JMIR mHealth and uHealth

Impact Factor (2022): 5.0
Volume 5 (2017), Issue 8 ISSN 2291-5222 Editor in Chief: Laurie Buis, PhD, MSI

Contents

Original Papers

XSL•FO
RenderX

XSL•FO
RenderX

Original Paper

Consent Processes for Mobile App Mediated Research: Systematic Review

Sarah Moore[1], BS; Anne-Marie Tassé[2], LLM, MA, PhD; Adrian Thorogood[3], BCL, LLB; Ingrid Winship[4], MBChB, MD, FRACP, FACD; Ma'n Zawati[3], LLB, LLM; Megan Doerr[1], MS, LGC

[1]Sage Bionetworks, Seattle, WA, United States

[2]Public Population Project in Genomics and Society, Montreal, QC, Canada

[3]Center of Genomics and Policy, McGill University, Montreal, QC, Canada

[4]Melbourne Health, University of Melbourne, Melbourne, Australia

Corresponding Author:
Sarah Moore, BS
Sage Bionetworks
1100 Fairview Ave N
Seattle, WA, 98109-4433
United States
Phone: 1 2066677841
Fax: 1 2066677841
Email: sarah.moore@sagebase.org

Abstract

Background: Since the launch of ResearchKit on the iOS platform in March 2015 and ResearchStack on the Android platform in June 2016, many academic and commercial institutions around the world have adapted these frameworks to develop mobile app-based research studies. These studies cover a wide variety of subject areas including melanoma, cardiomyopathy, and autism. Additionally, these app-based studies target a variety of participant populations, including children and pregnant women.

Objective: The aim of this review was to document the variety of self-administered remote informed consent processes used in app-based research studies available between May and September 2016. Remote consent is defined as any consenting process with zero in-person steps, when a participant is able to join a study without ever seeing a member of the research team. This type of review has not been previously conducted. The research community would benefit from a rigorous interrogation of the types of consent taken as part of the seismic shift to entirely mobile meditated research studies.

Methods: This review examines both the process of information giving and specific content shared, with special attention to data privacy, aggregation, and sharing.

Results: Consistency across some elements of the app-based consent processes was found; for example, informing participants about how data will be curated from the phone. Variations in other elements were identified; for example, where specific information is shared and the level of detail disclosed. Additionally, several novel elements present in eConsent not typically seen in traditional consent for research were highlighted.

Conclusions: This review advocates the importance of participant informedness in a novel and largely unregulated research setting.

(JMIR Mhealth Uhealth 2017;5(8):e126) doi:10.2196/mhealth.7014

KEYWORDS

mHealth, informed consent, smartphone, cell phone, mobile applications, privacy, research ethics

Introduction

Patients want to share their health data to accelerate scientific discovery efforts [1]. This desire to share data has led to projects such as Open Humans [2], whereby people can upload data from many sources to share with researchers, and PatientsLikeMe [3], a popular website to find communities of patients clustered by disease or condition. In both of the aforementioned cases, the Internet serves as a powerful tool for remote connection and a central location for data.

XSL•FO
RenderX

Recent advances in technology, especially mobile platform development, offer further opportunity for data aggregation and sharing by activated patients [4]. Smartphones: portable, loaded with sensors, and virtually ubiquitous have the potential to revolutionize both the way in which individuals monitor their health and the way they share that data with researchers [5].

In March 2015, Apple launched ResearchKit [6] and in June 2016 Android launched ResearchStack [7]. Both open source frameworks can be used to create apps for research on mobile devices. Unsurprisingly, by offering a dynamic, customizable, and responsive platform for engaging participants in research and enabling rapidly scalable, longitudinal investigations, these devices are being heralded as a potential boon to human health researchers [8].

In order to fully capitalize on the promise of mobile technology to enable scalable research, approaches to informed consent must be adapted in parallel [9]. Novel approaches to informed consent must still ensure the core tenets of informedness, comprehension, and voluntariness [10-12]. Further, informed consent must address unique issues that arise from conducting research using mobile platforms, such as data security and transferability. Researchers are faced with a novel challenge of consenting participants in a completely self-administered setting with no required contact with the research team [9].

Research apps present new challenges and opportunities in informed consent [9]. In this review, a granular inventory of the informed consent processes of publically available research apps has been presented to serve as a foundation for a community-wide discussion on how best to uphold the tenets of informed consent in mobile research settings.

Methods

Establishing a Task Team

A task team was convened under the Global Alliance for Genomics and Health (GA4GH) [13]. The GA4GH, established in 2013, works to enable responsible and effective sharing of genomic and clinical data to advance understanding of human health. Aware of the developments in mobile device-facilitated human subjects' research, GA4GH convened an international task team in May 2016 to examine issues of informed consent within this novel setting.

Inclusion and Exclusion Criteria

First, we sought to establish a listing (inventory) of app-based research studies with entirely self-administered consent processes currently available worldwide (Multimedia Appendix 1). We excluded app-based research studies that had one or more mandatory in-person informed consent steps.

The inventory included research apps using the ResearchKit framework that are publically available on the Apple iOS platform. We do not include any ResearchStack apps (based on Android platform) in the inventory as none were publically available during our period of review.

Within the inventory (Multimedia Appendix 1), there are multiple sheets based on the consent information available for each app. The sheet of the app-based research study informed consent inventory labeled "Apps with no consent info" contains limited information of apps that claim to use the ResearchKit framework. However, these apps do not have a defined consent process, and are therefore excluded from the complete review.

Identifying Mobile Research Apps

Due to the emergent nature of the field, as there were few scientific publications and no centralized listings or public catalogs of research apps, we were unable to employ traditional review approaches. Instead, we used an automated continuous web search of the terms "ResearchKit" and "ResearchStack" over three months (May-July, 2016). The primary reviewer received email notification each time the search terms yielded a new result. Through this search, we were able to identify newly released apps via press releases, blog postings, and other web mentions (eg, social media). When possible, we corresponded with authors of web content about research apps to ensure we achieved saturation of the field. We created a listing of the names of individual apps and then searched for them in the respective app download site (iOS or Android). Additionally, we relied on contacts both in the US and in other countries aware of this project to inform us of new research apps. The list of research apps was continuously updated to include newly released study apps throughout the development of the inventory.

Identification and Refinement of Domains

We identified 3 components within app-mediated research where information is presented to a potential participant: the eConsent, long form consent (LFC), and privacy policy (PP). Across the apps surveyed, after meeting the eligibility criteria, prospective participants typically self-guide through the eConsent. The eConsent is a series of screens within the research app that a participant navigates prior to enrollment. It contains information traditionally disclosed during an informed consent process such as information about the study, study procedures, alternatives to participation, and risk and benefits. The LFC document is commonly interpreted to be required by US regulation. Participants are presented with the document on the phone prior to signing and joining the study. Participants receive a copy of the document following electronic signature and enrollment (Figure 1). Additionally, we found privacy policies, traditionally used within apps and by websites that alerted users on how data is gathered, used, managed, and potentially disclosed by the organization hosting the app or website, as a repository of study information critical to the informed consent process. Participants can view the PP from the app download site before downloading the app and, for some of the apps, review from within the app as well.

We began by surveying each app's informed consent materials for the 8 required elements of consent under the US Common Rule (description, risks, benefits, alternatives, confidentiality, compensation, contact information, and voluntary participation/withdrawal). We considered these domains to be predefined by regulatory requirements. Due to the nature of app-based research, we expanded our assessment of some core consent topics to include subtopics critical to app-based research. For example, "risks" needed to include not only general discomforts one may experience as being part of the study, but

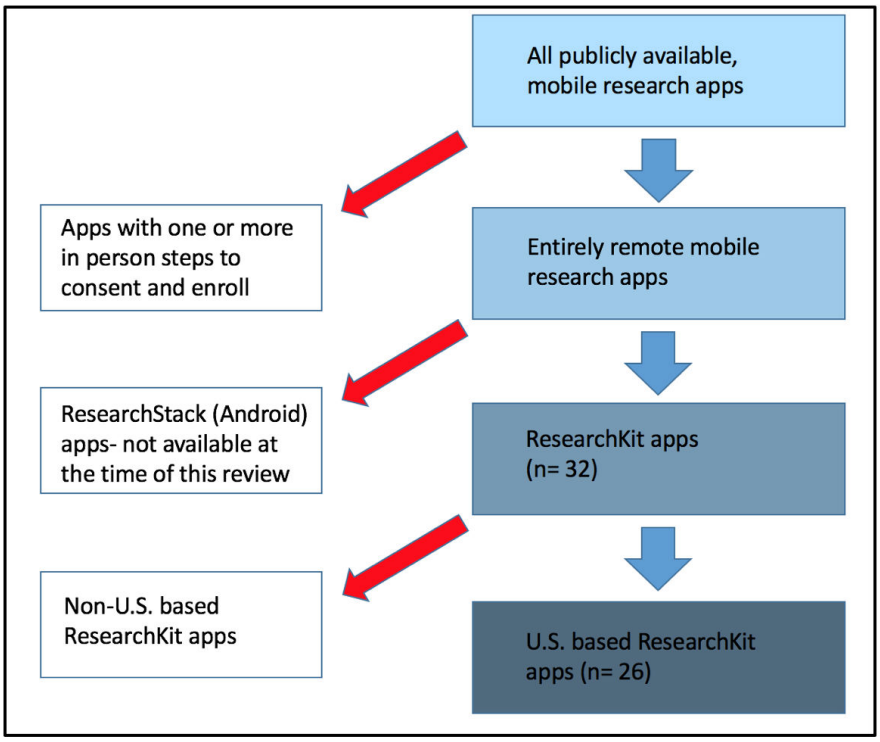
XSL•FO
RenderX

also address the risk to privacy. "Confidentiality" was expanded to include data collection, handling, and protection of participant privacy. We considered these domains to be emergent in reviewing the apps. Further, we chose to include domains based on thematic consistency. For example, if we saw that the majority of apps surveyed addressed a certain topic, we chose to include this topic in our review.

In reviewing eConsent processes and LFC documents for both predefined and emergent domains within "confidentiality," we concluded that we must expand our survey to include the PP as another possible source of this information in order to accurately assess disclosure. We expanded the inventory to include the source of information as well as its presence or absence.

The task team iteratively reviewed and discussed findings to refine domains included in the inventory via monthly video meetings as well as email discussions between meetings. Each of the 6 authors of this paper engaged in the discussions. Domains were finalized through group consensus based on a combination of information presented in the consent processes as well as calling attention to domains that have been omitted from the consent processes. After deciding on the domains, we determined how we would categorize the information gleaned from the informed consent materials to populate each domain (ie, Boolean, multi-option, or qualitative description).

Figure 1. Summary of methods.



Data Extraction and Coding

Due to mobile device country of registration app access restrictions, the primary reviewer downloaded and performed the initial assessment of only US-based apps. Non–US-based apps were reviewed by a member of the task team residing in the country of origin of the app. We used key word text searches and subsequent coding to identify data within the consent materials for each domain. For example, to determine whether a particular app disclosed information about encryption, we searched the text of the consent document and/or PP for the word "encrypt." If the word was found within the text, the domain would be filled with "YES" to indicate that the app disclosed information about data encryption. When needed, a second reviewer cross examined the original and extracted content.

Results

We surveyed a total of 34 app-based informed consent processes for apps available between May and August, 2016, of which 26 underwent complete review (see Multimedia Appendix 1). Some apps were closed during the period of review which limited the information we were able to obtain (3/26 apps closed; GlucoSuccess, Share the Journey and mTech).

Within the inventory, the sheet, "US-based apps with partial/complete consent," gives an overview of the apps inventoried, including the app name (column A), the version we reviewed (B), and the date of review (C). The study's focus is described in Subject/Disease Area (D), Targeted Population (E), Identifiable/Personal Data Collected (F), and Data Collection (G). Given that one of the key components of the informed consent is the ability for potential participants to ask questions and request clarification, we identify whether columns for the Sponsor, Principal Investigator, and Contact Information (H-J) are listed, as well as for the Institutional Review Board

XSL•FO
RenderX

(IRB) for the study, if any (K). However, one of the key criteria for the apps included in this review is that participants are not required to have contact with anyone on the study team prior to consent and enrolling in the study.

"Elements of Consent LFC" catalogues US Common Rule consent requirements as presented in the LFC with the exception of confidentiality, which is addressed in its own sheet due to the number of subcriteria identified. Under the Common Rule (45 CFR part 46 subpart A) the description of the research should include a statement indicating that the study is intended as research, the purpose of the study, the expected duration, and the procedures (column B). Additional Common Rule requirements are covered in "Risks" (C), "Benefits" (D), "Alternatives to Participation" (E), "Compensation" (F), and "Whom to Contact" (G). The voluntary nature of research (H) and any data retention after withdrawal (I) are described. It is important to note that all of the apps we surveyed were for observational studies, and none administered any treatment. For each domain, a checkmark indicates that all required subcriteria are met; missing subcriteria are listed by number. The sheet labeled "Elements of Consent eConsent" addresses whether or not these same US Common Rule consent requirements are addressed within the eConsent.

The theme of confidentiality is of critical importance in app-based research. We assessed 24 domains on the "Confidentiality" sheet. The first 4 columns address data collection. "Data Collection: Active/Passive" (column B) describes participant effort in data generation. Active tasks require deliberate action on the part of the participant, for example, by responding to surveys or doing a sensor-based task like a tapping test. Passive tasks are those in which the participant donates data without conscious effort, such as through the tracking and transmitting of GPS data. Some studies request permission access to data from other apps on the phone, for example HealthKit (C), or to phone features (D) like the camera and the microphone. Column E addresses whether the research app integrates any other data provided by an outside source. At present, this domain includes the integration of genomic data from testing companies or research sponsor organizations.

Columns labeled "Data Security" address the transmission of data from the participant phone to the backend collection including the app developer (F), whether or not the data will be encrypted (G), the name of the backend collector (H), and whether or not the data collected will be coded or pseudonymized (I). It is important to note that the information described in the table represents what is disclosed to participants. In reviewing the encryption of data, we found variation in which point in the process data is encrypted. Some apps appear to encrypt participant data on the participant phone and maintain this encryption during transmission, while others appear only to encrypt the data when it arrived to the backend server. The domain labeled "Backend Collection" is particularly important to note. In this column, we noted when a researcher has contracted out to a third party to provide data collection services (H).

We include 4 columns describing disclosure. "Required Disclosure" includes informing participants of possible sharing of data with US federal agencies, such as the US Department of Health and Human Services, and the institutional review board (IRB) or other ethics committee (J). "Commercial resale of data" addresses whether the participant is informed about the possibility that the data they have donated for research may be sold to third parties for advertising or other commercial endeavors (K). Additionally, we noted whether this information was found in the LFC or the PP. "Open data sharing for scientific discovery" (L) describes whether the research app has any kind of data sharing outside the primary investigator. If so, we have described what kind of data sharing the participant may consent to: with researchers within the institution of the primary research, with researchers at other institutions, or with qualified researchers worldwide including cross-border data transfer.

The remaining columns on the "Confidentiality" sheet address the option of using of the app without enrolling in research (M), if the recontact of participants is addressed (N), the return of the participant's own data (O), and incidental findings (P). "Data Preservation" addresses how long the data is authorized to be used by the researcher (Q). About half of the apps surveyed addressed this information either by giving a date that the data will expire or by giving a time period (eg, "6 years after the close of the study"). "Closure" addresses if information is given to participants about what will happen to their data after the period of authorization is over, or in the case of the close of databank or research institution (R). None of the apps reviewed address this domain.

The "Privacy Policy" sheet inventories the availability and content addressed within each app's PP. Column B addresses whether the PP in the iOS app store, prior to download, was reviewed. As it is an Apple requirement for any app collecting any user or usage data, including personal health information, all domains are marked "YES." "Privacy Policy presented as part of consent interaction" addresses whether the prospective participant is made aware of the PP before consenting to participate (C) (eg, Figure 2). In many cases, the PP addresses what personal identifiers the app collects (D) and states that it cannot fully guarantee privacy (E). Column F for "notes" includes whether the PP is tailored specifically to the app and whether it includes the date on which it was last updated. The "Unique to eConsent" sheet addresses elements novel to self-administered, entirely remote consent, including the use of an assessment to assess informedness (D) and approach to consent documentation. If an app uses a quiz at the end of the eConsent, we have indicated whether it is mandatory (E), how many questions are part of the assessment (F), and the information the questions are assessing (G). As eConsent may not meet US informed consent requirements, participants also have access to a LFC (H). We indicate whether finger signature is required (I) (Figure 3, mPower, Sage Bionetworks) and whether the participant would receive the signed LFC via email automatically (J).

The sheet, "Non–US-based apps," contains the same descriptive domains (eg, app name, version, and subject matter) as the overview "US-based apps" sheet. Due to the geolocation requirements of app download, we struggled to review these

XSL•FO
RenderX

apps completely. We were successful in completing a partial review of 2 non–US-based apps: ACL Rupture and Depressed, both from Germany. In the US based review, there was tremendous diversity in the detail and information provided by the apps about the nature of the research being conducted and elements of informed consent.

Figure 2. 6th Vital Sign, Duke University privacy policy embedded in consent process.

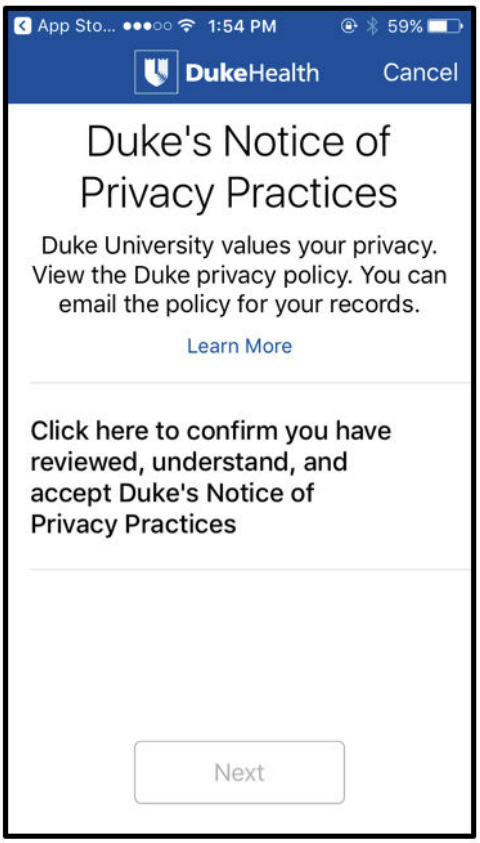


Figure 3. Participants use finger signature to authorize consent to research (image from mPower, Sage Bionetworks).

XSL•FO
RenderX

Discussion

Limitations to Methods

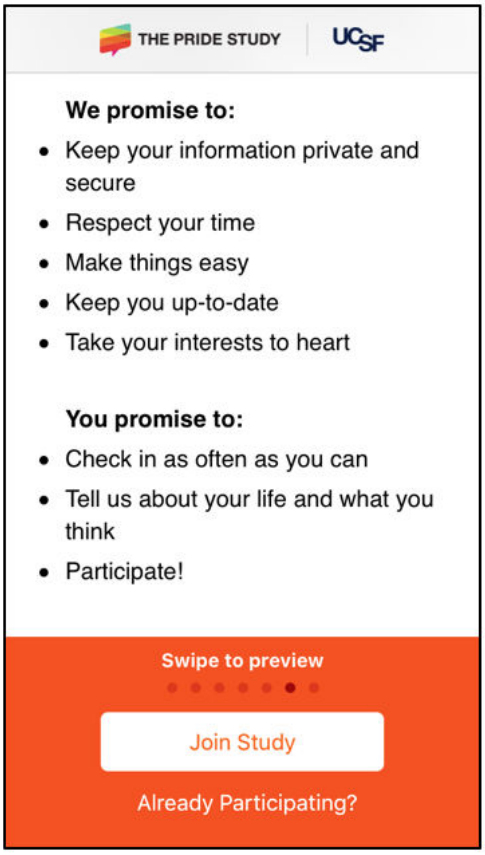
Due to the limitations of the approach, the listing may not be comprehensive. Any omissions are due to inability to find these apps through the methods described above rather than deliberate exclusion. Additionally, because of the duration of this project, 3 apps that were available the beginning of the study are no longer available (GlucoSuccess, MGH; mTech, University of the Pacific; Share the Journey, Sage Bionetworks). The inventory is current as of September 1, 2016.

Using mobile platforms for research creates new opportunities for both the types and volume of data collected within a highly scalable, delocalized framework. To harness the full potential of these platforms, informed consent must be similarly scalable without betraying the core principles of informedness, voluntariness, and comprehension [12].

In this novel research ecosystem, without a face-to-face interaction, participants must rely solely on eConsent, LFC, and PP documents to understand what researchers intend to do with their data and how they will protect it. Because of this, it is vital that researchers take an intentional approach to participant informedness, as seen in The Pride Study app (Figure 4).

In current practice, the eConsent, LFC, and PP are used in tandem with one another to inform the prospective participant, with key information potentially found in any of these locations. For example, the disclosure that researchers "cannot fully guarantee privacy," was found in either in the PP or the LFC, without a consistent standard across the apps surveyed. Additionally, it was found that in most apps, participants were only required to review the eConsent, but not the LFC and/or PP (See sheet "Privacy Policy," Column C) prior to participating.

Figure 4. The Pride Study, UCSF researcher and participant agreement.



Privacy, Security, and Data Use

Privacy is the ability to control the recording and sharing of personal information with others. This requires knowledge of what will be recorded, how it will be used and for how long, who will have access to this information, and what risks are of discovery and misuse by third parties [14].

The opportunity for researchers to gather more and different types of data through app-mediated research than they would be able to gather in a traditional clinical study poses a unique risk to privacy. For example, in mPower, a study on Parkinson Disease, researchers are also able to gather GPS data.

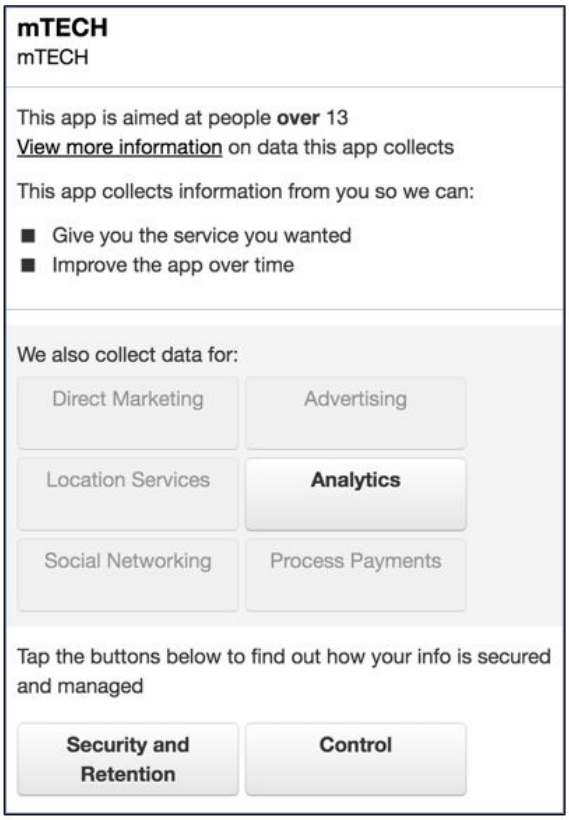
Because of the diversity and volume of data being collected, participants are potentially more easily identifiable. It may be impossible to deidentify an individual's mobile phone data, the standard way of protecting personal privacy in research [14]. Further, a number of the apps surveyed use a backend data collection service other than the primary research sponsor (See sheet "Confidentiality," Column K). Third party cloud and hosting services provide an economical solution for hosting and storing large amounts of collected data. In addition to an outside

XSL•FO
RenderX

group having access to the data collected, the use of a third party poses the threat via hacking of information sent over the Internet [15]. Furthermore, data collected through app-mediated research is collected in the context of a particular disease or condition, such as Parkinson Disease, Melanoma, Autism, or Cardiomyopathy, potentially compromising a participant's confidentiality through reidentification.

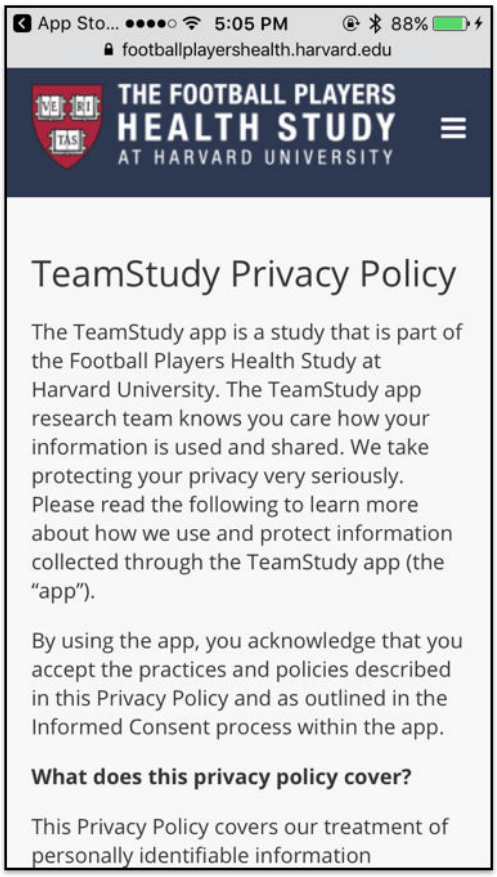
The multitude of potential mishandlings of data make a strong case for participants to be informed of these possibilities. Information for participants about the collection, transfer, and use of their data often resides in the app PP. Within its AppStore/iOS policies, Apple states that any app collecting any user or usage data, including personal health information must have a PP, however they provide no guidelines on what elements are required within that PP [16]. Unsurprisingly, we found PPs with a broad spectrum of detail and transparency from the very minimalist mTech, University of the Pacific (Figure 5) to the exhaustive Team Study, Harvard University (Figure 6) which is tailored specifically to the research app.

