displays the results.

The corpus of documents consisted of 1977 posts in this analysis. The TF, IDF, and TFIDF of the identified keywords are listed in Table 2, which shows that the list of keywords is generally consistent with those identified in Table 1. TFIDF enables us to recognize the contrast more clearly and facilitate interpretations. First, based on the first 2 keywords, the emphasis on "questions" was about 3 times stronger than that on "answers" (590.3 vs 199.8) in this forum. In terms of utilizing telemedicine, the emphasis was about 14 times stronger on "hospitals" than on "clinics" (660.6 vs 46.5). HC3i participants appeared to pay more attention to how the company can support the development of telemedicine and mHealth than to the market of the new technology, as indicated by a ratio of 8.5 (471.2 vs 55.7) between the keywords "company" and "market."

Table 2. Term frequency–inverse document frequency of selected keywords.

Number and themes	Term frequency–inverse document frequency = (Term frequency × Inverse document frequency)	Term frequency	Inverse document frequency
1			
Question	590.3	150.0	3.9
Answer	199.8	38.0	5.3
2			
Hospital	660.6	173.0	3.8
Clinic	46.5	7.0	6.6
3			
Market	55.7	9.0	6.2
Company	471.2	112.0	4.2
4			
Doctor	458.6	106.0	4.3
Nurse	340.0	73.0	4.7
5			
Family	28.8	4.0	7.2
Patient	203.7	39.0	5.2
6			
iPad	331.4	74.0	4.5
Tablet	433.4	103.0	4.2
7			
System	397.7	91.0	4.4
App	84.3	14.0	6.0
8			
Security	103.6	18.0	5.8
Caregiving	305.7	68.0	4.5

In terms of health providers, the emphasis on “doctor” and “nurse” was similar (458.6 vs 340.0). However, in terms of service recipients, the emphasis was 7 times stronger on “patient” than on “family” (203.7 vs 28.8). It did not seem to matter much for HC3i participants whether the device to deliver telemedicine and mHealth was an “iPad” or “tablet” (331.4 vs 433.4). Participants’ concern for the “system” was much stronger than that for the “app,” as indicated by a ratio of 4.7 (397.7 vs 84.3). Finally, there was stronger emphasis on “caregiving and nursing” than on “safety and security,” as indicated by a ratio of 3.0 (305.7/103.6).

Qualitative Analysis

As mentioned above, we organized the most prominent 16 keywords from 571 threads into 8 themes (see Table 1). These themes often represented how HC3i participants made a practical choice or put greater emphasis between two related items (eg, whether iPad or Android-based tablet is more appropriate for providing mHealth services in China) when they considered telemedicine and mHealth. The following discussions are based on a further collapsing of these themes into 2 higher-level categories regarding the perception of telemedicine and mHealth in China: (1) perceived opportunities and (2) perceived barriers. We selected exemplary quotes and phrases from HC3i participants to support the analysis.

Perceived Opportunities

In terms of opportunities brought forth by telemedicine and mHealth technology, many HC3i participants believed that the new technology would expand services that can be delivered by health professionals, particularly physicians and nurses. Increasing access to health services for patients and families from underserved areas was mentioned by a number of HC3i participants. One participant commented: “You can now stay at home and shop [online]. [And] won’t [it] be great to stay at home and [still can] ‘see’ a doctor?”

Between “quality” and “cost,” participants paid slightly greater attention to quality. New technological possibilities were believed to lead to new lifestyles to remain healthy. For example, a participant suggested that telemedicine and mobile devices could “digitalize the function of human bodies.” Similarly, other HC3i participants suggested that patients or healthy individuals could utilize wearable technologies to record heartbeat, body temperature, sleep conditions, mood, weight

change, and other vital information. Records of this information could help both patients and caregivers better monitor the body, including the recovery progress of individuals suffering from chronic diseases.

In this forum, many participants believed that telemedicine and mHealth technology could help certain companies expand their businesses, but participants were relatively less concerned with the larger market. These participants recognized that telemedicine and mobile devices had a very promising economic prospect if combined with business apps. Several participants pointed out that China—with a population of 1.3 billion—has a huge market for health products. As one participant put it: "Each disease is worth billions [of dollars]." This participant also mentioned that the mobile app Weixin—which has already enrolled millions of Chinese users—could be further explored to build new apps. For this forum participant, "Only if a small proportion of [Weixin] users adopt telemedicine or mobile apps for health purposes, there will be enormous market opportunities [returns]."

Perceived Barriers

Although excellent opportunities were associated with telemedicine, forum participants raised concerns regarding barriers. First, participants were eager to understand how telemedicine might be compatible with other existing technologies. Some participants saw the use of telemedicine and mHealth technology as primarily for diagnosis and treatment within hospitals. In this sense, telemedicine and mHealth technology could improve health care access to Chinese patients only if they could be admitted to a hospital or registered as a hospital outpatient. However, it is difficult for Chinese patients from some regions to find a hospital in the neighborhood of their residence; thus, there can be strong barriers to be admitted to a hospital or registered as an outpatient [31].

Moreover, some forum participants were skeptical about how health professionals could actually apply telemedicine or mobile devices in practice. Some labeled telemedicine and mobile devices as "an idea only," "toys for doctors," or used other unfavorable descriptions. The slightly more optimistic participants called this new technology "a plausible model of health delivery," but they still remained worried that health professionals might not feel comfortable with different computing interfaces. For them, if physicians and nurses did not fully embrace telemedicine, the delivery of health care cannot be satisfactory. One of these participants wrote: "There is some use with a tablet to provide health care, but mostly for high-end medical applications. For a nurse, this is not really useful. What they [nurses] need is a movable cart for medical devices with remote capabilities, not an entire mHealth system. Many [mHealth systems] are just products of engineers' imagination."

Other threads indicated forum participants' concerns regarding interoperability among telemedicine, mHealth devices, and existing health information technologys. According to these participants, interoperability issues could generate large startup costs to lay the infrastructure for implementing telemedicine; thus, the new technology might not see quick returns to investment. For example, using new mobile devices required barcode scanning, interfacing using radiofrequency identification, existing electronic health records, and the like. One participant was concerned whether a convenient mobile device could be built: "putting everything together makes [the device] ugly [and cumbersome]."

Another group of participants was concerned about standardizing telemedicine and mHealth. In this respect, participants pointed out that the concepts of telemedicine and mHealth are too broad and there is no universal standard to govern this new technology. Among other things, it made it difficult for health professionals and patients to ensure data security. One participant saw the application of telemedicine devices as going between different health care settings—such as between a physician's office and a patients' room. When a physician is too busy, he or she could leave the telemedicine-related tablet at insecure places; thus, sensitive information about patients, such as their diagnosis and treatment plans, might fall into inappropriate hands.

Finally, participants raised concerns about the amount and type of human resources needed to support telemedicine devices and systems. For example, some participants had worked in blood banks within hospitals where some mobile devices were used. According to these participants, in such a crowded and busy environment, it was unclear who was accountable for operating the mobile devices properly. The use of these mobile devices was, therefore, prone to frequent errors and could generate undue stress among staff.

Discussion

Examining social media data is a relatively new methodology in the growing field of health analytics [32-36]. In this paper, we have identified a number of interesting issues from a social media platform on telemedicine, mHealth, and related technology. Several observations are worthy of further investigation: first, many participants of the selected social media platform still saw telemedicine as something used primarily within hospitals [23]. Although some participants mentioned that patients could receive consultation at home by physicians from afar, this did not represent the general view of forum participants. Therefore, developers need to increase the technical soundness of new devices to convince patients and physicians that telemedicine and mHealth can be used "outside" hospitals and that new devices can really deliver the same level of high-quality health care as face-to-face consultations. In future research, it will be useful to compare China with other developed and less developed countries.

As our findings show, there was a strong concern with respect to interoperability among new telemedicine, mHealth, and existing technology. Understandably, health providers did not want to "start everything anew." Forum participants often linked telemedicine and mHealth to familiar and established technology firms, such as Google, Microsoft, and Apple. Further research can examine whether these firms have a significant market advantage over smaller companies when promoting telemedicine and mHealth products in China [8]. If that were the case, smaller companies, hospitals, and clinics may need to incur significant costs to develop telemedicine and mHealth before they can

realize cost savings for patients in the current Chinese market [37,38].

Finally, we acknowledge the limitations of this research; furthermore, the generalizability of our findings may be limited. First, the delimitation of Chinese sentences differs from the English ones in several ways. One approach is to segment the sentence into words in Chinese language processing by character-based sequence labeling. Existing algorithms such as Viterbi might be further explored [20]. Although our approach of utilizing a predefined set of words is feasible, it is not necessarily the most efficient and precise detection method. As mentioned, it is important to examine whether bots are used to generate contents on any social media platform [39], and more sophisticated research is needed in this regard.

Conflicts of Interest

None declared.

References

1. Wang Z, Gu H. A review of telemedicine in China. J Telemed Telecare 2009;15(1):23-27. [doi: 10.1258/jtt.2008.080508] [Medline: 19139216]
2. Yu X, Han H, Du J, Wei L, Li C, Zhang X, et al. mHealth in China and the United States: How Mobile Technology is Transforming Healthcare in the World's Two Largest Economies. Washington, DC: Center for Technology Innovation at Brookings; 2014.
3. American Telemedicine Association. Telemedicine defined. 2018. URL: http://www.americantelemed.org/main/about/about-telemedicine# [accessed 2018-08-31] [WebCite Cache ID 725Q73kc9]
4. Leung R. Determinants of Telemedicine Utilization in Rural America: Application of the Dynamic Capability Theory. Journal of Information Technology Research 2013;6(2):46-59.
5. Kay M, Santos J, Takane M. mHealth: New horizons for health through mobile technologies. Geneva, Switzerland: World Health Organization; 2011.
6. Anker SD, Koehler F, Abraham WT. Telemedicine and remote management of patients with heart failure. Lancet 2011 Aug 20;378(9792):731-739. [doi: 10.1016/S0140-6736(11)61229-4] [Medline: 21856487]
7. Tachakra S, Wang XH, Istepanian RSH, Song YH. Mobile e-health: the unwired evolution of telemedicine. Telemed J E Health 2003;9(3):247-257. [doi: 10.1089/153056203322502632] [Medline: 14611692]
8. Xu Z, Chen Z, Nie H. Handheld Computersmartphone-Centric Wireless Applications. Microwave Magazine, IEEE 2014;15:36-44.
9. Van Velthoven MH, Li Y, Wang W, Du X, Wu Q, Chen L, et al. mHealth Series: mHealth project in Zhao County, rural China-Description of objectives, field site and methods. Journal of global health 2013;3(2):1-22.
10. Yu X, Han H, Du J. mHealth in Chinathe United States. Washington, D. C. Washington, DC: Center for Technology Innovation, Brookings Institute; 2014.
11. Yip W, Hsiao W, Chen W, Hu S, Ma J, Maynard A. Early appraisal of China's huge and complex health-care reforms. The Lancet 2012;379:833-842.
12. Liu Y. Development of the rural health insurance system in China. Health Policy Plan 2004 May;19(3):159-165. [Medline: 15070864]
13. Simpson P. China's Urban Population Exceeds Rural for First Time Ever. The Telegraph 2012 Jan 17 [FREE Full text]
14. Corpman D. Mobile health in China: a review of research and programs in medical care, health education, and public health. Journal of health communication 2013;18(11):1345-1367.
15. Guo Y, Shibuya K, Cheng G, Rao K, Lee L, Tang S. Tracking China's health reform. Lancet 2010 Mar 27;375(9720):1056-1058. [doi: 10.1016/S0140-6736(10)60397-2] [Medline: 20346797]
16. Guo X, Sun Y, Wang N, Peng Z, Yan Z. The dark side of elderly acceptance of preventive mobile health services in China. Electronic Markets 2013;23:49-61.
17. Brown CL. Health Care Data Protection and Biometric Authentication Policies: Comparative Culture and Technology Acceptance in China and in the United States. Review of Policy Research 2012;29:141-159.
18. Xue Y, Liang H. Analysis of telemedicine diffusion: the case of China. IEEE Trans Inf Technol Biomed 2007 Mar;11(2):231-233. [Medline: 17390993]
19. Mao Y, Zhang Y, Zhai S. Mobile phone text messaging for pharmaceutical care in a hospital in China. J Telemed Telecare 2008;14(8):410-414. [doi: 10.1258/jtt.2008.080406] [Medline: 19047450]
20. Kaplan A. If you love something, let it go mobile: Mobile marketing and mobile social media 4x4. Business horizons 2012;55(2):129-139.
21. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151.
22. Provost F, Fawcett T. Representing and Mining Text. In: Data Science for Business: What you need to know about data mining and data-analytic thinking. Sebastopol, CA: O'Reilly Media, Inc; 2013:251-278.
23. Peng F, Feng F, McCallum A. Chinese segmentation and new word detection using conditional random fields. Computer Science Department Faculty Publication Series 2004;92.

24. Lin S, Ho J. Discovering informative content blocks from Web documents. In: Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining. 2002 Presented at: ACM SIGKDD international conference on knowledge discovery and data mining; 2002; Edmonton, AB, Canada p. 588-593.
25. Berry M. Survey of text mining. Computing Reviews 2004;45(9):548.
26. Yu C, Jannasch-Pennell A, DiGangi S. Compatibility between text mining and qualitative research in the perspectives of grounded theory, content analysis, and reliability. The Qualitative Report 2011;16(3):730.
27. Emerson R, Fretz R, Shaw L. Writing ethnographic fieldnotes. Chicago, IL: University of Chicago Press; 2011.
28. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
29. Peng F, Kurnia S. Understanding Hospital Information Systems Adoption in China. In: PACIS 2010 Proceedings. 2010 Presented at: PACIS 2010 Proceedings; July 9-12, 2010; Taipei, Taiwan p. 66.
30. Tang C, Zhang Y, Chen L, Lin Y. The growth of private hospitals and their health workforce in China: a comparison with public hospitals. Health policy and planning 2014;29(1):30-41.
31. Jing W, Otten H, Sullivan L, Lovell-Simons L, Granek-Catarivas M, Fritzsche K. Improving the doctor-patient relationship in China: the role of balint groups. The International Journal of Psychiatry in Medicine 2013;46(4):417-427.
32. Barbosa L, Feng J, Torres V. Systems and Methods for Social Media Data Mining. Mountain View, CA: Google Patents; 2012. URL: https://patents.google.com/patent/US20120047219A1/en [accessed 2018-08-31] [WebCite Cache ID 725TIBJwl]
33. Zeng D, Chen H, Lusch R, Li SH. Social media analytics and intelligence. IEEE Intelligent Systems 2010;25(6):13.
34. Leung R, Pasupathy K. The economics of social computingome preliminary findings on healthcare organizations. Journal of Computational Science 2011;2:253-261.
35. Hansen D, Shneiderman B, Smith M. Analyzing social media networks with NodeXL: Insights from a connected world. Burlington, MA: Morgan Kaufmann; 2010.
36. Kaplan A. If you love something, let it go mobile: Mobile marketing and mobile social media 4x4. Business Horizons 2012;55:129-139.
37. Men L, Tsai WHS. How companies cultivate relationships with publics on social network sites: Evidence from China and the United States. Public Relations Review 2012;38(5):723-730.
38. Frew SE, Sammut SM, Shore AF, Ramjist JK, Al-Bader S, Rezaie R, et al. Chinese health biotech and the billion-patient market. Nat Biotechnol 2008 Jan;26(1):37-53. [doi: 10.1038/nbt0108-37] [Medline: 18183014]
39. Davis C, Varol O, Ferrara E, Flammini A, Menczer F. BotOrNot: A system to evaluate social bots. In: Proceedings of the 25th International Conference Companion on World Wide Web. 2016 Presented at: 25th International Conference Companion on World Wide Web; 2016; Montreal, Canada p. 273-274.

Abbreviations

IDF: inverse document frequency
RMB: Chinese Renminbi
TF: term frequency
TFIDF: term frequency–inverse document frequency

Edited by G Eysenbach; submitted 01.03.17; peer-reviewed by E Cheng, E Sezgin; comments to author 09.08.17; revised version received 01.05.18; accepted 21.06.18; published 25.09.18.

Please cite as:
Leung R, Guo H, Pan X
Social Media Users' Perception of Telemedicine and mHealth in China: Exploratory Study
JMIR Mhealth Uhealth 2018;6(9):e181
URL: http://mhealth.jmir.org/2018/9/e181/
doi:10.2196/mhealth.7623
PMID:30274969

Original Paper

An mHealth App for Self-Management of Chronic Lower Back Pain (Limbr): Pilot Study

Aliza Selter[1], MA; Christina Tsangouri[1], BEng; Sana B Ali[2], MS, MPH; Diana Freed[1], MA; Adrian Vatchinsky[1], MEng; James Kizer[1], MEng; Arnaud Sahuguet[1], PhD; Deneen Vojta[3], MD; Vijay Vad[4], MD; JP Pollak[1], PhD; Deborah Estrin[1], PhD

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

[2]Healthix Inc, New York, NY, United States

[3]UnitedHealth Group Research & Development, Minnetonka, MN, United States

[4]Hospital for Special Surgery, New York, NY, United States

Corresponding Author:
JP Pollak, PhD
Cornell Tech
2 W Loop Rd
New York, NY,
United States
Phone: 1 646 971 3716
Email: jpp9@cornell.edu

Abstract

Background: Although mobile health (mHealth) interventions can help improve outcomes among patients with chronic lower back pain (CLBP), many available mHealth apps offer content that is not evidence based. Limbr was designed to enhance self-management of CLBP by packaging self-directed rehabilitation tutorial videos, visual self-report tools, remote health coach support, and activity tracking into a suite of mobile phone apps, including Your Activities of Daily Living, an image-based tool for quantifying pain-related disability.

Objective: The aim is to (1) describe patient engagement with the Limbr program, (2) describe patient-perceived utility of the Limbr program, and (3) assess the validity of the Your Activities of Daily Living module for quantifying functional status among patients with CLBP.

Methods: This was a single-arm trial utilizing a convenience sample of 93 adult patients with discogenic back pain who visited a single physiatrist from January 2016 to February 2017. Eligible patients were enrolled in 3-month physical therapy program and received the Limbr mobile phone app suite for iOS or Android. The program included three daily visual self-reports to assess pain, activity level, and medication/coping mechanisms; rehabilitation video tutorials; passive activity-level measurement; and chat-based health coaching. Patient characteristics, patient engagement, and perceived utility were analyzed descriptively. Associations between participant characteristics and program interaction were analyzed using multiple linear regression. Associations between Your Activities of Daily Living and Oswestry Disability Index (ODI) assessments were examined using Pearson correlation and hierarchical linear modeling.

Results: A total of 93 participants were enrolled; of these, 35 (38%) completed the program (age: mean 46, SD 16 years; female: 22/35, 63%). More than half of completers finished assessments at least every 3 days and 70% (19/27) used the rehabilitation component at least once a week. Among respondents to a Web-based feedback survey, 76% (16/21) found the daily notifications helped them remember to complete their exercises, 81% (17/21) found the system easy to use, and 62% (13/21) rated their overall experience good or excellent. Baseline Your Activities of Daily Living score was a significant predictor of baseline ODI score, with ODI increasing by 0.30 units for every 1-unit increase in Your Activities of Daily Living (P<.001). Similarly, hierarchical linear modeling analysis indicated that Your Activities of Daily Living daily assessment scores were significant predictors of ODI scores over the course of the study (P=.01).

Conclusions: Engagement among participants who completed the Limbr program was high, and program utility was rated positively by most respondents. Your Activities of Daily Living was significantly associated with ODI scores, supporting the validity of this novel tool. Future studies should assess the effect of Limbr on clinical outcomes, evaluate its use among a wider patient sample, and explore strategies for reducing attrition.

Trial Registration: ClinicalTrials.gov NCT03040310; https://clinicaltrials.gov/ct2/show/NCT03040310 (Archived by WebCite at http://www.webcitation.org/722mEvAiv)

(JMIR Mhealth Uhealth 2018;6(9):e179) doi:10.2196/mhealth.8256

KEYWORDS

low back pain; chronic disease; self-assessment; telemedicine; self-management; activities of daily living; pain; rehabilitation

Introduction

Management of chronic conditions places a considerable burden on patients, communities, and health care systems worldwide [1], but evidence indicates that symptom management in chronic disease can be significantly improved through self-management interventions [2,3]. Given that mobile phone usage in the United States has become widespread in recent years and is still on the rise [4], advancements in mobile technology can be leveraged to deliver mobile health (mHealth) apps that support patients in effective self-management of chronic conditions. In a recent US study of mHealth use among primary care patients, 55% of respondents reported having a mobile phone and 70% of these had used mHealth apps for management of health conditions [5].

Conditions for which exercise therapy has been shown to be effective, such as chronic lower back pain (CLBP) [6], stand to benefit greatly from mHealth integration because sustained adherence to exercise-based rehabilitation is vital for recovery [7-9]. The effectiveness of mobile phone-based interventions for measuring and influencing physical activity has been explored in a number of studies, and there is increasing evidence that mHealth interventions that are adaptive to user preference while supplementing standard care with disease monitoring, self-reporting, education, and promoting physical therapy adherence have the potential to improve health outcomes among those living with chronic diseases [1,10,11]. Support from a health coach has also been shown to help drive mHealth app use [10,12], and remote health coaching in the form of text messages can effectively improve self-management of symptoms and promote long-term behavior change retention [13,14], including increased compliance with physical therapy [15]. Ecological momentary assessment, or "experience sampling"—including self-report surveys and sensor-assisted reminders—are effective tools for collecting in situ user data [16,17] and can be used to enhance mHealth interventions for the self-management of CLBP.

Limbr is a compliance enhancement intervention that was developed to incorporate many of these elements by packaging self-directed rehabilitation tutorial videos, personalizable visual self-report tools, health coach support, and sensor-assisted, passive activity-level tracking into a suite of mobile phone-based apps for patients with CLBP. The Limbr program aims to promote adherence to the Back Rx exercise rehabilitation regimen [18], increase engagement in self-directed management of pain (including pain, medication, and exercise tracking), and improve self-reported outcomes for pain. One novel aspect of Limbr is its use of Your Activities of Daily Living, an image-based tool for characterizing functional status [19]. Although a recent preliminary evaluation of Your Activities of Daily Living conducted among a small number of patients with arthritis suggested promise of its utility [19], it has not yet been evaluated among a larger patient group.

Despite the existence of numerous patient-targeted mobile phone apps for pain tracking, self-management, and exercise training, the implementation of mHealth technology for chronic conditions remains an important area for further research. In particular, many currently available mobile phone apps targeted at low back pain (LBP) management are low in quality, offering content that is not based on current research and has not been reviewed or tested by health care providers [20,21]. This study was performed to (1) describe patient engagement with the Limbr program, (2) describe the patient-perceived utility of the Limbr program, and (3) assess the validity of the Your Activities of Daily Living module as a quantifier of pain and disability level among patients with CLBP.

Methods

Study Design

This was a single-arm trial utilizing a convenience sample of 93 adult patients who visited physiatrist Vijay Vad, MD (New York, NY, USA), from January 2016 through February 2017 and were diagnosed with discogenic back pain. Included patients were required to be English speaking and to have a diagnosis of LBP with predominantly axial symptoms, persistence of symptoms for at least 3 months, lumbar intervertebral disk pathology evident on magnetic resonance imaging, and possession of a mobile phone device (iPhone models 5S and later or Android models 2.3 and later). Patients were excluded if they had a history of trauma, a history of lumbar spine surgery or severe lumbar disk degeneration prior to the beginning of the study, or concurrent pathology that could have contributed to axial low back symptoms (eg, spondylolysis, spondylolisthesis, facet arthropathy), or if their case involved a legal claim. Informed consent was obtained from patients at the initial doctor visit (onboarding) after the nature of the study had been explained. The study is registered on ClinicalTrials.gov (identifier NCT03040310), was approved by the institutional review board of the Hospital for Special Surgery (New York, NY), and was conducted in accordance with all applicable regulations.

Intervention

Eligible patients were enrolled in an mHealth-based 3-month physical therapy program (Limbr) and received a mobile phone app suite free of charge to monitor and manage their CLBP. The program included three daily visual self-reports to assess pain, medication/coping mechanisms, and affect; self-directed rehabilitation via Back Rx video tutorials personalized for patients with discogenic back pain; and passive measurement

of activity levels. At onboarding, patients underwent baseline assessments (including a Your Activities of Daily Living full assessment and an Oswestry Disability Index [ODI] [22] assessment) and received assistance with installation and setup of the mobile phone app suite. For the duration of the program, patients received remote support from a health coach available in real time, and several patient engagement methods were utilized to improve user compliance. Elements of Limbr are described in further detail subsequently.