Figure 5. Privacy policies: mTech, University of the Pacific and Team Study, Harvard University.



XSL•FO
RenderX

Figure 6. Privacy Policy, Team Study, Harvard University.



Redistribution of Data

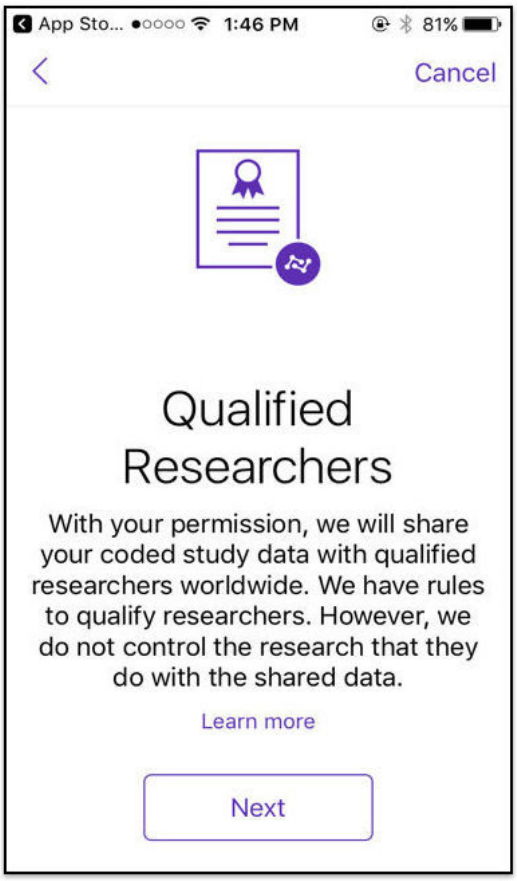
Because of its highly structured, electronic nature, data collected through app-mediated research can be easily redistributed. There are 16 apps that specifically engage participants on the topic of data reuse for additional research as part of the eConsent. In the consent processes of these apps, participants designate if their data will be available in aggregate for reuse in future independent research. For those who advocate for open data sharing, app-mediated research has the potential to revolutionize how data is shared and analyzed by many researchers at one time, and thus maximize the scientific value of participant data donation (Figure 7). Of the apps included in this review, many use the nomenclature of "Qualified Researchers Worldwide" advocated by Wilbanks and Friend of Sage Bionetworks [17]. This nomenclature was first used by Sage Bionetworks as a process to qualify researchers to access open data from the ResearchKit app and mobile study mPower [18]. Even though the process and requirements to access such data may be different from the process used by Sage Bionetworks, this standard nomenclature sets a precedent for open data sharing in app-mediated research.

Data is easily redistributed for commercial use. Of the apps surveyed, 18 of 26 disclose to participants that they will not sell or share participant data for commercial purposes (See sheet "Confidentiality, Column K"). But perhaps more interesting to this review are the 8 apps that do not explicitly state that participant data will not be sold. As per the Apple developer guidelines, "Apps may not use or disclose to third parties data gathered in the health, fitness, and medical research context—including from the HealthKit API, Motion and Fitness, or health-related human subject research—for advertising or other use-based data mining purposes other than improving health management, or for the purpose of health research, and then only with permission" [16].

However, it is unclear whether these apps intend to disclose data to third parties, or simply fail to state that they will not disclose this data to participants.

XSL•FO
RenderX

Figure 7. Participants have the option to share data broadly with other qualified researchers. Image from mPower, Sage Bionetworks.



Integration of Genomic Data

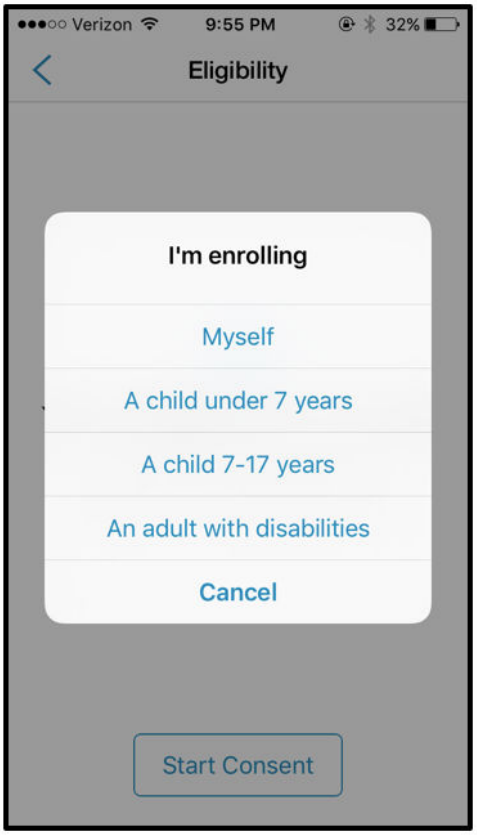
Recently, researchers have expanded the scope of their app-mediated research to include genomic data from outside sources. Participants in the MyHeart Counts, Stanford and Asthma Health, Mount Sinai studies may elect to integrate their 23andMe genomic analysis into the app. However, at the time of this review, there was no update to any of the primary informed consent documentation to reflect this optional expansion of scope. Rather, a second consent specific to genomic information integration is contained within the app under the 23andMe module. PPD ACT, University of North Carolina (UNC) at Chapel Hill has elected to integrate genomic data through independent sequencing, in which a subset of participants will receive a "spit kit" in the mail to be returned to the researchers. This optional research activity is disclosed in Phase I of the consent and was included in the initial scope of the research. In Phase I consent, participants agree to answering survey questions contained within the app. Participants go through Phase II of the consent process if they have been chosen and elected to donate genomic data. The reviewers were unable to include analysis of the consent to donate genomic data, as it is only available to participants who have been selected by researchers to participate in Phase II of the study.

Vulnerable Populations

App-mediated research may include vulnerable populations, most notably children (Autism and Beyond, Duke; FeverPrints, Boston Children's Hospital) and pregnant women (Yale EPV, Yale). The consent for app-mediated research targeting vulnerable populations differ from those enrolling only populations considered not vulnerable. Both Autism and Beyond and FeverPrints modify the consent process to reflect the assent of children and consent by parents or guardians. FeverPrints requires that the person going through the consent process on the phone be 18 years of age or above. Then they are asked who they are enrolling (see Figure 8). However, beyond this designation, there is no difference in the eConsent or lexicon used to reflect differences in who is being enrolled. Traditionally, pregnant women have been considered vulnerable populations when they cannot expect benefit from participation in research or when the fetus might be adversely impacted by the mother's participation [19]. In the case of the Yale EPV app, participants are expected to receive benefit (tracking of placental development) without any suspicion of risk to the fetus. In this way, pregnant women may not be deemed vulnerable within this research study.

XSL•FO
RenderX

Figure 8. Enrolling children. Image from Autism and Beyond, Duke University.



Recontact

About half (12/26, 46%) of apps surveyed address recontacting participants, primarily in the context of participating in future research studies (See sheet "Confidentiality," Column N). Fourteen studies also address the return of the data the participant shares with the study to the participant (See sheet "Confidentiality," Column O). Mobile research studies ease the return of data and potential results to participants and may facilitate a deeper reciprocal relationship between researchers and participants, although the risk of return of data and/or results are just beginning to be explored.

Informedness in Remote Consent

According to the Nuremberg Code, a human subject consenting to research "should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision" [10]. Thus, it is an ethical requirement of human subjects research, and one of many IRBs, that researchers ensure participants are adequately informed about the research before participation.

Lexicon used within the eConsent slides and the LFC document to explain potentially complex concepts to participants is central to participants' understanding of consent. Many have advocated for the use of simple language to promote participant comprehension and encourage informedness. Within clinical care, the National Quality Forum advocates for the use of universal symbols and pictures to improve comprehension in informed consent and specifies that written informed consent documents be at a US fifth-grade reading level (age 11) or lower [20]. Of the LFC documents reviewed, the average reading level is 10.9 (age 15-16) using the Flesh Kinkaid scale. Within the eConsent, another potential factor effecting comprehension was noted: a deviation in nomenclature from traditional consent documents, including the app's own LFCs. "Risks" in the LFC document (risk to privacy, risk of general discomforts) are referred to as "Issues to Consider" in the eConsent.

As consent processes reviewed are administered entirely remotely, this creates a new challenge for researchers in assessing informedness of participants. In a traditional research setting, a study coordinator or other personnel would sit face-to-face with potential participants to administer consent, with the opportunity to assess informedness in real time. Clearly a different strategy is needed for remote, self-administered consent. Perhaps the most common response among the apps surveyed is the use of an assessment. Some apps use this assessment as a summative evaluation and as a measure of participant ability to give informed consent. With these evaluations, if a prospective research participant does not answer enough questions correctly, they are sent back to the beginning of the consent process. However, more commonly researchers used the assessment as a formative evaluation, a chance to enhance participant understanding by prompting incorrect answers with the correct information about the study. About one-third of apps surveyed (9/26, 35%) have implemented a summative evaluation at the end of the consenting process to test participants' understanding of key concepts contained in the consent (See sheet "Unique to eConsent," Column D). Most commonly, these quizzes test participant understanding of the

XSL•FO
RenderX

purpose of the study, the fact that it is not medical care, that study data will be stored without direct identifiers, and participants' ability to withdraw at any time.

Conclusions

The consent processes presented in this review contain varying elements to contribute to participant informedness and transparency on the part of the research team. The new ecosystem of app-mediated research holds great promise to accelerate medical discovery through gathering potentially unprecedented amounts of highly structured data. However, app-mediated research also holds unique risks to participant data. The self-administered consent processes reviewed here present scalable approaches to informed consent to facilitate app-based research studies. The research community must continue to advocate for the importance of participant informedness, voluntariness, and comprehension in human subjects research. The variation, including strengths and gaps, observed in these informed consent processes have been highlighted to open a community dialogue about standards within this emerging field.

Acknowledgments

We would like to thank Julien Thérier, PhD of Shazino (Shazino blog) for his assistance with this project.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Consent toolkit inventory.

[XLSX File (Microsoft Excel File), 58KB - mhealth_v5i8e126_app1.xlsx]

References

1. University of California, San Deigo. RWJF. 2014. Personal data for the public good: new opportunities to enrich understanding of individual and population health URL: http://www.rwjf.org/en/library/research/2014/03/personal-data-for-the-public-good.html [accessed 2017-07-26] [WebCite Cache ID 6sF0KxwXR]
2. Open humans. 2017. URL: https://www.openhumans.org/ [accessed 2017-07-25] [WebCite Cache ID 6sDxUW6yX]
3. Patients Like Me. 2017. URL: https://www.patientslikeme.com/?format=html [accessed 2017-07-25] [WebCite Cache ID 6sDxjdLEv]
4. Trister AD, Dorsey ER, Friend SH. Smartphones as new tools in the management and understanding of Parkinson's disease. NPJ Parkinsons Dis 2016;2:16006 [FREE Full text] [doi: 10.1038/npjparkd.2016.6] [Medline: 28725694]
5. Friend SH. App-enabled trial participation: tectonic shift or tepid rumble? Sci Transl Med 2015 Jul 22;7(297):297ed10. [doi: 10.1126/scitranslmed.aab1206]
6. Apple. 2015. Apple introduces ResearchKit giving medical researchers the tools to revolutionize medical studies URL: https://www.apple.com/newsroom/2015/03/09Apple-Introduces-ResearchKit-Giving-Medical-Researchers-the-Tools-to-Revolutionize-Medical-Studies/ [accessed 2017-07-25] [WebCite Cache ID 6sDyCTIy9]
7. Wicklund E. mHealth Intelligence. 2016 Oct 7. ResearchStack goes live, opening mHealth studies to the android ecosystem URL: https://mhealthintelligence.com/news/researchstack-goes-live-opening-mhealth-studies-to-the-android-ecosystem [accessed 2017-07-25] [WebCite Cache ID 6sDyPAWHw]
8. Check HE. Mobile-phone health apps deliver data bounty. Nature 2016 Mar 24;531(7595):422-423. [doi: 10.1038/531422a]
9. Doerr M, Maguire TA, Bot B, Wilbanks J, Suver C, Mangravite LM. Formative evaluation of participant experience with mobile eConsent in the app-mediated Parkinson mPower study: a mixed methods study. JMIR Mhealth Uhealth 2017 Feb 16;5(2):e14 [FREE Full text] [doi: 10.2196/mhealth.6521] [Medline: 28209557]
10. The Nuremberg Code (1947). BMJ 1996 Dec 07;313(7070):1448-1448. [doi: 10.1136/bmj.313.7070.1448]
11. Declaration of Helsinki (1964). Br Med J 1996;313:1448-1449. [doi: 10.1136/bmj.313.7070.1448a]
12. Levine RA. Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. New Haven, CT: Yale University Press; 1986.
13. Genomicsandhealth. Regulatory and ethics working group, global alliance for genomics and health, 2016 URL: http://genomicsandhealth.org/working-groups/our-work/mobile-health-consent [accessed 2017-07-24] [WebCite Cache ID 6sCiS4e9j]
14. Carter A, Liddle J, Hall W, Chenery H. Mobile phones in research and treatment: ethical guidelines and future directions. JMIR Mhealth Uhealth 2015;3(4):e95 [FREE Full text] [doi: 10.2196/mhealth.4538] [Medline: 26474545]
15. He D, Naveed M, Gunter C, Nahrstedt K. Security concerns in android mHealth apps. AMIA Annu Symp Proc 2014;2014:645-654 [FREE Full text] [Medline: 25954370]

XSL•FO
RenderX

16. Apple. App store review guidelines - Apple developer URL: https://developer.apple.com/app-store/review/guidelines/#health-and-health-research [accessed 2017-07-24] [WebCite Cache ID 6sCgwTYuR]
17. Wilbanks J, Friend SH. First, design for data sharing. Nat Biotechnol 2016 Apr;34(4):377-379. [doi: 10.1038/nbt.3516] [Medline: 26939011]
18. Synapse. Accessing the mPower data URL: https://www.synapse.org/#!Synapse:syn4993293/wiki/247860 [accessed 2017-07-24] [WebCite Cache ID 6sCgrdgG8]
19. Blehar MC, Spong C, Grady C, Goldkind SF, Sahin L, Clayton JA. Enrolling pregnant women: issues in clinical research. Womens Health Issues 2013 Jan;23(1):e39-e45 [FREE Full text] [doi: 10.1016/j.whi.2012.10.003] [Medline: 23312713]
20. Qualityforum. 2005. NQF: Implementing a national voluntary consensus standard for informed consent: a user's guide for healthcare professionals URL: http://http://www.qualityforum.org/Publications/2005/09/Implementing_a_National_Voluntary_Consensus_Standard_for_Informed_Consent__A_User_s_Guide_for_Healthcare Professionals.aspx [accessed 2017-07-24] [WebCite Cache ID 6sCgpE2pn]

Abbreviations

GA4GH: global alliance for genomics and health
IRB: institutional review board
LFC: long form consent
PP: privacy policy
UNC: University of North Carolina

Edited by G Eysenbach; submitted 17.11.16; peer-reviewed by S MacLean, S Orlowski; comments to author 09.01.17; revised version received 26.05.17; accepted 13.06.17; published 30.08.17.

Please cite as:
Moore S, Tassé AM, Thorogood A, Winship I, Zawati M, Doerr M
Consent Processes for Mobile App Mediated Research: Systematic Review
JMIR Mhealth Uhealth 2017;5(8):e126
URL: http://mhealth.jmir.org/2017/8/e126/
doi:10.2196/mhealth.7014
PMID:28855147

©Sarah Moore, Anne-Marie Tassé, Adrian Thorogood, Ingrid Winship, Ma'n Zawati, Megan Doerr. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 30.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Validation of a Smartphone App for the Assessment of Sedentary and Active Behaviors

Meynard John Toledo[1], MS; Eric Hekler[1], PhD; Kevin Hollingshead[1], BS; Dana Epstein[2,3], PhD; Matthew Buman[1], PhD

[1]Arizona State University, School of Nutrition and Health Promotion, Phoenix, AZ, United States

[2]Phoenix Veterans Affairs Health Care System, Phoenix, AZ, United States

[3]Arizona State University, College of Nursing and Health Innovation, Phoenix, AZ, United States

Corresponding Author:
Matthew Buman, PhD
Arizona State University
School of Nutrition and Health Promotion
Arizona Biomedical Collaborative
550 N 3rd St
Phoenix, AZ, 85004
United States
Phone: 1 6028272289
Fax: 1 6028272253
Email: mbuman@asu.edu

Abstract

Background: Although current technological advancements have allowed for objective measurements of sedentary behavior via accelerometers, these devices do not provide the contextual information needed to identify targets for behavioral interventions and generate public health guidelines to reduce sedentary behavior. Thus, self-reports still remain an important method of measurement for physical activity and sedentary behaviors.

Objective: This study evaluated the reliability, validity, and sensitivity to change of a smartphone app in assessing sitting, light-intensity physical activity (LPA), and moderate-vigorous physical activity (MVPA).

Methods: Adults (N=28; 49.0 years old, standard deviation [SD] 8.9; 85% men; 73% Caucasian; body mass index=35.0, SD 8.3 kg/m2) reported their sitting, LPA, and MVPA over an 11-week behavioral intervention. During three separate 7-day periods, participants wore the activPAL3c accelerometer/inclinometer as a criterion measure. Intraclass correlation (ICC; 95% CI) and bias estimates (mean difference [δ] and root of mean square error [RMSE]) were used to compare app-based reported behaviors to measured sitting time (lying/seated position), LPA (standing or stepping at <100 steps/minute), and MVPA (stepping at >100 steps/minute).

Results: Test-retest results suggested moderate agreement with the criterion for sedentary time, LPA, and MVPA (ICC=0.65 [0.43-0.82], 0.67 [0.44-0.83] and 0.69 [0.48-0.84], respectively). The agreement between the two measures was poor (ICC=0.05-0.40). The app underestimated sedentary time (δ=-45.9 [-67.6, -24.2] minutes/day, RMSE=201.6) and overestimated LPA and MVPA (δ=18.8 [-1.30 to 38.9] minutes/day, RMSE=183; and δ=29.3 [25.3 to 33.2] minutes/day, RMSE=71.6, respectively). The app underestimated change in time spent during LPA and MVPA but overestimated change in sedentary time. Both measures showed similar directions in changed scores on sedentary time and LPA.

Conclusions: Despite its inaccuracy, the app may be useful as a self-monitoring tool in the context of a behavioral intervention. Future research may help to clarify reasons for under- or over-reporting of behaviors.

(JMIR Mhealth Uhealth 2017;5(8):e119) doi:10.2196/mhealth.6974

KEYWORDS

Sedentary and physical activity measurement; smartphone daily-log app; self-monitoring app; BeWell24

XSL•FO
RenderX

Introduction

Lack of physical activity (PA; <150 minutes of moderate-vigorous PA per week [1]) and sedentary behavior (ie, sitting/reclining with low energy expenditure while awake [2-4]) are associated with increased risks for cardiometabolic diseases, such as hypertension, dyslipidemia, diabetes, and coronary heart disease. As such, accurate assessment of these behaviors is integral in developing potent behavioral interventions aimed at improving overall cardiometabolic health.

Assessment of PA and sedentary behaviors have improved significantly in recent years. Technological advancements have led to the development of sensors that can objectively and accurately measure these behaviors [5-7]. Although these sensors have led to an unprecedented objectivity in PA measurement, the self-report method is still ubiquitous in PA research; this is because it offers advantages over objective methods [8]. Accelerometers do not capture the contextual information associated with these activities. Researchers studying specific domains of these behaviors (ie, leisure-time, occupational, or transportation) still rely on self-reports to isolate the behaviors that occur in each of these domains. Distinguishing which domains these behaviors occur in is necessary for developing and evaluating targeted interventions to modify these domain-specific behaviors [9]. Thus, self-reports remain an important method of measurement for PA and sedentary behaviors.

Various types of PA questionnaires have been developed, ranging from global questionnaires to detailed quantitative histories. Strath et al [8] classified PA questionnaires into three broad categories: global, short recalls, and quantitative history. Global PA questionnaires are usually short (2 to 4 items) and provide an overview of an individual's overall activity level. These questionnaires are primarily used to identify whether individuals meet the PA standard, or classify individuals according to their PA levels (eg, active vs inactive). In contrast, short recalls provide a measure of an individual's PA level, as classified by the dimension of intensity level or domain. Quantitative history questionnaires are detailed measures that are used to understand the types and intensity of PAs that contribute to mortality or morbidity. A systematic review of studies that evaluated the reliability and objective criterion-related validity of new and existing PA questionnaires examined 65 studies that looked at a total of 96 PA questionnaires [10]. The results revealed poor-to-moderate validity, with median validity coefficients ranging from 0.30-0.39 for existing (and from 0.25-0.41 for new) PA questionnaires. Furthermore, other studies have shown that although these questionnaires show acceptable agreement for structured vigorous-intensity PAs, they are less accurate for more prevalent lower-intensity activities [11-14]. Similar patterns of accuracy and reliability are also consistent with existing instruments that measure sedentary behavior. For example, habitual domain-specific sedentary behaviors tend to have higher correlations with criterion measures than overall sedentary time (0.14-0.83 vs 0.07-0.61) [15]. This pattern is mainly due to the high cognitive demands associated with reporting usual daily activities [14].

The PA diary is another type of self-report measure of PA and sedentary behavior that is often used to obtain a detailed hour-by-hour or activity-by-activity record [8]. An example of a daily diary is the Bouchard Physical Activity Record, which asks participants to categorize their activity every 15 minutes in 1 of 9 types of movement behaviors [16]. Using a PA diary is advantageous because it includes detailed information on dimensions and domains of PA and sedentary behavior, and is less subject to memory bias than other methods. However, the detail of information and frequency of reporting required causes significant participant burden, which limits its use in long-term studies. A possible solution to this problem is to leverage mobile technology as a platform to deliver these PA and sedentary behavior measurement tools. The ubiquitous use of mobile phones offers an opportunity to capture self-report behaviors in real-time and with minimal recall bias [17]. Utilizing mobile platforms to deliver these questionnaires can lead to an easy-to-use and a readily accessible version of PA and sedentary behavior assessment tool, which could significantly lessen the burden associated with traditional diaries.

The purpose of this paper is to evaluate the reliability, validity, and sensitivity to change of a smartphone-based app (*BeWell24*), which is designed to assess sedentary behavior, light-intensity physical activity (LPA), and moderate-vigorous physical activity (MVPA), as compared to an objective measure (activPAL3c; PAL Technologies Ltd, Scotland, UK). Evaluating the accuracy of the app is the first step in establishing its usefulness as a self-report tool to measure PA and sedentary behavior in free-living environments, and provide rationale for future studies to evaluate its capacity in measuring context-specific activities.

Methods

Recruitment

Participants were US veterans that were receiving care at a regional Veterans Health Administration (VHA) hospital, and employees from a university in the Southwestern United States. All participants were aged 35-60 years and were recruited to participate via flyers and targeted emails for a smartphone-based lifestyle behavior program. Inclusion criteria included: (1) insufficient PA (activity ranking category of <4 on the Stanford Leisure-Time Activity Categorical Item) [18], (2) excessive sitting (defined as >8 hours of sitting from the International Physical Activity Questionnaire) [19], (3) short sleep duration (<7 hours/night) or a mild-to-moderate sleep complaint (modified version of the insomnia Severity Index) [20], (4) fasting glucose level of <100 mg/dL, and (5) owning an Android smartphone. All study procedures were approved by the institutional review boards of the local VHA hospital and the university. Each participant underwent telephone screening for eligibility and those who were eligible provided written informed consent.

BeWell24 App Design

The smartphone app design has been discussed previously [21]. In brief, the smartphone app included four components: (1) a self-monitoring component to allow for rapid self-monitoring of sleep, sedentary time, LPA, and MVPA across the full 24-hour day; (2) a behavioral sleep component that included

XSL•FO
RenderX

sleep education, sleep hygiene, and stimulus control therapy; (3) a sedentary component that provided graphical feedback on total sitting time and time spent sitting at work, watching television, socializing, transportation, and other activities; and (4) a PA component that included goal setting, feedback, and problem-solving interventions.

Study Protocol

Eligible participants were randomized to receive some combination of the sleep, sedentary behavior, and PA components using a full factorial or multiphase optimization strategy [22] study design to optimize efficiency and explore potential synergies among behavioral outcomes (more study design details are provided elsewhere [23]). Briefly, we utilized a full-factorial 2x2x2 screening experiment in which participants were randomized to receive one of eight possible combinations (k) of the sleep, sedentary, and exercise components of the app: none (k=1), one of three app components (k=3), two of three app components (k=3), or all three app components (k=1). More relevant to the purpose of this study, all participants received access to the self-monitoring component of the app for the entire duration of the study. Using this self-monitoring component, participants were asked to log time spent across 24 hours into domains of sleep, sedentary behavior, LPA, and MVPA throughout the 3-week baseline period and 8-week intervention period. Participants simultaneously wore an activPAL3c accelerometer (criterion) on three 7-day time periods (week 3 as baseline, week 7, and week 11).

Measures

BeWell24 Self-Monitoring App

Self-reported time spent in sedentary behavior, LPA, and MVPA were assessed using the BeWell24 app. The app provided an interface for users to report sleep, sedentary behavior, LPA, and MVPA behaviors in 5-minute epochs across the 24 hours. Figure 1 displays this interface. Users were given standardized definitions of each behavior [21] by tapping the icon on top of each column. Participants were instructed to log any nap or main sleep period, including all time in bed for the purpose of sleep under “Sleep Activities”. Any sitting behaviors (eg, sitting at desk or watching television) were logged in “Sedentary Activities”. All MVPAs such as brisk walking, jogging/walking, and aerobic exercise were logged under “Exercise Activities”. Participants were also instructed to categorize all other activities not fitting into the previous categories (eg, household chores, light gardening, leisurely walking, and other activities of light intensity) under “Other Activities”. For this study’s purposes, all activities categorized under “Other Activities” were classified as LPAs. Users allocated their time into each behavior by dragging their finger down the column throughout the specified time. Although the app was designed to gather contextual information about each of the activities performed, this paper focused on evaluating total time spent on each activity category due to lack of a gold standard measure for these contexts. For the purposes of analyses, time spent in each behavior was summed over all 5-minute epochs after excluding reported sleep time for each day.

Figure 1. Self-monitoring component of the BeWell24 app.

XSL•FO
RenderX

Criterion Measure of Sedentary and More Active Behaviors

Objective measures were derived from the activPAL3c accelerometer. The devices were waterproofed using medical grade adhesive coverings. Participants wore the device on the midline of their right thigh using breathable hypoallergenic tape, and were instructed to continuously wear the device for 7 days. The validity and reliability of this device in measuring sedentary and PA behaviors have been previously evaluated [24-27]. Collected data were processed into events of sitting, standing, or stepping using the activPAL software version 7.2.32 (PAL Technologies Ltd, Scotland, UK). Self-logged sleep and wake times collected from the BeWell24 app were used to exclude sleep time from the analyses. We used the consensus definition of sedentary behavior as seated/lying positions with low energy expenditure [4]; therefore, all wake time measured by the activPAL3c as lying/seated were considered sedentary. LPA time was defined as time spent standing or stepping at <3 metabolic equivalents (METs). MVPA was defined as time spent stepping at >3 METs. Although the activPAL3c device was primarily developed for measuring sedentary behavior, a recent study has shown that the device is also accurate at classifying and estimating the time spent at these higher-intensity activities (mean bias=-2.6 [-5.8, 0.7] minutes, RMSE=8.4, ICC=0.98 [0.95, 0.99]) [28]. All behaviors were summed to total time spent in that category and expressed in minutes/day. Bouts of continuous sedentary behavior >10 hours from the activPAL3c were considered nonwear time and were excluded from the analyses.

Statistical Analyses

Data were summarized using means, standard deviations (SDs), frequencies, and percentages. Analyses were performed using SAS Enterprise Guide 6.1 (SAS Institute Inc., Cary, NC) and SPSS version 23 (SPSS, Inc., Chicago, IL).

Reliability

Reliability of the BeWell24 app was evaluated using the data collected during the last two weeks of the baseline period. The last two weeks of the baseline period were arbitrarily chosen because they represent a period when participants were well acquainted with the app but before the start of intervention. Days with >80% completion rates on the BeWell24 app were considered valid days, and weeks with less than 3 valid days were excluded from the analyses. The activities spent per day in the second and third baseline weeks were summarized into weekly averages and intraclass correlations (ICCs; 95% CI), with absolute agreement between the two weeks calculated using two-way random effects models [29]. Reliability was considered poor, moderate, or strong when correlation coefficients were <0.4, 0.4-0.8, or >0.8, respectively [21].

Validity

The validity of the BeWell24 app was evaluated by comparing the app-based reported values to activPAL3c-measured sedentary time, LPA, and MVPA. Days with <80% completion rate of the app or those with <8 hours of valid activPAL3c wear time were excluded from analyses. Agreements between the two measures were assessed using single-measure with absolute definition ICCs using the two-way random effects model. Validity coefficients were interpreted with the scale referenced in the reliability section. Bias estimates such as mean difference (δ) and root of mean square error (RMSE; Figure 2) were also used to determine the degree of over/underestimation of the time spent on each behavior.

Figure 2. Root mean square error equation.

$$RMSE = \sqrt{\sum(BeWell24 - activPAL3c)^2/n}$$

The Bland-Altman method was also used to estimate the mean bias and the 95% limits of agreement (2 SD of the difference) between the two measurement methods. The plots were visually inspected for the presence of heteroscedasticity. The degree of heteroscedasticity was then assessed by calculating the Kendall's tau (τ) correlation between the absolute differences and the corresponding means. When τ>.1, the data were denoted heteroscedastic. When τ<.1 or negative, the data were denoted homoscedastic. If heteroscedasticity was present, the data were transformed by logarithms to the base 10. The limits of agreement from the log transformed data were then transformed back into the original scale by taking antilogs, which were then expressed as a function of the mean in the Bland-Altman plot [22].

Responsiveness to Change

The responsiveness statistic (RS) quantifies the minimal clinically important difference, in relation to the variability in scores of stable participants [30,31]. Ideally, participants are measured multiple times during the baseline and postintervention period to calculate the amount of variability in scores over a stable period. In this analysis, we categorized participants as stable if they did not change their sedentary or LPA time by 30 minutes/day or MVPA by 10 minutes/day. The SDs of the differences in scores between weeks 3 and 7 of these stable participants were used as the denominator in the RS calculation. For each participant, we also used the mean change in time spent in each behavior from week 3 (baseline) to week 7 (Δ) as our estimate of the minimal clinically important difference. To supplement our results, we calculated the degree of over/underestimation of the mean change in behavior by the BeWell24 app using mean percentage error (MPE; Figure 3).

XSL•FO
RenderX

Figure 3.

$$MPE = \frac{\sum[(\Delta BeWell24 - \Delta activPAL3c)/\Delta activPAL3c] * 100}{n}$$

Data Exclusion

Uncategorized hours in the app and sleep time were excluded from the data analyses. A total of 595 participant days were gathered from participants. Days with <10 hours of activPAL3c wear time and days with >20% annotation (n=219 days, 36.8%) were excluded from the analyses.

Results

User Statistics

A total of 26 adults (age 49.0 years, SD 8.9; 85% men; 73% Caucasian; body mass index [BMI]=35.0 kg/m^2, SD 8.3) participated but only 17 participants completed all aspects of the study. Four subjects withdrew from the study during the 3-week run-in period due to an unrelated health concern (n=1), burdensome assessment protocol (n=2), and loss of contact (n=1). Four subjects were lost to follow-up after randomization (17/21, 81% retention) due to an unrelated health concern (n=1) and loss of contact (n=3). There were no differences in demographic characteristics between *withdrawn* and *lost to follow-up* participants among participants who completed the study (N=17). All available data, from both completers and noncompleters, were included in subsequent analyses. Participants were asked to report their sedentary behavior, LPA, and MVPA in the app over the 11-week study period but only days with matching activPAL3c data (weeks 3, 7, and 11) were included in the analyses. A total of 376 days from 21 participants were analyzed. Across all time points, participants spent an average of 695.5 (SD 139.3) minutes/day being sedentary, 144.9 (SD 112.4) minutes/day on LPAs, and 21.5 (SD 16.5) minutes/day on MVPAs, as measured by the activPAL3c.

Reliability

Reliability in this study refers to the consistency of the BeWell24 app in measuring a behavior over a stable 2-week period (baseline). Based on the test-retest data, the reliability of the BeWell24 app revealed moderate agreement between measures of total time spent in sedentary behavior, LPAs, and MVPAs (ICC=0.65 [0.43, 0.82], 0.67 [0.44, 0.83], and 0.69 [0.48, 0.84], respectively).

Validity

Table 1 shows the agreement between the BeWell24 app and activPAL3c in total minutes/day spent in each of the three behaviors. Overall, the agreements between the two measures were poor (ICC range=0.10-0.35). Figure 4 A shows the Bland-Altman plot for self-reported sedentary behavior and activPAL3c-derived total time spent in sedentary activity per day. The data were determined to be homoscedastic (τ=-.35, P<.001). Linear regression showed a significant positive bias (β=0.53, P<.001) with increasing sedentary time. On average, the BeWell24 app substantially underestimated total sedentary behavior by -160.4 (-179.8, -141.0) minutes/day, and RMSE of 249.5.

Table 1. Agreement between activPAL3c and the BeWell24 app.

	Intraclass correlation[a] (95% CI)	δ[b] (95% CI)	Root of mean square error
Sedentary Activity	0.35 (0.04, 0.56)	-160.4 (-179.8, -141.0)	249.5
Light-Intensity Physical Activity	0.20 (0.02, 0.36)	144.4 (125.4, 163.5)237.0)	237.0
Moderate-Vigorous Physical Activity	0.10 (-0.01, 0.17)	15.5 (8.8, 22.3)	68.2

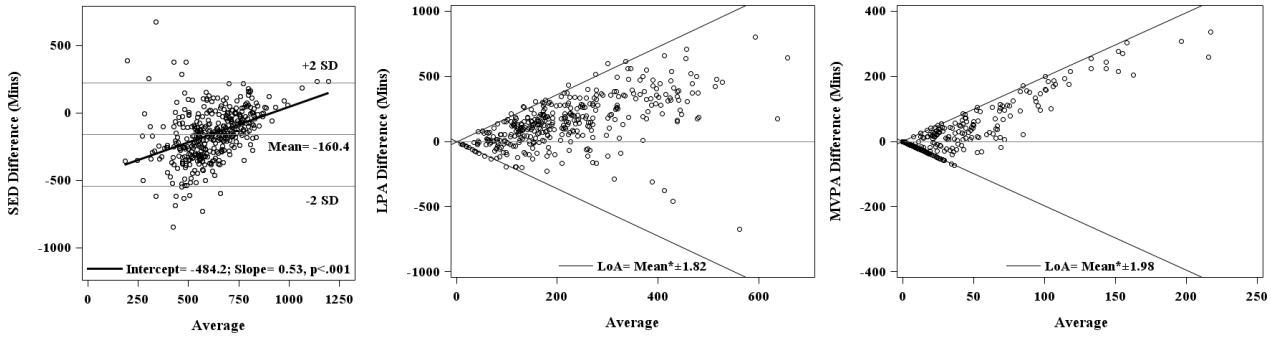
[a]Calculated using two-way random effects model with absolute agreement

[b]Mean difference between two measures calculated as δ= (BeWell24-activPAL3c)/n

Figure 4B and 4C show the Bland-Altman plot for self-reported and activPAL3c-derived time spent per day in LPA and MVPA activities, respectively. For both LPA and MVPA, visual inspection of the Bland-Altman plots on their original unit of measurement revealed heteroscedasticity, which were confirmed by Kendall's τ correlation coefficients (0.52 and 0.72, respectively). As such, data were analyzed on the log-transformed scale. There was significant negative bias (β=-0.56, P<.001 for LPA, and β=-1.1, P<.001 for MVPA). Furthermore, the BeWell24 app significantly overestimated time spent on LPA activities by 144.4 (125.4, 163.5) minutes/day with RMSE of 237.0 and slightly overestimated the time spent on MVPA by 15.5 (8.8, 22.3) minutes/day with RMSE of 68.2.

XSL•FO
RenderX

Figure 4. Bland-Altman plot of the BeWell24 app versus activPAL3c. (A) Total time spent sedentary (minutes/day); (B) total time spent in LPAs (minutes/day); (C) total time spent in MVPAs (minutes/day). The y-axes are the difference between the two measures (BeWell24 app – activPAL3c), and the x-axes are the average between the two measures ([BeWell24 app + activPAL3c]/2). SED: sedentary activity; LPA: light-intensity physical activity; MVPA: moderate-vigorous physical activity; LoA: Limits of agreement for heteroscedastic data (calculated as a function of the mean, LoA=±2Mean[10^SD-1]/[10^SD+1]).



Responsiveness to Change

Table 2 depicts the variability of each measure in each activity category in the preintervention period, the mean change from week 3 to week 7, and the RS of the BeWell24 app and activPAL3c. The mean change in sedentary behavior was greater in the BeWell24 app (-17.7 [-91.5, 56.0] minutes/day vs 2.16 [-49.3, 53.6]) minutes/day). However, the RS of the BeWell24 app was lower compared to the activPAL3c (0.20 vs 0.64) due to higher variability in changed scores among stable participants using the app. Similarly, the LPA and MVPA RS of the BeWell24 app was smaller (0.04 vs 2.0, and 0.19 vs 0.40) due to greater variability of changed LPA and MVPA time from the BeWell24 app.

Table 2. Responsiveness to change of the BeWell24 app and activPAL3c (n=20).

Variable		% Substantial Change (n/N)	Mean T1 and T2 Change (95% CI)	SD	Responsiveness Statistics
Sedentary Activity		85 (17/20)			
	BeWell24		-17.7 (-91.5, 56.0)	87.7	0.20
	activPAL3c		2.16 (-49.3, 53.6)	3.4	0.64
Light-Intensity Physical Activity		35 (7/20)			
	BeWell24		-3.50 (-44.9, 37.9)	79.7	0.04
	activPAL3c		-29.2 (-70.4, 11.9)	14.7	2.00
Moderate-Vigorous Physical Activity		10 (2/20)			
	BeWell24		-6.40 (-28.8, 15.9)	34.0	0.19
	activPAL3c		-2.00 (-5.00, 0.96)	5.00	0.40

[a]Percentage of participants who decreased by at least 30 minutes/day for sedentary activity, increased by at least 30 mins/day of LPA, or increased 10 minutes/day of MVPA based on activPAL3c

[b]A measure of variability in change score of stable participants

[b]Calculated as (mean change/SD), with direction of change removed

On average, the BeWell24 app slightly underreported change in sedentary behaviors and MVPAs (MPE, 95% CI=-2.0% [-161.7, 157.6.2] and -9% [-316.5, 298.5], respectively). The app significantly overestimated change in LPA time by 207.1% (-28.6 to 442.7). Nevertheless, both measures showed similar directions in change scores for all behaviors.

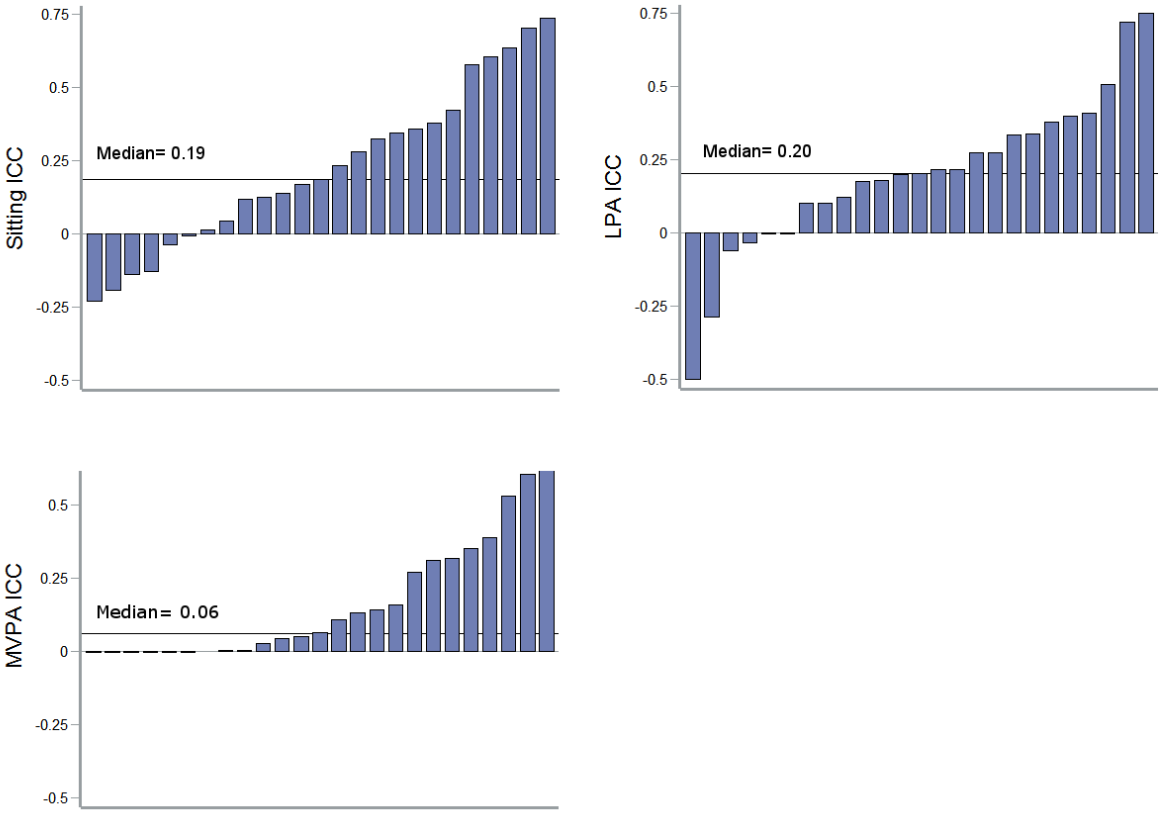
To understand the reasons for between-subject reporting error, we explored the agreement between the two measures within participants. We observed substantial variation in the agreement of the two measures between subjects (ICC range=-0.19 to 0.74, -0.85 to 0.93, and <-0.001 to 0.75 for sedentary behaviors, LPAs, and MVPAs, respectively; Figure 5). We then evaluated possible predictors of this observed variability using a linear regression model. Multiple variables (age, sex, BMI, activPAL3c-derived total time spent in sedentary behavior, LPA, and MVPA, and total percent of day categorized in the app) were evaluated for their associations with accuracy at reporting sedentary behaviors, LPAs, and MVPAs. We only found a significant positive association in ICC scores between BMI and sedentary behaviors and LPAs (β= 0.026, P<0.01 and β=0.025, P=.02, respectively), suggesting that participants with higher BMIs tended to be more accurate at reporting their sedentary and LPA behaviors. However, our analysis was limited in power due to a small number of observations within each subject.

We also explored the accuracy of the reported time spent in each behavior by hour (Figure 6) to understand whether subjects were more accurate in their reporting during morning, midday, or evening times. In general, hourly ICCs ranged from -0.26 to

XSL•FO
RenderX

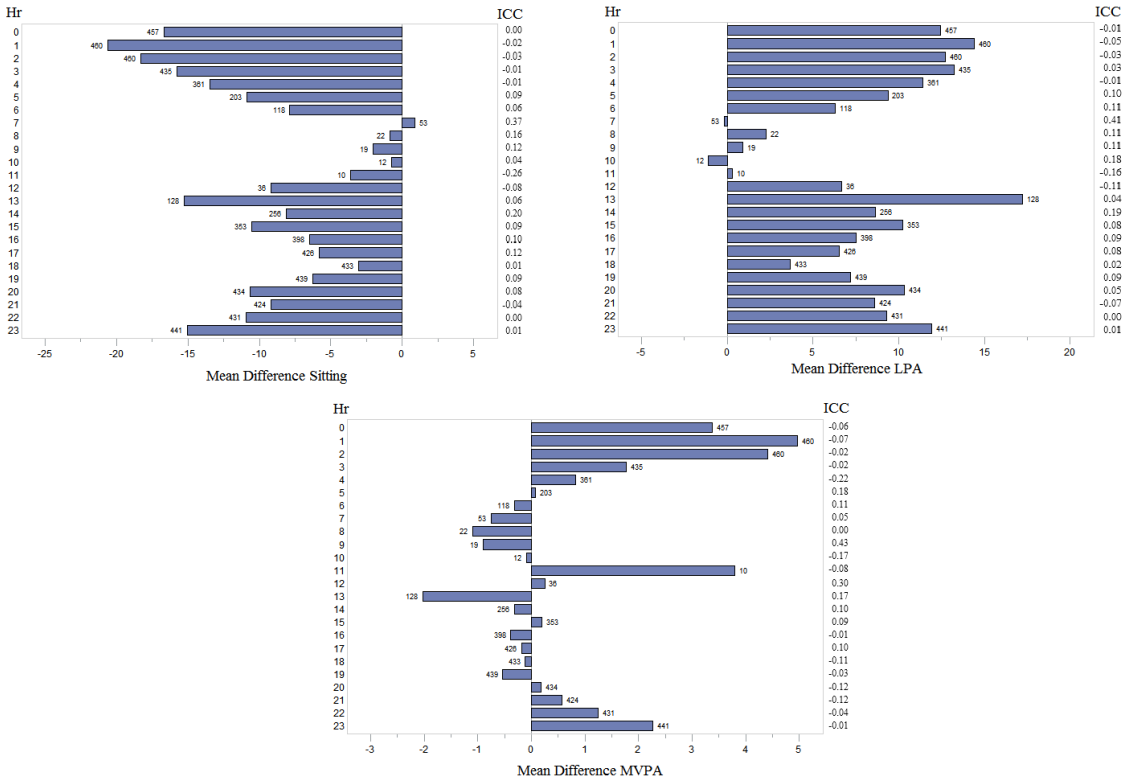
0.37, -0.16 to 0.41, and -0.22 to 0.43 (for sedentary behaviors, LPAs, and MVPAs, respectively). The overall pattern suggested greater accuracy during morning hours for sedentary and MVPAs, and uniform levels of accuracy for MVPAs during daytime hours with lower accuracy during early morning and late evening hours. The number of observations in each hour are presented beside each bar. Sleep time has been excluded from all analyses.

Figure 5. Individual variability in accuracy of self-reported behavior among participants. LPA: light-intensity physical activity; MVPA: moderate-vigorous physical activity; ICC: intraclass correlation.



XSL•FO
RenderX

Figure 6. Hourly variability in accuracy of self-reported behavior among participants. The number of observations in each hour is presented beside each bar. Sleep time has been excluded from all analyses. LPA: light-intensity physical activity; MVPA: moderate-vigorous physical activity; ICC: intraclass correlation.



Discussion

Principal Results

This study examined the utility of the BeWell24 app for assessing time spent in sedentary and more active behaviors. The results indicated that the app has moderate reliability and comparable RS to the activPAL3c when measuring sedentary and more active behaviors. This method of sedentary and PA behavior assessment leverages the mobile platform to deliver a more accessible and user-friendly assessment tool. This approach lowers the burden associated with daily diaries and allows for continuous reporting of sleep, sedentary behaviors, LPAs, and MVPAs.

However, our results also indicated poor absolute agreement between the app and the criterion measure for sedentary, LPA, and MVPA behaviors. These results were similar to other paper-based self-report measures of these behaviors (median Spearman of 0.23 for sedentary behavior and 0.30 for total activities) [10]. However, it must be noted that these PA questionnaires were evaluated using Spearman correlations, thus systematic differences between the two measurement methods were not taken into account [32]. Using ICCs to evaluate the reliability and validity of a measure is more suitable because it accounts for these systematic differences, which are considered to be an important element of overall measurement error [33].

Technological advances have led to the development of various mobile technologies that directly measure PA through sensors (ie, accelerometers) integrated into the mobile system. A review that included 10 studies that utilized mobile systems to measure PA found that these systems varied in their ability to classify behaviors (accuracy ranges from 52-100%) [34]. Although these mobile apps do measure behaviors passively and objectively, they are also limited in their ability to detect the context associated with these behaviors. A similar study that used mobile technology to deliver a self-report questionnaire to participants showed that total activity level (as assessed by the questionnaire) was moderately correlated with total counts from the Actigraph (r=.45, P<.05) [17]. However, the mobile app was only designed to measure total activity level of a person and does not measure the time spent in each activity category. A major advantage of our approach is that we allowed for 24-hour annotation of activity and did not just focus on MVPAs. This factor enabled us to critically evaluate the app for its ability to measure time spent in each activity category. This design will also allow future researchers to evaluate the independent health impacts of these behaviors.

The discrepancies in reported PA behaviors could be due to misclassification of MVPAs to LPAs by the activPAL3c. One limitation of the activPAL3c device, and with other accelerometer-based monitors, is that they are limited in their ability to measure the relative intensity of an activity. For example, a person walking leisurely at 3 miles per hour could be logged as an LPA in the activPAL3c device but could be at moderate intensity relative to the person's fitness level, which would be reported as MVPA in the app. The under-reporting of LPA time and over-reporting of MVPA time in the BeWell24 app supports this notion. It must also be emphasized that LPA time was calculated from the app using the total time classified

XSL•FO
RenderX

as “others”, in which participants were instructed to include behaviors that do not belong to other specified categories. This instruction could have led to some misclassification by the participants. However, participants were given clear and exhaustive instructions when classifying their activities, and any misclassification that occurred could also have randomly occurred in real-life settings, and should be treated as random errors.

Our results also included a responsiveness to change analysis to determine how well the app detects changes in behavior relative to the activPAL3c measure. This change is an important metric to evaluate, given its utility in the context of behavioral interventions, in which the absolute estimate of an activity may be less important than whether change in that activity has occurred. Notably, the app has consistently higher variability in changed scores among stable participants. This result also led the RS scores of the app to be consistently lower compared to the activPAL3c. Regardless, the mean change scores for both the activPAL3c and the BeWell24 app had similar trends in LPA and MVPA categories, suggesting that these measures agree on the overall direction of change for each of the behaviors.