Self-Reports

The daily visual self-reports—Your Activities of Daily Living [19], Medications of Daily Living, and the Photographic Affect Meter (PAM) [23]—used an experience sampling approach to collect in situ user data for the purpose of providing tailored content to users. Rather than relying on words or numbers, these questionnaires offer a variety of photographs from which the user selects those that best describe their mood/condition.

Your Activities of Daily Living (Figure 1) [19] is an image-based survey inspired by the PAM [23] that characterizes a patient's functional status using images representing activities of daily living (ADL) from the Western Ontario and McMaster Universities Arthritis Index [24] and Boston Activity Measure for Post-Acute Care [25], which are validated clinical measures. To complete the Your Activities of Daily Living assessment, patients used the app to select images of activities during which they recently experienced LBP-induced difficulty. The full assessment conducted at onboarding included 47 images and was intended as a substitute for a conventional, clinician-administered long-form ADL questionnaire (eg, the ODI). The Your Activities of Daily Living daily assessment was intended to provide interim reports between time points at which a long-form assessment would typically be administered and included only the images selected by the patient during the baseline full assessment.

Medications of Daily Living (Figure 2) is an app-based medication log with a visual interface. Similar to Your Activities of Daily Living, patients completing the daily Medications of Daily Living assessment by choosing images that characterized any LBP medication or coping strategies used over the past 24 hours. All medications and coping strategies included in Medications of Daily Living had been confirmed by the study physician as relating to LBP.

PAM (Figure 3), used to assess patients' daily affect, is a rigorously validated tool for measuring emotion through a series of images [23]. Photos in the PAM are arranged in a grid from low arousal and negative valence in the bottom left, to high arousal and positive valence in the top right. To complete the daily PAM assessment, patients used the app to choose the image that best represented their emotion at the time of assessment.

Self-Directed Rehabilitation

The self-directed rehabilitation component of Limbr was administered via Force Therapeutics [26], an app providing a series of exercise videos tailored to patients with LBP. Patients were requested to watch three times per week. Patients used the app to view videos and to indicate if they watched the videos.

Activity-Level Measurement

Activity levels, including personal location and activity classification information (eg, minutes active per day, hours out of the house), were monitored via Moves [27], an app that utilizes mobile device sensors for passive collection of activity-level data.

Figure 1. Screenshots from the Your Activities of Daily Living app: (A) daily assessment, (B) daily assessment reminders, and (C) when there was nothing to report, patients selected "Today was a good day!".

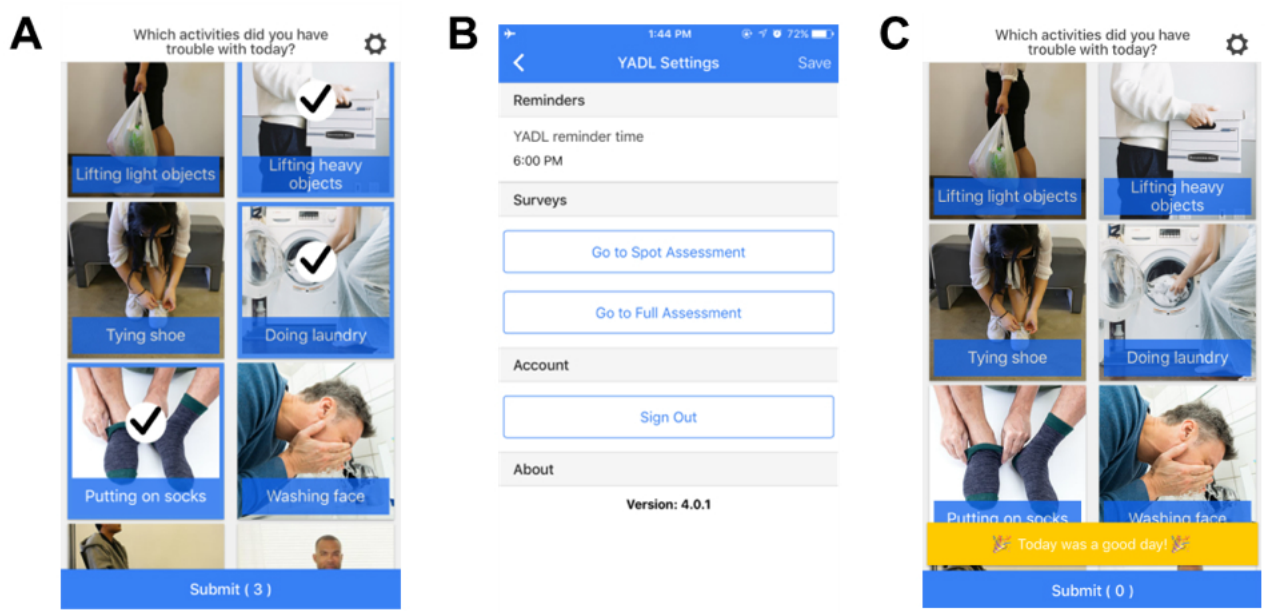


Figure 2. Screenshots from the Medications of Daily Living app: (A) daily assessment, (B) daily assessment reminders, and (C) when there was nothing to report, patients selected "Today was a good day!".

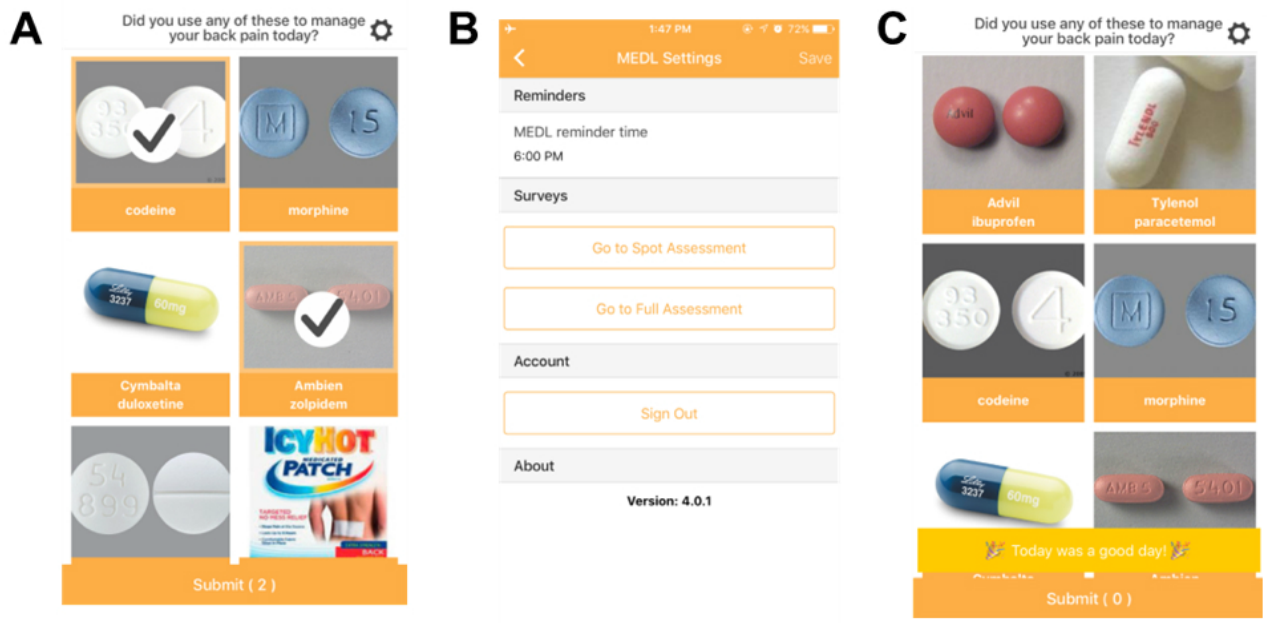
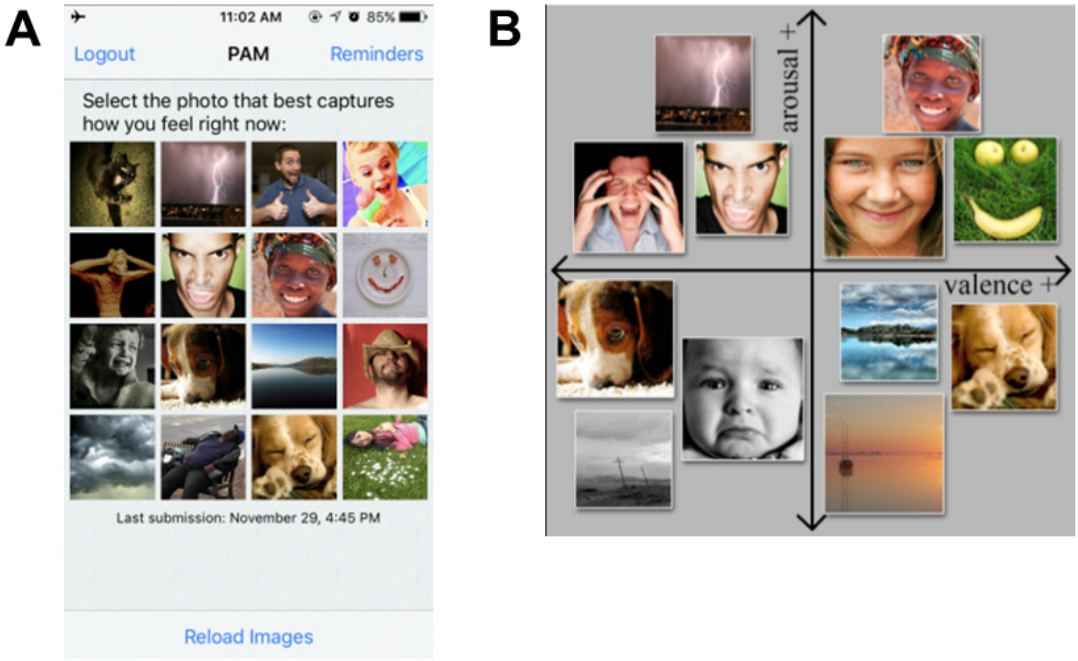


Figure 3. The Photographic Affect Meter (PAM) app. (A) Screenshot of the PAM visual interface, and (B) how images are arranged from low arousal and negative valence in the bottom left to high arousal and positive valence in the top right.



Health Coach Support

Data-informed health coaching with a certified health coach was made available via Limbr Chat, a text-messaging app. (Standard iOS and Android messaging apps could not be used to maintain compliance with the Health Insurance Portability and Accountability Act of 1996 [28].) The coach in this study was familiar with the Force Therapeutics exercises and advised patients about exercise, technical issues, and personal support. The coach monitored participant data from the Limbr suite, including daily self-reports, indicators of participant compliance, and activity levels; identified trends in participants' progress to provide personalized care; and used the Limbr Chat app to send responses and other messages, including support messages and reminders to interact with the program. The coach used casual language and abbreviations typical of text messaging (eg, "u" instead of "you") to promote an informal relationship and reduce intimidation on the part of participants, and any patient messages were responded to within 24 hours.

Patient Engagement

During the study, Limbr participants were categorized according to their level of engagement with the program for the purpose of tracking and improving compliance. Patient categorization

was updated three times a week on the basis of the frequency and quality of engagement with the interactive components of the system (watching videos or completing the visual self-reports). Patients were categorized as "frequently interacting" (>2 interactive components/week), "infrequently interacting" (<1 interactive component/week), and "unproductive-active" (1-2 interactive components/week).

To promote sustained engagement, all frequently interacting and unproductive-active participants were sent weekly summary emails (Figure 4) consisting of a visual feedback regarding their interactions across the Your Activities of Daily Living and Medications of Daily Living visual self-reports, self-directed rehabilitation (Force Therapeutics), and passively collected mobility data. To encourage engagement among infrequently interacting participants, an email was sent reminding them to engage with the Limbr components and log their activities. Unproductive-active patients were sent personalized messages from the Limbr health coach, who worked with the patients to construct a new care plan according to individual patient needs. For example, the coach might suggest that a patient reduce interaction with the daily assessments from daily to three times per week. Finally, the Limbr health coach would check in weekly with all frequently interacting, unproductive-active, and infrequently interacting patients, inquire about their progress, and send personalized motivational messages. Examples of messages that might be sent to encourage patient engagement and/or check in regarding a patient's progress are:

> *Good morning, I know it can seem like we are asking u to track a lot of things and data. Honestly, nothing is more important to your healthy outcome than the FORCE exercises. I do them myself. Please don't forget to do them—1 set, 3 times per week AND mark them all done in the FORCE app. Thank you and good health.*

> *I am glad the nights are getting better. Sometimes the lack of movement can stiffen the body. Series B is best thought of as the next level up shall we say on the exercises. You should keep doing the exercises you have been doing until they seem too easy, then contact Dr Vad to get his permission to move to Series B.*

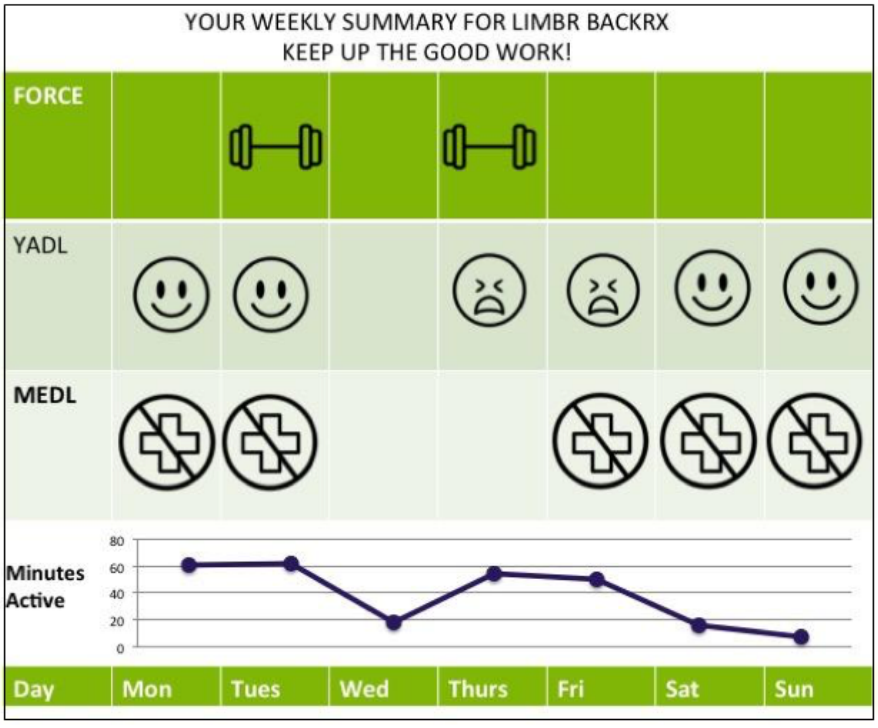
Patients were categorized as inactive if they had had no interaction with the interactive components for 4 successive weeks. Participants were said to have completed the program if they remained active for 12 weeks.

Outcomes

Patient Engagement

Patient engagement was assessed using three outcome variables: (1) the frequency of interactions across the visual self-reports, (2) a binary outcome representing at least one viewing of the physical therapy videos versus none watched, and (3) the frequency of messages to the health coach. For outcome analysis, an interaction was defined as an instance of using Your Activities of Daily Living, Medications of Daily Living, PAM, or the Limbr Chat app during the study period. The percentage of physical therapy videos watched was automatically collected by the Force Therapeutics app. Frequency of messages was computed as the number of times a participant sent a message to the coach using the Limbr Chat app during the study period.

Figure 4. Example of a weekly summary email.



Patient-Perceived Utility of Limbr

The overall utility of the Limbr program was assessed using a Web-based feedback survey administered to all participants at the completion of the study. The feedback survey consisted of 13 questions, presented on a 5-point Likert scale, and divided into three sections that assessed the perceived helpfulness of (1) the patient engagement features (Limbr Chat app and weekly summary emails), (2) the app notifications reminding users to complete the daily self-reports and Back Rx exercises, and (3) the visual self-reports for Your Activities of Daily Living, Medications of Daily Living, and PAM. The response options for each section ranged from "strongly disagree/not useful" (5 points) to "strongly agree/very useful" (1 point).

Association of Your Activities of Daily Living with Conventional Pain Assessment

To determine whether the Your Activities of Daily Living visual self-report could serve as a proxy for a more traditional pain index, outcomes from the Your Activities of Daily Living assessment were compared at baseline with those from the ODI, a questionnaire that measures levels of disability in ADLs among patients rehabilitating from LBP [22]. Participants were directed to complete the ODI at onboarding (baseline) and at 2 weeks, 6 weeks, and 3 months after enrollment. The ODI was completed via Ohmage [29], a mobile survey app utilized for recording, analyzing, and visualizing participant data and administering clinical surveys.

Statistical Analysis

Baseline patient characteristics, patient engagement, and patient-perceived utility of the Limbr program were analyzed descriptively; means and standard deviations were provided for continuous variables, and numbers and percentages were provided for discrete variables. Associations between participant characteristics and level of interaction with Limbr (total interactions and interactions per week) were analyzed using multiple linear regression to determine whether characteristics (either collectively or individually) had any effect on use of the Limbr system.

To analyze the association between Your Activities of Daily Living and ODI assessment results at baseline, Pearson correlation coefficient was calculated after confirming normality using the Shapiro-Wilk test. In addition, hierarchical linear modeling (HLM) was used to analyze the ability of Your Activities of Daily Living daily self-reports to predict ODI scores among participants who both entered multiple ODI scores and completed the full Limbr program. For the HLM analysis, the outcome was the ODI score reported on a particular day and the predictor variable was the Your Activities of Daily Living score reported closest in time to that ODI score.

Results

Patient Characteristics

A total of 93 participants were enrolled from January 2016 through February 2017, of which 13 dropped out after completing the onboarding session and before interacting with the Limbr components. Of the remaining 80 participants, an additional 45 dropped out before completing the 3-month study duration. This left 35 participants (age: mean 46, SD 16 years; female: 22/35, 63%; duration of symptoms: mean 19.6, SD 7.4 months; Table 1) who remained enrolled and used the Limbr interactive components throughout the study.

Patient characteristics did not have a significant collective effect on the total number of program interactions ($F_{6,27}$=2.183, P=.08, R^2=.177) or the number of program interactions per week ($F_{6,27}$=2.337, P=.06, R^2=.1956). However, age and duration of symptoms were each individual predictors of total interactions (t_{27}=3.128, P=.004 and t_{27}=–2.258, P=.03, respectively) and interactions per week (t_{27}=3.159, P=.004 and t_{27}=–2.355, P=.03, respectively; Table 2).

Outcomes

Patient Engagement

The 35 participants who completed the program averaged 96 total interactions with the three daily self-reports over the 12-week study, roughly evenly distributed among Your Activities of Daily Living, Medications of Daily Living, and PAM. The number of interactions per week ranged from 1 to 29, with a mean of 8 (SD 7). On average, participants who interacted with the daily assessments between daily and every 3 days comprised slightly over 50% of the study group. Median participant interaction frequency across assessments is depicted in Figure 5. Participants were instructed to watch the Force Therapeutics instructional videos only 1 to 3 times a week, as opposed to daily. Interaction data shows that 70% (19/27) of participants interacted with Force Therapeutics a median of at least once a week (Figure 6).

Table 1. Patient characteristics (N=35).

Characteristic	Completing participants
Age (years), mean (SD)	46 (16)
Female, n (%)	22 (63)
Body mass index, mean (SD)	25.4 (4.0)
Duration of symptoms (months), mean (SD)	19.6 (7.4)
Mobile phone operating system, n (%)	
Android	4 (11)
iOS	31 (89)

Table 2. Associations between participant characteristics and program interaction measures (N=35).

Variable	Total interactions[a]		Interactions per week[a]	
	t_{27}	*P* value	t_{27}	*P* value
Age, years	3.128	.004	3.159	.004
Male sex	–1.167	.25	–0.997	.33
Body mass index	0.509	.63	0.702	.49
Duration of symptoms	–2.258	.03	–2.355	.03
iOS mobile phone operating system	1.197	.24	1.279	.21

[a]An interaction was defined as an instance of using Your Activities of Daily Living, Medications of Daily Living, PAM, or the Limbr Chat app during the study period.

Figure 5. Median interaction frequency across daily self-reports for Your Activities of Daily Living (YADL), Medications of Daily Living (MEDL), and the Photographic Affect Meter (PAM).

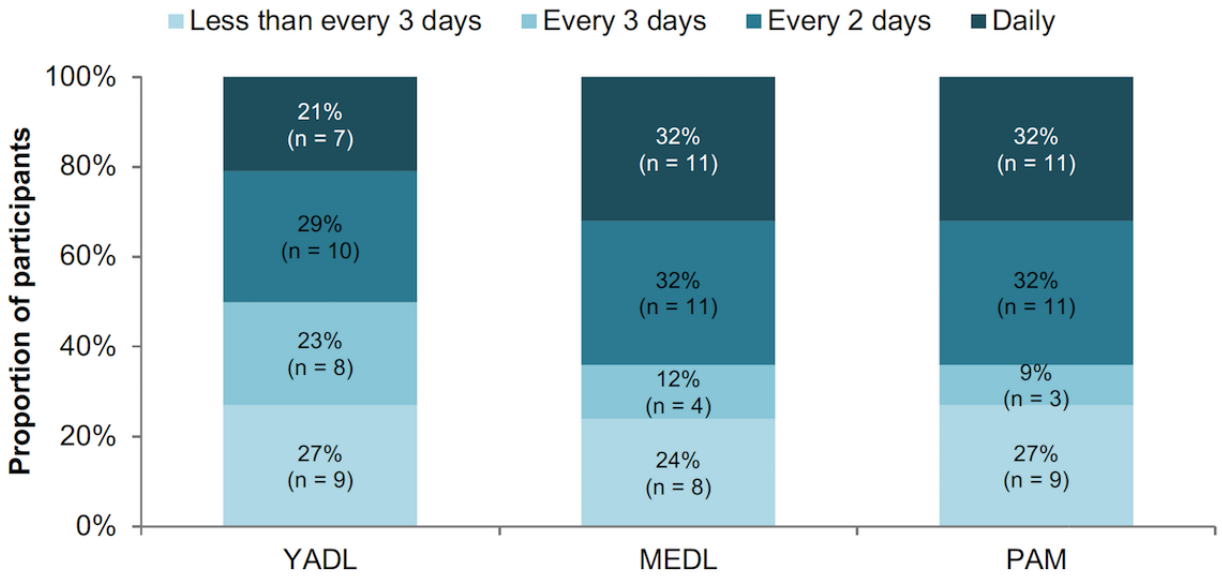


Figure 6. Median interaction frequency for daily self-reports for Force Therapeutics. Note there were different frequencies of data reported from Force Therapeutics versus the other assessments.

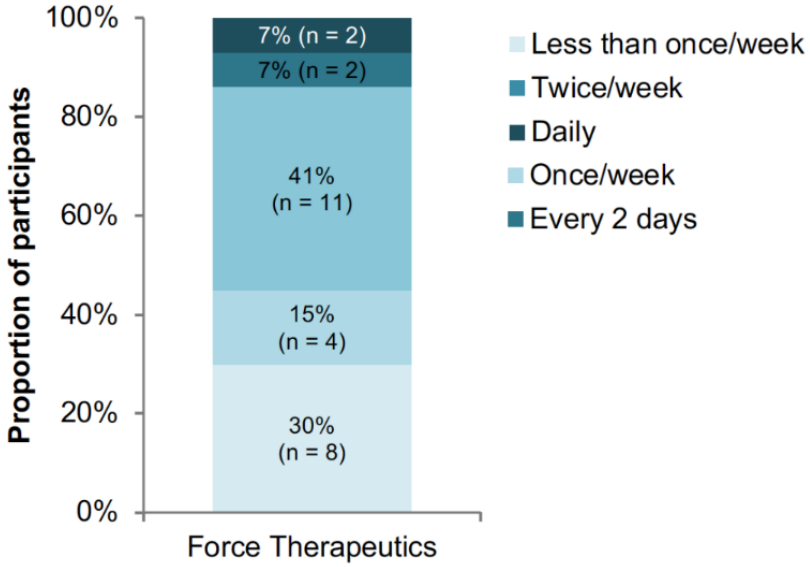
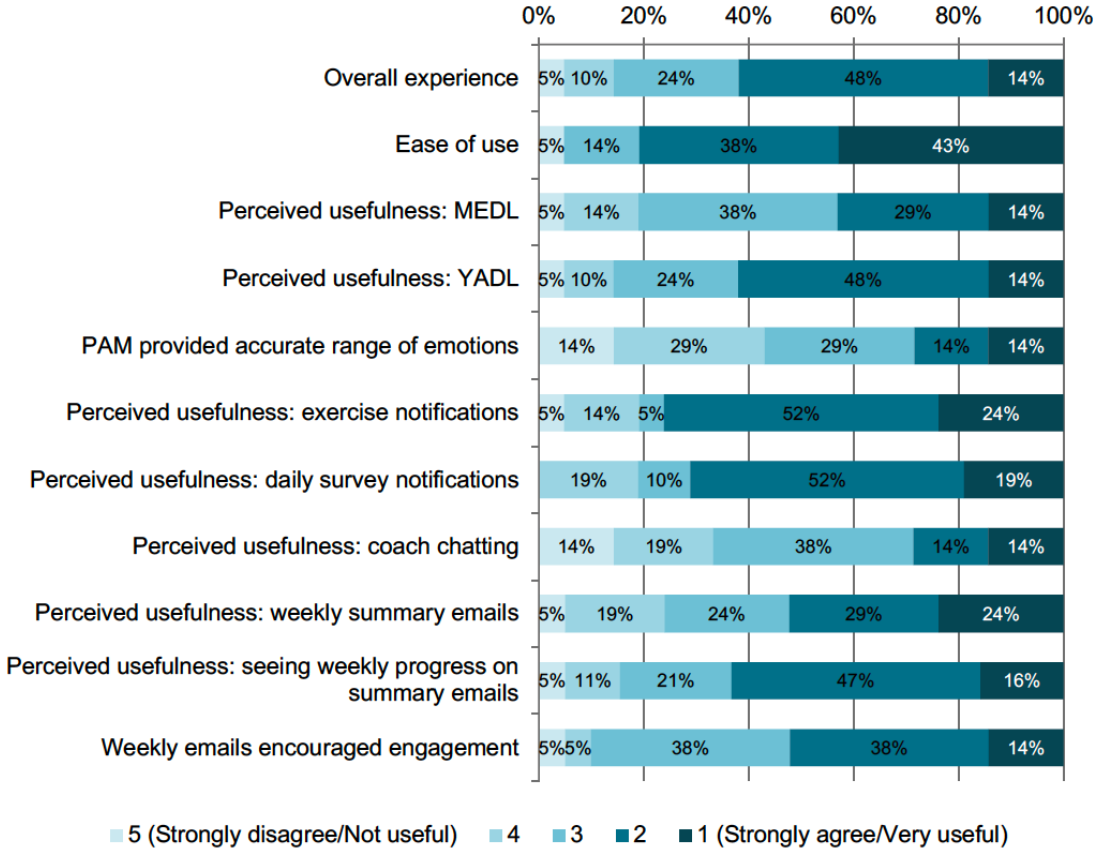


Figure 7. Feedback survey results. Surveys were scored on a 5-point Likert scale with response options ranging from "strongly disagree/not useful" (5 points) to "strongly agree/very useful" (1 point). Percentages may not sum to 100% because of rounding. YADL: Your Activities of Daily Living; MEDL: Medications of Daily Living; PAM: Photographic Affect Meter.



A total of 147 messages were sent from participants to the coach using the Limbr Chat app. The majority of these (73/147, 49.7%) were tech support messages (eg, "I cannot seem to log in to force. Can you help?"), with the next most common type (47/147, 32.0%) comprising messages about the Force Therapeutics exercises (eg, "Yes my knee reaches. Not the hip flexor. The hip flexor is tight and I feel pain. So was wondering if that's normal and if not then I shouldn't stretch that much."). A smaller percentage (23/147, 15.6%) were medical messages (eg, "Thank you. I have had extreme back pain for the last 2 days. I hurts to sit, sleep, and bend at the moment"), and 2.7% of messages (4/147) recorded participant criticisms of the system.

Patient-Perceived Utility of Limbr

Feedback surveys were returned by 21 participants; a question-by-question breakdown of participant responses in presented in Figure 7. Among respondents, 11 of 21 (52%) found the daily self-reports to be helpful in tracking pain-related ADL functionality, medication use, and affect. In particular, 13 of 21 (62%) found that the Your Activities of Daily Living daily assessment helped them track the activities of daily living affected by their back pain; 16 of 21 (76%) and 15 of 21 (71%) agreed that the daily notifications were helpful in reminding them to complete the Force Therapeutics exercises and daily surveys, respectively; 17 of 21 (81%) found the Limbr system easy to use; and 13 of 21 (62%) rated their overall experience as either good or excellent.

Association of Your Activities of Daily Living With Conventional Pain Assessment

Baseline Your Activities of Daily Living and ODI scores were found to be significantly associated (Pearson correlation coefficient=.551, P<.001). Linear regression modeling further revealed that the baseline Your Activities of Daily Living score was a significant predictor of baseline ODI score, with ODI increasing by 0.30 units for every 1-unit increase in Your Activities of Daily Living (P<.001). Similarly, HLM analysis (among the 14 patients with multiple ODI scores who completed the full Limbr program) indicated that Your Activities of Daily Living daily assessment scores were significant predictors of ODI scores recorded over the course of the study, with ODI increasing by 0.33 units for every 1-unit increase in Your Activities of Daily Living (P=.01).

Discussion

In this pilot study conducted among patients with CLBP, engagement with the Limbr compliance enhancement intervention was high among those who finished the program; the majority of completers interacted with the daily self-reports multiple times per week and 70% used the self-directed rehabilitation component as directed. Approximately half of feedback survey respondents found the daily self-report components of Limbr to be helpful, and more than 70% indicated that the daily notifications helped them remember to perform their rehabilitation exercises. Moreover, the Your

Activities of Daily Living assessment was found to be significantly associated with conventional pain assessment scores, thereby validating its utility as a novel quantifier of pain and disability level. These findings suggest that Limbr has substantial potential as an approach to promoting patient engagement and self-directed rehabilitation adherence for CLBP management.

Although US data regarding mHealth interventions for patients with CLBP are limited, evidence exists that Web- or mobile-based strategies can be effective for reducing pain and improving self-management in this population [30-32]. In this study, participants who completed the Limbr program exhibited a high level of engagement throughout the trial, frequently interacting with the self-report modules as well as with Force Therapeutics. It is particularly encouraging that most respondents found Limbr helpful in remembering to engage with self-directed rehabilitation, as rehabilitation adherence is critical for maximum improvement in physical function among patients with CLBP [7-9]. In addition, the sustained use of Your Activities of Daily Living, Medications of Daily Living, and PAM observed throughout the trial suggests that patients may find these visual assessments—which provide a simple and intuitive interface, can be completed quickly, and are readily adapted to mobile devices—to be easier to use than conventional reporting methods such as text-based surveys [19,23].

The significant association of the Your Activities of Daily Living assessment with the ODI scores is a key finding of this study. As a visual survey, Your Activities of Daily Living leverages the inherent ambiguities of images to mitigate some of the limitations of conventional pain assessments [19]. For example, although it is not possible for a standardized survey to contain a comprehensive list of every ADL that could be relevant for every patient, the images in Your Activities of Daily Living are open to individual interpretation and, therefore, can be used by different patients to express a wider range of ADL experiences [19]. Furthermore, standard ADL assessments are typically performed in an office setting in association with a clinical encounter, limiting their ability to reflect day-to-day variability in ADL performance. In contrast, Your Activities of Daily Living is mobile-friendly and can be completed by patients at any time without assistance, greatly increasing the scope and granularity of the ADL information that can be captured. Validation of Your Activities of Daily Living opens the door to the creation of mHealth apps that are capable of reliably measuring patient-reported pain outcomes on a day-to-day basis. The relationship between Your Activities of Daily Living daily assessments and pain-related disability as assessed by the ODI should be explored in further studies.

Participant attrition in this study was high; of 93 patients enrolled, only 35 (38%) completed the 3-month program. Although this level of attrition is substantial, it is not unusual among mHealth interventions, for which the challenge of many participants discontinuing the intervention and/or being lost to follow-up is widely acknowledged [33]. For example, only 32 of 180 (18%) study enrollees were retained after 12 weeks in one recent analysis of a multidisciplinary LBP pain treatment app, despite the fact that those who completed the program experienced significant reductions in pain [34]. Similarly, only 25% of participants in a 1-year internet-mediated exercise intervention for patients with CLBP maintained at least 80% compliance with required data uploads for the duration of the study [32]. Excellent participant retention was observed in the 4-month FitBack randomized controlled trial, which demonstrated greater pain reduction among program users versus those in comparison groups. Of 597 initial enrollees, 580 (97%) submitted assessments at all three designated time points [30]. It is noteworthy, however, that FitBack participants received cash rewards for submitting assessments, and the degree to which those who submitted all three assessments engaged with the intervention program on a daily or weekly basis (eg, tracking pain and pain-management activities or watching instructional videos) is unknown [30]. Furthermore, FitBack is a comparatively simple Web-based program in contrast to Limbr, which required installation, maintenance, and utilization of seven separate component apps. The complexity of the combined interactions and maintenance tasks required of Limbr participants may have been a factor in the high dropout rate. As nearly half of the chat messages sent by participants in this study were categorized as tech support-related, the possibility that technical difficulties contributed to attrition also cannot be discounted.

Considerable effort was made to improve patient engagement and reduce attrition during the course of this study. First, patient input from feedback surveys and calls was used to improve usability and enhance the user experience of the daily self-reports; for instance, two of the reports (Your Activities of Daily Living and Medications of Daily Living) were updated and moved from beta testing to the app store. Before this change was made, beta testing expirations frequently required patients to manually update their apps, and there was a significant difference in device usage between iOS and Android phones. After the update, however, there was no difference in usage between operating systems (data not shown). This example highlights the importance of making the patient experience as seamless as possible to promote engagement with the intervention.

The other major effort to enhance patient interaction comprised personalized Limbr Chat messages from the health coach, which were tailored according to participant data collected via the Limbr suite. A post hoc analysis revealed that 408 health coach messages were sent, with the largest proportion categorized as engagement (169/408, 41.4%), followed by technical support (113/408, 27.6%), and medical/exercise-related messages (60/408, 14.7%). Findings from previous studies suggest that interactions with a health coach, including two-way messaging systems similar to that used in Limbr [35], tend to increase patient engagement with the intervention [35] and promote improved self-management of chronic pain conditions [36]. However, of the messages sent from patients to the coach, only 74 of 147 were unrelated to technical support. Although we expected more patient messaging, Limbr did little to encourage patients to engage in a two-way exchange or spontaneously send messages to the coach. Messages from the coach rarely prompted the patient to respond, as the coach was not instructed to do so, and the system itself did not actively guide the patient to the chat app unless there was a new message waiting. The

low number of patient-sent messages is not necessarily an indicator of poor engagement, because patients who were highly engaged in the exercising and reporting may have felt little need for communication with the coach. Nevertheless, the effectiveness of various types of health coach messaging is an important area for future research.

Despite the relatively high attrition rate, the overall utility of the Limbr system was scored positively by the majority of respondents and some individual components were widely found to be useful; for example, the weekly summary emails were met with optimum positive feedback. On the other hand, reception of other Limbr components varied considerably across participants. Daily notifications and recurrent coach messages were particularly polarizing, seen as vexatious by some participants but highly motivating by others. This finding underscores the need for mHealth interventions to take a personalized approach to engagement rather than relying on a single method of promoting compliance. Additional studies to characterize which engagement efforts are most effective both overall and for particular patient populations could help enable the design of highly personalizable interventions in the future. Other changes that could reduce attrition and enhance engagement in a future iteration of the Limbr system include combining the disparate components into a single app designed with user-first principles and fewer technical barriers, conducting more formal user testing and employing analytical techniques such as conversion funnel optimization prior to considering a larger trial and emphasizing two-way messaging between patient and coach.

Our results should be considered in light of several limitations. Because the trial was conducted among a small convenience sample of patients with CLBP, the applicability of the study findings to other patient populations is unknown and should be assessed in future studies. In addition, because there was no comparison group, the study outcomes cannot be definitively attributed to use of the Limbr suite. Finally, although the reasons underlying the high dropout rate of the trial warrant further exploration, this study was not designed to analyze the causes of patient attrition or the types of engagement efforts that were most helpful in promoting compliance with the intervention.

The findings of this pilot study suggest that the Limbr program shows promise as an approach to enhancing patient self-management and adherence to self-directed rehabilitation for CLBP. Engagement among participants who completed the program was high, and the utility of the program was rated positively by the majority of respondents. Our results also support the validity of the Your Activities of Daily Living visual self-assessment for quantifying pain and disability level. Future studies should assess the effect of Limbr on clinical outcomes, evaluate its use among a wider patient sample, and explore strategies for reducing attrition.

Acknowledgments

This research was supported with funds from The National Science Foundation (award #1344587), The National Institutes of Health MD2K (award #5-40312), and gift funding from UnitedHealth Group. Medical writing support was provided by Yvette Edmonds, PhD, an employee of Optum (Eden Prairie, MN). The authors would like to thank the Foundry at Cornell Tech for the development and upkeep of the apps used in the intervention; Jiangtao Gou, PhD, for assistance with the HLM analysis; Force Therapeutics for providing the mobile Back Rx program; and Hospital for Special Services for recruitment of participants. The authors also wish to acknowledge the following for their contributions to this study: Sasinda Rukshan, Sreshta Paranji, Nick Payne, Glenn Emanuel, and Longqi Yang.

Conflicts of Interest

None declared.

References

1. 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;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
2. Coleman MT, Newton KS. Supporting self-management in patients with chronic illness. Am Fam Physician 2005 Oct 15;72(8):1503-1510 [FREE Full text] [Medline: 16273817]
3. Turner A, Anderson JK, Wallace LM, Bourne C. An evaluation of a self-management program for patients with long-term conditions. Patient Educ Couns 2015 Feb;98(2):213-219 [FREE Full text] [doi: 10.1016/j.pec.2014.08.022] [Medline: 25441096]
4. Pew Research Center. 2018 Feb 05. Mobile fact sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [WebCite Cache ID 70B4RLXQz]
5. Bauer AM, Rue T, Keppel GA, Cole AM, Baldwin L, Katon W. Use of mobile health (mHealth) tools by primary care patients in the WWAMI region Practice and Research Network (WPRN). J Am Board Fam Med 2014;27(6):780-788 [FREE Full text] [doi: 10.3122/jabfm.2014.06.140108] [Medline: 25381075]
6. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, Lancet Low Back Pain Series Working Group. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018 Jun 09;391(10137):2368-2383. [doi: 10.1016/S0140-6736(18)30489-6] [Medline: 29573872]

7. Mannion AF, Helbling D, Pulkovski N, Sprott H. Spinal segmental stabilisation exercises for chronic low back pain: programme adherence and its influence on clinical outcome. Eur Spine J 2009 Dec;18(12):1881-1891 [FREE Full text] [doi: 10.1007/s00586-009-1093-7] [Medline: 19609785]
8. World Health Organization. 2003. Adherence to long-term therapies: evidence for action URL: http://www.who.int/chp/knowledge/publications/adherence_full_report.pdf?ua=1 [WebCite Cache ID 72Qb4kWhD]
9. Cecchi F, Pasquini G, Paperini A, Boni R, Castagnoli C, Pistritto S, et al. Predictors of response to exercise therapy for chronic low back pain: result of a prospective study with one year follow-up. Eur J Phys Rehabil Med 2014 Apr;50(2):143-151 [FREE Full text] [Medline: 24429917]
10. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
11. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
12. Alley S, Jennings C, Plotnikoff RC, Vandelanotte C. Web-based video-coaching to assist an automated computer-tailored physical activity intervention for inactive adults: a randomized controlled trial. J Med Internet Res 2016 Aug 12;18(8):e223 [FREE Full text] [doi: 10.2196/jmir.5664] [Medline: 27520283]
13. Job JR, Spark LC, Fjeldsoe BS, Eakin EG, Reeves MM. Women's perceptions of participation in an extended contact text message-based weight loss intervention: an explorative study. JMIR Mhealth Uhealth 2017 Feb 27;5(2):e21 [FREE Full text] [doi: 10.2196/mhealth.6325] [Medline: 28242595]
14. Spark LC, Fjeldsoe BS, Eakin EG, Reeves MM. Efficacy of a text message-delivered extended contact intervention on maintenance of weight loss, physical activity, and dietary behavior change. JMIR Mhealth Uhealth 2015;3(3):e88 [FREE Full text] [doi: 10.2196/mhealth.4114] [Medline: 26373696]
15. Lilje SC, Olander E, Berglund J, Skillgate E, Anderberg P. Experiences of older adults with mobile phone text messaging as reminders of home exercises after specialized manual therapy for recurrent low back pain: a qualitative study. JMIR Mhealth Uhealth 2017 Mar 30;5(3):e39 [FREE Full text] [doi: 10.2196/mhealth.7184] [Medline: 28360026]
16. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [Medline: 12173681]
17. Froehlich J, Chen M, Consolvo S, Harrison B, Landay J. MyExperience: a system for in situ tracing and capturing of user feedack on mobile phones. 2007 Presented at: 5th International Conference on Mobile Systems, Applications and Services; June 11-13, 2007; San Juan, Puerto Rico.
18. Vad VB, Bhat AL, Tarabichi Y. The role of the Back Rx exercise program in diskogenic low back pain: a prospective randomized trial. Arch Phys Med Rehabil 2007 May;88(5):577-582. [doi: 10.1016/j.apmr.2007.02.008] [Medline: 17466725]
19. Yang L, Freed D, Wu A, Wu J, Pollak J, Estrin D. Your Activities of Daily Living (YADL): An image-based survey technique for patients with arthritis. 2016 Presented at: The EAI International Conference on Pervasive Computing Technologies for Healthcare; May 16-19, 2016; Cancun, Mexico.
20. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a pain app for that": review of patient-targeted smartphone applications for pain management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
21. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016 Dec;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]
22. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine (Phila Pa 1976) 2000 Nov 15;25(22):2940-2952; discussion 2952. [Medline: 11074683]
23. Pollak J, Adams P, Gay G. PAM: a photographic affect meter for frequent, in situ measurement of affect. 2011 Presented at: ACM CHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC.
24. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis Rheum 2001 Oct;45(5):453-461. [Medline: 11642645]
25. Jette DU, Stilphen M, Ranganathan VK, Passek S, Frost FS, Jette AM. Interrater Reliability of AM-PAC "6-Clicks" Basic Mobility and Daily Activity Short Forms. Phys Ther 2015 May;95(5):758-766. [doi: 10.2522/ptj.20140174] [Medline: 25504489]
26. Force Therapeutics. 2017. URL: https://www.forcetherapeutics.com/ [WebCite Cache ID 70B4kwg5Z]
27. ProtoGeo Oy. 2017. Moves: activity diary of your life URL: http://moves-app.com/ [accessed 2018-06-14] [WebCite Cache ID 70B4s1WX2]
28. HHS.gov. 2013 Jul 26. Summary of the HIPAA privacy rule URL: https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html [accessed 2018-06-14] [WebCite Cache ID 70B4wtnbL]
29. Ramanathan N, Alquaddoomi F, Falaki H, George D, Hsieh C, Jenkins J, et al. Ohmage: an open mobile system for activity and experience sampling. 2012 Presented at: 6th International Conference on Pervasive Computing Technologies for Healthcare; May 21-24, 2012; San Diego, CA.

30. Irvine AB, Russell H, Manocchia M, Mino DE, Cox GT, Morgan R, et al. Mobile-Web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
31. Chiauzzi E, Pujol LA, Wood M, Bond K, Black R, Yiu E, et al. painACTION-back pain: a self-management website for people with chronic back pain. Pain Med 2010 Jul;11(7):1044-1058. [doi: 10.1111/j.1526-4637.2010.00879.x] [Medline: 20545873]
32. Krein SL, Kadri R, Hughes M, Kerr EA, Piette JD, Holleman R, et al. Pedometer-based internet-mediated intervention for adults with chronic low back pain: randomized controlled trial. J Med Internet Res 2013 Aug 19;15(8):e181 [FREE Full text] [doi: 10.2196/jmir.2605] [Medline: 23969029]
33. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
34. Huber S, Priebe JA, Baumann K, Plidschun A, Schiessl C, Tölle TR. Treatment of low back pain with a digital multidisciplinary pain treatment app: short-term results. JMIR Rehabil Assist Technol 2017 Dec 04;4(2):e11 [FREE Full text] [doi: 10.2196/rehab.9032] [Medline: 29203460]
35. Jamison RN, Jurcik DC, Edwards RR, Huang C, Ross EL. A pilot comparison of a smartphone app with or without 2-way messaging among chronic pain patients: who benefits from a pain app? Clin J Pain 2017 Aug;33(8):676-686 [FREE Full text] [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]
36. Allen KD, Oddone EZ, Coffman CJ, Datta SK, Juntilla KA, Lindquist JH, et al. Telephone-based self-management of osteoarthritis: a randomized trial. Ann Intern Med 2010 Nov 2;153(9):570-579. [doi: 10.7326/0003-4819-153-9-201011020-00006] [Medline: 21041576]

Abbreviations

ADL: activities of daily living
CLBP: chronic low back pain
HLM: hierarchical linear modeling
LBP: low back pain
mHealth: mobile health
ODI: Oswestry Disability Index
PAM: Photographic Affect Meter

Edited by G Eysenbach; submitted 21.06.17; peer-reviewed by L Whitehead, JH Lee, P Anderberg; comments to author 22.11.17; revised version received 15.07.18; accepted 14.08.18; published 17.09.18.

Please cite as:
Selter A, Tsangouri C, Ali SB, Freed D, Vatchinsky A, Kizer J, Sahuguet A, Vojta D, Vad V, Pollak JP, Estrin D
An mHealth App for Self-Management of Chronic Lower Back Pain (Limbr): Pilot Study
JMIR Mhealth Uhealth 2018;6(9):e179
URL: http://mhealth.jmir.org/2018/9/e179/
doi:10.2196/mhealth.8256
PMID:30224333

Original Paper

The SPLENDID Eating Detection Sensor: Development and Feasibility Study

Janet van den Boer[1], PhD; Annemiek van der Lee[1], MSc; Lingchuan Zhou[2], PhD; Vasileios Papapanagiotou[3], MSc; Christos Diou[3], PhD; Anastasios Delopoulos[3], PhD; Monica Mars[1], PhD

[1]Sensory Science and Eating Behaviour Chair Group, Division of Human Nutrition, Wageningen University, Wageningen, Netherlands

[2]Electronics & Firmware, Systems Division, Centre Suisse d'Electronique et de Microtechnique, Neuchâtel, Switzerland

[3]Multimedia Understanding Group, Department of Electrical and Computer Engineering, Aristotle University, Thessaloniki, Greece

Corresponding Author:
Monica Mars, PhD
Sensory Science and Eating Behaviour Chair Group
Division of Human Nutrition
Wageningen University
Stippeneng 4
Wageningen, 6708WE
Netherlands
Phone: 31 317 485 340
Email: monica.mars@wur.nl

Abstract

Background: The available methods for monitoring food intake—which for a great part rely on self-report—often provide biased and incomplete data. Currently, no good technological solutions are available. Hence, the SPLENDID eating detection sensor (an ear-worn device with an air microphone and a photoplethysmogram [PPG] sensor) was developed to enable complete and objective measurements of eating events. The technical performance of this device has been described before. To date, literature is lacking a description of how such a device is perceived and experienced by potential users.

Objective: The objective of our study was to explore how potential users perceive and experience the SPLENDID eating detection sensor.

Methods: Potential users evaluated the eating detection sensor at different stages of its development: (1) At the start, 12 health professionals (eg, dieticians, personal trainers) were interviewed and a focus group was held with 5 potential end users to find out their thoughts on the concept of the eating detection sensor. (2) Then, preliminary prototypes of the eating detection sensor were tested in a laboratory setting where 23 young adults reported their experiences. (3) Next, the first wearable version of the eating detection sensor was tested in a semicontrolled study where 22 young, overweight adults used the sensor on 2 separate days (from lunch till dinner) and reported their experiences. (4) The final version of the sensor was tested in a 4-week feasibility study by 20 young, overweight adults who reported their experiences.

Results: Throughout all the development stages, most individuals were enthusiastic about the eating detection sensor. However, it was stressed multiple times that it was critical that the device be discreet and comfortable to wear for a longer period. In the final study, the eating detection sensor received an average grade of 3.7 for wearer comfort on a scale of 1 to 10. Moreover, experienced discomfort was the main reason for wearing the eating detection sensor <2 hours a day. The participants reported having used the eating detection sensor on 19/28 instructed days on average.