Strengths and Limitations

One strength of this study is the use of the activPAL3c device, which allowed us to compare the self-report measure to a more valid measure of sedentary behavior and LPA. We also reported the RS of both measures, which enabled us to determine the ability of our instrument to detect changes in our target behavior over time. One key feature of this app that may be useful in future studies is the ability to capture the contextual information of these sedentary and more active behaviors, which greatly improves its utility in interventions targeting domain-specific behaviors. However, this feature was not evaluated in this study due to lack of objective criterion measures. The results from this study will provide useful information for future studies that would evaluate the validity of the app to measure these domain-specific behaviors.

A limitation of our study was the lack of a control group. Due to our study design, all participants received the at least one component of the app (self-monitoring component and/or a combination of the sleep, sedentary, or exercise component). This factor limited our ability to determine whether the observed change in behaviors were due to actual change caused by the intervention, or due to other causes, such as systematic misreporting of the behavior. As pointed out by Gardiner et al [35], a larger RS in our study suggests a greater magnitude of reported change in behavior during the intervention period, and not necessarily a better ability to detect a minimal clinically meaningful change. Furthermore, the activPAL3c device is primarily aimed at measuring sedentary activities and may be limited in its usefulness in measuring MVPAs. The lack of a more accurate measure of active behaviors may have led to the lower validity of the app in measuring MVPA. However, recent studies have shown that although the activPAL3c does overestimate and underestimate the energy expenditure of higher intensity activities, it does a satisfactory job of estimating time spent in these activities [28,36]. We were also limited in our ability to generalize the findings due to our small sample size, which was slightly older and had one or more morbid conditions compared to the general population.

Conclusions

Our results suggest that the BeWell24 app is reliable and sensitive to change in both sedentary behavior and more active behaviors after an intervention. The analysis showed poor agreement with the activPAL3c when measuring these behaviors. However, this finding is not unexpected, given that using self-reports for absolute measurement of these forms of activities have traditionally been difficult. Despite this limitation, the app is still useful in studies aimed at evaluating interventions targeted at changing these specific behaviors. In addition, the app could be used as a tool to capture context-specific forms of sedentary, LPA, and MVPA behaviors. Further study is needed to evaluate possible correlates for the large amount of between-subject variability in the accuracy of participants when reporting sedentary and more active behaviors.

Acknowledgments

This study was funded in part by the Virginia G. Piper Charitable Trust. This research was partially supported by resources and the use of facilities at the Phoenix Veterans Affairs Health Care System. The authors would like to thank Sean Branscome, Joseph Brinkman, Sarah Bryant, Rickman Fish, Monica Gutierrez, Christine Herb, Peter Reaven, and Justin Zeien for their assistance in app development and study execution. The contents of this article do not represent the views of the Department of Veterans Affairs or the United States Government.

Conflicts of Interest

None declared.

Authors' Contributions

MB, DE, and EH designed the study and supervised data collection. KH was involved with the BeWell24 app design and data management. MT was involved with the data analysis. MT and MB wrote the first draft of the manuscript. All authors have read and accepted the manuscript.

References

XSL•FO
RenderX

1. US Department of Health and Human Services. 2008 Physical Activity Guidelines Advisory for American. 2008. URL: https://health.gov/paguidelines/pdf/paguide.pdf [accessed 2017-07-25] [WebCite Cache ID 6sDu68cTm]
2. Healy GN, Wijndaele K, Dunstan DW, Shaw JE, Salmon J, Zimmet PZ, et al. Objectively measured sedentary time, physical activity, and metabolic risk: the Australian Diabetes, Obesity and Lifestyle Study (AusDiab). Diabetes Care 2008 Feb;31(2):369-371. [doi: 10.2337/dc07-1795] [Medline: 18000181]
3. Katzmarzyk PT, Church TS, Craig CL, Bouchard C. Sitting time and mortality from all causes, cardiovascular disease, and cancer. Med Sci Sports Exerc 2009 May;41(5):998-1005. [doi: 10.1249/MSS.0b013e3181930355] [Medline: 19346988]
4. Sedentary Behavior Research Network. Letter to the editor: standardized use of the terms "sedentary" and "sedentary behaviours". Appl Physiol Nutr Metab 2012 Jun;37(3):540-542. [doi: 10.1139/h2012-024] [Medline: 22540258]
5. Lamonte MJ, Ainsworth BE. Quantifying energy expenditure and physical activity in the context of dose response. Med Sci Sports Exerc 2001 Jun;33(6 Suppl):S370-8; discussion S419. [Medline: 11427762]
6. Plasqui G, Westerterp KR. Physical activity assessment with accelerometers: an evaluation against doubly labeled water. Obesity (Silver Spring) 2007 Oct;15(10):2371-2379 [FREE Full text] [doi: 10.1038/oby.2007.281] [Medline: 17925461]
7. Bonomi AG, Westerterp KR. Advances in physical activity monitoring and lifestyle interventions in obesity: a review. Int J Obes (Lond) 2012 Feb;36(2):167-177. [doi: 10.1038/ijo.2011.99] [Medline: 21587199]
8. Strath SJ, Kaminsky LA, Ainsworth BE, Ekelund U, Freedson PS, Gary RA, American Heart Association Physical Activity Committee of the Council on Lifestyle and Cardiometabolic Health and Cardiovascular, Exercise, Cardiac Rehabilitation and Prevention Committee of the Council on Clinical Cardiology, and Council. Guide to the assessment of physical activity: clinical and research applications: a scientific statement from the American Heart Association. Circulation 2013 Nov 12;128(20):2259-2279 [FREE Full text] [doi: 10.1161/01.cir.0000435708.67487.da] [Medline: 24126387]
9. Chastin S, Schwarz U, Skelton D, Skelton DA. Development of a consensus taxonomy of sedentary behaviors (SIT): report of Delphi Round 1. PLoS One 2013;8(12):e82313 [FREE Full text] [doi: 10.1371/journal.pone.0082313] [Medline: 24312653]
10. Helmerhorst HJ, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. Int J Behav Nutr Phys Act 2012;9:103 [FREE Full text] [doi: 10.1186/1479-5868-9-103] [Medline: 22938557]
11. Ainsworth BE, Richardson MT, Jacobs DR, Leon AS, Sternfeld B. Accuracy of recall of occupational physical activity by questionnaire. J Clin Epidemiol 1999 Mar;52(3):219-227. [Medline: 10210239]
12. Jacobs DR, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. Med Sci Sports Exerc 1993 Jan;25(1):81-91. [Medline: 8423759]
13. Strath SJ, Bassett DR, Swartz AM. Comparison of the college alumnus questionnaire physical activity index with objective monitoring. Ann Epidemiol 2004 Jul;14(6):409-415. [doi: 10.1016/j.annepidem.2003.07.001] [Medline: 15246329]
14. Matthews CE, Moore SC, George SM, Sampson J, Bowles HR. Improving self-reports of active and sedentary behaviors in large epidemiologic studies. Exerc Sport Sci Rev 2012 Jul;40(3):118-126 [FREE Full text] [doi: 10.1097/JES.0b013e31825b34a0] [Medline: 22653275]
15. Healy GN, Clark BK, Winkler EAH, Gardiner PA, Brown WJ, Matthews CE. Measurement of adults' sedentary time in population-based studies. Am J Prev Med 2011 Aug;41(2):216-227 [FREE Full text] [doi: 10.1016/j.amepre.2011.05.005] [Medline: 21767730]
16. Bouchard C, Tremblay A, Leblanc C, Lortie G, Savard R, Thériault G. A method to assess energy expenditure in children and adults. Am J Clin Nutr 1983 Mar;37(3):461-467. [Medline: 6829488]
17. Pfaeffli L, Maddison R, Jiang Y, Dalleck L, Löf M. Measuring physical activity in a cardiac rehabilitation population using a smartphone-based questionnaire. J Med Internet Res 2013;15(3):e61 [FREE Full text] [doi: 10.2196/jmir.2419] [Medline: 23524251]
18. Kiernan M, Schoffman DE, Lee K, Brown SD, Fair JM, Perri MG, et al. The Stanford Leisure-Time Activity Categorical Item (L-Cat): a single categorical item sensitive to physical activity changes in overweight/obese women. Int J Obes (Lond) 2013 Dec;37(12):1597-1602 [FREE Full text] [doi: 10.1038/ijo.2013.36] [Medline: 23588625]
19. Hagströmer M, Oja P, Sjöström M. The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. Public Health Nutr 2006 Sep;9(6):755-762. [Medline: 16925881]
20. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [Medline: 11438246]
21. Buman MP, Epstein DR, Gutierrez M, Herb C, Hollingshead K, Huberty JL, et al. BeWell24: development and process evaluation of a smartphone "app" to improve sleep, sedentary, and active behaviors in US Veterans with increased metabolic risk. Transl Behav Med 2016 Sep;6(3):438-448 [FREE Full text] [doi: 10.1007/s13142-015-0359-3] [Medline: 27528532]
22. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
23. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. Ann Behav Med 2005 Aug;30(1):65-73. [doi: 10.1207/s15324796abm3001_8] [Medline: 16097907]

XSL•FO
RenderX

24. Aminian S, Hinckson EA. Examining the validity of the ActivPAL monitor in measuring posture and ambulatory movement in children. Int J Behav Nutr Phys Act 2012 Oct 02;9:119 [FREE Full text] [doi: 10.1186/1479-5868-9-119] [Medline: 23031188]
25. Dowd KP, Harrington DM, Donnelly AE. Criterion and concurrent validity of the activPAL™ professional physical activity monitor in adolescent females. PLoS One 2012;7(10):e47633 [FREE Full text] [doi: 10.1371/journal.pone.0047633] [Medline: 23094069]
26. Ryan CG, Grant PM, Tigbe WW, Granat MH. The validity and reliability of a novel activity monitor as a measure of walking. Br J Sports Med 2006 Sep;40(9):779-784 [FREE Full text] [doi: 10.1136/bjsm.2006.027276] [Medline: 16825270]
27. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. Med Sci Sports Exerc 2011 Aug;43(8):1561-1567. [doi: 10.1249/MSS.0b013e31820ce174] [Medline: 21233777]
28. Lyden K, Keadle SK, Staudenmayer J, Freedson PS. The activPAL TM accurately classifies activity intensity categories in healthy adults. Med Sci Sports Exerc 2016 Dec 09. [doi: 10.1249/MSS.0000000000001177] [Medline: 27941421]
29. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline: 18839484]
30. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40(2):171-178. [Medline: 3818871]
31. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000 May;53(5):459-468. [Medline: 10812317]
32. Streiner D, Norman G, Cairney J. Health measurement scales Internet. In: Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 2015.
33. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
34. 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]
35. Gardiner PA, Clark BK, Healy GN, Eakin EG, Winkler EAH, Owen N. Measuring older adults' sedentary time: reliability, validity, and responsiveness. Med Sci Sports Exerc 2011 Nov;43(11):2127-2133. [doi: 10.1249/MSS.0b013e31821b94f7] [Medline: 21448077]
36. Montoye AH, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Evaluation of the activPAL accelerometer for physical activity and energy expenditure estimation in a semi-structured setting. J Sci Med Sport 2017 Apr 21. [doi: 10.1016/j.jsams.2017.04.011] [Medline: 28483558]

Abbreviations

BMI: body mass index
ICC: intraclass correlation
LPA: light-intensity physical activity
MET: metabolic equivalent
MPE: mean percentage error
MVPA: moderate-vigorous physical activity
PA: physical activity
RMSE: root of mean square error
RS: responsiveness statistic
SD: standard deviation
VHA: Veterans Health Administration
δ: mean difference between the app and activPAL3c
Δ: mean change from week 3 to week 7

XSL•FO
RenderX

Edited by G Eysenbach; submitted 09.11.16; peer-reviewed by J Bort-Roig, D Kerr, X Jiang, F Dutheil; comments to author 29.01.17; revised version received 14.04.17; accepted 10.06.17; published 09.08.17.

Please cite as:
Toledo MJ, Hekler E, Hollingshead K, Epstein D, Buman M
Validation of a Smartphone App for the Assessment of Sedentary and Active Behaviors
JMIR Mhealth Uhealth 2017;5(8):e119
URL: http://mhealth.jmir.org/2017/8/e119/
doi:10.2196/mhealth.6974
PMID:28793982

©Meynard John Toledo, Eric Hekler, Kevin Hollingshead, Dana Epstein, Matthew Buman. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 09.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Validation of a Smartphone-Based Approach to In Situ Cognitive Fatigue Assessment

Edward Price[1*], BSc; George Moore[1*], PhD; Leo Galway[1*], PhD; Mark Linden[2], PhD

[1]Computer Science Research Institute, School of Computing, Ulster University, Newtownabbey, United Kingdom

[2]School of Nursing and Midwifery, Queen's University Belfast, Belfast, United Kingdom

[*]these authors contributed equally

Corresponding Author:
George Moore, PhD
Computer Science Research Institute
School of Computing
Ulster University
Jordanstown Campus
Shore Road
Newtownabbey, BT37 0QB
United Kingdom
Phone: 44 02890366584
Fax: 44 02890366216
Email: g.moore@ulster.ac.uk

Abstract

Background: Acquired Brain Injuries (ABIs) can result in multiple detrimental cognitive effects, such as reduced memory capability, concentration, and planning. These effects can lead to cognitive fatigue, which can exacerbate the symptoms of ABIs and hinder management and recovery. Assessing cognitive fatigue is difficult due to the largely subjective nature of the condition and existing assessment approaches. Traditional methods of assessment use self-assessment questionnaires delivered in a medical setting, but recent work has attempted to employ more objective cognitive tests as a way of evaluating cognitive fatigue. However, these tests are still predominantly delivered within a medical environment, limiting their utility and efficacy.

Objective: The aim of this research was to investigate how cognitive fatigue can be accurately assessed in situ, during the quotidian activities of life. It was hypothesized that this assessment could be achieved through the use of mobile assistive technology to assess working memory, sustained attention, information processing speed, reaction time, and cognitive throughput.

Methods: The study used a bespoke smartphone app to track daily cognitive performance, in order to assess potential levels of cognitive fatigue. Twenty-one participants with no prior reported brain injuries took place in a two-week study, resulting in 81 individual testing instances being collected. The smartphone app delivered three cognitive tests on a daily basis: (1) Spatial Span to measure visuospatial working memory; (2) Psychomotor Vigilance Task (PVT) to measure sustained attention, information processing speed, and reaction time; and (3) a Mental Arithmetic Test to measure cognitive throughput. A smartphone-optimized version of the Mental Fatigue Scale (MFS) self-assessment questionnaire was used as a baseline to assess the validity of the three cognitive tests, as the questionnaire has already been validated in multiple peer-reviewed studies.

Results: The most highly correlated results were from the PVT, which showed a positive correlation with those from the prevalidated MFS, measuring 0.342 (P<.008). Scores from the cognitive tests were entered into a regression model and showed that only reaction time in the PVT was a significant predictor of fatigue (P=.016, F=2.682, 95% CI 9.0-84.2). Higher scores on the MFS were related to increases in reaction time during our mobile variant of the PVT.

Conclusions: The results show that the PVT mobile cognitive test developed for this study could be used as a valid and reliable method for measuring cognitive fatigue in situ. This test would remove the subjectivity associated with established self-assessment approaches and the need for assessments to be performed in a medical setting. Based on our findings, future work could explore delivering a small set of tests with increased duration to further improve measurement reliability. Moreover, as the smartphone assessment tool can be used as part of everyday life, additional sources of data relating to physiological, psychological, and environmental context could be included within the analysis to improve the nature and precision of the assessment process.

(JMIR Mhealth Uhealth 2017;5(8):e125) doi:10.2196/mhealth.6333

XSL•FO
RenderX

KEYWORDS

mental fatigue; fatigue; acquired brain injury; cognitive tests; assistive technology; smartphone

Introduction

Cognitive fatigue can be caused by multiple different conditions, including Acquired Brain Injuries (ABIs) [1], Parkinson's disease [2], stroke [3], heart failure [3], and many more. For those suffering from an ABI, cognitive fatigue can be triggered by carrying out simple everyday tasks, and is often much more prevalent [1]. Correct, timely, and accurate information is important to inform long-term management and rehabilitation from cognitive fatigue. Correspondingly, one of the primary challenges related to cognitive fatigue and its associated assessment is the subjective nature of the condition, coupled with a lack of identifiable biological markers. Consequently, there is a clear lack of technology-based approaches to assist with assessment and recovery management. Diagnosis and assessment typically relies upon assessment conducted within a clinical environment, meaning that incidents of cognitive fatigue can potentially go unrecognized, as they are most likely to happen when no professional or clinical supervision is present.

Traditional Fatigue Assessment Approaches

Evaluation of cognitive fatigue has traditionally been performed using self-assessment in the form of easy-to-comprehend questionnaires that are often delivered within a clinical setting [4]. Accordingly, a number of self-assessment questionnaires have been designed to specifically assess cognitive ability and its relationship to fatigue, including the Mental Fatigue Scale (MFS) [5], Fatigue Severity Scale (FSS) [6], and Visual Analogue Scale for Fatigue (VAS-F) [7]. All of these scales use a visual analogue representation or targeted questions to aid a participant in self-evaluation. Upon completion, a clinician must then calculate the resulting score to categorize the level of cognitive fatigue. The MFS is a multidimensional questionnaire containing 15 questions developed by Johansson and Rönnbäck [8] (adapted from Rödholm et al [9]) which utilizes a range of questions that cover the most common symptoms occurring after an ABI [10,11]. The questions concern fatigue in general, lack of initiative, mental fatigue, mental recovery, concentration difficulties, memory problems, slowness of thinking, sensitivity to stress, increased tendency to become emotional, irritability, sensitivity to light and noise, and decreased or increased duration of sleep (as well as 24-hour variations). A resulting rating is calculated based on intensity, frequency, and duration of cognitive fatigue. The scale has been clinically evaluated and has adequate internal consistency with a Cronbach alpha of 0.94 [12]. Similarly, the FSS is a 9-item questionnaire that assesses the effect of general fatigue on daily living, in which each item is rated on a 7-point Likert scale [6]. Originally designed to assess general fatigue in patients with multiple sclerosis and systemic lupus erythematosus, the FSS has been successfully adapted to the measurement of fatigue in relation to other conditions [13]. By contrast, the VAS-F developed by Lee et al [4] employs an 18-item visual analogue scale to allow participants to determine their own rating regarding each statement in the scale. However, of these self-assessment questionnaires, the MFS is the only one that measures cognitive fatigue irrespective of neurological conditions [14].

Computerized Fatigue Assessment Approaches

More recently, research into computerized cognitive fatigue assessment has sought to adopt cognitive testing methods by either adapting paper-based methods or repurposing cognitive testing batteries [15-17]. While cognitive testing methods have traditionally been used to determine relative cognitive ability, they have also been shown to indicate discrepancies in performance in relation to cognitive fatigue [16,18]. Van Dongen et al [19] defined three computerized methods for assessing fatigue: (1) a Mental Arithmetic Test to assess cognitive throughput; (2) the Psychomotor Vigilance Task (PVT) to assess reaction time; and (3) a Digit-Symbol Coding test to assess working memory. This study focused on measuring cognitive performance after periods of restricted sleep, with results indicating that even a relatively small amount of sleep restriction can seriously impair cognitive performance on everyday tasks [19].

Originally designed for a static desktop computer-based evaluation, PVT has since been modified for use on a mobile-based platform (to improve the utility of on-the-go assessment [16,20]) and has been shown to be an accurate predictor of vigilance due to fatigue and sleep loss [21,22], which is a direct predictor of cognitive fatigue. A shorter 3-minute version of PVT has been shown to be equally successful in evaluating cognitive fatigue as longer-administered versions [15]. Drawbacks observed from lengthier administered adaptations include an increased resource demand on participants due to the time on task required; this could potentially be harmful to participants already suffering from cognitive fatigue due to an ABI. The design of some of the PVTs mentioned above allowed for a degree of learning or preempting that occurred in these studies [20,21]. This problem needs to be addressed for tests to be equally taxing over time, so that results are continually obtained under the same level of effort. Johansson et al [18] compared the utility of the MFS to a suite of cognitive tests to determine if there is a direct correlation in their ability to subjectively and objectively measure cognitive fatigue. Neuropsychological tests that were employed included Digit Symbol Coding, Digit Span, Spatial Span [23], and Trail Making [24], which were used to measure processing speed, cognitive attention, working memory, verbal fluency, and reading speed. It was concluded that subjective cognitive fatigue following brain injury mainly correlates with objectively measured information processing speed [18]. In particular, specific cognitive tests from the Wechsler Adult Intelligence Scale, such as Spatial Span, have been shown to be effective when used to evaluate cognitive fatigue in terms of cognitive attention and working memory [8]. Consequently, these computerized tests allow for a detailed level of accuracy in testing, thus providing a more comprehensive view of cognitive ability than can be found using paper-based tests [16,18]. Work by Kay and Rector [16] and Gartenberg and McGarry [20] investigated the efficiency of using short

XSL•FO
RenderX

mobile-based tests, along with the potential usability issues that arise from test delivery on a mobile platform. These studies concluded that mobile-based assessment approaches were just as effective as desktop computer-based approaches [16,20]. Due to the simple design of tests such as the PVT, there were no usability issues noted when delivering it using a mobile platform. By contrast, Swendeman et al [25] conducted a study examining the validity of using a self-assessment-based approach on a smartphone; the study aimed to evaluate emotional and behavioral self-reports daily over a 6-week period. Daily completion rates were reported to be 50%, with 70% of participants completing three follow-up surveys after the 6-week period. However, adherence to daily assessment was observed to be low, which was attributed to errors in data that subsequently had to be excluded from the evaluation results. To resolve this issue, it was suggested that prior training with the technology should be provided for participants, as it is often the case that brain-injured individuals find technology difficult to use due to their condition [26].

The study presented in this paper aimed to address the research question, *" How can cognitive fatigue be accurately assessed through the use of a mobile assistive technology?"*. As such, the study sought to evaluate the correlations between subjective and objective measures of cognitive fatigue, with a future view to supplement traditional clinical evaluations with in situ cognitive assessment. From this approach, the following hypotheses were formulated: (1) a correlation will be observed between objective cognitive testing performance and the subjective measure reported by the MFS; (2) participants who are cognitively fatigued would exhibit a reduced level of performance in Spatial Span, PVT, and Mental Arithmetic Tests; and (3) accurate cognitive fatigue assessment can be carried out on a mobile app for use in everyday life.

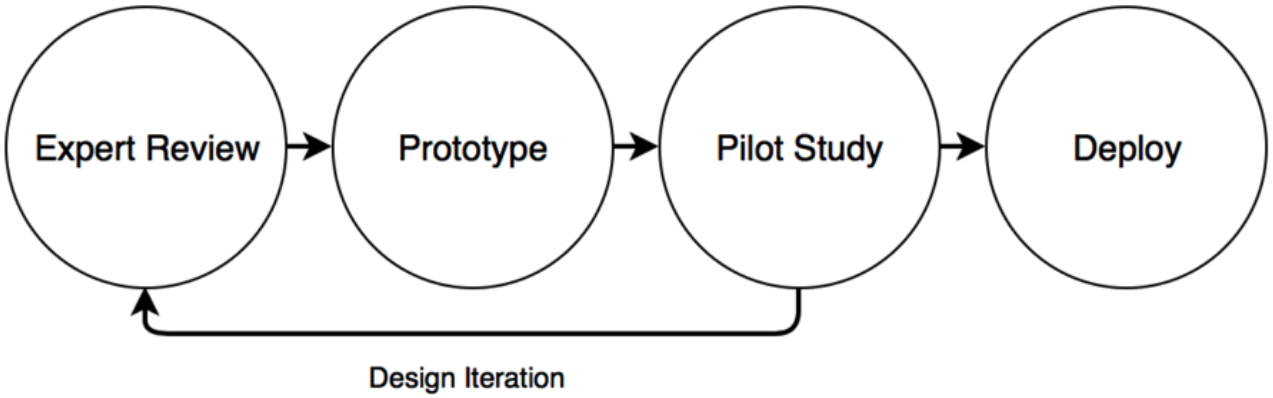
Methods

As highlighted, the assessment of cognitive fatigue commonly takes place within a clinical environment, which does not allow for accurate in situ assessment. This study primarily explored the adaptation of cognitive fatigue tests for delivery using a smartphone. During the study, the validity of the smartphone-based tests was compared to self-reported cognitive fatigue levels, as measured using the MFS. Accordingly, the MFS version employed had to be adapted to allow for delivery on a smartphone. Three cognitive tests were chosen from those identified in the literature. The Spatial Span Test was selected for inclusion to assess working memory. The PVT was also selected, as it has been proven as a cognitive fatigue evaluation method [27]; moreover, the nature of the test allows for small discrepancies in performance to be determined due to the timing of reactions being measured to approximately 1 millisecond. A Mental Arithmetic Test was also included, as simple addition and subtraction questions have been shown to successfully measure cognitive throughput [28], and are effective in assessing the characteristics associated with cognitive fatigue [19]. The adaptation and implementation of these cognitive tests are detailed later in this paper.

Smartphone App Design Process

A multi-disciplinary, iterative approach (as illustrated in Figure 1) was employed to systematically inform and evaluate each aspect of the smartphone app's design. The research team included experts in the fields of ABI, psychology, and interaction design, which informed initial design decisions and the proposed functionality of the app. This *Expert Review* informed the development and design process to establish a *Prototype* app. In turn, this Prototype was then used in a *Pilot Study* that focused on user-centered design, to test our three hypothesized assumptions and to refine the protocol for eventual deployment during the target study. This process permitted the design and development of the app to be informed by clinical theory and practice, as well as commercial design theory and participant feedback. Table 1 shows the different development stages of the study, in addition to the requirements within each stage, that facilitated the desired outcome. Accordingly, each outcome was met before moving on to the next stage.