Conclusions: The SPLENDID eating detection sensor, which uses an air microphone and a PPG sensor, is a promising new device that can facilitate the collection of reliable food intake data, as shown by its technical potential. Potential users are enthusiastic, but to be successful wearer comfort and discreetness of the device need to be improved.

(JMIR Mhealth Uhealth 2018;6(9):e170) doi:10.2196/mhealth.9781

KEYWORDS

chewing sensor; weight management; obesity prevention; overweight; PPG sensor; in-ear microphone; mobile phone

Introduction

Background

The available methods for monitoring food intake—which for a great part rely on self-report—often provide biased and incomplete data [1-5]. Depending on the exact method used, they require people to eat consciously, be knowledgeable about what they eat, be able to estimate portion size, and remember all that information. As a result, these methods are prone to underreporting. It is common for people to report an unrealistically low energy intake, that is, an energy intake that is too low to sustain their body at a low level of physical activity [6-9]. Current technological advances have enabled the development of tools that can facilitate the collection of reliable food intake data.

Currently, some devices are available that can be used to increase the reliability of food intake monitoring. The Mandometer, for example, can be used to measure the size of meals. It is a weighing scale that is placed underneath the plate during a meal [10]. Furthermore, a number of wearable devices have been developed that can automatically detect eating [11-14]. These are mostly ear- and neck-worn devices. They use sensors (eg, a microphone or strain sensor) to collect signals that contain information on whether or not a person is eating. Pattern-recognition algorithms are used to extract this information.

In particular, devices that can detect eating events have the potential to reduce underreporting. Such a device can take away the need for people to be conscious about their eating. Moreover, this information can be used to prompt people to report what they are eating at the moment they are eating it. It can, therefore, also take away the need for people to remember what they ate. However, there is not yet a device for the automatic detection of eating that is practical for everyday use, despite the progress made in this area. Such devices, for example, require people to accurately position a sensor on the body with tape or require people to wear items like glasses or a hat to carry the functional parts [13,15,16].

Development of the SPLENDID Eating Detection Sensor

Within the context of SPLENDID, an information and communications technology project funded by the European Union [17,18], we aimed to take the next step and develop a device for the automatic detection of eating events that is practical for everyday use. It was decided to create an ear-worn device as this was expected to be acceptable for young, overweight adults, which was our primary target group. In the future, such a device could be incorporated into other devices the target group is already using, such as earphones used for listening to music. Moreover, such a device could be appropriate for a wider population.

The eating detection sensor was built using an iterative, incremental development approach. At each iteration (ie, development stage), we introduced design modifications, added new functionalities, and evaluated the resulting prototype. The development of the eating detection sensor consisted of 3 stages. These are briefly described below.

Development of Preliminary Prototypes

During the development of the eating detection sensor, different options for signal collection were considered:

- An air microphone placed at the beginning of the ear canal that measures sounds produced by chewing [19-21].
- A bone conduction microphone placed on the cheekbone just in front of the ear that measures the vibrations in the bone produced by chewing [22,23].
- A photoplethysmogram (PPG) sensor placed on the ear that measures the blood volume in the tissue of the ear, which is affected by chewing activity [20,21,24]. This technique has never before been used for this application.

For all three options, a prototype was developed, and these prototypes were tested in a laboratory study [19].

Development of the First Wearable Version

Based on the results of the laboratory study, we decided to continue with a combination of the air microphone and PPG sensor without the bone conduction microphone. Overall, the air microphone had shown the best results, but the PPG sensor was better at detecting soft foods [19]. These two sensors were combined for more accurate detection of eating events over a wide range of foods. Furthermore, due to its low sampling rate (21.33 Hz for our prototype), the PPG sensor has low battery requirements and is computationally efficient.

To make the new version of the eating detection sensor wearable, another device the “datalogger” was added to it [21]. It houses a data acquisition system, a battery, and an accelerometer. It is connected via a cable to the eating detection sensor and is worn in the trouser pocket or on a belt.

This first wearable version of the eating detection sensor was tested in a semicontrolled study [20,21,24]. The results obtained with the eating detection sensor were promising and were further improved by the addition of an accelerometer in the datalogger. Algorithms using signals from the air microphone, PPG sensor, and accelerometer achieved an accuracy of 0.938, a precision of 0.794, and a recall of 0.807 [21].

Development of the Integrated Version

Finally, the wearable eating detection sensor was integrated into a larger system for added functionality (Figure 1). This system includes, among others, a smartphone app and a webtool. The smartphone app can prompt the user to report the detected eating events. The webtool can provide an overview of the recorded eating events. Furthermore, goals regarding a healthy eating pattern can be entered into this webtool. Consequently, the smartphone app can help the end user achieve these goals by providing real-time feedback when the eating detection sensor is worn. The integrated version of the eating detection sensor was tested in a 4-week feasibility study.

Figure 1. The SPLENDID eating detection sensor integrated into the full SPLENDID system. This system combines the eating detection sensor with a datalogger (including an accelerometer), the Mandometer, a smartphone app, and a webtool, functioning together as a "wearable personal coach.".

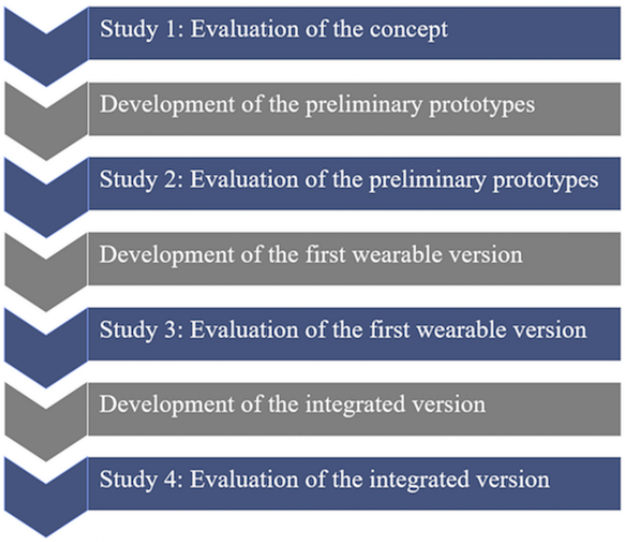
This Study

In this study, we aimed to explore how potential users perceive and experience the SPLENDID eating detection sensor. This will offer insight into its feasibility from a user's perspective. Furthermore, this will provide directions for the further development of the SPLENDID eating detection sensor and the development of similar devices. During the development of such devices, the primary focus is usually on their technical performance, but for these devices to be successful, they also need to be acceptable to the users.

Potential users evaluated the SPLENDID eating detection sensor at the different stages of its development (Figure 2).

- Study 1, evaluation of the concept of SPLENDID eating detection sensor: Before any prototypes of the eating detection sensor were developed, health professionals (n=12) were interviewed, and a focus group was held with potential end users (n=5) to find out their thoughts on the concept.
- Study 2, evaluation of the preliminary prototypes of the eating detection sensor: Young, normal-weight adults reported their experiences with the three preliminary prototypes of the eating detection sensor during the laboratory study.
- Study 3, evaluation of the first wearable version of the eating detection sensor: Young, overweight adults reported their experiences with the subsequent version of the eating detection sensor during the semicontrolled study where they used the sensor on 2 separate days (from lunch till dinner).
- Study 4, evaluation of the integrated version of the eating detection sensor: Finally, young, overweight adults reported their experiences with the eating detection sensor during a 4-week feasibility study where they used the eating detection sensor in combination with other devices (see Figure 1).

Figure 2. Flowchart indicating how the evaluation studies (blue) relate to the development stages of the SPLENDID eating detection sensor (gray).



Methods

Study 1: Evaluation of the SPLENDID Eating Detection Sensor Concept

Before any prototypes of the eating detection sensor were developed, potential users were asked about their thoughts on the concept of the eating detection sensor. Health professionals were interviewed, and a focus group was held with potential end users.

Study 1a: Interviews with Health Professionals

We conducted semistructured, in-depth, face-to face interviews with 12 health professionals who deal with weight management professionally (eg, dieticians, personal trainers). First, the concept was explained to them, and subsequently, they were asked about their views on different aspects of the concept. All interviews were recorded and later transcribed and systematically analyzed.

Study 1b: Focus Group With Potential End Users

A focus group was held with 5 young women (mean age 22 [SD 2] years; mean body mass index [BMI] 22.5 [SD 1.9] kg/m^2) interested in weight management. First, the concept was explained to them, and subsequently, they were asked open-ended questions to facilitate discussion. The focus group was recorded and later transcribed and systematically analyzed.

Study 2: Evaluation of the Preliminary Prototypes of the Eating Detection Sensor

The preliminary prototypes of the eating detection sensor were tested in a laboratory setting [19]. The pictures of these prototypes are shown in Figure 3. With these prototypes, it was yet not possible to move around freely.

The prototypes were tested by 23 healthy, young adults (13 men and 10 women; age 23 [SD 3] years; mean BMI 22.6 [SD 3] kg/m^2). They visited the university for a test session of approximately 1.5 hours. During this session, all three prototypes were worn simultaneously by the participants while they consumed a variety of foods and performed other activities such as talking. The air microphone was worn on the left ear, and the bone conduction microphone and PPG sensor were worn on the right ear. Afterward, the participants received a questionnaire concerning their experiences with the sensors. This included both closed and open-ended questions.

This study was approved by the Medical Ethical Committee of the Wageningen University (NL 48839.081.14).

Study 3: Evaluation of the First Version of the Wearable Eating Detection Sensor

One year later, the first wearable version of the eating detection sensor was tested in a semicontrolled study [20,21,24]. It comprised a commercial earhook in which both the air microphone and PPG sensor were incorporated (Figure 4). Also, a magnet was included to ensure that the PPG sensor was positioned properly. Furthermore, as described in the introduction, the datalogger was added to the eating detection sensor to make it wearable (Figure 4). The combination of the air microphone, PPG sensor, and accelerometer (incorporated into the datalogger) enables more accurate detection of eating events [21].

Twenty-two overweight, young adults (3 men and 19 women; mean age 23 [SD 2] years; mean BMI 28.0 [SD 2.3] kg/m^2) tested the wearable eating detection sensor. They participated for 2 testing days. They arrived just before lunch (11 am) and left after they had dinner (around 6 pm). At these testing days, they performed common, daily-life activities (including snacking) while wearing the eating detection sensor. Furthermore, the participants completed questionnaires on user comfort, which included both closed and open-ended questions.

This study was approved by the Medical Ethical Committee of the Wageningen University (NL52100.081.15).

Study 4: Evaluation of the Integrated Version of the Eating Detection Sensor

Finally, the integrated version of the eating detection sensor was tested by young, overweight adults in a 4-week feasibility study (Figures 1 and 5). To increase wearer comfort, the size of the datalogger and plug was reduced in this version (Figure 6). The eating detection sensor was virtually unchanged, and because of known issues with wearer comfort, the participants only had to wear it for 2 hours per day.

In total, 20 overweight, young adults (4 men and 16 women; mean age 25 [SD 2] years; mean BMI 28.8 [SD 2.8] kg/m^2) motivated to adopt healthier behavior participated in the 4-week feasibility study. During the first week, the participants used the system to assess their baseline eating behavior. Based on the observed behavior, personal goals were set for the following 3 weeks regarding the number of snacks. During these 3 weeks, the participants received personalized feedback through the smartphone app to help them achieve these goals. Afterward, they completed a questionnaire on their experiences, which included both closed and open-ended questions.

This study was approved by the Medical Ethical Committee of the Wageningen University (NL56853.081.16).

Figure 3. Preliminary prototypes of the eating detection sensor: air microphone (left), bone conduction microphone (middle), photoplethysmogram sensor (right).

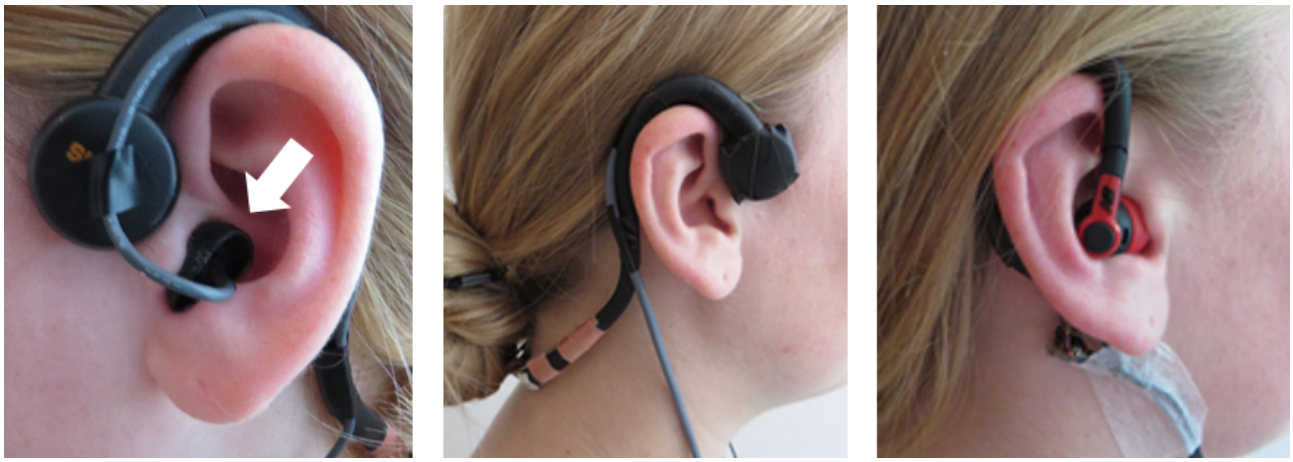


Figure 4. First wearable version of the eating detection sensor (left: the eating detection sensor; right: the eating detection sensor and the datalogger).

Figure 5. Integrated version of the SPLENDID eating detection sensor with its datalogger.

Figure 6. Old version and new, smaller version of the datalogger (left) and plug (right).

Results

Study 1: Evaluation of the SPLENDID Eating Detection Sensor Concept

Study 1a: Interviews With Health Professionals

A device like the eating detection sensor was new to all health professionals (n=12), but some already had experience with an app (n=5) or accelerometer (n=4). In general, the health professionals were enthusiastic about the eating detection sensor. Some were a bit skeptical at first (n=4), but after talking and thinking about it a little more, they thought that the sensor could be very useful to gain insight into the users' eating pattern. The users, however, need to forget that they are wearing the eating detection sensor.

> *The first thing I thought was: this is a bit excessive…But when thinking about myself when I am for example cooking, I unconsciously eat some food. People forget to write that down, so this could be very useful.*
>
> *An important thing is that end users should "forget" that they are wearing it. Then you will get a good overview of their eating patterns.*

Furthermore, they stressed that the eating detection sensor should be reliable and accurate, not cost them too much time, and come with a clear protocol on how to work with it.

Study 1b: Focus Group With Potential End Users

The participants were already familiar with all kinds of smartphone apps to record food intake. They were enthusiastic about what the eating detection sensor had to add. One of the participants mentioned that it will help her when she "secretly" eats something, and this will give a good insight into her eating pattern. However, the participants also had some concerns regarding the eating detection sensor. It should be ensured that it is comfortable to wear for a long time, and it should not be too noticeable.

Study 2: Evaluation of the Preliminary Prototypes of the Eating Detection Sensor

For wearer comfort, the air microphone received an average grade of 6.7 (range: 2-9) on a scale of 1 to 10, the bone conduction microphone 5.8 (range: 2-10), and the PPG sensor 6.7 (range: 4-9). The grade for wearer comfort did not differ significantly between the prototypes (analysis of variance [ANOVA], P=.13). The participants indicated they would be able to wear the air microphone for an average of 5.7 (range: 0-24) hours per day, the bone conduction microphone for 5.6 (range: 0.5-24) hours per day, and the PPG sensor for 5.4 (range: 4-9) hours per day. This value did not differ significantly between the prototypes (ANOVA, P=.99).

The open-ended questions provide an explanation for these results. The most frequently mentioned remarks regarding the air microphone were that the sensor was comfortable to wear (n=18), but that it lowered the users' hearing ability (n=10) and that they would get tired of the sensor after wearing it for a longer period (n=15). Regarding the bone conduction microphone, the participants most frequently mentioned that it remained unnoticed while wearing (n=10), that the sensor could be annoying during exercise (n=9), and that the sensor put pressure on their head and neck (n=5). Regarding the PPG sensor, the participants most frequently mentioned that they did not notice that they were wearing the sensor (n=11), the sensor lowered their hearing ability (n=6), and the sensor cable was pulling and annoying (n=4). Regarding the prototypes in general, the most frequently mentioned barrier for wearing them in real life was that they were very noticeable and oddly shaped (n=8). In turn, the most often mentioned wishes were that the prototypes should be as invisible as possible (n=13) and that they should be comfortable to wear (n=10).

Study 3: Evaluation of the First Wearable Version of the Eating Detection Sensor

The participants graded the wearer comfort of the chewing sensor an average of 3.8 (range: 2-7) on a scale of 1 to 10.

Furthermore, participants indicated that they would be able or willing to wear the chewing sensor for 3.9 (range: 2-7 h) hours per day. Some participants, however, mentioned that they would be able to wear it for a longer time if there were breaks in between.

There was large variation in the answers of the participants regarding how the chewing sensor affected eating, moving, and talking. Most participants agreed with the statement that the chewing sensor was bothering them: 19 out of the 20 participants scored higher than 5 on a 9-point Likert scale (1=totally disagree, 5=neutral, 9=totally agree).

The open-ended questions provide an explanation for these results. The most frequently mentioned remarks regarding the wearer comfort of the eating detection sensor were: “the chewing sensor was painful to the ear” (n=16), “the cable was annoying or hindering” (n=14), “the sensor reduced hearing” (n=8), and “internal noises were heard better” (n=5). Three participants experienced no or only little discomfort.

Study 4: Evaluation of the Integrated Version of the Eating Detection Sensor

Of the 20 participants, 19 experienced discomfort from the eating detection sensor; they started experiencing discomfort after an average of 1 hour and 20 minutes. The participants graded the average wearer comfort of the eating detection sensor at 3.7 (range: 1-7) on a scale of 1 to 10. Moreover, they scored the statement “The sensor bothered me” an average of 5.5 (range: 4-7) on a scale of 1 (totally disagree) to 7 (totally agree).

The participants reported having used the eating detection sensor on an average of 19 out of the intended 28 days, of which they used it for at least 2 hours on 17 days. During the first week, compliance was highest, with the eating detection sensor being used for an average of 6 days. The most frequently mentioned reasons for wearing the sensor <2 hours per day (open-ended question) were discomfort (n=14) and technical issues, such as broken sensor (n=8; Table 1). Furthermore, if the participants used the sensor, it was for an average of 1.9 hours (range: 1-4 hours).

Regarding reactions from the social environment, the participants gave mixed results. They scored the statements “People in my environment noticed the sensor” and “I did not like it when people noticed the sensor” an average of a 3.4 (range: 1-7) on a scale of 1 (totally disagree) to 7 (totally agree).

Table 1. Reasons mentioned for wearing the eating detection sensor <2 hours and their frequency.

Reason	Frequency
Discomfort	14
Technical issues (eg, broken sensor)	8
Reduced hearing	6
Impractical (eg, with sports)	6
Inappropriate (eg, at work)	3
Noticeable	1
Forgotten	1
Not enough time	1

Table 2. Additional remarks regarding the eating detection sensor and their frequency.

Additional remarks	Frequency
Cable is not practical	7
The eating detection sensor got noticed	7
The eating detection sensor reduced hearing	5
The eating detection sensor was uncomfortable	4
Experienced technical issues with the eating detection sensor	4
Had to explain what the eating detection sensor is	3
Inappropriate to use in certain situations	3
Added value of eating detection sensor unclear	3
Received no reactions from environment	2
Received positive reactions from environment	2
Experienced no problems	2
Looks like listening to music	2
Not practical	1

When the participants were asked whether they had any additional remarks (ie, open-ended question), they most frequently mentioned that the cable was not practical (n=7), the sensor got noticed (n=7), and the sensor lowered their hearing (n=5; Table 2). Furthermore, some participants indicated that they did not see the added value of the sensor because they believed that they did not need it to remind them to note the foods consumed and that the detections were not always accurate.

Discussion

Principal Results

The current paper explores how potential users perceive and experience the SPLENDID eating detection sensor. Across the different stages of development, the potential users were enthusiastic about the concept. They especially liked that it provided objective information on their eating patterns. However, they stressed that it needed to be comfortable to wear and discreet. The latest version of the eating detection sensor did not yet meet these requirements.

For the eating detection sensor to meet the user requirements, further improvements need to be made. In particular, the wearer comfort of the sensor requires attention. After wearing the sensor for a while (ie, on average, after 80 minutes) the potential users started experiencing discomfort. As a result, they graded the wearer comfort of the sensor at 3.7 on a scale of 1 to 10 in the final study. Moreover, the experienced discomfort was the main reason for the participants to wear the sensor for <2 hours.

One option would be to offer different shapes and sizes of the eating detection sensor so that the users can find a sensor with a good fit. This would also improve the ability of the device to detect eating events. The current eating detection sensor fits some people better than others, which is reflected in the wide range in the grades for wearer comfort (1-7 for the last version, on a scale from 1 to 10). Another option would be to reduce the size of the eating detection sensor and to make it more like a hearing aid; these are made to be worn throughout the day, unlike earphones. However, the technically feasibility needs to be investigated.

By resolving the issues related to wearer comfort, the visibility of the eating detection sensor is likely to be reduced as well. Furthermore, the visibility of the current version of the eating detection sensor was already acceptable to some of the participants. They mentioned that even though people in the environment noticed the eating detection sensor, they did not recognize it as such because it looks like a device for listening to music. This is a major advantage of the ear-worn devices over some of the other devices that are being developed for the detection of eating events (eg, neck-worn devices or a device mounted onto eyeglasses)[14,25-27].

It would be interesting to repeat the feasibility study once the eating detection sensor has been improved for wearer comfort and visibility. The SPLENDID eating detection sensor is a device with great potential as shown by its technical performance [19-21,24]. It could help provide a more complete picture of food intake, which is a major issue with the current methods for monitoring food intake [1-4,6-9].

Limitations

In the feasibility study, due to the issues with wearer comfort, the participants were asked to wear the eating detection sensor for at least 2 hours, while it is intended to be used throughout the day. This will affect the user experience. As was mentioned by the health professionals, people need to forget that they are wearing the eating detection sensor. Because the participants only used the eating detection sensor for an average of 1.9 hours per day and started to experience some discomfort after a while, they might not have been able to forget that they were wearing the eating detection sensor.

If the participants in the feasibility study had been less conscious about the fact that they were wearing the eating detection sensor, they probably would also have been less conscious about their eating, and then, the added value of the eating detection sensor would have been more evident. For 15% (3/20) of the participants, the added value of the eating detection sensor was unclear. They did not feel that they needed such a sensor to remind them to report the foods consumed.

Comparison With Prior Works

To our knowledge, this is the first paper to describe how an ear-worn device for the detection of eating events is received by potential users and to describe their experiences with such a device in real life. It shows that ear-worn devices for the detection of eating events need to meet high standards to be acceptable for everyday use.

When tested in a laboratory setting, the eating detection sensor received a sufficient grade for wearer comfort, while it received an insufficient grade when it was tested in real life. Moreover, the participants did not experience discomfort as soon as they started wearing the sensor; they started experiencing discomfort only after 80 minutes of wearing it. It is important to keep this in mind when interpreting the results from the laboratory studies.

Conclusions

The SPLENDID eating detection sensor is a promising new device that can facilitate the collection of reliable food intake data as shown by its technical potential, which has been described before. Furthermore, potential users are enthusiastic about it. They especially like that it provides objective information on their eating patterns. However, to be successful, the wearer comfort and discreetness of the device need to be improved. Therefore, further development of the device should mainly focus on the design of the hardware.

Acknowledgments

AD, LZ, MM, and CD obtained the funding for the project within which the eating detection sensor was developed. JvdB, LZ, VP, CD, AD, and MM were involved in the development of the eating detection sensor. JvdB and MM obtained ethical approval

for the studies described. JvdB and AvdL performed the studies. JvdB and AvdL analyzed the results. JvdB wrote the manuscript. All authors were involved in revising the paper and approved the final version of the manuscript. The work leading to these results has received funding from the European Community's ICT Programme under Grant Agreement No. 610746, 01/10/2013-30/09/2016. Furthermore, we would like to thank the research dieticians, Els Siebelink and Renske Hubers-Geers, for their contribution.

Conflicts of Interest

None declared.

References

1. Kroke A, Klipstein-Grobusch K, Voss S, Möseneder J, Thielecke F, Noack R, et al. Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. Am J Clin Nutr 1999 Oct;70(4):439-447. [doi: 10.1093/ajcn/70.4.439] [Medline: 10500011]
2. Hutchesson MJ, Rollo ME, Callister R, Collins CE. Self-monitoring of dietary intake by young women: online food records completed on computer or smartphone are as accurate as paper-based food records but more acceptable. J Acad Nutr Diet 2015 Jan;115(1):87-94. [doi: 10.1016/j.jand.2014.07.036] [Medline: 25262244]
3. Schoeller DA. Limitations in the assessment of dietary energy intake by self-report. Metabolism 1995 Feb;44:18-22. [doi: 10.1016/0026-0495(95)90204-X]
4. Trijsburg L, Geelen A, Hollman PC, Hulshof PJ, Feskens EJ, van't Veer P, et al. BMI was found to be a consistent determinant related to misreporting of energy, protein and potassium intake using self-report and duplicate portion methods. Public Health Nutr 2016 Oct 11;20(04):598-607. [doi: 10.1017/s1368980016002743]
5. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 2015 May;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]
6. Siebelink E, Geelen A, de VJHM. Self-reported energy intake by FFQ compared with actual energy intake to maintain body weight in 516 adults. Br J Nutr 2011 Jul;106(2):274-281. [doi: 10.1017/S0007114511000067] [Medline: 21338536]
7. van den Boer JHW, Kranendonk J, van DWA, Feskens EJM, Geelen A, Mars M. Self-reported eating rate is associated with weight status in a Dutch population: a validation study and a cross-sectional study. Int J Behav Nutr Phys Act 2017 Sep 08;14(1):121 [FREE Full text] [doi: 10.1186/s12966-017-0580-1] [Medline: 28886719]
8. Johansson G, Wikman Å, Åhrén A, Hallmans G, Johansson I. Underreporting of energy intake in repeated 24-hour recalls related to gender, age, weight status, day of interview, educational level, reported food intake, smoking habits and area of living. PHN 2007 Jan 2;4(04). [doi: 10.1079/phn2001124]
9. Johansson L, Solvoll K, Bjørneboe GE, Drevon CA. Under- and overreporting of energy intake related to weight status and lifestyle in a nationwide sample. Am J Clin Nutr 1998 Aug;68(2):266-274. [Medline: 9701182]
10. Ford AL, Bergh C, Södersten P, Sabin MA, Hollinghurst S, Hunt LP, et al. Treatment of childhood obesity by retraining eating behaviour: randomised controlled trial. BMJ 2009 Jan 05;340:b5388. [Medline: 20051465]
11. Doulah A, Farooq M, Yang X, Parton J, McCrory MA, Higgins JA, et al. Meal Microstructure Characterization from Sensor-Based Food Intake Detection. Front Nutr 2017 Jul;4:31 [FREE Full text] [doi: 10.3389/fnut.2017.00031] [Medline: 28770206]
12. Amft O. Automatic dietary monitoring using on-body sensors: Detection of eating and drinking behaviour in healthy individuals. Zürich: ETH Zürich; 2008:1-224.
13. Farooq M, McCrory MA, Sazonov E. Reduction of energy intake using just-in-time feedback from a wearable sensor system. Obesity (Silver Spring) 2017 Apr;25(4):676-681. [doi: 10.1002/oby.21788] [Medline: 28233942]
14. Rahman T, Adams A, Zhang M, Cherry E, Zhou B, Peng H. BodyBeat: a mobile system for sensing non-speech body sounds. BodyBeat; 2014 Presented at: MobiSys; June 16 - 19, 2014; Bretton Woods p. 2-13. [doi: 10.1145/2594368.2594386]
15. Bedri A, Verlekar A, Thomaz E, Avva V, Starner T. A wearable system for detecting eating activities with proximity sensors in the outer ear. 2015 Presented at: ISWC '15; September 07 - 11, 2015; Osaka p. 91-92.
16. Fontana JM, Farooq M, Sazonov E. Automatic Ingestion Monitor: A Novel Wearable Device for Monitoring of Ingestive Behavior. IEEE Trans. Biomed. Eng 2014 Jun;61(6):1772-1779. [doi: 10.1109/tbme.2014.2306773]
17. SPLENDID. SPLENDID; Personalised Guide for Eating and Activity Behaviour for the Prevention of Obesity and Eating Disorders URL: https://splendid-program.eu/ [accessed 2017-11-28] [WebCite Cache ID 6vJMc7VL8]
18. Maramis C, Diou C, Ioakeimidis I, Lekka I, Dudnik G, Mars M. Preventing obesity and eating disorders through behavioural modifications: The SPLENDID vision. Preventing obesity and eating disorders through behavioural modifications; 2014 Presented at: MobiHealth '14; November 3-5, 2014; Athens p. 7-10.
19. Papapanagiotou V, Diou C, Lingchuan Z, van den Boer J, Mars M, Delopoulos A. Fractal Nature of Chewing Sounds. 2015 Presented at: ICIAP 2015; September 7 - 11, 2015; Genoa p. 401-408.

20. Papapanagiotou V, Diou C, Zhou L, van den Boer J, Mars M, Delopoulos A. The SPLENDID chewing detection challenge. 2017 Presented at: EMBC '17; July 11-15, 2017; Seogwipo p. 817-820.
21. Papapanagiotou V, Diou C, Zhou L, van den Boer J, Mars M, Delopoulos A. A Novel Chewing Detection System Based on PPG, Audio, and Accelerometry. IEEE J Biomed Health Inform 2017 May;21(3):607-618. [doi: 10.1109/JBHI.2016.2625271] [Medline: 27834659]
22. Sazonov E, Schuckers S, Lopez-Meyer P, Makeyev O, Sazonova N, Melanson EL, et al. Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior. Physiol Meas 2008 May;29(5):525-541 [FREE Full text] [doi: 10.1088/0967-3334/29/5/001] [Medline: 18427161]
23. Shuzo M, Komori S, Takashima T, Lopez G, Tatsuta S, Yanagimoto S, et al. Wearable eating habit sensing system using internal body sound. JAMDSM 2010;4(1):158-166. [doi: 10.1299/jamdsm.4.158]
24. Papapanagiotou V, Diou C, Zhou L, van den Boer J, Mars M, Delopoulos A. A novel approach for chewing detection based on a wearable PPG sensor. 2016 Presented at: EMBC '16; August 16-20, 2016; Orlando.
25. Rahman T, Czerwinski M, Gilad-Bachrach R, Johns P. Predicting “About-to-Eat” Moments for Just-in-Time Eating Intervention. Proceedings of the 6th International Conference on Digital Health Conference; Montreal, Quebec, Canada: ACM; 2016 Presented at: DH'16; April 11 - 13, 2016; Montreal p. 141-150.
26. Farooq M, Fontana JM, Sazonov E. A novel approach for food intake detection using electroglottography. Physiol Meas 2014 May;35(5):739-751 [FREE Full text] [doi: 10.1088/0967-3334/35/5/739] [Medline: 24671094]
27. Farooq M, Sazonov E. A Novel Wearable Device for Food Intake and Physical Activity Recognition. Sensors 2016 Jul 11;16(7):1067. [doi: 10.3390/S16071067]

Abbreviations

PPG: photoplethysmogram

Edited by G Eysenbach; submitted 05.01.18; peer-reviewed by B Egberg Mikkelsen, W Jia; comments to author 15.02.18; revised version received 12.04.18; accepted 08.05.18; published 04.09.18.

Please cite as:
van den Boer J, van der Lee A, Zhou L, Papapanagiotou V, Diou C, Delopoulos A, Mars M
The SPLENDID Eating Detection Sensor: Development and Feasibility Study
JMIR Mhealth Uhealth 2018;6(9):e170
URL: http://mhealth.jmir.org/2018/9/e170/
doi:10.2196/mhealth.9781
PMID:30181111

Original Paper

Medical-Grade Physical Activity Monitoring for Measuring Step Count and Moderate-to-Vigorous Physical Activity: Validity and Reliability Study

Myles William O'Brien[1,2], BKinH; William Robert Wojcik[1], BKinH; Jonathon Richard Fowles[1], PhD

[1]Centre of Lifestyle Studies, School of Kinesiology, Acadia University, Wolfville, NS, Canada

[2]Division of Kinesiology, Dalhousie University, Halifax, NS, Canada

Corresponding Author:
Jonathon Richard Fowles, PhD
Centre of Lifestyle Studies, School of Kinesiology
Acadia University
550 Main Street
Wolfville, NS,
Canada
Phone: 1 9025851560
Fax: 1 902 585 1702
Email: jonathon.fowles@acadiau.ca

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2019/1/e12576/

Abstract

Background: The use of physical activity (PA) monitors is commonly associated with an increase in habitual PA level in healthy and clinical populations. The PiezoRx is a medical-grade PA monitor that uses adjustable step rate thresholds to estimate moderate-to-vigorous physical activity (MVPA) and is a valid indicator of free-living PA in adults. Laboratory validation of step count derived MVPA in adults is needed to justify the use of these monitors in clinical practice to track individuals' progress toward meeting PA guidelines that are based on MVPA, not steps.

Objective: The objective of our study was to assess the validity and interinstrument reliability of the PiezoRx to derive step count and MVPA in a laboratory setting compared with criterion measures and other frequently used PA monitors in a diverse sample of adults.

Methods: The adult participants (n=43; 39.4 years, SD 15.2) wore an Omron HJ-320 pedometer, an ActiGraph GT3X accelerometer, and four PiezoRx monitors during a progressive treadmill protocol conducted for 6 minutes at speeds of 2.4, 3.2, 4.0, 5.6, 6.4, and 7.2 km/hour, respectively. The four PiezoRx monitors were set at different MVPA step rate thresholds (MPA in steps/minute/VPA in steps/minute) 100/120, 110/130, height adjusted, and height+fitness adjusted.

Results: The PiezoRx was more correlated (intraclass correlation, ICC=.97; P<.001) to manual step counting than the ActiGraph (ICC=.72; P<.001) and Omron (ICC=.62; P<.001). The PiezoRxs absolute percent error in measuring steps was 2.2% (ActiGraph=15.9%; Omron=15.0%). Compared with indirect calorimetry, the height-adjusted PiezoRx and ActiGraph were accurate measures of the time spent in MVPA (both ICC=.76; P<.001).

Conclusions: The PiezoRx PA monitor appears to be a valid and reliable measure of step count and MVPA in this diverse sample of adults. The device's ability to measure MVPA may be improved when anthropometric differences are considered, performing at par or better than a research grade accelerometer.

(JMIR Mhealth Uhealth 2018;6(9):e10706) doi:10.2196/10706

KEYWORDS

pedometer; accelerometry; exercise prescription; validity; reliability

Introduction

Regular physical activity (PA) and exercise are associated with a reduced risk of developing cardiovascular diseases, cancer, and diabetes mellitus [1]. However, most Canadians do not engage in a sufficient level of daily activity with only 15% meeting the PA guidelines of 150 minutes of moderate-to-vigorous aerobic physical activity (MVPA) per week and 35% achieving 10,000 steps per day [2]. The use of PA monitors is one way of assisting individuals in increasing their PA; meta-analyses of randomized controlled trials have shown that the use of PA monitors increased the daily activity of the healthy and diseased populations by 1800-2500 steps per day [3,4]. Moreover, step count prescriptions provided by physicians increased the daily step count by 20% and improved the glycemic control of patients with type 2 diabetes and hypertension after 1 year compared with the control group [5]. Several primary care providers identify that the lack of tangible aids (ie, pedometers and resources to support PA) is a major barrier in prescribing exercise [6]. Health care providers may not use PA monitors as a measurement tool because of their accuracy and reliability, particularly owing to the different capabilities of popular commercial grade PA monitors in measuring step count and MVPA [7]. Considering the accumulating evidence on the benefits of PA and the clinical outcomes of PA monitoring, the validity and reliability of such patient-focused PA monitors must be evaluated. A greater confidence in the measurement of accuracy may lead to an increase in the use of PA monitors among health care professionals.

Recently, a new piezoelectric PA monitor (PiezoRx, StepsCount Inc, Deep River, ON, Canada) has been approved as a class 1 medical device by Health Canada, and this device may be prescribed by health care providers to their patients for monitoring PA. Moreover, the device uses adjustable step rate thresholds to determine the time spent performing MVPA. Factory settings of 100 steps per minute (spm) and 120 spm were identified to correspond to moderate physical activity (MPA; 3 metabolic equivalents, METs) and vigorous physical activity (VPA; 6 METs), respectively, based on previous indications for these thresholds from the literature [8-10]. However, the MVPA step rate thresholds differ when considering individual anthropometric (ie, height) and aerobic fitness differences [10-12]. Whether adjusting the MVPA step rate thresholds according to factors known to influence step rate bioenergetics improves the validity of the PiezoRx in measuring MVPA is not known. A recent pilot study has shown that the height-adjusted PiezoRx can accurately measure step count and MVPA in a diverse sample of adults in free-living conditions compared with accelerometry [13]. However, the use of the ActiGraph as a criterion measure of MVPA may produce some errors compared with direct counting of steps and indirect calorimetry. In addition, only a single-adjusted PiezoRx was used, and whether other MVPA thresholds are more accurate is unclear. Therefore, the accuracy of multiple PiezoRx devices set at individualized MVPA thresholds must be compared with that of indirect calorimetry in a controlled laboratory setting.

Previous laboratory validation studies have used the older versions of the device (SC-StepMX and SC-StepRx), and results have shown that the device was better than research grade PA monitors in measuring the step count of healthy children and young adults [14], adults [15], and older adults [16]. When the MVPA step rate thresholds (MPA/VPA) were increased to 110/130, the PiezoRx and a research grade accelerometer had similar measurements for MVPA compared with indirect calorimetry in children and young adults [14], although no study to date has compared the accuracy of the devices and indirect calorimetry in measuring MVPA in adult populations.

Considering the potential of a medical-grade PA monitor to assist health care providers in prescribing exercise and monitoring the PA of their patients, the accuracy of such devices in measuring step count and MVPA must be evaluated. Therefore, this study aimed to assess the validity and interinstrument reliability of these devices in measuring steps and MVPA during a progressive treadmill walking protocol in a diverse sample of adults.

Methods

Participants

A sample of 43 adults (25 women) aged between 20 and 64 years (mean age 39.4 years, SD 15.2) volunteered to participate in this study. The average body mass index (BMI) and aerobic fitness of the participants were 27.9 kg/m^2 (SD 6.1) and 41.2 mL/kg/min (SD 10.2), respectively. The majority (n=24) of participants answered “yes” to at least one question in the Physical Activity Readiness Questionnaire Plus (PAR-Q+), and they were cleared for MVPA [17]. Moreover, they completed the Canadian Society for Exercise Physiology-Physical Activity and Sedentary Behavior Questionnaire (CSEP-PASB-Q), which is a valid and reliable measurement tool for weekly MVPA [18]. The participants self-reported their aerobic fitness levels based on PASB-Q as poor (n=1), fair (n=6), good (n=12), very good (n=13), and excellent (n=11). All participants were recruited via a community-wide email and by word of mouth, and a written informed consent was obtained from the participants. The study was conducted in Wolfville, Nova Scotia, from May 2015 to September 2015, and it was approved by the research ethics board of Acadia University (REB# 15-20).

Experimental Design

After the prescreening procedures, anthropometric (ie, height and weight) measurements and aerobic fitness were evaluated by a Canadian Society for Exercise Physiology-Certified Personal Trainer according to published guidelines [19,20]. Aerobic fitness was predicted using the submaximal Ebbeling protocol [21], as described in more detail below. After the submaximal aerobic test, a resting period of 20-30 minutes was allotted to ensure that the participants returned to a rested state; then, four PiezoRxs, one Omron HJ-320, and one ActiGraph GT3X accelerometer were placed around the waist of the participants according to manufactures’ recommendations. Thereafter, the participants were asked to complete the multistage treadmill walking protocol to assess and compare the validity of the ActiGraph, Omron, and PiezoRx with finite

step rate thresholds (ie, 100/120 and 110/130) and individualized step rate thresholds (ie, height adjusted and height+fitness adjusted) with that of the criterion measures of manual step counting and indirect calorimetry.

Aerobic Fitness

Aerobic fitness was estimated using the Ebbeling walking treadmill protocol [21]. The Ebbeling consists of two 4-min walking stages. The first stage is designed to reach a speed that elicits 60% of the participants' maximum estimated heart rate (ie, 220–age). The second stage included the increase in the incline by 5% and maintenance of the previously established speed. Treadmill speed and steady-state heart rate were used to estimate $VO_{2\,max}$ using a prediction equation [20]. A submaximal test was chosen over a maximal test for safety reasons and a minimal influence after the walking assessment because it was most practical to complete the fitness testing and step assessment in a single session.

Physical Activity Monitors

Each participant used four PiezoRx pedometers (StepsCount, ON, Canada), one ActiGraph GT3X accelerometer (ActiGraph, FLA, the USA), and one Omron HJ-320 pedometer (Omron Healthcare, Kyoto, Japan) that were attached on an adjustable leather belt around their waist. All devices were worn in accordance with the manufactures' recommendations. In particular, two PiezoRx and ActiGraph were fitted in line with the left thigh, whereas the other two PiezoRx and Omron were fitted in line with the right thigh. The ActiGraph accelerometer was sampled at 30 Hz and was initialized to collect data using 15-second epochs. The ActiGraph cut point for MVPA was >1952 counts [22]. The PiezoRx thresholds for monitoring MVPA were set as follows: 100/120, 110/130, adjusted for height, and adjusted for height+fitness (Table 1). The PiezoRxs thresholds that were set at 100/120 and 110/130 were based on the current literature that consistently showed that 100-110 spm is equal to ~3 METs (ie, 10.5 mL/min/kg) and that 120-130 spm is equal to ~6 METs (ie, 21 mL/min/kg) in adults [8-11,23,24].

Adjustments recommended for height and fitness were based on previous literature and unpublished results in our laboratory indicating height and fitness related changes in step rate bioenergetics [10-12]. For these adjustments, baseline step thresholds were chosen based on the available literature indicating that 100 and 130 spm are heuristic cadence-intensity thresholds for MPA and VPA, respectively [24].

An adjustment of 5 spm is appropriate for every 10 cm increase (ie, 5 spm lower) or 10 cm decrease in height (ie, 5 spm higher) based on the premise that shorter individuals must take more steps to cover the same distance and therefore similar external work. In addition, considering the known impact of cardiorespiratory fitness on step rate bioenergetics [12], aerobic fitness was proportionally associated with 10 spm adjustments to step rate thresholds. For example, a $VO_{2\,max}$ of "Excellent" [19] resulted in an increase in MPA and VPA step rate thresholds by 10 spm (eg, 110/140; see Table 1).

The distributions of MVPA step rate thresholds for the height-adjusted PiezoRx were as follows: 90/120 (n=1), 95/125 (n=11), 100/130 (n=10), 105/135 (n=17), and 110/140 (n=4). The distributions of MVPA step rate thresholds for the height+fitness-adjusted PiezoRx were as follows: 90/120 (n=2), 95/125 (n=13), 100/130 (n=6), 105/135 (n=13), 110/140 (n=7), and 115/145 (n=2).

Exercise Protocol

Once the participants were equipped with the activity monitors, they were fitted with a headpiece, two-way nonrebreathing valve (Hans Rudolph, Inc, KS, the USA), noseclip, and mouthpiece. The metabolic cart (TrueOne 2400; Parvo Medics, UT, USA) was calibrated using nitrogen and two primary standard gas mixtures to an error rate of 0.01%. The pneumotachometer was calibrated using a 3-L syringe that delivered fixed volumes at different flow rates. Volume calibration was <0.1 L. They proceeded to complete a 6-stage treadmill protocol at predetermined speeds of 2.4, 3.2, 4.0, 5.6, 6.4, and 7.2 km/h at 0% grade, which correlated to a metabolic intensity average of 2.6 (SD 0.4), 2.8 (SD 0.4), 3.1 (SD 0.4), 4.3 (SD 0.4), 5.3 (SD 0.4), and 6.8 (SD 0.7) METs for each stage, respectively. Each stage consisted of 6 min of walking to achieve a metabolic steady state, followed by 4 min of rest. The steady-state VO_2 was used as stage VO_2 in the analysis to limit the variability introduced by oxygen kinetics during the onset of exercise in each stage. Steps were manually counted by two instructors for 2-3 and 4-5 min during each stage to determine the step count criterion. A video camera was used to film the feet of the participant in case the testers recorded >1 step difference during a stage. The 6-stage protocol corresponded to stepping cadences of 86 (SD 8), 97 (SD 7), 106 (SD 7), 120 (SD 7), 128 (SD 7), and 139 (SD 8) spm for each stage, respectively. Data from the PiezoRx and Omron monitors were extracted during the 4-min rest phase with the participant straddling the treadmill, and the devices were reset prior to the next stage. Data from the ActiGraph were extracted at the end of the exercise. At the end of each stage, the participants immediately straddled the treadmill. Immediately prior to the start of the next stage, the treadmill belt speed was increased to the necessary speed to avoid differences in step count and MVPA that may occur during treadmill acceleration or deceleration. The test was terminated because of volitional fatigue if the participant reached 85% of his or her estimated maximum heart rate (220 – age) or if they finished all 6 stages of the protocol. An appropriate cool down was administered by the instructor while monitoring their heart rate recovery.

Data and Statistical Analyses

Data were analyzed using the Statistical Package for Social Sciences software version 23.0 (IBM, NY, USA). A *P* value <.05 was considered statistically significant. Manually counted steps and MVPA obtained via indirect calorimetry were assessed for normality using the Shapiro-Wilk test. Manually counted steps were normally distributed (*P*>.05). However, MVPA obtained via indirect calorimetry were not normal owing to the number of walking stages composed of either 0 or 360 sec of MVPA. Only the PiezoRx set at factory settings was used for the validity of the step count analysis. The steps counted within 2-3 and 4-5 min for each stage were averaged and multiplied by a factor of 6 to determine the number of steps for each 6-min

stage. In relation to this, the steps counted within 2-3 and 4-5 min were always within 2 step/min.

A two-way mixed-model intraclass correlation coefficients (ICC) with measures of consistency were utilized to measure the strength and consistency of each device's capability to measure steps with those of the criterion measure of manual step counting. The Lin's concordance coefficients (LCC) and Bland-Altman plots (BAPs) were utilized to assess the agreement between monitors and the criterion measures [25,26].

Table 1. Algorithm used to determine the PiezoRx height- and fitness-adjusted moderate-to-vigorous physical activity step rate thresholds.

Height, cm (ft in)	Height adjusted (MPA/VPA[a] thresholds)	Fitness classification[b] (MPA/VPA thresholds)		
		Poor/fair	Good/very good	Excellent
140 (4'8")	115/145	105/135	115/145	125/155
150 (5'0")	110/140	100/130	110/140	120/150
160 (5'3")	105/135	95/125	105/135	115/145
170 (5'7")	100/130	90/120	100/130	110/140
180 (5'11")	95/125	85/115	95/125	105/135
190 (6'3")	90/120	80/110	90/120	100/130
200 (6'6")	85/115	75/105	85/115	95/125

[a]MPA/VPA: moderate physical activity/vigorous physical activity.

[b]The ratings for aerobic fitness were based on normative values from the Canadian Society for Exercise Physiology (CSEP) physical activity and training for health.

BAPs were also generated based on the steps measured using the PiezoRx, ActiGraph, and Omron to determine the mean bias (ie, overpredicted or underpredicted values) and the limits of agreement (LoA; SD 95% of mean bias). A 4×6 (device×stage) repeated measure analysis of variance (RM-ANOVA) was used to assess the differences between the PiezoRx, Omron, and ActiGraph steps as well as manual step counting across treadmill stages. The absolute percent error (APE) of the PiezoRx, ActiGraph, and Omron was calculated using the equation: (|manually counted − device measured|)/(manually counted) × 100%. The reliability of the PiezoRx step count function was assessed using ICC and RM-ANOVA (device×stage) using all four devices. Bonferroni post hoc testing was carried out for all statistically significant RM-ANOVAs.