Figure 1. Smartphone app iterative design process.



XSL•FO
RenderX

Table 1. Requirements for each Development Stage.

Development Stage	Targeted Outcome	Specific Processes	Modifications
Expert Review	Overall workflow of app finalized. Requirements list and features identified.	Research into appropriate cognitive testing methods was carried out alongside mobile workflow and design research. Expert opinion was used for each requirement inclusion.	Sequencing of cognitive tests were discussed and finalized based upon perceived enjoyment and difficulty.
Prototype	Prototype app finalized, based upon requirements set up from Expert Review.	Storyboarding of smartphone app screens to create a structured guide for design. Color scheme and visual design was finalized. Front and backend development finalized to create a stable, secure, and usable app.	Through storyboarding and visual planning, complexity of the cognitive tests was reduced to improve their usability. The Mental Fatigue Scale was adapted for easier use on a smartphone.
Pilot Study	Deployment of the prototype app to a small group of participants in a pilot study. Gather usability feedback from participants.	Carried out a two-week pilot study using the developed app. Participants did not have any prior acquired brain injuries (to validate the app on a healthy population first, as using a vulnerable population is unethical).	Feedback from the pilot study informed the team of usability issues and bugs, and participants suggested improvements to make the app more enjoyable.
Deploy	Deployment of the finalized app to a larger number of participants. Data collection to allow for analysis of results.	Targeted study was used to validate the cognitive fatigue measures that were used within the app.	Targeted study deployment allowed for data collection through participant use of the finalized app. This data was then analyzed to establish the accuracy and validity of proposed cognitive testing methods.

Expert Review

The initial phase of the design process primarily focused on using the research team's multidisciplinary expertise to inform the overall nature and workflow of the app. The order in which the cognitive tests would be delivered was considered an important design decision, as retaining participant engagement (particularly when cognitively fatigued) is crucial. Through Expert Review, the order of the selected tests was based on a combination of the perceived level of difficulty and expected level of participant enjoyment. These considerations subsequently lead to potentially less-engaging tests being sequenced earlier in the workflow of the app, whereas more engaging tests were sequenced later. Accordingly, the cognitive tests presented to participants would become increasingly stimulating over the course of a session. As a result, the ordering of the tests presented by the app was: MFS, Spatial Span Test, PVT, and Mental Arithmetic Test. In addition, the Expert Review also helped to ensure that any changes made to the MFS, specifically its redesign for optimal presentation on a smartphone screen, did not compromise the validity of the corresponding results.

Pilot Study

Following the initial Expert Review stage, a Pilot Study was delivered to a small set of participants (n=5), using a first-iteration prototype app. This stage facilitated an evaluation of the efficacy of each of the selected cognitive tests to help ensure that they would challenge participants sufficiently, and facilitate collection of a dataset with appropriate variability to permit subsequent analyses. The Pilot Study also provided an opportunity to gain feedback on information design and visual design choices made during the Expert Review, which might affect the usability of the app.

Iterative Improvements to Mobile App Cognitive Test Design

The iterative nature of the design process allowed for further Expert Review to inform a response to findings from the Pilot Study, which helped to further refine and optimize the design of the app and the planned larger-scale study. A notable instance of this process is the design of the Spatial Span Test, which was revised from a 5x5 grid to a 4x4 grid layout, as the original design proved difficult to accurately tap when displayed on a smartphone screen. Additionally, a countdown timer and progress bar were added to inform participants of both the remaining time during the test and their ongoing progress.

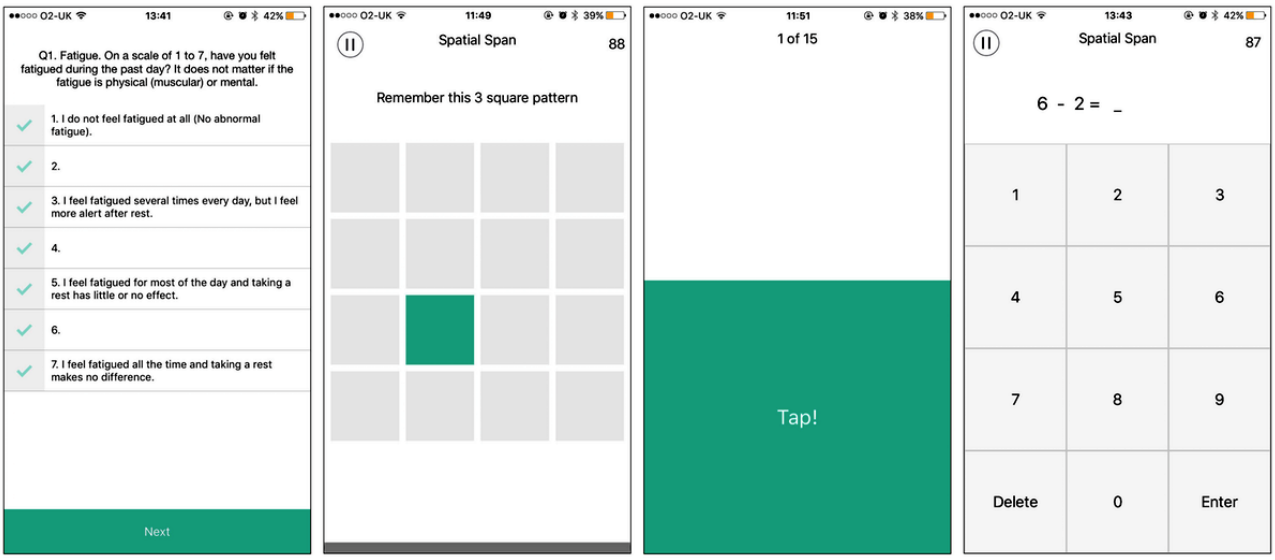
The design of the PVT was also revised to introduce a randomly positioned stimulus, to prevent participants from preempting a response. Immediate visual feedback was also incorporated to encourage concentration and participation during the test. In terms of the Mental Arithmetic Test, initial concerns regarding the layout of the on-screen number keypad used for participant's responses was validated, with a telephone style layout proving acceptable to participants. Moreover, changes to the MFS element of the app required much more care, as the purpose of including it was to provide an established cognitive fatigue assessment tool to help evaluate how effective the mobile cognitive tests were at assessing cognitive fatigue. To maintain the validity of the MFS, it was important that it differed as little as possible from the original paper-based questionnaire. However, the presentation of each question as a separate element marks a departure from the original paper-based version; this change was made in response to feedback from the design process that found a single continuous onscreen set of questions to be awkward from a user-interaction viewpoint.

XSL•FO
RenderX

Target Study Delivery

Upon completion of different iterations of Expert Review, Pilot Study, and subsequent iterations of Prototype development, the smartphone app was ready to *Deploy* within the target study. At the start of the study, guidance was given to each participant on the nature of the study and informed consent was obtained, for the purpose of conducting the study in an ethical manner. Subsequently, a brief tutorial on how to use the app was presented. Once the study had commenced, participants received a push notification daily at 3 o'clock p.m. as a reminder to carry out a session with the app. Upon launching the app, the participant was initially presented with brief instructions regarding the purpose and usage of the app, which was subsequently followed by delivery of the MFS and the set of cognitive tests in the predefined order: (1) MFS; (2) Spatial Span Test; (3) PVT; and (4) Mental Arithmetic Test. Completion of both the MFS and set of cognitive tests took approximately 5 minutes in total. Upon completion, the data collected was initially stored on the mobile device, before subsequent transmission to an online database. The set of tests presented to participants are illustrated in Figure 2. The individual tests and sequencing within the app are detailed in Textbox 1.

Figure 2. Screenshots of the smartphone app's four main screens from left to right: Mental Fatigue Scale, Spatial Span Test, Psychomotor Vigilance Task, and Mental Arithmetic Test.



Textbox 1. Sequencing of measures within the smartphone application.

- As shown on the left side of Figure 2, the first test presented was the MFS, which displayed 15 individual questions to participants, along with an associated 7-point response scale, each on a separate screen with a button at the bottom to progress to the next question.
- The second test presented was the Spatial Span Test, which required each participant to recreate a sequence of flashing boxes that appeared on the screen one after the other in a grid formation. If the participant correctly repeated the presented sequence, they were then presented with a new sequence that contained one additional box. Conversely, if the participant entered an incorrect sequence, then the next sequence was one box shorter, with a minimum sequence of three boxes being presented. The use of adaptive difficulty in this manner was intended to help maintain engagement regardless of the participant's level of performance. The Spatial Span Test lasted for a total of 90 seconds, during which time participants attempted to complete as many sequences as possible.
- Upon completion, the PVT was presente: the mobile implementation of PVT required the participant to attempt to achieve the quickest reaction times possible in response to the presentation of onscreen stimuli 20 times; each time a large block of color was randomly displayed on one half of the smartphone screen.
- The final test presented was the Mental Arithmetic Test, as illustrated on the right side of Figure 2, which required participants to make a single attempt at solving a number of serial addition and subtraction tasks within a predefined time limit of 90 seconds. Although the components of the individual arithmetic operations were randomly generated, they were restricted to ensure that the problem numbers were in the range 0-9 and the result was a positive number in the range 0-18. Immediate visual feedback was provided on the correctness of the answer before presenting the next addition or subtraction task.

XSL•FO
RenderX

Table 2. Data collected from mobile fatigue assessment app.

Test	Factors Measured	Data Collected	Type of Data
Mental Fatigue Scale	Overall fatigue assessment	Date and time test was started	Date/time
		Questionnaire results	Numerical array
Spatial Span Test	Cognitive attention	Total number of sequences complete	Numerical
	Working memory	Longest sequence achieved	Numerical
		Total number of correct sequences	Numerical
		Total number of incorrect sequences	Numerical
		Time to complete each full sequence	Numerical array
Psychomotor Vigilance Test	Alertness	Reaction times	Numerical array
	Reaction time	Total number of premature reactions (wrong)	Numerical
		Total number of timely reactions (right)	Numerical
Mental Arithmetic Test	Cognitive throughput	Total number of correct answers	Numerical
		Total number of incorrect answers	Numerical
		Total number of questions presented	Numerical array
		Correct answer	Numerical array
		User's answer	Numerical array

Data Collection

Each specific test captured a range of different data points to assess relative performance. These data included correct and incorrect responses and time it took to respond during each task. Data that was collected via the app is detailed in Table 2. All collected data were stored locally on the participant's device before being transmitted to a secure central server.

Participant Feedback

Upon conclusion of the study, all participants were invited to provide feedback on the smartphone app using an online version of the System Usability Scale (SUS) [29]. For the SUS, participants were asked to score 10 items with one of five responses ranging from *Strongly Agree* to *Strongly Disagree*. Additionally, participants were asked to freely comment on any aspect of the app or study to provide further feedback.

Participant Recruitment

Participants (n=21) were recruited within Ulster University to undertake the study over a two-week period. Participants with no prior brain injuries were explicitly chosen, as it was considered unethical to test a newly developed method on a clinical population without first understanding how it might be of benefit. Moreover, the experience of cognitive fatigue, while a characteristic of ABI, is not limited to the condition. The mean age of participants recruited was 22 years (standard deviation=4). In addition, participants were required to own an iPhone to ensure that they were familiar with using iOS-based smartphone apps, which removed the need for additional training in device use prior to undertaking the study. Participants were required to use the app once daily to self-assess their cognitive fatigue levels. During the study, 81 individual testing instances were recorded, resulting in an overall daily adherence by participants of 28%.

Target Study Statistical Analysis

Data were analyzed using IBM SPSS version 22. Descriptive statistics were used to provide information on average reaction time, MFS questionnaire total score, total number of timely reactions, total number of correct Spatial Span sequences, total number of correct answers to Mental Arithmetic questions, and longest sequence achieved in the Spatial Span Test (Table 3). Pearson's correlations explored the relationships between the MFS and the three cognitive tests. Finally, multiple linear regression was utilized to investigate variables thought to be predictive of fatigue. Although prompting notifications were issued at 3 o'clock p.m. every day, participants could take the test at any time they chose. This flexibility facilitated a range of testing instances throughout the day. Data obtained from the overall set of results was grouped into three episodic epochs and analyzed separately: (1) morning (midnight to midday); (2) afternoon (midday to 6 o'clock p.m.); and (3) evening (6 o'clock p.m. to midnight). Subsequently, the groupings were utilized to assess if there was a predominant time of day when fatigue levels were higher, based on self-reported figures.

XSL•FO
RenderX

Table 3. Correlations of testing variables to the self-reporting MFS scale.

Mental Fatigue Scale Score	Average Psychomotor Vigilance Task Reaction Time	Average Psychomotor Vigilance Task Reactions Correct	Average Spatial Span Correct	Average Arithmetic Questions Correct	Average Highest Spatial Score Reached
Correlation	.342[a]	.159	-.141	-.016	-.064
Significance (P-value)	.008	.157	.209	.884	.568

[a]Correlation is significant at the 0.01 level (2-tailed).

Data Exclusion

To identify outliers of extreme and incorrect testing instances, data were analyzed using box plots. Analysis of outliers used the standard measure of 1.5 times the interquartile range to find outliers that were viewed to be too far from the central values to be reasonable. A likely cause of outliers is inaccurate self-assessment (eg, assessment scores were abnormally high or low in comparison to testing scores), in conjunction with a simplistic pattern of responses observed (ie, all responses were the lowest or highest possible choices). Removal of outlying instances of this type helped to ensure only reliably accurate tests were included during analysis.

Results

Overall, 81 individual testing instances were recorded by the participant population, resulting in a daily adherence by participants of 28%. Daily reminder notifications resulted in a participant adherence rate of 23% within the first two hours of receiving the notification; this finding indicates that a large proportion of app usage instances occurred due to the daily reminders.

Correlation and Regression Analysis

Initial analysis of the results obtained indicated the average reaction time during the PVT had the strongest correlation with the MFS for evaluating cognitive fatigue, measuring 0.342 (*P*=.008). By comparison, the average Spatial Span sequences correct had a correlation of -0.141 (*P*=.209), and the total Mental Arithmetic questions correct had a correlation of -0.016 (*P*=.884). All correlations are detailed in Table 4.

The best correlation to the MFS was from the average reaction times in the PVT. The significance of this correlation shows that the PVT may be a valid way to assess cognitive fatigue as a replacement for self-reporting. Average reaction time, total number of correct reactions, total number of correct Spatial Span sequences, and total number of correct Mental Arithmetic questions were entered into a multiple linear regression model to determine which factors were predictive of fatigue (see Table 5). Correspondingly, a statistically significant model emerged (*P*=.038, F=2.682), which explained 8% (adjusted R-square=.078) of the variance observed in fatigue scores. The only statistically significant variable to predict fatigue was PVT average reaction time.

Table 4. Correlations of testing variables against the self-reported MFS scale.

Mental Fatigue Scale Score	Average Psychomotor Vigilance Task Reaction Time	Average Psychomotor Vigilance Task Reactions Correct	Average Spatial Span Correct	Average Arithmetic Questions Correct	Average Highest Spatial Score Reached
Correlation	.342[a]	.159	-.141	-.016	-.064
Significance (P-value)	.008	.157	.209	.884	.568

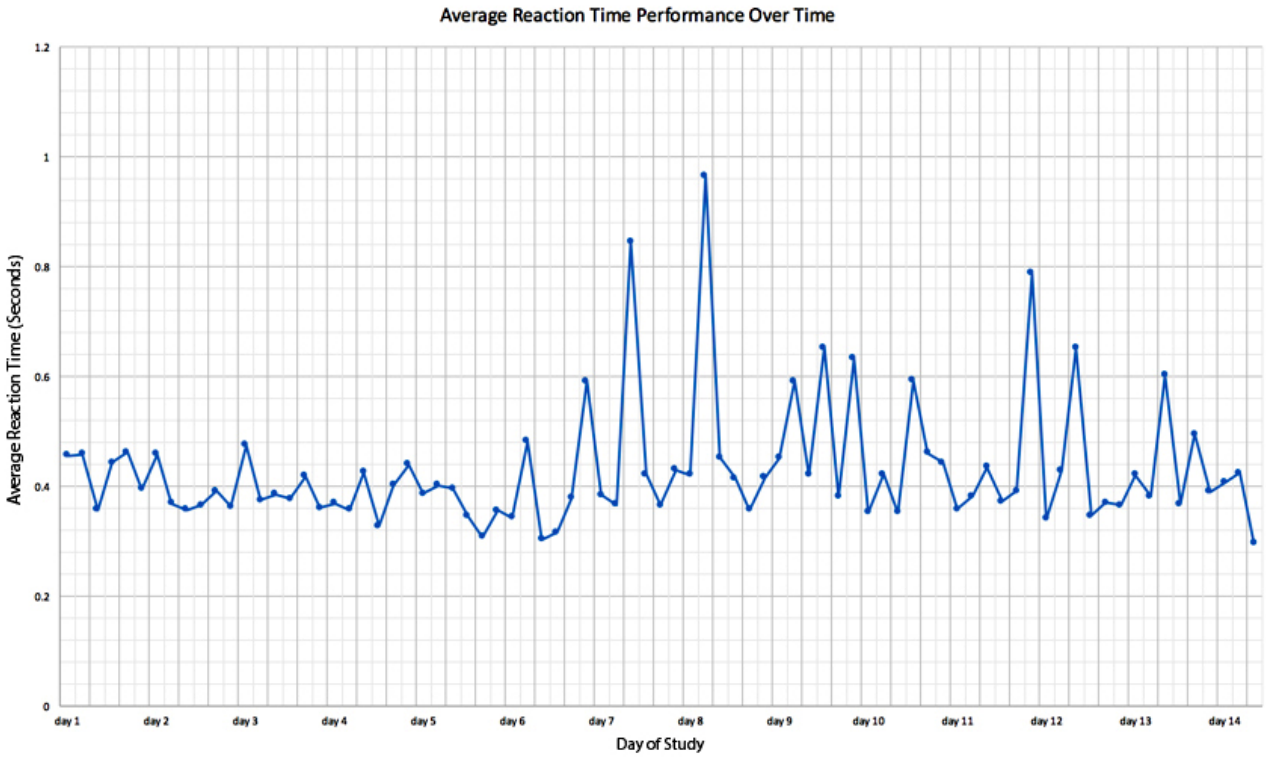
[a]Correlation is significant at the 0.01 level (2-tailed).

Table 5. Standardized and unstandardized regression coefficients for variables entered into the model.

Variable	B	Standard Error	Beta Coefficient	T-statistic	P-value	95% CI
Mental Fatigue Scale	-6.990	27.042	N/A	-0.258	.797	-60.848 to 46.869
Average Reaction Time	46.610	18.898	.312	2.466	.016	8.971 to 84.249
Average Reactions Correct	1.442	1.654	.099	0.872	.386	-1.853 to 4.737
Average Spatial Span Correct	-3.505	2.991	-.138	-1.172	.245	-9.462 to 2.452
Total Mental Arithmetic Correct	0.696	0.488	.179	1.427	.158	-.276 to 1.668

XSL•FO
RenderX

Figure 3. Average reaction time performance over the duration of the study.



Time Series Trend Analysis

Self-reported cognitive fatigue measured by the MFS was also shown to correlate with a slower average reaction time, which supports the hypothesis that information processing speed is a predictor of fatigue. From Figure 3, it can be observed that there is no significant increase or decrease in performance in the PVT over the two-week testing period. Average reaction time stayed consistent apart from several outliers above the upper control limit, which could be attributed to lapses in concentration. If a significant performance increase were observed, it would suggest a level of training and learning over time; however, this is not the case from the overall results obtained.

As previously mentioned, results data were grouped into three epochs throughout the day (*morning*, *afternoon*, *evening*). Analysis of the results obtained from the MFS demonstrates that the morning epoch produced the highest self-reported levels of cognitive fatigue, which were 30.18% higher than during the afternoon epoch, and 28.21% higher than during the evening epoch. Although there was no statistical significance observed, there is an interesting correlation with the metrics outlined in Textbox 2 all being lower, which also indicates higher levels of cognitive fatigue.

It may be appropriate for future work to further analyze the time of day that tests were undertaken. Consequently, this data could support the hypothesis that a reduced level of performance would be seen in participants with higher self-reported fatigue (see Table 6).

After completion of the study, participants were further invited to evaluate the smartphone app using an online version of the SUS and to comment on changes to the app that could encourage higher levels of participation. Subsequently, results obtained from this evaluation indicated that the primary reason for nonadherence was the MFS, which was considered *boring* and *strenuous* in terms of having to answer multiple questions. By comparison, the cognitive tests were perceived to be *fun* and *enjoyable*. This finding supports the study in trying to determine an accurate and more engaging method of in situ fatigue evaluation than is currently available.

Textbox 2. Observed lower scores during morning epoch.

- Mean amount of correctly answered arithmetic questions was 4.34% lower during the morning epoch than the afternoon epoch and 4.58% lower than the evening epoch
- The longest sequence achieved in the Spatial Span test was 1.05% lower in the morning epoch than the afternoon epoch and 1.57% lower than the evening epoch
- The total number of correct Spatial Span sequences achieved was 0.41% lower in the morning epoch than the afternoon epoch and 1.63% lower than the evening epoch

XSL•FO
RenderX

Table 6. Descriptive statistics for morning, afternoon, and evening epochs.

Epoch	N	Minimum	Maximum	Mean	Standard Deviation
Morning					
Reaction time (seconds)	15	0.34846	0.60071	0.41786[a]	0.07409
Mental Fatigue Scale result	15	24	84	40.93	17.123
Total number of timely reactions	15	12	15	14.53	0.834
Total number of correct Spatial Span sequences	15	4	5	4.87	0.352
Total number of correct Mental Arithmetic answers	15	15	28	23.47	3.461
Longest Spatial Score sequence reached	15	4	7	5.67	0.9
Afternoon					
Reaction time (seconds)	45	0.2950	0.65199	0.386[a]	0.05733
Mental Fatigue Scale result	45	14	78	30.2	12.836
Total number of timely reactions	45	12	15	13.89	0.959
Total number of correct Spatial Span sequences	45	4	6	4.89	0.573
Total number of correct Mental Arithmetic answers	45	14	31	24.51	3.501
Longest Spatial Score sequence reached	45	3	9	5.73	1.372
Evening					
Reaction time (seconds)	21	0.3069	0.7600	0.37417[a]	0.1443
Mental Fatigue Scale result	21	14	53	30.81	10.731
Total number of timely reactions	21	12	15	14	0.894
Total number of correct Spatial Span sequences	21	4	6	4.95	0.590
Total number of correct Mental Arithmetic answers	21	19	31	24.57	3.682
Longest Spatial Score sequence reached	21	3	9	5.76	1.411

[a]Denotes the median reaction time

Discussion

While it is clear from the results obtained that all cognitive fatigue tests showed a degradation in the participant's performance when there is a higher level of self-reported fatigue (as hypothesized prior to the study), analysis indicates that only the PVT is significant enough to make claims on its validity. Subsequently, both the Spatial Span Test and Mental Arithmetic Test do not provide a strong correlation with the MFS for measuring cognitive fatigue. Although a correlation may be observed from all three of the tests as hypothesized, currently only the PVT has the potential to accurately assess cognitive fatigue using a mobile device, as it is able to determine the current level of mental fatigue concordant with the levels reported using the MFS. Furthermore, PVT is less demanding for participants to undertake than the other cognitive tests. Feedback from a usability viewpoint also suggested that the PVT was considered by participants to be more engaging than the MFS. From a participant perspective, assessment of cognitive fatigue is feasible via a smartphone app; however, adherence to regular testing is crucial to gaining an understanding of the condition over time.

Adaptation of the smartphone app to increase the overall time on the cognitive tests could potentially lead to more accurate evaluation results across all tests, although this is only possible if participants are willing to take part in a longer daily task. Tailoring time-on-task to promote continued adherence, while also collecting sufficient data to accurately assess the condition, is a critical balance that requires further research. Such studies will enable a greater understanding of the testing duration required to permit accurate mobile-based cognitive fatigue evaluation. Future work may benefit from increased time-on-task for each of the three cognitive tests.

When considering the scores of the PVT from individual participants, it can be observed that there was no improvement in scores over time. It is important that the participants find the test equally taxing over time, so that the same indicators of fatigue are present, rather than a participant learning from previous sessions and becoming increasingly expert in undertaking the assessment. Such learning would consequentially mask any underlying cognitive fatigue.

Principal Findings

The approach of interpreting fatigue using a mobile device has been demonstrated to have validity. However, a number of improvements could be made to the process, in addition to changes to the specific data that is collected. Such changes could include collecting data on environmental factors such as the quantity and quality of sleep, daily exercise levels, current location and social environment, and current emotional status.

XSL•FO
RenderX

Based on the study performed, participant feedback indicated that carrying out the MFS was one of the least enjoyable parts of the overall process, which potentially lead to a reduction in engagement. The use of PVT has been shown to provide a similar capability in the assessment of cognitive fatigue, meaning that future work can potentially exclude the use of MFS. Subsequently, this exclusion may further increase user participation rates, which in turn may potentially increase the accuracy of the cognitive fatigue evaluation. An abridged variation of PVT was employed within the smartphone app; however, it is anticipated that a higher degree of accuracy could be achieved by using a longer test session. Accordingly, future work should increase the length of the PVT utilized, which could produce a more precise indicator of fatigue.