For each device and the metabolic cart, the amount of time spent in performing MVPA for each walking stage was summed up for full 6 min for analysis. The absolute MVPA difference per stage (metabolic cart-device measured) and Spearman correlation coefficients (SCCs) were used to assess and compare the validity of the four PiezoRx devices and the ActiGraph accelerometer in measuring MVPA with that of indirect calorimetry. BAP was used to assess the agreement between the height-adjusted PiezoRx and both indirect calorimetry and the ActiGraph in obtaining MVPA measurement. The difference in the absolute MVPA per stage (sec/stage) was calculated over an absolute percent change owing to the dichotomous nature of exercise stages (either the absence of time spent in MVPA or comprised entirely of time spent in MVPA). As such, the sensitivity and specificity rates of each device were calculated to identify if the stage has moderate-to-vigorous intensity compared with the metabolic cart. A stage was considered to have a moderate-to-vigorous intensity if at least 80% of the stage (ie, 288 sec) was greater than 3 METs (ie, 10.5 mL/kg/min).

The reliability of MVPA measurement obtained using the PiezoRx was assessed using the absolute MVPA difference per stage (|metabolic cart − device measured|) and SCCs for each stage where the height-adjusted and height+fitness-adjusted MVPA thresholds were equal (see Table 1; n=25).

Results

Validity and Reliability of Step Count Measurement

Compared with manual step counting, the PiezoRx had a higher intraclass correlation (ICC=.97; *P*<.001) and lower percent error (2.2%, SD 5.4), than both the ActiGraph (ICC=.72, *P*<.001; APE=15.9%, SD 26.8) and the Omron (ICC=.62, *P*<.001; APE=15.0%, SD 29.0), as shown in Table 2, Table 3, and Table 4. As shown in Figure 1, Bland-Altman analysis revealed a fixed bias of −6.4 steps (*P*=.001) with LoA of −63 and 51 steps for the PiezoRx and manually counted steps. Step count differences were observed between manual step counting and both the ActiGraph and Omron during the first stage (2.4 km/h; both *P*<.001) and second stage (3.2 km/h; ActiGraph: *P*<.001; Omron: *P*=.04) but only the Omron during the fourth stage (5.6 km/h; Omron: *P*=.046).

ICC using all four devices showed an ICC interinstrument reliability of 0.88 (*P*<.001), as shown in Table 5. All four PiezoRx devices were included in the reliability of step count analysis. Cases in which the height- and height/fitness-adjusted step rate thresholds were identical were used for the MVPA reliability analysis. The correlation was stronger (ICC>.90) during the fastest walking stages (6.4-7.2 km/h) and was weakest (ICC<.50) during the middle speed stages (4.0-5.6 km/h). RM-ANOVA revealed no significant (*P*>.05) differences between all four PiezoRx devices for measuring step count.

Validity and Reliability of Moderate-to-Vigorous Physical Activity

The devices with the highest Spearman correlations in measuring MVPA across all stages were the height-adjusted PiezoRx (SCC=.80; *P*<.001) and the ActiGraph (SCC=.80; *P*<.001), which also had similar absolute time differences per 6-minute walking stage (height-adjusted PiezoRx: 53±93 sec/stage; ActiGraph: 53±103 sec/stage). MVPA SCC of the PiezoRx (100/120), PiezoRx (110/130), and PiezoRx (height+fitness adjusted) were 0.74 (*P*<.001), 0.76 (*P*<.001), and 0.77 (*P*<.001), respectively.

Table 2. Two-way mixed-model intraclass correlation with measure of consistency.

Walking speed (km/h)	n	PiezoRx		Actigraph		Omron	
		Intraclass correlation (95% CI)[a]	*P* value	Intraclass correlation (95% CI)[a]	*P* value	Intraclass correlation (95% CI)[a]	*P* value
2.4	43	.71 (0.53 to 0.83)	<.001	.16 (−0.15 to 0.44)	.88	.14 (−0.16 to 0.42)	0.92
3.2	43	.80 (0.67 to 0.89)	<.001	.22 (−0.09 to 0.49)	.36	.28 (−0.02 to 0.53)[b]	.04
4.0	42	.98 (0.97 to 0.99)	<.001	.70 (0.50 to 0.83)	<.001	.94 (0.90 to 0.97)	<.001
5.6	41	.99 (0.976 to 0.993)	<.001	.96 (0.93 to 0.98)	<.001	.39 (0.10 to 0.62)[b]	.03
6.4	38	.95 (0.91 to 0.98)	<.001	.93 (0.86 to 0.96)	<.001	.53 (0.26 to 0.72)[b]	.01
7.2	30	.99 (0.97 to 0.992)	<.001	.98 (0.95 to 0.99)	<.001	.66 (0.43 to 0.81)[b]	<.001
Overall	N/A[c]	.97 (0.965 to 0.978)	<.001	.72 (0.65 to 0.77)	<.001	.62 (0.54 to 0.69)[b]	<.001

[a]Three decimal places were used for clarity when necessary.

[b]n=42.

[c]N/A: not applicable.

Table 3. Lin's concordance coefficient.

Walking speed (km/h)	n	Lin's concordance coefficient (95% CI)[a]		
		PiezoRx	Actigraph	Omron
2.4	43	0.69 (0.51 to 0.81)	0.02 (−0.02 to 0.05)	0.05 (−0.02 to 0.13)
3.2	43	0.80 (0.67 to 0.89)	0.12 (−0.01 to 0.25)	0.21 (0.09 to 0.32)
4.0	42	0.98 (0.96 to 0.99)	0.67 (0.49 to 0.80)	0.92 (0.86 to 0.95)
5.6	41	0.99 (0.97 to 0.991)	0.95 (0.90 to 0.97)	0.26 (0.09 to 0.41)
6.4	38	0.95 (0.91 to 0.97)	0.92 (0.85 to 0.96)	0.53 (0.34 to 0.68)
7.2	30	0.98 (0.96 to 0.99)	0.96 (0.93 to 0.98)	0.66 (0.45 to 0.80)
Overall	N/A[b]	0.97 (0.96 to 0.98)	0.64 (0.59 to 0.68)	0.61 (0.55 to 0.66)

[a]Three decimal places were used for clarity when necessary.

[b]N/A: not applicable.

Table 4. Absolute step difference between manually counted steps and steps measured using each device during treadmill walking.

Walking speed (km/h)	n	PiezoRx		Actigraph		Omron	
		Absolute step difference, % (SD)[a]	*P* value	Absolute step difference, % (SD)[a]	*P* value	Absolute step difference, % (SD)[a]	*P* value
2.4	43	6.2 (10.6)	.62	66.6 (22.3)	<.001	61.3 (37.1)	<.001
3.2	43	2.5 (5.4)	.81	18.0 (17.2)	<.001	19.2 (27.7)	.04
4.0	42	1.0 (1.0)	1.00	3.0 (6.5)	.20	1.8 (2.0)	1.00
5.6	41	0.8 (0.6)	1.00	1.3 (1.5)	.98	4.6 (16.3)	.05
6.4	38	1.0 (1.9)	1.00	1.5 (2.4)	.84	2.7 (8.6)	.30
7.2	30	1.1 (1.2)	.86	1.4 (1.4)	.98	2.8 (7.6)	.28
Overall	N/A[b]	2.2 (5.4)	.87	15.9 (26.8)	.50	15.0 (29.0)	.48

[a]Three decimal places were used for clarity when necessary.

[b]N/A: not applicable.

Figure 1. Bland-Altman plot analysis for manual step counting and the PiezoRx-determined step count.

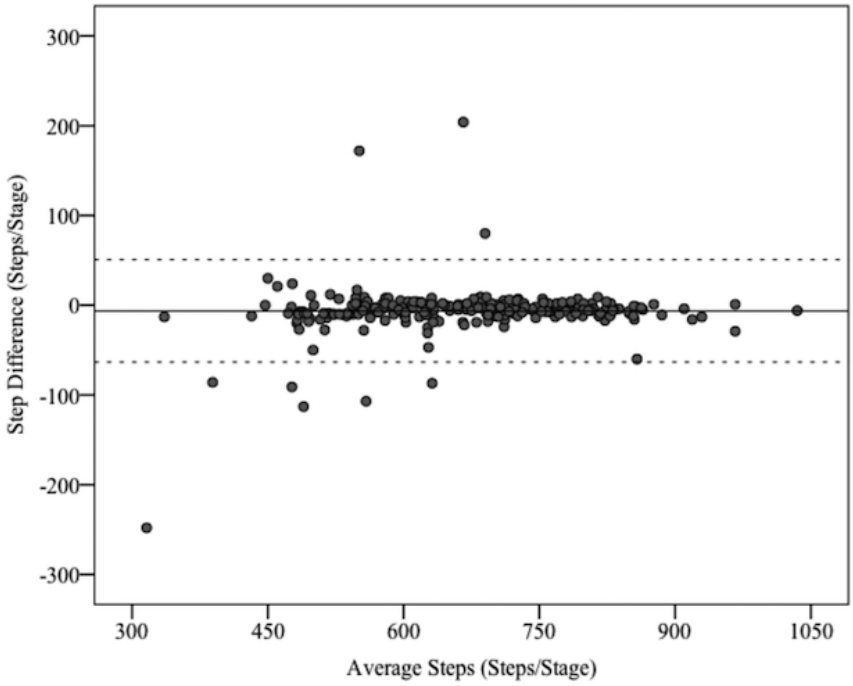


Table 5. Reliability of the PiezoRx in measuring step count and moderate-to-vigorous physical activity.

Walking speed (km/h)	n	Step count		Moderate-to-vigorous physical activity		
		Intraclass correlation (95% CI)	*P* value	Spearman correlation coefficient	*P* value	Absolute difference (sec/stage), % (SD)
2.4	25	.62 (.48 to.75)	<.001	.90	<.001	6.1 (17.3)
3.2	23	.59 (0.45 to 0.72)	<.001	.93	<.001	23.8 (55.1)
4.0	25	.33 (0.17 to 0.50)	.01	.92	<.001	22.8 (70.6)
5.6	23	.27 (0.12 to 0.49)	.02	.53	.02	2.4 (3.7)
6.4	23	.993 (0.988 to.996)	<.001	.78	.002	1.1 (2.4)
7.2	20	.98 (0.97 to 0.99)	<.001	.58	.01	3.0 (3.8)
Overall	N/A	.88 (0.86 to 0.90)	<.001	.95	<.001	10.4 (39.3)

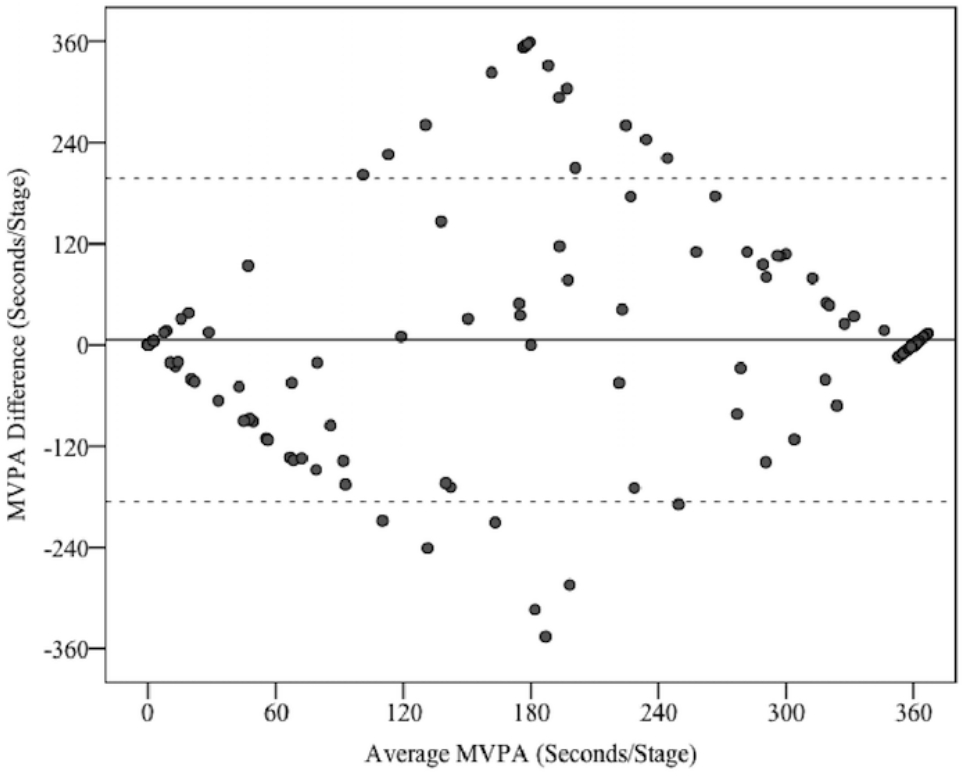
Table 6. Sensitivity and specificity rate of each activity monitor in measuring moderate-to-vigorous physical activity.

Monitor	Sensitivity (%)	Specificity (%)
PiezoRx (100/120)	70.9	91.7
PiezoRx (110/130)	90.3	79.2
PiezoRx (height adjusted)	82.1	94.3
PiezoRx (height and fitness adjusted)	81.6	91.0
ActiGraph	83.0	89.6

The PiezoRx (110/130) had the highest sensitivity percentage, whereas the height-adjusted PiezoRx had the highest specificity percentage (see Table 6). As shown in Figure 2, BAPs revealed a mean bias of 6.0 seconds (95% LoA: −186, 198) for the height-adjusted PiezoRx in measuring MVPA relative to the metabolic cart.

In the sample of individuals in which the height-adjusted and height+fitness-adjusted PiezoRxs were set at the same step rate thresholds (n=25), SCC of the devices in measuring MVPA was 0.95 (P<.001), as shown in Table 5. The interinstrument reliability was highest during the slowest stages (2.4-4.0 km/h; SCC>.90). However, the absolute MVPA differences were lowest during the faster stages (5.6-7.2 km/h; ≤3.0 sec per stage).

Figure 2. Bland-Altman plot analysis for indirect calorimetry and the height-adjusted PiezoRx-determined moderate-to-vigorous-intensity physical activity (MVPA).



Discussion

The objective of this study was to assess the validity and interinstrument reliability of a medical-grade PA monitor in measuring both step count and MVPA in a laboratory setting. Compared with manual step counting, the PiezoRx had the strongest correlations and a lower percent error than commonly used research grade monitors. When the PiezoRxs MVPA step rate thresholds were adjusted for height, the performance of the device in measuring intensity-related PA was similar to that of a research grade accelerometer.

The previous laboratory-based PiezoRx step count validation studies have shown that the device was accurate when used in a variety of populations [14-16]. Specifically, the 4-stage (1.4-14.1 km/h) step count evaluation of the SC-StepMX (previous version of the PiezoRx) by Colley et al [15] showed an excellent agreement with the manual step counting (r^2=0.97; APE=−0.2%). The findings of this study were in accordance with such findings for step count (APE=2.2%). Moreover, the PiezoRxs was found as a reliable measurement tool for step count across a variety of walking speeds (overall ICC=.88). ICCs were higher in the earlier (1 and 2) and later (5 and 6) stages but were lowest (ICC<.50) during stages 3 and 4 (4.0

and 5.6 km/h). However, the middle stages had the lowest APE (~1.0%) when the counted steps obtained using the PiezoRx were compared with the manually counted steps, and such result indicates that the low ICC may be due to the similarity in the number of steps in these stages (ie, less variance) with an average of 120±7 and 128±7 spm, respectively. Overall, the device is a reliable measurement tool for step count, particularly at slower walking cadence in which the PA monitoring devices are more likely to have difficulty measuring.

The only laboratory-based investigation on the validity of the PiezoRx device in measuring MVPA was conducted in children and young adults [14]. However, the PiezoRx has previously been used to measure MVPA in older adults [27], and it will be used in an upcoming randomized control trial aimed at increasing the number of steps per day of hospital employees [28]. Saunders et al [14] have shown that the device is a valid measurement tool for MVPA (r^2=0.64), and it was even more accurate when the step rate thresholds were increased from 100/120 to 110/130 (r^2=0.82), and this was identical to the correlation between indirect calorimetry and the ActiGraph. Such results are consistent with those of this study in which the height-adjusted PiezoRx and the ActiGraph were similarly correlated to indirect calorimetry (SCC=.80). BAP comparing MVPA measured via indirect calorimetry and the PiezoRx (height adjusted) resembles a diamond shape with a large cluster of data points with a mean difference of 0 (see Figure 2). The diamond shape is due to the dichotomous nature of measuring MVPA through walking stages in which the PiezoRx may incorrectly identify a stage as 360 sec of MVPA, whereas the metabolic cart did not detect any MVPA (average MVPA of 180 sec). In relation to this, the capability of the height-adjusted PiezoRx and ActiGraph in distinguishing each walking stage as MVPA or not was similarly good, as shown by their sensitivity rates (PiezoRx, H: 82%; ActiGraph: 83%) and specificity values (PiezoRx, H: 94%; ActiGraph: 90%). However, the PiezoRx devices set at 100/120, 110/130, and height+fitness adjusted were not significantly different (SCC: 0.74-0.80) than the height-adjusted PiezoRx.

The Yamax Digiwalker has been used as a criterion for comparing the capability of other devices in counting steps in free-living environments [29,30]. A previous research in adults has shown that the error rate of the Yamax Digiwalker (20.5%) was remarkably higher than that of the SC-StepMX (APE=0.2%) [15]. In addition, the Omron HJ-720 is more accurate than the ActiGraph GT3X+ and the Polar Active Accelerometer in laboratory-based and free-living conditions [29]. This study showed that the Omron and ActiGraph GT3X had similar validity in measuring step count (~15% APE).

As previously mentioned, the PiezoRx had a similar or higher accuracy rate than accelerometry in measuring intensity-related PA. Accelerometers may be costly and require proprietary software programs to access specific device data. Although accelerometers may be convenient for researchers, their usability among the general population is limited. After an explanation of the device's features and a 2-week monitoring, the performance of the PiezoRx was good among a diverse sample of adults and older adults [13]. More recently, the PiezoRxD has been developed which uses Bluetooth technology for uploading individual device data through a patient-focused mobile phone app that serves as a platform for health care providers and exercise professionals to access and provide their respective behavior changes or behavior maintenance approaches. In relation to this, the technology behind the determination of step count and MVPA is identical between the PiezoRx and PiezoRxD. Accurate devices that permit real-time step and MVPA monitoring may assist the general public in increasing their activity levels and may allow health care professionals and exercise professionals to objectively monitor and prescribe PA to their patients or clients with a goal of having more Canadians adhere to the PA guidelines [31].

The use of a submaximal assessment of aerobic fitness may be considered a limitation of this study. However, the single-stage treadmill protocol has been previously validated [21], and its use was to classify the aerobic fitness of the participants according to different $VO_{2\ max}$ ranges (excellent, very good, good, fair, or poor), which would be conducted in a clinical setting. As such, we believe that this limitation is minor relative to the validation of this device for clinical use. One limitation of PA studies is that they primarily appeal to physically active populations, potentially biasing the testing cohort to have a higher fitness compared with the general public. The population of this study had a wide range of BMI (17.9-42.5 kg/m^2) and predicted $VO_{2\ max}$ (20.4-58.9 mL/kg/min); therefore, our results are generalizable to the general public. Although the sample of the study was heterogeneous in nature and was designed to be representative of a typical patient population (56% answered "yes" to at least one question on the PAR-Q+), the participants were aged between 20 and 64 years; thus, the results may not be extrapolated to young or older adults. Hence, future studies should investigate the validity and reliability of the PiezoRx in measuring step count and MVPA in these populations who typically have shorter stride lengths and higher stride rates at a given walking velocity. In addition, the accuracy of the PiezoRx should be compared across a variety of movements, such as running, incline walking, stair walking, etc, that may affect the accuracy of this device.

The PiezoRx medical-grade PA monitor is a valid and reliable tool for measuring step count and intensity-related PA among a diverse sample of adults in a laboratory setting. The accuracy of the device in measuring MVPA may be similar to that of a research grade monitor, and it may be even more accurate than other frequently used PA monitors in measuring step counts. The PiezoRx may be a cost-effective alternative to research grade monitors that are used by primary care providers and exercise professionals in providing step-based exercise prescriptions.

Acknowledgments

MWO and WRW were supported by the Acadia University Research Fund.

Conflicts of Interest

Support was provided by the Acadia Research Support Fund and an unrestricted research grant from StepsCount/Diabeters Inc Grant #12-054. StepsCount had no influence on the design of the study, the analysis of the results or the writing of the paper.

References

1. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
2. Colley RC, Garriguet D, Janssen I, Craig CL, Clarke J, Tremblay MS. Physical activity of Canadian adults: accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. Health Rep 2011 Mar;22(1):7-14 [FREE Full text] [Medline: 21510585]
3. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
4. Qiu S, Cai X, Chen X, Yang B, Sun Z. Step counter use in type 2 diabetes: a meta-analysis of randomized controlled trials. BMC Med 2014 Feb 27;12:36 [FREE Full text] [doi: 10.1186/1741-7015-12-36] [Medline: 24571580]
5. Dasgupta K, Rosenberg E, Joseph L, Cooke AB, Trudeau L, Bacon SL, SMARTER Trial Group. Physician step prescription and monitoring to improve ARTERial health (SMARTER): A randomized controlled trial in patients with type 2 diabetes and hypertension. Diabetes Obes Metab 2017 Dec;19(5):695-704 [FREE Full text] [doi: 10.1111/dom.12874] [Medline: 28074635]
6. O'Brien MW, Shields CA, Oh PI, Fowles JR. Health care provider confidence and exercise prescription practices of Exercise is Medicine Canada workshop attendees. Appl Physiol Nutr Metab 2017 Apr;42(4):384-390 [FREE Full text] [doi: 10.1139/apnm-2016-0413] [Medline: 28177736]
7. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
8. Abel M, Hannon J, Mullineaux D, Beighle A. Determination of step rate thresholds corresponding to physical activity intensity classifications in adults. J Phys Act Health 2011 Jan;8(1):45-51. [Medline: 21297184]
9. Marshall SJ, Levy SS, Tudor-Locke CE, Kolkhorst FW, Wooten KM, Ji M, et al. Translating physical activity recommendations into a pedometer-based step goal: 3000 steps in 30 minutes. Am J Prev Med 2009 May;36(5):410-415. [doi: 10.1016/j.amepre.2009.01.021] [Medline: 19362695]
10. Beets MW, Agiovlasitis S, Fahs CA, Ranadive SM, Fernhall B. Adjusting step count recommendations for anthropometric variations in leg length. J Sci Med Sport 2010 Sep;13(5):509-512. [doi: 10.1016/j.jsams.2009.11.002] [Medline: 20096631]
11. Rowe DA, Welk GJ, Heil DP, Mahar MT, Kemble CD, Calabró MA, et al. Stride rate recommendations for moderate-intensity walking. Med Sci Sports Exerc 2011 Feb;43(2):312-318. [doi: 10.1249/MSS.0b013e3181e9d99a] [Medline: 20543754]
12. Sawyer BJ, Blessinger JR, Irving BA, Weltman A, Patrie JT, Gaesser GA. Walking and running economy: inverse association with peak oxygen uptake. Med Sci Sports Exerc 2010 Nov;42(11):2122-2127 [FREE Full text] [doi: 10.1249/MSS.0b013e3181de2da7] [Medline: 20351592]
13. O'Brien MW, Wojcik WR, D'Entremont L, Fowles JR. Validation of the PiezoRx® Step Count and Moderate to Vigorous Physical Activity Times in Free Living Conditions in Adults: A Pilot Study. Int J Exerc Sci 2018;11(7):541-551 [FREE Full text] [Medline: 29541342]
14. Saunders TJ, Gray CE, Borghese MM, McFarlane A, Mbonu A, Ferraro ZM, et al. Validity of SC-StepRx pedometer-derived moderate and vigorous physical activity during treadmill walking and running in a heterogeneous sample of children and youth. BMC Public Health 2014;14:519 [FREE Full text] [doi: 10.1186/1471-2458-14-519] [Medline: 24886319]
15. Colley RC, Barnes JD, Leblanc AG, Borghese M, Boyer C, Tremblay MS. Validity of the SC-StepMX pedometer during treadmill walking and running. Appl Physiol Nutr Metab 2013 May;38(5):520-524. [doi: 10.1139/apnm-2012-0321] [Medline: 23668759]
16. Webber SC, Magill SM, Schafer JL, Wilson KCS. GT3X+ accelerometer, Yamax pedometer and SC-StepMX pedometer step count accuracy in community-dwelling older adults. J Aging Phys Act 2014 Jul;22(3):334-341. [doi: 10.1123/japa.2013-0002] [Medline: 23921227]
17. Warburton D, Bredin S, Jamnik V, Gledhill N. Validation of the PAR-Q+ and ePARmed-X+. Health Fitness J Can 2011;4(2):38-46.
18. Fowles JR, O'Brien MW, Wojcik WR, d'Entremont L, Shields CA. A pilot study: Validity and reliability of the CSEP-PATH PASB-Q and a new leisure time physical activity questionnaire to assess physical activity and sedentary behaviours. Appl Physiol Nutr Metab 2017 Jun;42(6):677-680 [FREE Full text] [doi: 10.1139/apnm-2016-0412] [Medline: 28264170]
19. CSEP. Canadian Society for Exercise Physiology?Physical Activity Training for Health (CSEP-PATH). CSEP, Ottawa, Ont 2013.

20. Heyward V, Gibson A. Advanced Fitness Assesment and Exercise Prescription. Human Kinetics, Champaign, Ill 2014.
21. Ebbeling CB, Ward A, Puleo EM, Widrick J, Rippe JM. Development of a single-stage submaximal treadmill walking test. Med Sci Sports Exerc 1991 Aug;23(8):966-973. [Medline: 1956273]
22. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [Medline: 9588623]
23. Slaght J, Sénéchal M, Hrubeniuk TJ, Mayo A, Bouchard DR. Walking Cadence to Exercise at Moderate Intensity for Adults: A Systematic Review. J Sports Med (Hindawi Publ Corp) 2017;2017:4641203 [FREE Full text] [doi: 10.1155/2017/4641203] [Medline: 28459099]
24. Tudor-Locke C, Han H, Aguiar EJ, Barreira TV, Schuna JM, Kang M, et al. How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. Br J Sports Med 2018 Jun;52(12):776-788 [FREE Full text] [doi: 10.1136/bjsports-2017-097628] [Medline: 29858465]
25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
26. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989 Mar;45(1):255-268. [Medline: 2720055]
27. McLellan AG, Slaght J, Craig CM, Mayo A, Sénéchal M, Bouchard DR. Can older adults improve the identification of moderate intensity using walking cadence? Aging Clin Exp Res 2018 Jan;30(1):89-92. [doi: 10.1007/s40520-017-0746-3] [Medline: 28378230]
28. Mitchell M, White L, Oh P, Kwan M, Gove P, Leahey T, et al. Examining Incentives to Promote Physical Activity Maintenance Among Hospital Employees Not Achieving 10,000 Daily Steps: A Web-Based Randomized Controlled Trial Protocol. JMIR Res Protoc 2016 Dec 12;5(4):e231 [FREE Full text] [doi: 10.2196/resprot.6285] [Medline: 27956377]
29. Lee JA, Williams SM, Brown DD, Laurson KR. Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. J Sports Sci 2015;33(10):991-1000. [doi: 10.1080/02640414.2014.981848] [Medline: 25517396]
30. Tully MA, McBride C, Heron L, Hunter RF. The validation of Fibit Zip™ physical activity monitor as a measure of free-living physical activity. BMC Res Notes 2014;7:952 [FREE Full text] [doi: 10.1186/1756-0500-7-952] [Medline: 25539733]
31. Tremblay MS, Warburton DER, Janssen I, Paterson DH, Latimer AE, Rhodes RE, et al. New Canadian physical activity guidelines. Appl Physiol Nutr Metab 2011 Feb;36(1):36-46; 47 [FREE Full text] [doi: 10.1139/H11-009] [Medline: 21326376]

Abbreviations

APE: absolute percent error
BAP: Bland-Altman plots
BMI: body mass index
CSEP: Canadian Society for Exercise Physiology
ICC: intraclass correlation
LCC: Lin's concordance coefficient
LoA: limits of agreement
METs: metabolic equivalents
MPA: moderate-intensity aerobic physical activity
MVPA: moderate-to-vigorous-intensity physical activity
PA: physical activity
PAR-Q+: Physical Activity Readiness Questionnaire Plus
PASB-Q: Physical Activity and Sedentary Behaviour Questionnaire
RM-ANOVA: repeated measures analysis of variance
SCC: Spearman correlation coefficient
VPA: vigorous intensity physical activity

Edited by A Tsanas; submitted 05.04.18; peer-reviewed by A Triantafyllidis, S Webber; comments to author 29.04.18; revised version received 19.06.18; accepted 24.06.18; published 05.09.18.

Please cite as:
O'Brien MW, Wojcik WR, Fowles JR
Medical-Grade Physical Activity Monitoring for Measuring Step Count and Moderate-to-Vigorous Physical Activity: Validity and Reliability Study
JMIR Mhealth Uhealth 2018;6(9):e10706
URL: http://mhealth.jmir.org/2018/9/e10706/
doi:10.2196/10706
PMID:30185406

Original Paper

Assessment of Google Glass for Photographic Documentation in Veterinary Forensic Pathology: Usability Study

Giuseppe Piegari[1], DVM; Valentina Iovane[2], DVM, PhD; Vincenzo Carletti[3], PhD; Rosario Fico[4], DVM; Alessandro Costagliola[1], DVM, PhD; Davide De Biase[1], DVM; Francesco Prisco[1], DVM; Orlando Paciello[1], DVM, PhD

[1]Department of Veterinary Medicine and Animal Production, University of Naples Federico II, Naples, Italy

[2]Department of Pharmacy, University of Salerno, Fisciano, Italy

[3]Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, Fisciano, Italy

[4]Istituto Zooprofilattico Sperimentale delle Regioni Lazio e Toscana, National Center for the Forensic Veterinary Medicine, Grosseto, Italy

Corresponding Author:
Giuseppe Piegari, DVM
Department of Veterinary Medicine and Animal Production
University of Naples Federico II
Via F Delpino
Naples,
Italy
Phone: 39 0812536466
Email: giuseppe.piegari@unina.it

Abstract

Background: Google Glass is a head-mounted device designed in the shape of a pair of eyeglasses equipped with a 5.0-megapixel integrated camera and capable of taking pictures with simple voice commands.

Objective: The objective of our study was to determine whether Google Glass is fit for veterinary forensic pathology purposes.

Methods: A total of 44 forensic necropsies of 2 different species (22 dogs and 22 cats) were performed by 2 pathologists; each pathologist conducted 11 necropsies of each species and, for each photographic acquisition, the images were taken with a Google Glass device and a Nikon D3200 digital single-lens reflex (DSLR) camera. The pictures were collected, divided into 3 groups (based on the external appearance of the animal, organs, and anatomical details), and evaluated by 5 forensic pathologists using a 5-point score system. The parameters assessed were overall color settings, region of interest, sharpness, and brightness. To evaluate the difference in mean duration between necropsies conduced with Google Glass and DSLR camera and to assess the battery consumption of the devices, an additional number of 16 necropsies were performed by the 2 pathologists. In these cases, Google Glass was used for photographic reports in 8 cases (4 dogs and 4 cats) and a Nikon D3200 reflex camera in the other 8 cases. Statistical evaluations were performed to assess the differences in ratings between the quality of the images taken with both devices.

Results: The images taken with Google Glass received significantly lower ratings than those acquired with reflex camera for all 4 assessed parameters (P<.001). In particular, for the pictures of Groups A and B taken with Google Glass, the sum of frequency of ratings 5 (very good) and 4 (good) was between 50% and 77% for all 4 assessed parameters. The lowest ratings were observed for the pictures of Group C, with a sum of frequency of ratings 5 and 4 of 21.1% (342/1602) for region of interest, 26% (421/1602) for sharpness, 35.5% (575/1602) for overall color settings, and 61.4% (995/1602) for brightness. Furthermore, we found a significant reduction in the mean execution time for necropsy conduced with the Google Glass with respect to the reflex group (P<.001). However, Google Glass drained the battery very quickly.

Conclusions: These findings suggest that Google Glass is usable in veterinary forensic pathology. In particular, the image quality of Groups A and B seemed adequate for forensic photographic documentation purposes, although the quality was lower than that with the reflex camera. However, in this step of development, the high frequency of poor ratings observed for the pictures of Group C suggest that the device is not suitable for taking pictures of small anatomical details or close-ups of the injuries.

(JMIR Mhealth Uhealth 2018;6(9):e180) doi:10.2196/mhealth.9975

KEYWORDS

Google Glass; necropsy; pictures; documentation; veterinary forensic pathology; mobile phone

Introduction

Background

Google Glass is a device that was released for the first time as Google Glass explorer edition in 2013. It is a head-mounted device designed in the shape of a pair of eyeglasses with a 5.0-megapixel integrated camera; wireless connection; and the ability to take pictures, record a video, and call people with simple voice commands or manually by touching the frame. A small prism placed on the right side of the device allows a display of information to the user [1-3]. As a whole, the multitasking capabilities of the device provide users a comfortable and multifunctional virtual experience. Although these advantages have not fully met the needs of private consumers, its voice control, wireless transmission capabilities, integrated camera, and app customization have attracted the interest of commercial industries and professionals from various fields, including the health care [1,2]. In human medicine, Google Glass has been tested in many nonsurgical fields such as on-demand data visualization and real-time analysis [4], clinical simulations [5], management of diabetes [6], and pediatric cardiopulmonary resuscitation [7]. Furthermore, in neuropsychiatry, the usability and acceptability of Google Glass has been tested in children with autism spectrum disorder [8]. Similarly, in surgical settings, the multitasking capabilities of the device have allowed Google Glass to be tested in many surgical subfields such as cardiac surgery [9], neurosurgery [10], orthopedics [11], general surgery [12], and plastic surgery [13]. In these studies, Google Glass has been used as a tool to monitor vital signs, as an education instrument, and for telemonitoring and audiovisual recording. In human forensic pathology, Google Glass has been tested as a hands-free image acquisition device to document autopsies and postmortem examinations [14]. However, to the best of our knowledge, despite the several publications in human medicine, no empirical evidence for using Google Glass in veterinary medicine setting is currently known. The aim of this study was to determine the suitability of Google Glass in veterinary forensics pathology by assessing (1) the difference in mean duration between necropsies conduced with Google Glass and a digital single-lens reflex (DSLR) camera (Nikon D3200, lens AF-S DX Nikon 18-55 mm f/3.5-5.6G VR), (2) the battery consumption during the necropsies, (3) the usability aspects, and (4) the quality of the photographic documentation of the Google Glass compared with a DSLR camera.

Veterinary Forensic Pathology

Over the last years, forensic necropsies in veterinary medicine have rapidly increased due to the increasing demand for investigations of crimes against animals. For these reasons, the subfield of veterinary forensic pathology has emerged as a distinct discipline, essentially based on a transverse, multiorgan approach that includes necropsy, histological examination, immunohistochemistry, and collateral examinations such as laboratory analysis and diagnostic imaging to resolve obscure fatalities [15-17]. The range of interests of veterinary forensic pathology is very broad and includes unlawful killing and animal abuse, diagnosis of drowning, nonaccidental injuries, violation of wildlife laws, malpractice or disciplinary procedures, and support to human forensic pathology [18-25]. Although there is no rigid scheme for veterinary forensic necropsies, the Italian Group of Veterinary Forensic Pathologists standardized a procedure for forensic autopsy cases, which is a useful guide that should be followed but, at present, is nonbinding. On the basis of this protocol, the forensic necropsy starts with victim's identification and thanatological examination of the cadaver. Subsequently, a systematic evaluation of the external appearance of the animal is performed (outer necropsy). For this purpose, the head, mouth, mucous membranes, thorax, abdomen, perianal region, outer genitalia, hair coat, tail, as well as the front and hind limbs are surveyed. After that, the inner necropsy can start. During this step, a full skinning of the animal, inspection of the muscles and subcutis, as well as opening of all body cavities (skull, chest, abdomen, and pelvis) must be performed. Finally, all organs must be examined and dissected.

Image Acquisition in Veterinary Forensic Pathology

Photography is an important component of documentation in forensic pathology [17,26,27]. A correct and complete photographic documentation is also expressly required by the guidelines for forensic veterinary autopsies issued by the Italian Group of Forensic Veterinary Pathology. During the forensic necropsies, photography is important to document both the presence (positive photograph) and absence (negative photograph) of injuries [28]; the main aim is the acquisition of images useful for legal purposes. Since necropsy is an unrepeatable procedure, the forensic photographic documentation should not only be accurate and detailed but also produce a minimal delay in the execution time of the necropsy [26]. Many DSLR cameras and mobile phones with photographic capabilities can be used for this purpose. However, these devices need to be used by qualified personnel with knowledge of photography and basics of veterinary forensic pathology in order to take clear and understandable pictures and minimize distortion and misleading information [14,26,29]. Usually, this assignment is delegated to veterinary forensic pathologists themselves because there are not professional figures suited to this purpose. Furthermore, during autopsies, pathologists are forced to a continuous replacement of gloves so as to use cameras for documentary purposes. These limitations result in an excessive workload for the pathologists, with consequent lengthening of the time required to perform the necropsy. In this context, it is easy to imagine the advantage of having a device that allows taking hands-free pictures.

Methods

Forensic Necropsy Protocol and Image Acquisition

A total of 44 necropsies of 2 different species (22 dogs, 22 cats; dogs: medium-sized, age range 6-9 years, mean age 7.31 [SD 1.04] years; cats: age range 6-9 years, mean age 7.00 [SD 0.92] years) were performed by 2 pathologists (FP and GP) with training in forensic medicine. All necropsies were performed in the necropsy room of the Department of Veterinary Medicine and Animal Productions at the University of Naples Federico II, Naples, Italy, following a standard necropsy protocol summarized in Textbox 1.

Textbox 1. Forensic autopsy protocol.

1. Victim identification procedures
2. Evaluation of thanatological aspects and estimation of the time elapsed since death
3. External examination of the body (state of nutrition, mucous membranes, body orifices, general conformation, superficial lesions, hair coat, external parasites, and teeth)
4. Skinning with evaluation of subcutis and muscles
5. Opening and evaluation of body cavities (skull, thorax, abdomen, and pelvis)
6. Extraction and general macroscopic evaluation of organs
7. Dissection of all organs
8. Specific evaluation of wounds or injuries
9. Complete photographic documentation of external appearance of the animals, body cavities, organs, and injuries

Each pathologist conducted 11 necropsies of both species. For each photographic acquisition, the images were taken using two different devices: Google Glass and a Nikon D3200 DSLR camera. During the external inspection of the body, in accordance with the guidelines for forensic veterinary autopsies, pathologists were asked to take pictures of the external appearance of the animals from many different angles. Furthermore, during the necropsies, pictures of organs before and after extraction and any other detail useful for documentation purposes was acquired. All images were acquired under standard lighting conditions and without using the internal flash of the camera. In addition, a standard background (blue table of 90 × 70 cm) was used to acquire pictures of organs and small anatomical details. Finally, during image acquisition, a photomacrographic scale (American Board of Forensic Odontology No. 2 Standard Reference Scale) placed near the injuries was used to provide a geometrical reference in the forensic photographic documentation of the evidences.

Evaluation of the Time of Necropsy and Battery Performance

To evaluate the differences in time of necropsy between autopsies conduced with Google Glass and DSLR camera and the battery performance of the devices, an additional number of 16 necropsies of 2 different species (8 cats and 8 dogs; dogs: medium-large dogs, age range 8-10 years, mean age 8.75 [SD 0.88] years; cats: age range 7-9 years, mean age 8.0 [SD 1.19] years) were performed by the 2 pathologists (FP and GP). In these cases, each pathologist conducted 4 necropsies of each species, with half of them conducted using Google Glass and another half using the Nikon D3200 DSLR camera. For each postmortem examination, we measured the time required to perform the necropsy using the stopwatch functionality available on a smartphone iPhone 6s Plus. To standardize the measurement, for all 16 forensic necropsies, the timer started at the first photographic acquisition and ended when the pathologist declared that he had acquired all pictures useful for the documentation purpose. Furthermore, each forensic examination began with devices (DSLR camera and Google Glass) charged to 100%, and at the end of each necropsy, the remaining battery power was noted.

Usability Aspect

At the end of each necropsy performed with the Google Glass, pathologists were interviewed to acquire information about the user experience. The questions were designed to obtain information about the usability aspects, general experiences, and the main positive and negative features of the device.

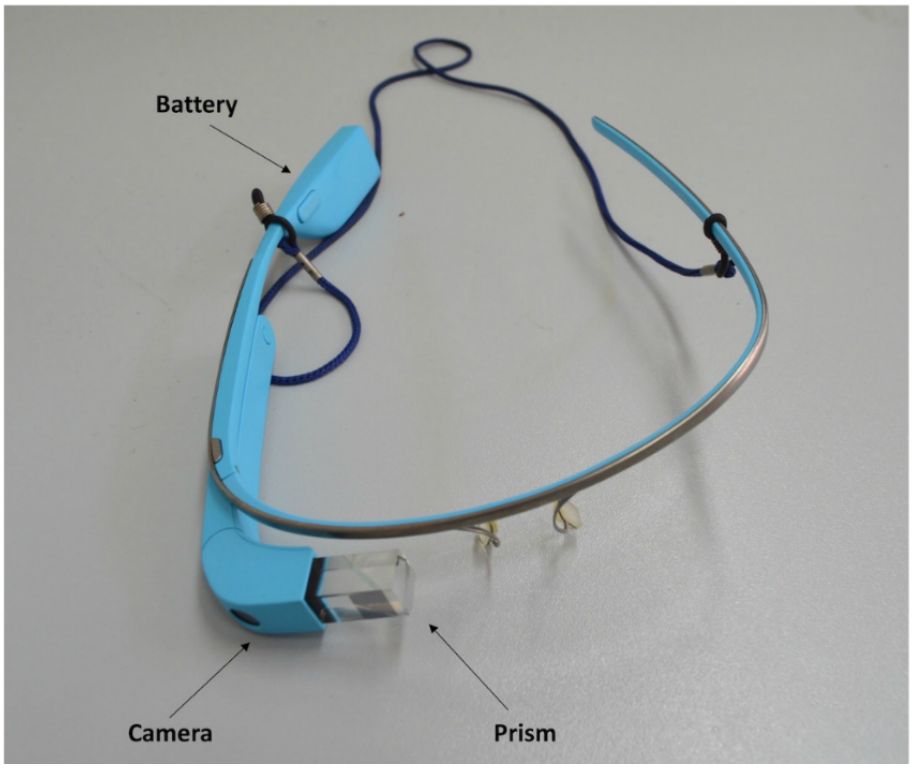
Google Glass

The device—a Google Glass explorer edition—available during our study, ran on Android 4.4.2. specifications of the available developer explorer unit that included Texas Instrument OMAP 4430 SoC 1.2 GHz Dual core (ARMv7) processor, 2 GB of RAM and 12 GB of usable storage space, a 640 × 360 display, 802.11b/g Wi-Fi, Bluetooth, and a 5-megapixel camera [3,14]. It also had a 3-axis gyroscope, a 3-axis accelerometer, a 3-axis magnetometer, ambient light sensor, proximity sensor, bone conduction audio transducer, and 2 omnidirectional microphones (Figure 1).

Software Setup

For image acquisition, we used the preinstalled camera app for two reasons. First, we did not know whether the use of an app other than the preinstalled one would increase the battery consumption or decrease the photographic quality. Second, the voice commands and gestures performed on running the preinstalled app were easy to perform, intuitive, and precise; thus, we were not inclined to use an accessory app. However, to properly use the Google Glass camera during the necropsies, both pathologists followed a training course that lasted approximately 15 minutes. At the end of it, the pathologists declared to be able to use the devise correctly. The voice commands used during our study were as follows: “show viewfinder,” to frame the anatomical reason of interest correctly, and “take a picture,” to acquire the images. All accepted images were stored in a folder on the device until the pictures were transmitted via USB to an iOS-based laptop (MacBook Air 13").

Figure 1. Google Glass device.



Digital Quality Image Assessment

A total of 5 forensic pathologists (AC, OP, RF, DDB, and VI; 4 males, 1 female; age range 33-58 years, mean age 42.60 [SD 9.37] years) with a mean work experience of 19 (SD 7.41) years in the field of veterinary pathology, both diagnostic and teaching, were selected to evaluate the quality of the images taken with both devices. To avoid compromising the evaluation of the images by the memories of the necropsies, the pathologists selected for the evaluation of the images were different from those who physically performed them. Furthermore, before the beginning of the evaluation, the pictures were divided into 3 groups: Group A, pictures of the external appearance of the animals; Group B, pictures of the organs; and Group C, pictures of small anatomical details or close-ups of the injuries. Each group included images taken with both devices. However, the device used for the acquisition was known (DSLR camera or Glass) only to the coordinators of the study. All 5 pathologists separately evaluated each group. In addition, all pictures were presented on a same computer (MacBook Air 13") with fixed display settings and similar environmental lighting conditions to avoid differences of evaluation due to external variables. The pathologists gave their opinions individually about the quality of the images analyzing 4 parameters: (1) overall color setting; (2) sharpness; (3) region of interest; and (4) brightness. Each one of these 4 parameters was separately evaluated on a 5-point score system according to Albrecht et al [14] (Table 1).

Statistical Analysis

Statistics were computed using SPSS Version 22.0 (IBM Corporation, 2014, Armonk, NY, United States). Student's *t* test was used to evaluate the difference in the mean duration between the necropsies conducted with Google Glass and the DSLR camera. The descriptive statistics for the ratings consisted of the tabulation of the frequency and percentages of scale items for each group for each item per device. To evaluate the differences between the ratings obtained for each group for both devices, we calculated an unpaired rank sum test (2-sided Mann-Whitney *U* test, with Cronbach alpha=.05) [14,30]. The same test was used to detect differences among the groups of the same device. To evaluate interrater reliability, we calculated intraclass correlation coefficients (ICCs) for the 5 pathologists for the items region of interest, sharpness, brightness, and color discrimination [31].

Table 1. Scoring system for image quality.

Parameter	Very poor	Poor	Average	Good	Very good
Sharpness	1	2	3	4	5
Overall color settings	1	2	3	4	5
Region of interest	1	2	3	4	5
Brightness	1	2	3	4	5

Results

Digital Image Quality Assessment

During the 44 forensic autopsies, the pathologists took 985 pictures with Google Glass and 985 pictures with the D3200 DSLR camera (504 photos of dogs and 481 photos of cats with each device). Each picture was evaluated by 5 pathologists, resulting in 4925 single evaluations each for Google Glass and the DSLR camera. Table 2 summarizes the results of the absolute frequencies and percentages of the ratings obtained per group for both devices for each of the 4 assessed parameters.

The ratings of the images taken with Google Glass during necropsies were significantly lower than those of the images acquired with the DSLR camera (Table 3). In particular, considering the percentage values, most ratings of the images taken with DSLR camera were high (good or very good) for all 4 parameters assessed. In contrast, for the images of Group A taken with Google Glass, the sum of frequency of ratings 5 (very good) and 4 (good) was 77.3% (1390/1800), 66.4% (1195/1800), 70.4% (1268/1800), and 71.7% (1290/1800) for region of interest, sharpness, brightness, and overall color settings, respectively. Furthermore, the images of Group B taken with Google Glass received a sum of frequency of ratings 5 and 4 of 54.7% (823/1505), 55.7% (838/1505), 65.8% (990/1505), and 54.0% (813/1505) for region of interest, sharpness, brightness, and overall color settings, respectively.

Table 2. Frequencies and percentages of evaluations given by 5 pathologists for images taken during forensic necropsies with Google Glass and Nikon D3200 reflex camera stratified for each group.