Limitations

A limitation of the study is that daily participation was not enforced. This enforcement could have increased the amount of data obtained, thereby improving the confidence of the findings. Given the real-world approach employed throughout the study, the only available (and appropriate) participation encouragement mechanism was seen to be mobile device notifications. Future work will aim to improve participation rates, in conjunction with increasing the amount of data collected to potentially obtain a more accurate assessment of daily fatigue levels. Furthermore, all cognitive test metrics that were employed alongside the MFS were constrained to a relatively short timeframe, in order to aid participation rates. However, this limitation may have had an adverse impact on the effectiveness of the tests in identifying small discrepancies in performance. Other limitations of the current study include the short duration of the overall study; future work will seek to address all of these limitations.

Comparison with Prior Work

The validity of assessing fatigue using both the Spatial Span Test and Mental Arithmetic Test has been proven to be effective by Van Dongen et al [19] and Johansson et al [18]. In both cases, longer forms of the tests were utilized; however, a static nonmobile platform was employed for test delivery. Aspects of the research carried out by Johansson et al [18] directly corresponds with the research discussed in this paper. In their comparative study on the utility of self-assessment and cognitive tests, Johansson et al found that both the questionnaires and cognitive tests produced the same results when measuring fatigue in an individual [18]. In particular, significant decreases in information processing speed were measured using the Digit Symbol Coding and Trail Making Tests [18]. Furthermore, the cognitive tests that were used proved to be quicker than filling out a self-assessment questionnaire. Accordingly, the results obtained from the research carried out in this paper concur that information processing speed, measured through cognitive tests such as PVT, correlates to subjective measurements of fatigue.

Work carried out by Thomson et al [30] demonstrates visual reaction time using a smartphone is a simple and repeatable method for objectively monitoring the effects of sedation. Visual reaction times were found to be considerably less variable than patient-assessed sedation scores and could be used to identify impending over-sedation [30]. Again, this finding shows the validity of a reaction time test as a valid measure of attention and wakefulness. Work carried out in this paper shows its adaptability and provides evidence that it can be translated for use in measuring cognitive fatigue.

However, in terms of the efficacy of mobile-based self-assessment, Swendeman et al [25] carried out a study on the validity of self-reporting via a smartphone and found relatively low compliance with daily assessment. Consequently, these findings indicated the potential unreliability of self-assessment with respect to assessment completion rates [25]. A more playful approach to consider would be the integration of game-based testing, which may potentially improve the adherence to completion, as suggested by the results obtained from the research presented herein.

One key concern with cognitive tests (such as PVT) that have been previously adapted for a mobile platform [16] is general usability, particularly for individuals with brain injuries, due to their specific cognitive impairments. Although this concern could potentially be alleviated with initial training, it does not allow for immediate assistance if issues are encountered when used outside of a clinical environment. Kay and Rector initially adapted PVT for use with a modern touchscreen device to assess reaction time as a means to evaluate fatigue in everyday life [16]. PVT requires accurate evaluation of reaction time (down to approximately 1 millisecond), making accurate input sensitivity crucial. Within the work conducted by Kay and Rector, they first evaluated different input methods including *touchdown*, *swipe*, and *touch-up*, and concluded that participants preferred *touchdown* due to it similarities to a button press [16]. The study was then carried out with 20 participants to assess everyday fatigue, and results were comparable to those gathered from the desktop computer-based implementation of PVT [16]. Kay and Rector concluded that future work in the area should aim to add situational context to the app so that assessment could be moved away from the clinical environment, thereby reducing cost and increasing contextual evaluation when needed.

Conclusions

Previous studies have confirmed the use of the MFS and PVT as a feasible way to assess cognitive fatigue in a clinical setting. This study demonstrated that smartphone-based adaptations of these proven methods are internally compatible methods of assessing cognitive fatigue in situ. The smartphone app presented in this research provides a potentially effective tool for the individual evaluation of cognitive fatigue levels in situations where formal intervention and assessment approaches are neither feasible nor available. Furthermore, the use of smartphone app-based fatigue assessment permits evaluation to be carried out on a continual daily basis.

All three of the cognitive tests employed within the smartphone app produced positive participant feedback, with some participants even indicating that they would like personal scores, as it would further encourage them to participate more frequently. Consequently, by introducing a competitive aspect to the cognitive tests, participant effort and daily participation rates could potentially be improved. Future work may additionally permit the provision of real-time data to relevant medical professionals, so effective and timely interventions can

XSL•FO
RenderX

be arranged if extreme fatigue becomes apparent. Correspondingly, there exists three main areas that in situ fatigue assessment could benefit from sensor data and contextual factors: (1) to improve notifications so that daily participation can be increased; (2) to provide additional data that may give a more precise indication of the occurrence of cognitive fatigue; and (3) to track a participant's daily activity and advise them on appropriate steps to further combat cognitive fatigue.

Based on the three episodic epochs, results from both the MFS and PVT indicated that the morning epoch produced higher levels of fatigue. Accordingly, this knowledge could be employed on an individual basis to help tailor the time of delivery of the test session, to pinpoint higher fatigue levels throughout the day. Adaptation of notifications (based on collected data over time) would further facilitate assessment when fatigue is usually at its highest. Correspondingly, identifying the time of day when fatigue levels are high offers an increased ability to prevent mental fatigue by preempting its arrival and suggesting steps to take to reduce its severity. Assessing the time of an individual's participation could help inform notifications for the most likely time of adherence.

Acknowledgments

This research was funded and supported by the Department of Learning and Employment, Northern Ireland. This publication was supported by the Smart Environments Research Group, within the Computer Science Research Institute, Ulster University.

Conflicts of Interest

None declared.

References

1. Taylor CA, Bell JM, Breiding MJ, Xu L. Traumatic brain injury-related emergency department visits, hospitalizations, and deaths - United States, 2007 and 2013. MMWR Surveill Summ 2017 Mar 17;66(9):1-16. [doi: 10.15585/mmwr.ss6609a1] [Medline: 28301451]
2. Thurman DJ, Alverson C, Dunn KA, Guerrero J, Sniezek JE. Traumatic brain injury in the United States: a public health perspective. J Head Trauma Rehabil 1999 Dec;14(6):602-615. [Medline: 10671706]
3. Tagliaferri F, Compagnone C, Korsic M, Servadei F, Kraus J. A systematic review of brain injury epidemiology in Europe. Acta Neurochir (Wien) 2006 Mar;148(3):255-268. [doi: 10.1007/s00701-005-0651-y] [Medline: 16311842]
4. Lee KA, Hicks G, Nino-Murcia G. Validity and reliability of a scale to assess fatigue. Psychiatry Res 1991 Mar;36(3):291-298. [Medline: 2062970]
5. Johansson B, Starmark A, Berglund P, Rödholm M, Rönnbäck L. A self-assessment questionnaire for mental fatigue and related symptoms after neurological disorders and injuries. Brain Inj 2010 Jan;24(1):2-12. [doi: 10.3109/02699050903452961] [Medline: 20001478]
6. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol 1989 Oct;46(10):1121-1123. [Medline: 2803071]
7. Shahid A, Wilkinson K, Marcu S, Shapiro C. Visual Analogue Scale to Evaluate Fatigue Severity (VAS-F). In: STOP, THAT and One Hundred Other Sleep Scales. New York: Springer; 2011:399-402.
8. Johansson B, Rönnbäck L. Mental fatigue; a common long term consequence after a brain injury. In: Brain Injury - Functional Aspects, Rehabilitation and Prevention. Rijeka: InTech; Mar 2012:2009.
9. Rödholm M, Starmark JE, Svensson E, Von Essen C. Astheno-emotional disorder after aneurysmal SAH: reliability, symptomatology and relation to outcome. Acta Neurol Scand 2001 Jun;103(6):379-385. [Medline: 11421850]
10. King NS, Crawford S, Wenden FJ, Moss NE, Wade DT. The Rivermead Post Concussion Symptoms Questionnaire: a measure of symptoms commonly experienced after head injury and its reliability. J Neurol 1995 Sep;242(9):587-592. [Medline: 8551320]
11. van Zomeren AH, van den Burg W. Residual complaints of patients two years after severe head injury. J Neurol Neurosurg Psychiatry 1985 Jan;48(1):21-28 [FREE Full text] [Medline: 3973618]
12. Johansson B, Rönnbäck L. Mental fatigue and cognitive impairment after an almost neurological recovered stroke. ISRN Psychiatry 2012;2012:1-7 [FREE Full text] [doi: 10.5402/2012/686425] [Medline: 23738208]
13. Herlofson K, Larsen JP. Measuring fatigue in patients with Parkinson's disease - the Fatigue Severity Scale. Eur J Neurol 2002 Nov;9(6):595-600. [Medline: 12453074]
14. Johansson B, Ronnback L. Evaluation of the Mental Fatigue Scale and its relation to cognitive and emotional functioning after traumatic brain injury or stroke. Int J Phys Med Rehabil 2014;02(01). [doi: 10.4172/2329-9096.1000182]
15. Grant DA, Honn KA, Layton ME, Riedy SM, Van Dongen HP. 3-minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation. Behav Res Methods 2017 Jun;49(3):1020-1029. [doi: 10.3758/s13428-016-0763-8] [Medline: 27325169]
16. Kay M, Rector K, Consolvo S, Greenstein B, Wobbrock JO, Watson NF, et al. PVT-touch: adapting a reaction time test for touchscreen devices. In: Pervasive Computing Technologies for Healthcare (PervasiveHealth). 2013 May 5 Presented

XSL•FO
RenderX

at: 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops; 5-8 May 2013; Venice p. 248-251.
17. Timmers C, Maeghs A, Vestjens M, Bonnemayer C, Hamers H, Blokland A. Ambulant cognitive assessment using a smartphone. Appl Neuropsychol Adult 2014;21(2):136-142. [doi: 10.1080/09084282.2013.778261] [Medline: 24826507]
18. Johansson B, Berglund P, Rönnbäck L. Mental fatigue and impaired information processing after mild and moderate traumatic brain injury. Brain Inj 2009 Dec;23(13-14):1027-1040. [doi: 10.3109/02699050903421099] [Medline: 19909051]
19. Van Dongen HP, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 2003 Mar 15;26(2):117-126. [Medline: 12683469]
20. Gartenberg D, McGarry R, Pfannenstiel D, Cisler D, Shaw T, Parasuraman R. Advances in Cognitive Engineering and Neuroergonomics. Florida: CRC Press; 2012. Development of a neuroergonomic application to evaluate arousal URL: http://www.crcnetbase.com/doi/pdfplus/10.1201/b12313-32 [accessed 2017-08-06] [WebCite Cache ID 6sWbIdpt4]
21. Ferguson SA, Smith BP, Browne M, Rockloff MJ. Fatigue in emergency services operations: assessment of the optimal objective and subjective measures using a simulated wildfire deployment. Int J Environ Res Public Health 2016 Jan 29;13(2):171 [FREE Full text] [doi: 10.3390/ijerph13020171] [Medline: 26840323]
22. Liu K, Li B, Qian S, Jiang Q, Li L, Wang W, et al. Mental fatigue after mild traumatic brain injury: a 3D-ASL perfusion study. Brain Imaging Behav 2016 Sep;10(3):857-868. [doi: 10.1007/s11682-015-9492-3] [Medline: 26781483]
23. Wechsler D. Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV). San Antonio, TX: NCS Pearson; 2008. URL: https://pdfs.semanticscholar.org/2858/f906a462c4424192f80361f689bdec24c16d.pdf [accessed 2017-07-27] [WebCite Cache ID 6sGxCShOA]
24. Reitan RM, Wolfson D. The Halstead-Reitan neuropsychological test battery: theory and clinical interpretation. Tucson, AZ: Neuropsychology Press; 1985.
25. Swendeman D, Comulada WS, Ramanathan N, Lazar M, Estrin D. Reliability and validity of daily self-monitoring by smartphone application for health-related quality-of-life, antiretroviral adherence, substance use, and sexual behaviors among people living with HIV. AIDS Behav 2015 Feb;19(2):330-340 [FREE Full text] [doi: 10.1007/s10461-014-0923-8] [Medline: 25331266]
26. Engström AL, Lexell J, Lund ML. Difficulties in using everyday technology after acquired brain injury: a qualitative analysis. Scand J Occup Ther 2010 Sep;17(3):233-243. [doi: 10.1080/11038120903191806] [Medline: 19707949]
27. Drummond SP, Bischoff-Grethe A, Dinges DF, Ayalon L, Mednick SC, Meloy MJ. The neural basis of the psychomotor vigilance task. Sleep 2005 Sep;28(9):1059-1068. [Medline: 16268374]
28. Thorne DR, Genser SG, Sing HC, Hegge FW. The Walter Reed performance assessment battery. Neurobehav Toxicol Teratol 1985;7(4):415-418. [Medline: 3840579]
29. Brooke J. SUS-A quick and dirty usability scale. In: Usability evaluation in industry. London: Taylor & Francis; 1996.
30. Thomson AJ, Nimmo AF, Tiplady B, Glen JB. Evaluation of a new method of assessing depth of sedation using two-choice visual reaction time testing on a mobile phone. Anaesthesia 2009 Jan;64(1):32-38 [FREE Full text] [doi: 10.1111/j.1365-2044.2008.05683.x] [Medline: 19087003]

Abbreviations

ABI: Acquired Brain Injury
FSS: Fatigue Severity Scale
MFS: Mental Fatigue Scale
PVT: Psychomotor Vigilance Task
SUS: System Usability Scale
VAS-F: Visual Analogue Scale for Fatigue

Edited by G Eysenbach; submitted 03.08.16; peer-reviewed by R Band, DA Grant, A Blokland, F Seoane; comments to author 17.12.16; revised version received 28.01.17; accepted 27.03.17; published 17.08.17.

Please cite as:
Price E, Moore G, Galway L, Linden M
Validation of a Smartphone-Based Approach to In Situ Cognitive Fatigue Assessment
JMIR Mhealth Uhealth 2017;5(8):e125
URL: http://mhealth.jmir.org/2017/8/e125/
doi:10.2196/mhealth.6333
PMID:28818818

XSL•FO
RenderX

©Edward Price, George Moore, Leo Galway, Mark Linden. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 17.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Usability Testing of the BRANCH Smartphone App Designed to Reduce Harmful Drinking in Young Adults

Joanna Milward[1], BSc, MSc; Paolo Deluca[1], PhD; Colin Drummond[1], MD, FRCPsych; Rod Watson[2], MPH; Jacklyn Dunne[1], BSc, MSc; Andreas Kimergård[1], PhD

[1]Addictions Department, King's College London, London, United Kingdom

[2]Health Innovation Network, London, United Kingdom

Corresponding Author:
Joanna Milward, BSc, MSc
Addictions Department
King's College London
4 Windsor Walk
Denmark Hill
London, SE58BB
United Kingdom
Phone: 44 7590829898
Email: joanna.milward@kcl.ac.uk

Abstract

Background: Electronic screening and brief intervention (eSBI) apps demonstrate potential to reduce harmful drinking. However, low user engagement rates with eSBI reduce overall effectiveness of interventions. As “Digital Natives,” young adults have high expectations of app quality. Ensuring that the design, content, and functionality of an eSBI app are acceptable to young adults is an integral stage to the development process.

Objective: The objective of this study was to identify usability barriers and enablers for an app, BRANCH, targeting harmful drinking in young adults.

Methods: The BRANCH app contains a drinking diary, alcohol reduction goal setting functions, normative drinking feedback, and information on risks and advice for cutting down. The app includes a social feature personalized to motivate cutting down and to promote engagement with a point-based system for usage. Three focus groups were conducted with 20 users who had tested the app for 1 week. A detailed thematic analysis was undertaken.

Results: The first theme, “Functionality” referred to how users wanted an easy-to-use interface, with minimum required user-input. Poor functionality was considered a major usability barrier. The second theme, “Design” described how an aesthetic with minimum text, clearly distinguishable tabs and buttons and appealing infographics was integral to the level of usability. The final theme, “Content” described how participants wanted all aspects of the app to be automatically personalized to them, as well as providing them with opportunities to personalize the app themselves, with increased options for social connectivity.

Conclusions: There are high demands for apps such as BRANCH that target skilled technology users including young adults. Key areas to optimize eSBI app development that emerged from testing BRANCH with representative users include high-quality functionality, appealing aesthetics, and improved personalization.

(JMIR Mhealth Uhealth 2017;5(8):e109) doi:10.2196/mhealth.7836

KEYWORDS

alcohol; drinking; young adults; mHealth; brief intervention; apps; usability testing; user experience; focus group

Introduction

Electronic screening and brief intervention (eSBI), delivered through devices such as computers, tablets and smartphones, is an increasingly popular method to deliver alcohol brief interventions [1-3]. Meta-analyses demonstrate eSBI to be effective in reducing alcohol consumption by 1 to 2 drinks per week after 6 months compared to controls [4,5]. However the majority of eSBIs that have been evaluated are Web-based, as opposed to app-based programs. A recent systematic review of mobile interventions for alcohol and substance use reported that while mobile delivery of alcohol interventions is an acceptable

XSL•FO
RenderX

and effective communication channel, targeting of interventions to specific populations is required [6]. Another enduring challenge for eSBI development is usability. A recent review of online feedback for existing alcohol apps reported that nonintuitive functionality, software malfunctions, and lack of personalization of content are frequently cited criticisms of existing alcohol apps [3,7]. While many eSBI apps are available, the majority are not evidence-based [2,8]. Indeed eSBI is still in its infancy and should be subjected to multiple stages of rigorous development and testing [9].

The app "BRANCH" targets harmful drinking in young adults (18-30 year olds). This study reports the second stage of development of the BRANCH app that aims to evaluate the usability of the prototype app to improve its functioning, content, and design. Usability testing is a crucial stage in the app design process as it provides end user feedback about what does and what does not work in the program. This is an important step as the effectiveness of an electronic intervention has been shown to be associated with its level of usability [10]. Conversely, poor usability is associated with nonusage [11].

Usability testing is widely used in digital health intervention design, and more recently in the development of digital programs for alcohol harm reduction. A range of methods for usability testing exist including questionnaires [12,13], think-aloud observation, and interview-based techniques [14]. For alcohol field, the interview-based techniques are more widely used. Crane et al [7] conducted usability testing on an app designed to reduce alcohol consumption using think aloud testing and semistructured interviews to determine if the features in the app were acceptable and feasible to users and also determine what could be improved. Similarly, Davies [15] assessed the feasibility and acceptability of a digital alcohol harm reduction tool for adolescents, and Dulin [16] conducted usability testing via questionnaires and interviews of a smartphone intervention for adults. These studies highlighted issues of ease-of-use, clarity of information, and appealing design as integral to improvement of the digital tools.

However, the BRANCH app is different from other alcohol harm reduction apps in that it targets young adults specifically. Young adults' preferences for the usability of apps may differ from other groups who have lower usage rates of apps, such as older population. Indeed, young adults have the highest level of smartphone ownership out of all age groups [17]. Referred to as "digital natives," many are proficient in technology use having grown up being exposed to computers, smartphones, and the Internet [18] . It is therefore critical to optimize usability for apps targeting this age group. Moreover, usability data and experiences may be of use in the development of other apps designed to influence substance use among young adults.

The aim of the current study was to explore experiences of app usability, in terms of content, functionality, and design of the BRANCH app to improve user experience.

Methods

Design and Setting

Qualitative interviews were chosen as most appropriate to meet the aim of the current project. While other methods such as questionnaires can provide useful overviews of app functioning, the level of fine-grained detail provided can be limited. Qualitative methods provide additional insights as they encourage participants to think about ways to improve usability and identify unanticipated challenges [19]. Focus groups were chosen as the most appropriate method of data collection instead of 1:1 interviews as they allow for views to be developed and discussed, and also for individual opinions to be expressed. As the aim of the study was to identify specific usability issues, the intention was that participants in the focus groups would remind and prompt each other about specific issues they experienced, hence yielding higher identification of usability issues compared to 1:1 interviewing. Ethical approval was obtained from the University Ethics Committee (ref. number HR14/150453).

Participants

Young adults, aged 18-30 years, who lived in South London and scored 16+ on the alcohol use disorders identification test (AUDIT), indicating harmful drinking (AUDIT score between 16-19) or probable dependence (AUDIT score ≥20) [20].

Materials

The app "BRANCH" targeted harmful drinking in young adults. Prototype development was informed by 3 studies: (1) a systematic review of engagement promoting strategies for online substance use interventions, (2) a review of user-reviews of existing alcohol eSBI apps available on iTunes and Google Play stores [3], and (3) focus groups with young adults drinking at harmful levels and residing in South London exploring their preferences for content features and style for an alcohol brief intervention app [3]. The prototype was designed iteratively, using a user-centered design approach (UCD) [21,22]. This involved collaboration between the program developers, research team, and target population. The core functions of BRANCH were based on the FRAMES model (feedback, responsibility, advice, menu of options, empathy, and self-efficacy) of alcohol brief interventions [23], which has been previously adapted for eSBI [24]. The FRAMES model is based upon the principles of motivational interviewing, an established and evidence based method to reduce alcohol harm [25]. The core functions included a drinking diary for recording alcohol consumption (see Multimedia Appendix 1) and a goal setting function where users could set weekly goals based on cost, calories, and alcohol units as well as setting a drink free day (see Multimedia Appendix 2). Users monitored their drinking over time and received feedback on it, both descriptively and graphically. Information on drinking risks and cutting down was available to users (see Multimedia Appendix 2).

In addition to the core components, several strategies to optimize engagement with the app were incorporated. In order to tailor the app to young adults, these strategies were developed in collaboration with a user group recruited from the target

XSL•FO
RenderX

population. The two main components young adults requested were social features and tailoring of the app to broader wellbeing issues associated with alcohol use [3]. Consequently, the app included a Twitter-style newsfeed enabling interaction between app users, as well as providing personalized notifications, motivational messaging, and reminders based on goals (see Multimedia Appendix 3). The research team could also upload relevant material for young adults, such as links to Web-based articles, YouTube videos, and photos. There was also a personalization feature in which the app users selected their motivations for cutting down drinking when signing up. These motivations were chosen by the user group and included options such as mental health, sugar intake, appearance, and weight. Personalized feedback and targeted information was delivered to users based on their selection of motivators. For example, if a user selected "fitness" as their primary motivator, they would receive tailored messaging on their newsfeed on this topic as well as feedback on how much exercise would be needed to burn off the alcohol calories they had consumed over the last week. Additionally, users were allocated to a team based on these motivators. Users could compare their progress against other users in their team and were awarded points for engaging with the app (see Multimedia Appendix 3). The app was Web-based, and could therefore be accessed from all devices.

Recruitment

Participants were recruited through Gumtree, an online classified and social community website. Potential participants were invited to take part in a focus group interview to test the app designed to support young adults reducing their alcohol use. A link was provided to an online screening questionnaire where they completed the AUDIT and provided their contact and basic demographic data. Eligible participants were invited to take part. Participants who attended a focus group interview were compensated £30 for their time.

Data Collection

The focus group participants were provided with the BRANCH app 1 week before and instructed to use it daily over the course of the week to monitor their drinking. Participants were provided with the following specific tasks to complete: (1) set up the app, (2) fill out a weeks' worth of drinking in the drinks diary, (3) set a goal and 3 drink free days, and (4) join a team and review team feedback (See Multimedia Appendices 1-).

A topic guide was designed to explore the extent to which the participants found the different features acceptable, in terms of content, features, and design. Participants were asked to give their views on their experience of using the different features in the app, focusing around what did and what did not work well. Written informed consent was obtained before commencement of the focus group. The focus group was facilitated by 3 researchers (JM, JD, and RD). They were introduced to participants as researchers who were developing an app to reduce harmful drinking. JM led the group; JD and RD took notes of key themes.

Data Analysis

Focus group interviews were recorded and transcribed verbatim by a professional transcription company. All data were coded using NVivo qualitative data analysis software (QSR International Pty Ltd. Version 10, 2012). JM coded the transcripts and JD double-coded. Any discrepancies that arose were discussed until a consensus was reached.

A thematic analysis was undertaken as outlined by Braun and Clarke [26]. A deductive approach was used. Usability issues were coded into categories of (1) App content, (2) Functionality, or (3) Design. Each of these categories was considered in terms of being a barrier or an enabler to use or a suggested improvement. Themes were systematically refined by going back and forth between the data and the coding framework.

Results

A total of 70 people completed the online screening survey. Of these, 32 (46%) scored 16 or more on the AUDIT, were between 18-30 years of age, and lived in South London.

Participant Characteristics

A total of 20 participants attended 1 of the 3 focus groups with 6 to 7 participants in each over a 1-month period in August 2016. These numbers in each group allowed for meaningful discussions to take place between participants. Of the 20 participants, 18 were female (90%), 10 (50%) were employed, 1 (5%) was unemployed, and 9 (45%) were students. The mean age was 23 years (SD 3.9). The mean AUDIT score across the participants was 21 (SD 5.7).

Usability Analyses

Functionality

This theme reflected how all the participants wanted simple and fast functionality, with features that would minimize the amount of effort, input, and time required from them. Features that had efficient and automated functionality were praised as enablers to usability. For example, being able to quickly complete functions in the app such as posting a newsfeed message, setting up the app, or adding data to the drinking diary via a guided walkthrough.