Parameters assessed and score[a]	Group A, n (%)		Group B, n (%)		Group C, n (%)	
	Glass (n=1800)	DSLR[b] camera (n=1800)	Glass (n=1505)	DSLR camera (n=1505)	Glass (n=1602)	DSLR camera (n=1602)
Region of interest						
5	426 (23.7)	933 (51.8)	173 (11.5)	522 (34.7)	55 (3.4)	1027 (63.4)
4	964 (53.6)	803 (44.6)	650 (43.2)	898 (59.7)	287 (17.7)	519 (32.0)
3	393 (21.8)	64 (3.6)	622 (41.3)	85 (5.6)	827(51.1)	74 (4.6)
2	17 (0.9)	0 (0)	60 (4)	0 (0)	451 (27.8)	0 (0)
1	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Sharpness						
5	353 (19.6)	1005 (55.8)	37 (2.5)	434 (28.8)	125 (7.7)	762 (47)
4	842 (46.8)	708 (39.4)	801 (53.2)	929 (61.7)	296(18.3)	718 (44.4)
3	580 (32.2)	87 (4.8)	597 (39.7)	142 (9.4)	735 (45.4)	140 (8.6)
2	25 (1.4)	0 (0)	70 (4.6)	0 (0)	464 (28.6)	0 (0)
1	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Overall color settings						
5	268 (14.9)	912 (50.7)	100 (6.6)	852 (56.6)	2 (0.1)	813 (50.2)
4	1022 (56.8)	783 (43.5)	713 (47.4)	619 (41.1)	573 (35.4)	742 (45.8)
3	507 (28.1)	105 (5.8)	582 (38.7)	34 (2.3)	770 (47.5)	65 (4.0)
2	3 (0.2)	0 (0)	110 (7.3)	0 (0)	275 (17)	0 (0)
1	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Brightness						
5	285 (15.8)	997 (55.4)	204 (13.6)	680 (45.2)	125 (7.7)	880 (54.3)
4	983(54.6)	688 (38.2)	786(52.2)	806(53.6)	870 (53.7)	645 (39.8)
3	482 (26.8)	115 (6.4)	494 (32.8)	19 (1.3)	513 (31.7)	95 (5.9)
2	50 (2.8)	0 (0)	21 (1.4)	0 (0)	112 (6.9)	0 (0)
1	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

[a]Parameters were assessed on a scale of 1-5 (1, very poor; 2, poor; 3, average; 4, good, 5, very good).

[b]DSLR: digital single-lens reflex.

Table 3. Unpaired rank sum, 2-sided Mann-Whitney *U* (Cronbach alpha=.05) for ratings of images taken with Google Glass and Nikon D3200 reflex camera for each group for each of the 4 assessed parameters.

Assessed parameters	Group A		Group B		Group C	
	z score	*P* value	z score	*P* value	z score	*P* value
Region of interest	−20.211	<.001	−26.771	<.001	−44.398	<.001
Sharpness	−26.468	<.001	−29.412	<.001	−37.705	<.001
Overall color settings	−25.177	<.001	−38.475	<.001	−40.931	<.001
Brightness	−26.786	<.001	−28.283	<.001	−31.452	<.001

The lowest ratings were observed in the pictures of Group C taken with Google Glass, with a sum of frequency of ratings 5 and 4 of 21.1% (342/1602) for region of interest, 26% (421/1602) for sharpness, 35.5% (575/1602) for overall color settings, and 61.4% (995/1602) for brightness. In this group, the differences between devices were particularly noticeable for region of interest, overall color settings, and sharpness. With regard to region of interest, the sum of frequency of ratings 2 and 3 amounted to 4.6% (74/1602) for images acquired using the DSLR camera versus 78.9% (1278/1602) for those acquired using Google Glass. Similarly, for overall color settings, this sum was 4% (65/1602) for images acquired using the DSLR camera versus 64.5% (1045/1602) for those acquired using Google Glass. Finally, for sharpness, the sum was 8.6% (140/1602) for images taken with DSLR versus 74% (1199/1602) for those taken with Google Glass. Furthermore, with regard to the pictures taken with Google Glass, statistical differences were observed in the distribution of ratings between Group A and Group B and between Group C versus B and A for all 4 assessed parameters (Table 4).

Evaluation of Battery Performance and Time of Necropsy

Of the 16 necropsies conducted with Google Glass (8 necropsies) and DSLR camera (8 necropsies), we observed a reduction in the time of necropsy with Google Glass compared with that with the DSLR camera group. The mean duration of a single postmortem examination in the DSLR camera group was 126.38 (SD 3.46) minutes for the dogs group and 68.90 (SD 2.30) minutes for the cats group, whereas for the Google Glass, the mean duration was 111.11 (SD 3.29) minutes for dogs group and 55.5 (SD 2.06) for cats group. The differences were significant (*P*<.001). Furthermore, at the end of the necropsies conducted with Google Glass, the average percentage of battery power was 47% and 60% for the dogs and cats group, respectively (Table 5). For the DSLR camera, it was not possible to monitor the battery level because the display showed an icon with a crude scale in the unit of 33% and not the battery percentage. In any case, a significant battery consumption was not detected.

Usability

Based on interviews conducted at the end of the postmortem examination, we obtained subjective assessments about the user experience of Google Glass. As a positive aspect, the voice control was reported as useful, particularly in cases where both hands were occupied. The use of voice control led to increased saving of rubber gloves because the pathologists were not forced to take them off whenever they needed to take some pictures. In addition, the pathologists agreed about the ergonomics of the device and its lightness, which makes it comfortable to wear. As negative aspects, they reported the short battery life and difficulty to capture the desired regions of interest, especially for the close-ups. During use, the device was placed on the head and there was no zoom function available. For these reasons, the pathologists were forced to place themselves too close to the dissection table to be able to take pictures of small anatomical details or close-ups of the injuries correctly (Figure 2).

Interrater Reliability

The interrater reliability was high. The ICC for the ratings obtained based on the forensic necropsy pictures indicated a good positive relationship for overall color settings (0.815, 95% CI 0.771-0.853; *P*<.001), region of interest (0.787, 95% CI 0.750-0.819; *P*<.001), sharpness (0.711, 95% CI 0.632-0.775, *P*<.001), and brightness (0.822, 95% CI 0.777-0.860; *P*<.001).

Table 4. Unpaired rank sum, 2-sided Mann-Whitney *U* (Cronbach alpha=.05) for ratings of the images of the 3 groups taken with Google Glass.

Assessed parameters	Group A versus B		Group B versus C		Group C versus A	
	z score	*P* value	z score	*P* value	z score	*P* value
Region of interest	−14.577	<.001	−22.603	<.001	−34.101	<.001
Sharpness	−11.396	<.001	−18.256	<.001	−26.095	<.001
Overall color settings	−12.740	<.001	−12.515	<.001	−25.412	<.001
Brightness	−2.737	.006	−5.052	<.001	−8.005	<.001

Table 5. Descriptive statistics of loss of battery consumption during forensic necropsy stratified by the device used to acquire the images: a Nikon D3200 reflex camera and the Google Glass device.

Device	Mean (SD)
Digital single-lens reflex camera	N/A[a]
Glass (dogs group)	47 (2.6)
Glass (cats group)	60 (2.9)

[a]N/A: not applicable.

Figure 2. The forensic pathologist wears Google Glass and takes pictures of small anatomical details in a hands-free manner.

Discussion

Principal Findings

Potentially disruptive technologies such as Google Glass create excitement for the possible applications that they can have in health care; however, the use of new tools should be thoroughly evaluated and validated before applying them in the medical or biomedical fields. Google Glass has been tested in many medical fields such as clinical simulations [5], surgery [9-13], and neuropsychiatry [8], but to the authors' knowledge, this is the first study to assess the potential of Google Glass in veterinary medicine. Both devices tested in our study achieved the set goals and both allowed a complete photographic documentation. However, differences in efficiency between the two devices were observed. Our study showed a reduction in the necropsy time in forensic examinations conducted with Google Glass compared with those conducted with the Nikon D3200 DSLR camera; this was because Google Glass allowed hands-free operation, avoiding the continuous replacement of gloves that is necessary during necropsies performed with a reflex camera. Although glove saving was not a parameter directly evaluated in this study, this aspect was highlighted by both pathologists during the interviews conducted at the end of the necropsies. In addition, we observed a rapid reduction in the battery life of the Google Glass. In our opinion, power consumption was not a limiting factor because at the end of the necropsies, the average battery percentage was 47% and 60% for the dogs and cats groups, respectively. However, problems could arise in the case of more autopsies being performed on the same day. In these cases, pauses to allow battery charging should be considered. The difference in average Google Glass power consumption observed between the groups could be due to the different mean duration of a single postmortem examination of the dogs group (111.11 minutes) compared with the lower duration of the cats group (55.5 minutes) and, consequently, the mean duration of Google Glass use. Regarding the DSLR camera, it was not possible to monitor the battery level because the display showed an icon with a crude scale in units of 33% and not the battery percentage. In any case, significant loss of battery power was not detected. This is understandable considering the high capacity of the batteries commonly used in the modern DSLR cameras. Finally, images taken with Google Glass received significantly lower ratings for all 4 assessed parameters than those taken with the DSLR reflex camera. Most ratings for the images taken with the DSLR camera were high (good or very good) for all 4 assessed parameters. In contrast, in the pictures of the Group A taken with Google Glass, the sum of the frequency of ratings 5 (very good) and 4 (good) was 77.3% (1390/1800), 66.4% (1195/1800), 70.4% (1268/1800), and 71.7% (1290/1800) for region of interest, sharpness, brightness, and overall color settings, respectively; furthermore, the images of Group B taken with Google Glass received a sum of the frequency of ratings 5 and 4 of 54.7% (823/1505), 55.7%

(838/1505), 65.8% (990/1505), and 54.0% (813/1505), respectively. The lowest ratings were observed for the pictures of group C with a sum of the frequency of ratings 5 and 4 of 21.1% (342/1602) for region of interest, 26.0% (421/1602) for sharpness, 35.5% (575/1602) for overall color settings, and 61.4% (995/1602) for brightness. This finding could be explained considering the lower camera resolution of Google Glass (5-megapixel camera with a fixed focal length of 3 mm) compared with the DSLR camera (24 megapixel) [32]. Hashimoto et al [12], in a recent study, reported that the video quality of iPhone 5 was greater than that of Google Glass during human surgical telementoring sessions. In contrast, some eye surgeons have reported a good video quality of the Google Glass device during scleral buckling surgeries [33]. These findings suggest that the camera quality of Google Glass is evaluated differently depending on the medical field of application and, consequently, on the photographic and recording quality required. Specifically, in our study, the gradual reduction in good or very good ratings observed among Groups A, B, and C could reflect the progressive increase in photographic quality required for the evaluation of anatomical details compared with that required for external examination for the evaluation of the body or organs. Albrecht et al [14], in a study conducted to assess the quality of the photographic documentation of the Google Glass device in human forensic pathology, showed a lower picture quality of the images taken with Google Glass than of those taken with a DSLR camera. In particular, the authors found that the differences between the devices were particularly noticeable only for region of interest and sharpness, whereas brightness and overall color settings showed similar distributions of ratings, with results slightly in favor of the pictures taken with the DSLR camera. These results appear in apparent contradiction with those obtained in our study. However, in this previous study, the images were evaluated as a whole and without being divided by type (external appearance, organs, and anatomical details). Furthermore, in the same study, an external application was used for image acquisition and a lower number of pictures was evaluated. However, excluding the methodological differences, the different results obtained in this study could be explained considering the different size of the organs of pet animals compared with human anatomy. These differences suggest a greater difficulty in the evaluation of images of organs and injuries of pet animals than that of humans. Similarly, for the external examination of the body, the difference in the size of animals compared with human anatomy and the presence of hair, common in all species of veterinary interest but absent in humans, makes the qualitative camera differences between devices more evident in animals than in humans. In addition, in our study, regarding the images taken with Google Glass, statistical analysis showed differences between the groups for all 4 assessed parameters. However, the highest frequency of rating 2 (poor) was observed for the pictures of Group C. The lower ratings observed for region of interest in Group C compared with those in Groups A and B was the most important aspect, and it could be explained considering that the Google Glass device had a wide-angle lens but not a zoom function [12,32]. During the forensic necropsy, the field of view of the Google Glass camera was too large, which forced the pathologists to place themselves too close to the dissection table to acquire pictures of small anatomical details. This led to the acquisition of poorly framed images that were unsuitable for forensic documentary purposes. Similarly, Moshtaghi et al [34], in a study conducted to assess the feasibility of using Google Glass during otorhinolaryngological procedures, found that the image quality was inadequate for viewing small and deep-seated anatomical structures. However, in our opinion, although this aspect has already been observed in human medicine, it could be even more relevant in pet animals than humans because of the differences in size between these species.

Limitations

A few limitations of the study should be noted. First, we were able to test the device with only 2 pathologists. However, this allowed a more accurate assessment of the necropsy execution times. In our opinion, a greater number of pathologists would have determined a high variability in the time of necropsy due to the different levels of manual dexterity of each pathologist. Second, for the evaluation of the images, this study was based on subjective opinions of raters and not on objective and reproducible parameters. However, to reduce this limitation, a high number of highly qualified pathologists was selected for the evaluation of the images. In addition, the ICCs were evaluated for the 5 raters for the items region of interest, sharpness, brightness, and color discrimination. Furthermore, this study was conducted only on cats and dogs. We decided to test the device on these animals because although diagnostic necropsies are commonly performed on a broad range of animals, at present, they are the main species of forensic interest in the veterinary field [35,36]. However, this limitation makes these findings nonreproducible on other species of veterinary interest such as mice, rats, rabbits, or zoo and farm animals. Finally, the joint evaluation of the acquired images from 2 different species could be a further limitation of this study. However, considering the similar morphology and size of these animals (medium-sized dogs vs adult cats), we do not believe this to be a limitation.

Conclusions

These findings suggest that Google Glass is usable in the veterinary forensic pathology of pet animals, but its image quality is lower than that of a reflex camera. In particular, the image quality of Groups A and B seemed adequate for forensic photographic documentation purposes. However, in this step of development, the high frequency of poor ratings observed for the pictures of Group C suggest that the device is not suitable for taking pictures of small anatomical details or close-ups of the injuries. In our opinion, the combined use of the two devices, reflex camera for capturing images of small anatomical details and Google Glass for capturing images of the external appearance of the animals and organs, could reduce the execution times of the necropsy, lead to considerable saving of gloves, and allow acquisition of pictures useful for forensic documentation purposes. However, further studies will be needed to evaluate the application of this device to other species of veterinary interest such as wildlife or farm animals. In some of these species, the greater volume of the organs than that in pet animals could make the qualitative differences between

Google Glass and the reflex camera less evident but, above all, could make the absence of the photographic zoom in Google Glass less limiting.

Authors' Contributions

GP drafted the manuscript and contributed to the study concept, study design, and the analysis and interpretation of data. GP and FP conducted the necropsies and statistical analysis of data. OP, AC, RF, DDB and VI evaluated the quality of the images taken with both devices. RF, DDB, VI, VC, AC and OP revised the manuscript for content and contributed to the interpretation of data. OP also contributed to the study concept and design.

Conflicts of Interest

None declared.

References

1. Wei NJ, Dougherty B, Myers A, Badawy SM. Using Google Glass in Surgical Settings: Systematic Review. JMIR Mhealth Uhealth 2018 Mar 06;6(3):e54 [FREE Full text] [doi: 10.2196/mhealth.9409] [Medline: 29510969]
2. Dougherty B, Badawy SM. Using Google Glass in Nonsurgical Medical Settings: Systematic Review. JMIR Mhealth Uhealth 2017 Oct 19;5(10):e159 [FREE Full text] [doi: 10.2196/mhealth.8671] [Medline: 29051136]
3. Google. Glass: Frequently Asked Questions-Tech Specs URL: https://support.google.com/glass/answer/3064128?hl=en&ref_topic=3063354 [WebCite Cache ID 6MLscCTn0]
4. Zhang YS, Busignani F, Ribas J, Aleman J, Rodrigues TN, Shaegh SAM, et al. Google Glass-Directed Monitoring and Control of Microfluidic Biosensors and Actuators. Sci Rep 2016 Mar 01;6:22237 [FREE Full text] [doi: 10.1038/srep22237] [Medline: 26928456]
5. Chaballout B, Molloy M, Vaughn J, Brisson IR, Shaw R. Feasibility of Augmented Reality in Clinical Simulations: Using Google Glass With Manikins. JMIR Med Educ 2016 Mar 07;2(1):e2 [FREE Full text] [doi: 10.2196/mededu.5159] [Medline: 27731862]
6. Wall D, Ray W, Pathak RD, Lin SM. A google glass application to support shoppers with dietary management of diabetes. J Diabetes Sci Technol 2014 Nov;8(6):1245-1246 [FREE Full text] [doi: 10.1177/1932296814543288] [Medline: 25015954]
7. Siebert JN, Ehrler F, Gervaix A, Haddad K, Lacroix L, Schrurs P, et al. Adherence to AHA Guidelines When Adapted for Augmented Reality Glasses for Assisted Pediatric Cardiopulmonary Resuscitation: A Randomized Controlled Trial. J Med Internet Res 2017 May 29;19(5):e183 [FREE Full text] [doi: 10.2196/jmir.7379] [Medline: 28554878]
8. Sahin NT, Keshav NU, Salisbury JP, Vahabzadeh A. Second Version of Google Glass as a Wearable Socio-Affective Aid: Positive School Desirability, High Usability, and Theoretical Framework in a Sample of Children with Autism. JMIR Hum Factors 2018 Jan 04;5(1):e1 [FREE Full text] [doi: 10.2196/humanfactors.8785] [Medline: 29301738]
9. Schaer R, Salamin F, Jimenez DTOA, Atzori M, Muller H, Widmer A. Live ECG readings using Google Glass in emergency situations. Conf Proc IEEE Eng Med Biol Soc 2015;2015:315-318. [doi: 10.1109/EMBC.2015.7318363] [Medline: 26736263]
10. Nakhla J, Kobets A, De LGRR, Haranhalli N, Gelfand Y, Ammar A, et al. Use of Google Glass to Enhance Surgical Education of Neurosurgery Residents: “Proof-of-Concept” Study. World Neurosurg 2017 Feb;98:711-714. [doi: 10.1016/j.wneu.2016.11.122] [Medline: 27923752]
11. Armstrong DG, Rankin TM, Giovinco NA, Mills JL, Matsuoka Y. A heads-up display for diabetic limb salvage surgery: a view through the google looking glass. J Diabetes Sci Technol 2014 Sep;8(5):951-956 [FREE Full text] [doi: 10.1177/1932296814535561] [Medline: 24876445]
12. Hashimoto DA, Phitayakorn R, Fernandez-del CC, Meireles O. A blinded assessment of video quality in wearable technology for telementoring in open surgery: the Google Glass experience. Surg Endosc 2016 Jan;30(1):372-378. [doi: 10.1007/s00464-015-4178-x] [Medline: 25829065]
13. Sinkin JC, Rahman OF, Nahabedian MY. Google Glass in the Operating Room: The Plastic Surgeon's Perspective. Plast Reconstr Surg 2016 Jul;138(1):298-302. [doi: 10.1097/PRS.0000000000002307] [Medline: 27348661]
14. Albrecht U, von JU, Kuebler J, Zoeller C, Lacher M, Muensterer OJ, et al. Google Glass for documentation of medical findings: evaluation in forensic medicine. J Med Internet Res 2014;16(2):e53 [FREE Full text] [doi: 10.2196/jmir.3225] [Medline: 24521935]
15. Wobeser G. Forensic (medico-legal) necropsy of wildlife. J Wildl Dis 1996 Apr;32(2):240-249. [doi: 10.7589/0090-3558-32.2.240] [Medline: 8722261]
16. McDonough SP, Gerdin J, Wuenschmann A, McEwen BJ, Brooks JW. Illuminating dark cases: veterinary forensic pathology emerges. Vet Pathol 2015 Jan;52(1):5-6. [doi: 10.1177/0300985814551582] [Medline: 25540449]
17. Brownlie HWB, Munro R. The Veterinary Forensic Necropsy: A Review of Procedures and Protocols. Vet Pathol 2016 Dec;53(5):919-928. [doi: 10.1177/0300985816655851] [Medline: 27371536]
18. Millins C, Howie F, Everitt C, Shand M, Lamm C. Analysis of suspected wildlife crimes submitted for forensic examinations in Scotland. Forensic Sci Med Pathol 2014 Sep;10(3):357-362. [doi: 10.1007/s12024-014-9568-1] [Medline: 24935435]

19. Munro HM, Thrusfield MV. 'Battered pets': non-accidental physical injuries found in dogs and cats. J Small Anim Pract 2001 Jun;42(6):279-290. [Medline: 11440397]
20. Salem NH, Belhadj M, Aissaoui A, Mesrati MA, Chadly A. Multidisciplinary approach to fatal dog attacks: a forensic case study. J Forensic Leg Med 2013 Aug;20(6):763-766. [doi: 10.1016/j.jflm.2013.04.015] [Medline: 23910877]
21. Stern AW, Smith-Blackmore M. Veterinary Forensic Pathology of Animal Sexual Abuse. Vet Pathol 2016 Dec;53(5):1057-1066. [doi: 10.1177/0300985816643574] [Medline: 27169881]
22. McEwen BJ, Gerdin J. Veterinary Forensic Pathology: Drowning and Bodies Recovered From Water. Vet Pathol 2016 Dec;53(5):1049-1056. [doi: 10.1177/0300985815625757] [Medline: 26926081]
23. McEwen BJ. Trends in domestic animal medico-legal pathology cases submitted to a veterinary diagnostic laboratory 1998-2010. J Forensic Sci 2012 Sep;57(5):1231-1233. [doi: 10.1111/j.1556-4029.2012.02123.x] [Medline: 22458814]
24. Gerdin JA, McDonough SP. Forensic pathology of companion animal abuse and neglect. Vet Pathol 2013 Nov;50(6):994-1006. [doi: 10.1177/0300985813488895] [Medline: 23686766]
25. Aquila I, Di NC, Paciello O, Britti D, Pepe F, De LE, et al. An unusual pedestrian road trauma: from forensic pathology to forensic veterinary medicine. Forensic Sci Int 2014 Jan;234:e1-e4. [doi: 10.1016/j.forsciint.2013.08.024] [Medline: 24091032]
26. Henham AP, Lee KA. Photography in forensic medicine. J Audiov Media Med 1994 Jan;17(1):15-20. [Medline: 8034901]
27. Peterson GF, Clark SC, National Association of Medical Examiners. Forensic autopsy performance standards. Am J Forensic Med Pathol 2006 Sep;27(3):200-225. [doi: 10.1097/01.paf.0000243580.43150.3c] [Medline: 16936503]
28. Dolinak D, Matshes E, Lew E. Forensic Pathology: Principles and Practice. San Diego: Elsevier Science; 2005.
29. Senn D, Stimson P. Forensic Dentistry, Second Edition. Boca Raton: Taylor & Francis, Inc; 2010.
30. Mann H, Whitney D. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist 1947;18(1):50-60. [doi: 10.1214/aoms/1177730491]
31. Koch G. Intraclass correlation coefficient. In: Kotz S, Johnson NL. editors. Encyclopedia of Statistical Sciences, 4th ed. New York: John Wiley & Sons; 1982:213-217.
32. Aldaz G, Shluzas LA, Pickham D, Eris O, Sadler J, Joshi S, et al. Hands-free image capture, data tagging and transfer using Google Glass: a pilot study for improved wound care management. PLoS One 2015;10(4):e0121179 [FREE Full text] [doi: 10.1371/journal.pone.0121179] [Medline: 25902061]
33. Rahimy E, Garg SJ. Google Glass for Recording Scleral Buckling Surgery. JAMA Ophthalmol 2015 Jun;133(6):710-711. [doi: 10.1001/jamaophthalmol.2015.0465] [Medline: 25811488]
34. Moshtaghi O, Kelley KS, Armstrong WB, Ghavami Y, Gu J, Djalilian HR. Using Google Glass to solve communication and surgical education challenges in the operating room. Laryngoscope 2015 Oct;125(10):2295-2297. [doi: 10.1002/lary.25249] [Medline: 25775933]
35. Laricchiuta P, Russo V, Costagliola A, Piegari G, Capasso M, Silvestre P, et al. Histological and immunohistochemical characterization of uterine adenocarcinoma in an Asian elephant (Elephas Maximus). Folia Morphol (Warsz) 2018 Mar 23 [FREE Full text] [doi: 10.5603/FM.a2018.0025] [Medline: 29569705]
36. Frank C, Madden DJ, Duncan C. Field Necropsy of the Horse. Vet Clin North Am Equine Pract 2015 Aug;31(2):233-245. [doi: 10.1016/j.cveq.2015.04.002] [Medline: 26021522]

Abbreviations

DSLR: digital single-lens reflex
ICC: intraclass correlation coefficient

Edited by G Eysenbach; submitted 29.01.18; peer-reviewed by A Mavragani, UV Albrecht; comments to author 05.03.18; revised version received 03.05.18; accepted 08.05.18; published 21.09.18.

Please cite as:
Piegari G, Iovane V, Carletti V, Fico R, Costagliola A, De Biase D, Prisco F, Paciello O
Assessment of Google Glass for Photographic Documentation in Veterinary Forensic Pathology: Usability Study
JMIR Mhealth Uhealth 2018;6(9):e180
URL: http://mhealth.jmir.org/2018/9/e180/
doi:10.2196/mhealth.9975
PMID:30249586

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org