However, while participants typically appreciated features of the app that functioned well, some still expressed views that there were scopes for improvements. When discussing their experiences of using the app, several participants reported becoming quickly frustrated when a feature was hard-to-use or took too much time. Some expressed strong views that their time was precious and that they did not want to spend unnecessary time inputting data, such as when entering drinks into the drinking diary:

> *Generally when I'm having a cocktail I can't put in a brand so it's difficult...for example, a Long Island Iced Tea...has five different variants of alcohol in it...* [P7; Focus group 3]

> *And if you had a double, like a double gin and tonic, I had to put in a single and then a shot because you can't put in a double.* [P3; Focus group 3]

Many participants also wanted push notifications (ie, reminders) on their phone to prompt them to use the app. A few wanted reduced scrolling to find information on the newsfeed and a

XSL•FO
RenderX

help feature to facilitate app usage. Functionality requests were typically related to increasing the level of "automation" in the app, subsequently reducing the amount of time required to use it. For the majority of the participants, increasing automation seemed to improve usability and make the app more valued.

Another important finding that arose was related to the impact that the quality of the functionality of an app can have on a user's intentions to interact with it. There was a consensus for deleting apps that did not immediately function as expected:

> *The NHS Change for Life? Was it yellow?* [P1; Focus group 3]
>
> *Yes, it was that and I hated it. The notifications were awful, it just wouldn't let me do anything so I was like what's the point of you? Just delete.* [P7; Focus group 3]
>
> *I really hate apps that do that when they send you notifications as well and you still get emails.* [P5; Focus group 3]

This underlines a difficult balance between too little and too much output, both of which impact usability. As this quotation suggests, it is not a case of providing users with as much content as possible, instead carefully tailoring the content to the requirements of the individual user. Overall, what participants appeared to want from the functionality of BRANCH were features "at their fingertips," pressing as few buttons as possible, in a fast and seamless interaction. If this was not achieved then there was a risk of losing the user entirely.

Design

The design of the app was the usability issue most discussed in all the focus groups. Participants frequently commented on the need for the app to have a well-considered design, with short pieces of clearly presented text and features that were easily distinguishable from one another. An aesthetically displeasing app seemed to be considered to be a major usability barrier. For instance, many participants commented on how the newsfeed was too text heavy, with large blocks of text, not separated by pictures or colors, which made the information difficult to absorb and certain core features not clearly distinguishable:

> *Did anyone get any goal related messaging?* [JM; Focus group 1]
>
> *I got something like 'you went to a barbeque this weekend'...* [P7; Focus group 1]
>
> *No, it wouldn't have been that, it would have had a star next to it here on the newsfeed...* [P1; Focus group 1]
>
> *Oh yeah, look, oh dear...didn't achieve your goals.* [P3; Focus group 1]

This is a critical issue as participants were not aware that certain integral features aimed at reducing alcohol-related harm even existed, as the buttons and tabs through which they were advertised were not easily discernible from other newsfeed content.

While participants seemed to express the view that too much text and a poor design hindered usability, the use of multimedia in the "information about drinking" section was the preferred style of most of the participants. This section had been especially designed with an infographic style, which aims to present complex information in easy-to-digest, short components in a colorful, succinct presentation with lots of pictures (see Multimedia Appendix 2):

> *I like that, I think that would be better, that style, on the Newsfeed. Because I think I'd be more inclined to read it with the pictures and bits and bobs. Cause that looks more...like if you clicked on that and then there's something in different colours and...Yeah, looks really good.* [63; Focus group 2]

From this extract, it is apparent that the design of the app can impact the level of engagement a user has with it; in this case, motivating the user to read through information provided. Appealing designs seemed to promote usage while those which were difficult to digest or read through discouraged usage.

Another significant finding about the design of the app was the importance of consistency in style. While efforts were made to present a consistent and coherent theme throughout the design of BRANCH, many participants still expressed a desire for a greater level of consistency. However, it appeared from the views expressed in the focus groups that while participants unanimously agreed that a consistent and appealing design was integral to app usability, participants could not necessarily agree on the type of design which was the most appealing, with many different opinions being expressed. This identifies another usability barrier for BRANCH, ie, providing an aesthetic that is agreeable to all.

> *It looks a bit like it was maybe made on Word. So it just needs to be a bit more like corporate.* [P3; Focus group 2]
>
> *It was very green.* [P4; Focus group 2]
>
> *I didn't find it that exciting and fun to go into like other apps...they're colourful and, you know, didn't find it that...found it quite bland.* [P2; Focus group 2]

Overall, this theme highlights how important a well-considered design is to the usability of an app. Participants reported that to promote usability and subsequent motivation to engage with the app, information must be clearly and succinctly provided, with standout features and a consistent design throughout.

Content

Maximized personalization of app content emerged as an important usability issue. Personalization was presented as a two-fold concept where participants wanted the app to be both automatically personalized to them as much as possible, but they also wanted the autonomy to personalize the app by themselves. Providing autonomy gave participants a sense of empowerment over their interactions with the app, and made them feel in control of their drinking.

While participants commonly enjoyed the personalization of the motivators and feedback features, there were aspects of the app that some reported could be improved. A particular issue was the daily newsfeed messages. These messages were sent to the entire user-base each day and were generic and not

XSL•FO
RenderX

targeted to the individual. The majority said that these messages were not relatable (See Multimedia Appendix 4). Indeed, participants appeared to adeptly identify any area of the app that was not specifically tailored to each user. Surprisingly, this is in spite of the messages being written by the user-group of young adults with whom the app was developed. For the focus group participants, it seemed that personalization to every aspect of the individual, such as motivations for use, preferences for style and content, and even targeting geographical area, were usability enablers.

Equally, while participants wanted maximum automated personalization, they also wanted independence and autonomy to personalize the app by themselves. For example they wanted to be able to select whose posts were visible on their newsfeed and to be able to personalize the colors and content in the app. Having the option to personalize the app made them feel that the app was unique, belonging to them, and tailored specifically to their own preferences, which made the app more relatable and appeared to increase the usability of the app:

> *Cause if it was personalise...you could say that I don't like this kind of tip or I don't like certain things, and then you could kind of have it specific to you...* [P3; Focus group 2]

> *Or you [could] save it for later. 'Cause a lot of these things...could be something you might not need now but in a month's time you might think, oh, let me have a look on that app.* [P7; Focus group 2]

A novel feature of the app was the "social" component, where participants could post messages and interact with other users through the newsfeed and a "teams" page where users collected points for engaging with the app (see Multimedia Appendix 3). The majority of participants found the newsfeed feature a useful way to connect with other users, which made them feel like they were part of a community of like-minded others. Participants compared the newsfeed feature with other social media apps they enjoyed using, like Facebook, and reflected on how the newsfeed elevated using the app from an isolated, solitary activity to something that is shared with and connected to others. Participants also found it a useful tool to be able to compare their experiences with that of others, which enabled participants to normalize their experiences and motivated them to both continue using the app and to continue their drinking goals.

However, not all participants had a positive view on connecting with other users on the newsfeed. As the newsfeed connected users together who did not know each other, some participants reported that connecting with strangers was irrelevant to them. Participants held conflicting views on this issue and discussed how they had different ways of using online social tools, with no single model being suitable for everybody. Some users liked to be very active, while others did not want to be involved at all. In general, participants wanted the flexibility to be able to choose how involved they were with the social features:

> *You've got the danger of weirdoes and all that kind of stuff...* [P7; Focus group 1]

> *I mean I don't care what Steve from Birmingham has got...it's irrelevant for me...I would opt out of other people's comments.* [P3; Focus group 1]

> *I liked hearing other people's struggle and I liked hearing about other people's triumphs. I didn't feel so bad when I kind of, you know, fell off the wagon myself, okay, it's not just me.* [P5; Focus group 1]

While the social feature of the newsfeed was highly praised by the participants, the teams section of the app was one of the most criticized areas. Participants reported that the teams concept was underdeveloped and not engaging to use. In particular, most of the participants mentioned that the objective of joining a team and the benefits of the feature, (which provided points for using the app) were not clear. However, participants generally still thought the feature had merit but that it needed to be improved. One method participants suggested to improve the usability of the Teams section, was to make it more socially interactive, and have a live feed where users could interact specifically with people in their team. However, as mentioned above, some participants were hesitant about interacting with people they did not know, and that interactions with friends would be more meaningful:

> *I think what needs to be really clear is once you've picked a team and once you're in a team what can you do?* [P5; Focus group 3]

> *What do you get for winning? Otherwise it all just seems a bit pointless.* [P3; Focus group 3]

> *I'd just quite like to...go with my six or seven mates who all play football...and then your newsfeed is based around the team that you choose, so you see what your mates are saying.* [P6; Focus group 3]

Overall, the option for social connectivity within the app was a highly praised feature, which participants strongly believed improved the usability by fostering an engaging and interesting user experience that could be shared with others. However, participants also highlighted the need for improvement in the teams section, making the objectives and concepts clearer to understand and also increasing the level of social connectivity.

Discussion

Principal Findings

The aim of this study was to explore the experiences of enablers and barriers to usability for a prototype eSBI app called BRANCH targeting harmful drinking in young adults. The study found that an easy-to-use interface, with minimum required user-input and high levels of automated functioning were the most important usability requirements for participants. It also found that clear, consistent, and visually appealing design was integral to the level of usability. The option for social connectivity was important to participants, as were high levels of personalization. Poor functionality, text heavy content, high user-burden costs in time and effort, and unappealing design were considered major usability barriers. This study showed how focus group interviews can be used to get detailed feedback on the usability of an alcohol app, which can then be used to

XSL•FO
RenderX

improve its effectiveness and ultimately increase its potential for reducing drinking among users.

The findings of this study are consistent with previous research, which found that difficulties in inputting data were among the most frequently criticized functionality issues in existing eSBI alcohol apps [3]. High data entry burden costs, which were considered a usability barrier by participants using BRANCH, have been reported to be a primary reason why people stop using health apps [27]. Participants wanted features that reduced the amount of time and effort required from them. They suggested that reminders, guided walkthroughs, and reduced scrolling were all features that could improve usability. The level of data participants are expected to enter into self-monitoring apps should be carefully considered in future app development [27]. Indeed, frustration with poor performance is one of the most common complaints of app users and results in apps being deleted entirely [28].

A well-conceived visual design was integral to the usability of BRANCH. Participants wanted clear, concise presentation of information which was not text heavy, a finding which is consistent with previous research [7]. Poor design, such as features and buttons not standing out, inhibit use as users are unable to distinguish between features. This has implications not only for usability but also for the potential effectiveness of the intervention. If an eSBI user cannot identify that a particular feature is available, then the user will not be exposed to the targeted alcohol harm reduction intervention. Furthermore, visual design influences the credibility of an app and users are more likely to rate consumer health information on the web as credible if it is presented in an aesthetic style [29].

An issue closely related to usability was engagement. Engagement refers both to how a user interacts with a technology and their emotional response to it [30-32]. For BRANCH, participants often stated how positive experiences of usability made them engage more with the app, making them more likely to keep on using it. For example, participants praised the newsfeed feature as one which enhanced their experience of app usability, as it provided them with meaningful interactions with other users and a sense of community. Participants also praised how personalization made the app feel more tailored to their own needs, providing a positive user experience. This is consistent with the elaboration likelihood model (ELM), which proposes that people are more motivated to engage with and process information more thoroughly if the message is personally relevant and meaningful [33]. The theoretical model of user engagement by Short et al [34] proposes that an individual's characteristics and personal circumstances may influence their user experience of the app. It may be that future applications can enhance usability by targeting features and increasing personalization to target specific user characteristics.

Not improving usability would result in frustrating features and boredom is associated with disengagement with online programs [35]. The teams section in BRANCH, where participants were awarded points for engaging with the app, was criticized for being a major usability barrier because the objective of the feature came across as confusing to participants. While gamification methods (the use of gaming design in nongaming contexts) are popular and effective methods with which to improve engagement [36], it is apparent from the present study that the design of such features needs to be carefully considered in the context of the intervention, otherwise its relevance will be challenged.

Engagement is an ongoing issue for health app development. Issues such as low login rates and limited use of intervention features are consistently reported in literature [37,38]. Findings from the online intervention "Down Your Drink" reported that only 6% of users stayed with the program until the end of the 6 week program [39]. Enhancing the usability of engagement features is crucial to the effectiveness of an eSBI app. The more usable an app is, the more likely an individual is to revisit it and repeatedly use the program intervention features. Indeed, higher engagement through logins and repeated use is associated with better participant health outcomes [38,40,41].

Conclusions

Optimizing usability for eSBI apps is a critical step in the development process. Consumer expectations for digital products are high and if products do not meet their expectations, then they may cease to use them. A recent survey demonstrated that peoples' tolerance for poor performing apps has reduced, with approximately 50% of people reporting that they are less tolerant of problems in apps they use compared to a few years ago [28]. In case of young adults, the age group with the highest use of health apps [27], if an app does not function as they want it to, regardless of its objective, they will delete it. It is not good enough to have only a strong evidence-based core intervention, the whole package of delivery, including design, aesthetics, usability, and functionality needs to be iteratively refined and improved. As there are high demands on apps such as BRANCH that target skilled technology users such as young adults, the development of future eSBI apps that allows for usability testing with representative users may help support the effectiveness of eSBI to reduce harmful drinking. The BRANCH app is currently being evaluated as part of a randomized controlled trial with results expected in early 2018.

Limitations

Significant efforts were made to recruit a sample of young adults, both male and female; however, the majority of participants (90%) were female. This is consistent with a previous study for a Web-based alcohol intervention, which also reported a high sample of females (73%) [42]. The advertisement was designed for both men and women, however more females replied to the advert and requested to bring along more female friends, creating a multiplying effect. While potentially introducing bias, there is research to suggest that women are more motivated to use the Internet for seeking health information [43,44] and are more likely to use eHealth interventions as recommended [45]. Therefore, the sample in the current study may represent the type of individuals more likely to engage with the Web-based BRANCH app. Future studies may wish to explore these differences in more detail, examining how males and females engage with eSBI, informing how interventions can be tailored to gender. Participants were not screened on their intention to cut down their alcohol use.

XSL•FO
RenderX

Consequently, there may have been differences between the participants in the current study and some end-users in terms of how motivated they would be to engage with the app. However, BRANCH was not designed exclusively for those wishing to cut down, but also for those wanting to monitor their use or learn more about the risks of drinking. While this data was not collected, there were 20 participants in the present study who would have likely held a range of reasons for wanting to participate in the study, reflecting the target end-user of the app. Future studies should be further improved by specifically targeting at recruitment stage the types of end users the app is designed for. Focus groups can be subject to response bias, where participants provide answers based on what they think the researchers want to hear. However the findings in the current research reflect those in previous research both from the alcohol field [3,7] and from usability testing research in the computer science disciplines, suggesting that the results have meaning across different populations. Focus groups may also limit the full range of views due to convergence of ideas. Future research may wish to conduct both 1:1 as well as focus group interviews. The participants in the focus group scored high on the AUDIT. Alcohol is a topic seen as sensitive and stigmatizing, therefore participants may not have been comfortable sharing all of their experiences of using the app, as this may reveal details about their level of drinking.

Acknowledgments

The research was funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South London at King's College Hospital NHS Foundation Trust. The views expressed are those of the authors' and not necessarily those of the NHS, the NIHR, or the Department of Health. Colin Drummond is partly funded by the 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.

Conflicts of Interest

None declared.

Multimedia Appendix 1

BRANCH prototype app: drinking diary.

[PDF File (Adobe PDF File), 79KB - mhealth_v5i8e109_app1.pdf]

Multimedia Appendix 2

BRANCH prototype app: goals and information sections.

[PDF File (Adobe PDF File), 62KB - mhealth_v5i8e109_app2.pdf]

Multimedia Appendix 3

BRANCH prototype app: selecting a motivator, the Newsfeed, and Teams sections.

[PDF File (Adobe PDF File), 85KB - mhealth_v5i8e109_app3.pdf]

Multimedia Appendix 4

Automated messaging examples.

[PDF File (Adobe PDF File), 18KB - mhealth_v5i8e109_app4.pdf]

References

1. 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;1(1):e9 [FREE Full text] [doi: 10.2196/mhealth.2709] [Medline: 25100681]
2. 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;17(5):e118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
3. Milward J, Khadjesari Z, Fincham-Campbell S, Deluca P, Watson R, Drummond C. User preferences for content, features, and style for an app to reduce harmful drinking in young adults: analysis of user feedback in app stores and focus group interviews. JMIR Mhealth Uhealth 2016 May 24;4(2):e47 [FREE Full text] [doi: 10.2196/mhealth.5242] [Medline: 27220371]

XSL•FO
RenderX

4. Donoghue K, Patton R, Phillips T, Deluca P, Drummond C. The effectiveness of electronic screening and brief intervention for reducing levels of alcohol consumption: a systematic review and meta-analysis. J Med Internet Res 2014;16(6):e142 [FREE Full text] [doi: 10.2196/jmir.3193] [Medline: 24892426]
5. Dedert EA, McDuffie JR, Stein R, McNiel JM, Kosinski AS, Freiermuth CE, et al. Electronic interventions for alcohol misuse and alcohol use disorders: a systematic review. Ann Intern Med 2015 Aug 4;163(3):205-214. [doi: 10.7326/M15-0285] [Medline: 26237752]
6. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May;22(5):413-432. [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
7. Crane D, Garnett C, Brown J, West R, Michie S. Factors influencing usability of a smartphone app to reduce excessive alcohol consumption: think aloud and interview studies. Front Public Health 2017 Apr 03;5:39. [doi: 10.3389/fpubh.2017.00039]
8. Cohn AM, Hunter-Reel D, Hagman BT, Mitchell J. Promoting behavior change from alcohol use through mobile technology: the future of ecological momentary assessment. Alcohol Clin Exp Res 2011 Dec;35(12):2209-2215 [FREE Full text] [doi: 10.1111/j.1530-0277.2011.01571.x] [Medline: 21689119]
9. McCambridge J, Saitz R. Rethinking brief interventions for alcohol in general practice. BMJ 2017 Jan 20;22(5):j116-j112. [doi: 10.1136/bmj.j116]
10. Chiu TM, Eysenbach G. Stages of use: consideration, initiation, utilization, and outcomes of an internet-mediated intervention. BMC Med Inform Decis Mak 2010;10:73 [FREE Full text] [doi: 10.1186/1472-6947-10-73] [Medline: 21092275]
11. 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]
12. Bangor A, Kortum P, Miller J. An Empirical Evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
13. Stoyanov S, Hides L, Kavanagh D, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
14. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
15. Davies EL. Feasibility of the prototype willingness model as the basis for school-delivered alcohol misuse prevention: a qualitative think-aloud study to explore acceptability of 'The Alcohol Smart Quiz' with adolescents and teachers. J Health Psychol 2016 May 31:- Epub ahead of print. [doi: 10.1177/1359105316648481] [Medline: 27247092]
16. Dulin PL, Gonzalez VM, Campbell K. Results of a pilot test of a self-administered smartphone-based treatment system for alcohol use disorders: usability and early outcomes. Subst Abus 2014;35(2):168-175 [FREE Full text] [doi: 10.1080/08897077.2013.821437] [Medline: 24821354]
17. Pew Research Center. 2017. Mobile fact sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2017-04-06] [WebCite Cache ID 6pWZtFObW]
18. Prensky M. Digital natives, digital immigrants, part 1. On the Horizon 2001 Sep;9(5):1-6. [doi: 10.1108/10748120110424816]
19. Lyles C, Sarkar U, Osborn C. Getting a technology-based diabetes intervention ready for prime time: a review of usability testing studies. Curr Diab Rep 2014 Oct;14(10):534 [FREE Full text] [doi: 10.1007/s11892-014-0534-9] [Medline: 25173689]
20. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, 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]
21. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013;15(3):e51 [FREE Full text] [doi: 10.2196/jmir.2350] [Medline: 23475457]
22. 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;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
23. Howard G, Nathan P. Alcohol use and misuse by young adults. Notre Dame, Indiana: University of Notre Dame Press; 1994.
24. Deluca P, Coulton S, Alam MF, Cohen D, Donoghue K, Gilvarry E, et al. Linked randomised controlled trials of face-to-face and electronic brief intervention methods to prevent alcohol related harm in young people aged 14-17 years presenting to Emergency Departments (SIPS junior). BMC Public Health 2015 Apr 10;15:345 [FREE Full text] [doi: 10.1186/s12889-015-1679-4] [Medline: 25886178]
25. Nilsen P, Baird J, Mello MJ, Nirenberg T, Woolard R, Bendtsen P, et al. A systematic review of emergency care brief alcohol interventions for injury patients. J Subst Abuse Treat 2008 Sep;35(2):184-201. [doi: 10.1016/j.jsat.2007.09.008] [Medline: 18083321]
26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

XSL•FO
RenderX

27. 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]
28. Brauer C. Appdynamics. 2014. The app attention span URL: https://www.appdynamics.com/media/uploaded-files/1425406960/app-attention-span-research-report-1.pdf [accessed 2017-04-06] [WebCite Cache ID 6pWaevteI]
29. Robins D, Holmes J, Stansbury M. Consumer health information on the web: the relationship of visual design and perceptions of credibility. J Am Soc Inf Sci 2009 Oct 02;61(1):13-29. [doi: 10.1002/asi.21224]
30. Crutzen R, de NJ, Brouwer W, Oenema A, Brug J, de Vries NK. Strategies to facilitate exposure to internet-delivered health behavior change interventions aimed at adolescents or young adults: a systematic review. Health Educ Behav 2011 Feb;38(1):49-62. [doi: 10.1177/1090198110372878] [Medline: 21189422]
31. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
32. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2 Suppl):TC9-T20. [doi: 10.1097/NCN.0b013e3182155274] [Medline: 21372645]
33. DiClemente RJ, editor. The elaboration likelihood model of persuasion: health promotions that yeild sustained behavioral change. In: Emerging Theories in Health Promotion Practice and Research. San Francisco, CA: Jossey-Bass; 2009.
34. Short C, Rebar A, Plotnikoff R, Vandelanotte C. EHPS. 2015. Designing engaging online behaviour change interventions: a proposed model of user engagement URL: http://www.ehps.net/ehp/index.php/contents/article/viewFile/763/pdf_39 [WebCite Cache ID 6dlspECdB]
35. O'Brien HL, Toms EG. What is user engagement? a conceptual framework for defining user engagement with technology. J Am Soc Inf Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
36. Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? a systematic review. PLoS One 2017;12(3):e0173403 [FREE Full text] [doi: 10.1371/journal.pone.0173403] [Medline: 28362821]
37. 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 [FREE Full text] [Medline: 14691071]
38. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
39. Postel MG, de Haan HA, ter Huurne ED, Becker ES, de Jong CA. Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: randomized controlled trial. J Med Internet Res 2010 Jan;12(4):e68 [FREE Full text] [doi: 10.2196/jmir.1642] [Medline: 21163776]
40. Hutton HE, Wilson LM, Apelberg BJ, Tang EA, Odelola O, Bass EB, et al. A systematic review of randomized controlled trials: web-based interventions for smoking cessation among adolescents, college students, and adults. Nicotine Tob Res 2011 Apr;13(4):227-238. [doi: 10.1093/ntr/ntq252] [Medline: 21350042]
41. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
42. Schwartz K, Roe T, Northrup J, Meza J, Seifeldin R, Neale A. Family medicine patients' use of the internet for health information: a MetroNet study. J Am Board Fam Med 2006;19(1):39-45 [FREE Full text] [Medline: 16492004]
43. Fox S. Pew Research Center. Washington DC: Pew Internet & American Life Project; 2011. The social life of health information URL: http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2017-07-12] [WebCite Cache ID 6rthsJyWb]
44. Bidmon S, Terlutter R. Gender differences in searching for health information on the internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. J Med Internet Res 2015 Jun 22;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
45. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]

Abbreviations

AUDIT: Alcohol use disorder identification test
ELM: Elaboration likelihood model
eSBI: Electronic screening and brief intervention
FRAMES: Feedback, Responsibility, Advice, Menu of options, Empathy, Self-efficacy
UCD: User-centered design

XSL•FO
RenderX

Edited by G Eysenbach; submitted 11.04.17; peer-reviewed by P Dulin, M Lim, E Davies; comments to author 11.05.17; revised version received 30.05.17; accepted 02.06.17; published 08.08.17.

Please cite as:
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;5(8):e109
URL: http://mhealth.jmir.org/2017/8/e109/
doi:10.2196/mhealth.7836
PMID:28790022

©Joanna Milward, Paolo Deluca, Colin Drummond, Rod Watson, Jacklyn Dunne, Andreas Kimergård. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 08.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Developing and Evaluating JIApp: Acceptability and Usability of a Smartphone App System to Improve Self-Management in Young People With Juvenile Idiopathic Arthritis

Ran A Cai[1], BSc, PhD; Dominik Beste[2], MSc; Hema Chaplin[1], BSc; Socrates Varakliotis[2], PhD; Linda Suffield[1], BSc; Francesca Josephs[1], MSc; Debajit Sen[1,3], MD; Lucy R Wedderburn[1,4], MD, PhD; Yiannakis Ioannou[1,3], MD, PhD; Stephen Hailes[2], PhD; Despina Eleftheriou[1,4], MD, PhD

[1]Arthritis Research UK Centre for Adolescent Rheumatology, University College London, London, United Kingdom

[2]Department of Computer Science, University College London, London, United Kingdom

[3]University College London Hospitals NHS Foundation Trust, Adolescent Rheumatology, London, United Kingdom

[4]UCL Great Ormond Street Institute of Child Health, Infection, Immunity, Inflammation, and Physiological Medicine, London, United Kingdom

Corresponding Author:
Ran A Cai, BSc, PhD
Arthritis Research UK Centre for Adolescent Rheumatology
University College London
UCL Great Ormond Street Institute of Child Health
30 Guilford St
London, WC1N 1EH
United Kingdom
Phone: 44 0203 108 2420
Fax: 44 0203 108 2420
Email: a.cai@ucl.ac.uk

Abstract

Background: Flare-ups in juvenile idiopathic arthritis (JIA) are characterized by joint pain and swelling and often accompanied with fatigue, negative emotions, and reduced participation in activities. To minimize the impact of JIA on the physical and psychosocial development and well-being of young people (YP), it is essential to regularly monitor disease activity and side effects, as well as to support self-management such as adherence to treatment plans and engagement in general health-promoting behaviors. Smartphone technology has the potential to engage YP with their health care through convenient self-monitoring and easy access to information. In addition, having a more accurate summary of self-reported fluctuations in symptoms, behaviors, and psychosocial problems can help both YP and health care professionals (HCPs) better understand the patient's condition, identify barriers to self-management, and assess treatment effectiveness and additional health care needs. No comprehensive smartphone app has yet been developed in collaboration with YP with JIA, their parents, and HCPs involved in their care.

Objectives: The objective of this study was to design, develop, and evaluate the acceptability and usability of JIApp, a self-management smartphone app system for YP with JIA and HCPs.

Methods: We used a qualitative, user-centered design approach involving YP, parents, and HCPs from the rheumatology team. The study was conducted in three phases: (1) phase I focused on developing consensus on the features, content, and design of the app; (2) phase II was used for further refining and evaluating the app prototype; and (3) phase III focused on usability testing of the app. The interview transcripts were analyzed using qualitative content analysis.

Results: A total of 29 YP (aged 10-23, median age 17) with JIA, 7 parents, and 21 HCPs were interviewed. Major themes identified as the ones that helped inform app development in phase I were: (1) remote monitoring of symptoms, well-being, and activities; (2) treatment adherence; and (3) education and support. During phase II, three more themes emerged that informed further refinement of the app prototype. These included (4) adapting a reward system to motivate end users for using the app; (5) design of the app interface; and (6) clinical practice integration. The usability testing during phase III demonstrated high rates of overall satisfaction and further affirmed the content validity of the app.

Conclusions: We present the development and evaluation of a smartphone app to encourage self-management and engagement with health care for YP with JIA. The app was found to have high levels of acceptability and usability among YP and HCPs and

XSL•FO
RenderX

has the potential to improve health care and outcomes for this age group. Future feasibility testing in a prospective study will firmly establish the reliability, efficacy, and cost-effectiveness of such an app intervention for patients with arthritis.

(JMIR Mhealth Uhealth 2017;5(8):e121) doi:10.2196/mhealth.7229

KEYWORDS

juvenile idiopathic arthritis; self-management; adolescent; young adult; mobile applications; qualitative research; smartphone

Introduction

Juvenile idiopathic arthritis (JIA) is one of the most common chronic diseases with an early onset in childhood or adolescence. It causes articular inflammation, which in turn leads to pain, swelling, and stiffness of the joints [1]. The progression of JIA into adulthood is common in the majority of teenagers with JIA [2] and, if not treated effectively during the childhood and adolescent years, it can lead to disability and complications related to educational, psychosocial, and physical development in later life [3]. There are now effective therapies for JIA that have the potential to transform the quality of life for young people (aged 10-24 years; YP) [4] with the disease [1]. If optimal treatment is instigated, it can improve JIA prognosis and negate the need for further health care input such as joint replacement surgery, which was often required in YP before the advent of new therapies [5-8].

However, treating JIA requires patients to take regular medications and attend frequent hospital appointments for monitoring disease progression and side effects. In addition, successful treatment often relies on a holistic approach such as engaging YP in appropriate physical exercises and maintaining psychological well-being [9-11]. Health outcomes are thus highly dependent on self-management behaviors to facilitate adherence of YP to treatment protocols and engagement in behaviors that promote physical and emotional well-being [12-15]. As YP mature and transition from pediatric to adult services, they are expected to become more independent and responsible for their own health care. Unfortunately, health service provision often fails to offer the type of support and skills that YP need to manage their JIA, especially during times of significant change [16-18], which may explain the poor adherence to treatment and disengagement with services in YP [19-22]. It is therefore essential for health care professionals (HCPs) to adopt effective and innovative ways to promote self-management promptly and effectively [14,23-26].

According to a model known as the COM-B system [27] that emerged from a systematic review of behavior change theories, behavior (B) may depend on an interaction between three important components: capability (C), opportunity (O), and motivation (M). When one component is not fulfilled, it could impose barriers to self-management; this has been used to explain issues with adherence [28] and applied in several interventions [29,30]. The COM-B model can explain why providing educational information aimed at increasing skills and understanding alone, which was the main focus of previous self-management interventions, has not been sufficient in engaging YP and that additional intervention functions may be necessary [17,31]. These may require sending reminders, monitoring problems and achievements, or offering incentives in ways that are acceptable and appropriate for YP.

An ideal resource-efficient way of improving self-management is by incorporating multiple intervention functions to address all three sources of behaviors through smartphone technology. In general, YP are already familiar and comfortable using smartphone apps in their everyday lives. It is estimated that 75% of adolescents over the age of 13 years in the United Kingdom use smartphones on a daily basis, and smartphone ownership is highest among YP [32]. Capitalizing on the popularity of smartphone-based entertainment in adolescents, health apps may represent an important means of increasing self-management through remote monitoring and access to information [33-36]. For example, health management behaviors can be integrated with daily activities, using technologies that can track information “on the go.” Apps used to monitor symptoms remotely also have considerable potential in helping HCPs deliver safe and timely health care, as it provides more accurate and frequent reports of health status compared with using paper diaries [37-39]. Reliable and secure electronic data collection of symptoms and emotions not only benefits clinical assessments, but it is also an established method for validly collecting daily data in research as well [40-43] and can help advance knowledge on factors influencing disease progression.

However, recent systematic reviews of the literature exploring the effectiveness of mobile apps designed to support the management of chronic physical conditions by YP [44-46] found a limited number of apps in the scientific database. This is in stark contrast to the thousands of commercial apps available that have not been developed in close partnership with end users. Only one app [47-49] has been developed for JIA patients, but it did not involve parents and HCPs in the development process and its function is limited to monitoring symptoms. The app is also targeted for patients in pediatric clinics (<18 years), whose needs and preferences may differ from YP who have already transitioned to adult clinics (18-24 years). Moreover, previous reviews did not identify a single app for YP that designed a clinician interface jointly with the patient app portal, and none had a sound theoretical rationale.

Therefore, this study aimed to address issues with self-management and engagement with health care in YP by developing and evaluating a comprehensive smartphone-based app system for, and in collaboration with, (1) YP with JIA and (2) HCPs involved in their care. We followed a theory- and user-driven approach [50-52], where user feedback helped understand optimal ways of delivering intervention functions that have been identified in the COM-B model to overcome barriers to self-management in YP. This approach is in line with the United Kingdom Medical Research Council (MRC) guidelines, which highlight the importance of identifying

XSL•FO
RenderX

appropriate theory to inform the development of a complex intervention [50-52].

Methods

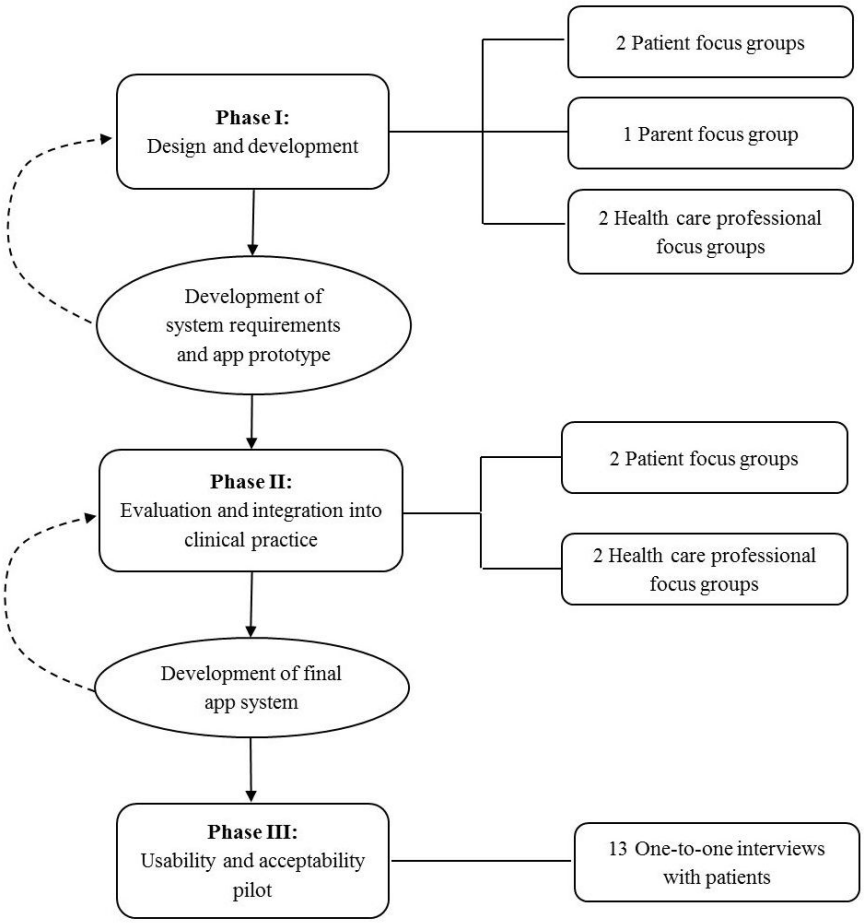
Participants

Patients were eligible for recruitment to the study when they fulfilled the following criteria: (1) a diagnosis of JIA, defined as arthritis of unknown etiology lasting more than 6 weeks and of onset at less than 16 years [53]; (2) were aged 10 to 24 years; (3) were able to read and speak English; and (4) were seen at the Great Ormond Street Hospital (GOSH) or University College London Hospital (UCLH) rheumatology clinics. Patients were excluded if they had severe cognitive impairments or major comorbid medical or psychiatric illnesses that would preclude their ability to participate in focus group discussions (FGDs). Parents of patients were eligible if they could read and understand English. Patient demographics and disease-related data collected were age at time of study, age at diagnosis, classification of JIA category, and current medication. The HCPs working in the multidisciplinary team at the two recruiting centers were eligible to participate if they had worked in pediatric/adolescent rheumatology for at least 6 months according to self-report. Participants were able to take part in all three phases of the study. Table 1 summarizes which patient participated multiple times.

Process

The overall project adopted a qualitative, user-centered approach as per the standard published guidance [54-56]. The study was conducted in three phases (see Figure 1). Phase I focused on developing consensus on the features and content of the app. This involved understanding barriers and enablers to self-management in YP, and, as suggested by the MRC guidance [50-52], identifying intervention functions from the COM-B model that should be incorporated in the app. Phase II was used to further refine and evaluate the functionalities and improve the design of the app prototype and the Web application for HCPs. Phase III focused on usability testing of the final app interface for YP. Purposive sampling was used to encompass variations in opinions because of demographic characteristics (age, gender, and ethnicity), disease duration, and JIA subtype. Both YP and parents were approached by a member of the research team in clinics before the rheumatology appointment. The families who agreed to participate gave informed consent or age-appropriate assent with parental consent for those aged under 16 years. Thereafter, YP were offered a group based on their age for phases I and II. The study was approved by the Local Ethics Committee (NRES Committee London – Queen Square, 15/LO/1288).

Figure 1. Study phases and iterative development cycle.



XSL•FO
RenderX

Table 1. Patient baseline characteristics, disease duration, and treatment received.

Phase	Patient ID	Age, in years	Gender	JIA[a] subclassification	Disease duration, in years	Current medication
I	1	21	F	Persistent oligoarticular JIA	17	Nil. Previously treated with intra-articular steroid injections
	2	18	M	ERA[b]	3	Methotrexate (15 mg/m^2 sc[c] weekly)
	3	17	M	Persistent oligoarticular JIA	4	Nil. Previously treated with intra-articular steroid injections
	4	18	M	ERA	7	Etanercept (0.8 mg/kg sc weekly) and methotrexate (15 mg/m^2 sc weekly)
	5	18	F	Polyarticular JIA	15	Methotrexate (15 mg/m^2 sc weekly) and adalimumab (40 mg sc fortnightly)
	6	10	M	Extended oligoarticular JIA	4	Methotrexate (15 mg/m^2 sc weekly)
	7	14	F	Systemic JIA	3	Methotrexate (15 mg/m^2 sc weekly)
	8	12	M	Extended oligoarticular JIA	7	Nil. Previously treated with intra-articular steroid injections
	9	11	M	Systemic JIA	2	Tocilizumab (8 mg/kg fortnightly) and methotrexate (15 mg/m^2 sc weekly)
	10	12	F	Psoriatic arthritis	0.4	Methotrexate (15 mg/m^2 sc weekly)
II	2	18	M	ERA	3	Methotrexate (20 mg/m^2 sc weekly) and humira (40 mg sc weekly)
	6	10	M	Extended oligoarticular JIA	5	Methotrexate (15 mg/m^2 sc weekly)
	7	14	F	Systemic JIA	3	Methotrexate (15 mg/m^2 sc weekly)
	8	12	M	Extended oligoarticular JIA	7	Nil
	9	12	M	Systemic JIA	3	Tocilizumab (8 mg/kg fortnightly) and methotrexate (15 mg/m^2 sc weekly)
	10	13	F	Psoriatic arthritis	0.8	Methotrexate (15 mg/m^2 sc weekly)
	11	23	F	Bilateral inflammatory hip arthritis	11	Methotrexate (10 mg/m^2 sc weekly)
	12	17	F	Systemic JIA	2	Hydroxychloroquine (400 mg orally daily)
	13	20	F	Persistent oligoarticular JIA	5	Methotrexate (20 mg orally weekly)
	14	16	F	Extended oligoarticular JIA	14	Methotrexate (15 mg/m^2 sc weekly)
	15	17	F	Polyarticular JIA	9	Etanercept (25 mg sc weekly)
	16	14	F	Psoriatic arthritis	12	Adalimumab (40 mg sc fortnightly)
	17	12	F	Persistent oligoarticular JIA	8	Methotrexate (15 mg/m^2 sc weekly)

XSL•FO
RenderX

Phase	Patient ID	Age, in years	Gender	JIA[a] subclassification	Disease duration, in years	Current medication
III	4	18	M	ERA	8	Methotrexate (15 mg/m^2 sc weekly)
	18	16	F	Oligoarticular JIA	12	Nil
	19	12	M	Polyarticular JIA	4	Methotrexate (25 mg orally weekly)
	20	14	F	Psoriatic arthritis	8	Methotrexate (12.5 mg/m^2 sc weekly)
	21	15	M	Extended oligoarticular JIA	8	Naproxen (500 mg twice a day)
	22	17	F	Extended oligoarticular JIA	15	Methotrexate (15 mg/m^2 sc weekly) and folic acid (5 mg orally weekly)
	23	13	F	Oligoarticular JIA	10	Methotrexate (15 mg/m^2 sc weekly) and humira (40 mg fortnightly)
	24	15	F	Systemic JIA	3	Methotrexate (27.5 mg/m^2 sc weekly), anakinra (150 mg sc daily), and folic acid (5 mg orally weekly)
	25	15	F	Oligoarticular JIA	4	Nil
	26	23	M	Extended oligoarticular JIA	21	Keppra (500 mg twice a day)
	27	15	F	ERA	3	Methotrexate (15 mg/m^2 sc weekly) and folic acid (5 mg orally weekly)
	28	16	M	Polyarticular JIA	1	Methotrexate (15 mg orally weekly) and folic acid (5 mg orally weekly)
	29	18	F	Polyarticular JIA	2	Enbrel (50 mg sc weekly) and methotrexate (7.5 mg orally weekly)

[a]JIA: juvenile idiopathic arthritis.

[b]ERA: enthesitis-related arthritis.

[c]sc: subcutaneously.

Phase I: Design and Development

For phase I, we used a nominal group technique to develop consensus on the app content and features. This included a number of separate FGDs for HCPs, younger patients (aged 10-15 years) and their parents, and older patients (aged 16-24 years). Following an introductory presentation on the overall goals of the meeting, YP and parents were asked what features and information will help them or their child better follow their treatment plans and cope with arthritis-related symptoms and consequences. The HCPs were asked what data can be collected remotely to help with understanding and assessing patient's condition and treatment effectiveness. All participants were also questioned regarding the design and aesthetics of the user interface. Lists of questions posed to each of the groups are summarized in Multimedia Appendix 1. Consensus was considered as having been achieved when 75% of the participants endorsed a given answer. Answers that did not reach 75% endorsement were discarded or reformulated through discussion.

Participants' answers to the questions posed during phase I FGDs were used to generate a catalog of initial system requirements. An app prototype was then developed in collaboration with the Computer Science Department at UCLH. The agreed set of features were iteratively built into the app codebase by a developer from the Computer Science Department. When clarification of the data generated was required, participants were queried by email and phone to provide further feedback (see Figure 1). The prototype was developed using open frameworks for cross-platform compatibility and runs on both the Google Android (version 4.4.1 and above) and the Apple iOS (version 8 and above) operating systems.

Phase II: Evaluation and Integration Into Clinical Practice

The FGDs were again conducted with HCPs and YP with JIA during phase II to vet and refine the requirements of the phase I app prototype. Participants were able to test and navigate through the app's various features using a smartphone to improve their understanding of the app and focus on their recommendations. Screenshots of important pages in the app were also shown to examine whether YP and HCPs liked or disliked the design and the question format. They were asked to provide further suggestions for improvement, and technical problems with the app were also recorded.

Both HCPs and YP were also asked whether the information collected would be useful during consultations, and how clinically relevant information reported by patients should be imported to, summarized, and displayed in a Web application for HCPs. Also, what data recorded by patients and information on app usage should be stored for research were discussed. Semistructured interview questions that were developed based on the Usefulness, Satisfaction, and Ease of Use questionnaire [57] were used to guide the FGDs (see Multimedia Appendix 1). The design of the app was modified and new screenshots

XSL•FO
RenderX

were generated and evaluated until no further changes were suggested. All FGDs were conducted with each participant group until point of data redundancy or point of no proposed new data [58] and lasted between 60 and 90 min. System analysis and final design were done in joint effort with the developer from the Computer Science Department.

Phase III: Usability and Acceptability Pilot

Following the development of the final app, the usability and acceptability of the user interface, as well as the registration and data extraction process were tested with YP during phase III. The interviews with YP took place on a one-to-one basis and YP were shown the app on a smartphone. The researcher explained that the purpose of the study was to pilot ease of use and evaluate its features and encouraged YP to input information in all sections of the app at least once. After navigating the various features of the app and using it for 10 to 15 min, YP answered a usability and acceptability questionnaire [57] where they rated on 5-point scales their general impressions of the app, its user-friendliness, clarity of information provided, and whether or not they found the app useful and would recommend it to other YP (see Multimedia Appendix 2). In addition, YP answered qualitative questions regarding the specific features of the app, and how it can help them improve self-management (see Multimedia Appendix 1). Each interview lasted between 25 and 45 min. Discussions were audio-recorded for all phases of the project, and field notes were made for phase III. At all FGDs, RAC and HC were present as interview leader and observer, respectively.

Data Analysis

All audio recordings were transcribed and qualitative content analysis [59,60] was applied. Qualitative data from transcripts and field notes were carefully reviewed to develop a coding system that reflected perceptions and suggestions for the smartphone-based JIA app. Data were then assigned with codes based on their content. Codes with similar content were grouped into meaningful categories. RAC and HC read and analyzed the transcripts separately and then compared the results. Differences from the separate analyses were discussed until consensus was reached. The coding system and analytic process were further discussed with DE, an experienced pediatric rheumatologist. The final analysis organized the categories together into overarching themes that were developed and refined by discussions between RAC, HC, and DE. Discussions were also held between the researchers on how each category mapped onto the COM-B model. In addition, content analysis was performed according to the age group of YP (10-15 years and 16-24 years) to identify any age-specific differences in opinion. All qualitative data were analyzed using NVivo software (NVivo version 10, QSR International, 2012) [61].

Results

Participant Characteristics

For phase I, a total of 10 patients were recruited and were separated into two groups: group A included 5 patients (median age 12 years, range 10-14 years, 2 females) and group B included 5 patients (median age 18 years, range 17-21 years, 2 females). A separate FGD was conducted with 7 parents of patients from group A. Two patient FGDs were again conducted for phase II: group A included 7 patients (median age 12 years, range 10-14 years; 4 females) and group B included 6 patients (median age 18 years, range 16-23 years, 5 females). Phase III interviewed 13 YP. Demographics and disease characteristics of YP included in the study are shown in Table 1. The participants had a variety of JIA categories with a median time since diagnosis of 4 years (range 5 months-17 years) for patients in phase I, 5 years (range 10 months-14 years) for patients in phase II, and 8 years (range 1-21 years) for patients in phase III. The participants were also diverse in ethnicity (62% white, 24% Asian, 10% black, 3% Hispanic).

Two FGDs with HCPs were also conducted for phases I and II. For phase I, a total of 19 HCPs with a median of 7 years (range 1-20 years) of experience in pediatric/adolescent rheumatology participated. These HCPs included 7 medical consultants, 4 clinical nurse specialists, 4 rheumatology trainees, 2 research nurses, 1 physiotherapist, and 1 clinical psychologist. For phase II, a total of 15 HCPs with a median of 8 years (range 1 to 20 years) of experience participated. The group included 6 medical consultants, 3 rheumatology trainees, 2 clinical nurse specialists, 2 research nurses, and 2 research fellows.

Phase I: Design and Development

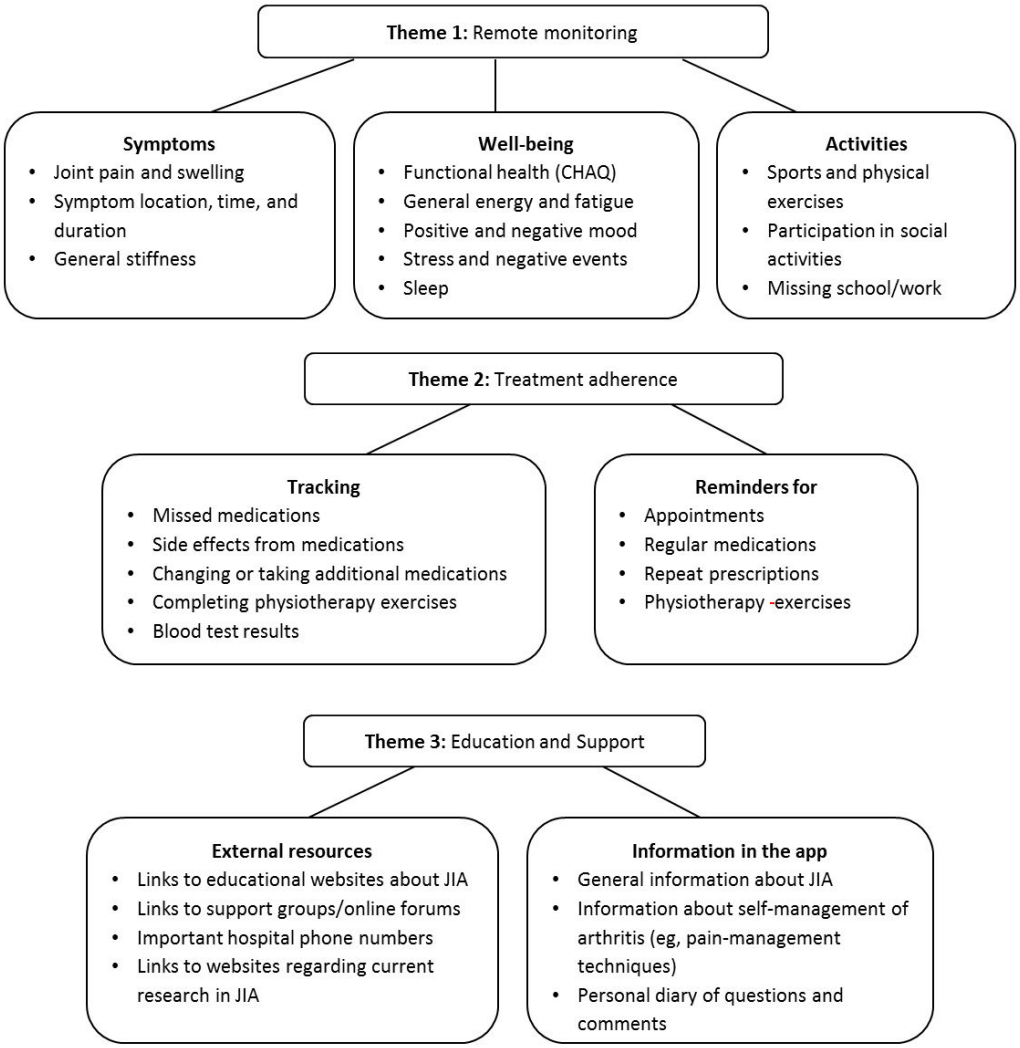
It was apparent that YP with JIA wanted a multifaceted way to help them manage their condition. Transcript analysis of data collected during phase I revealed three distinct themes that guided the development of this app (see Figure 2).

Theme 1: Self-Monitoring

The first theme was self-monitoring, as all participants thought that tracking certain information will (1) help them understand what influences their JIA symptoms and (2) provide better data for HCPs to use during assessments. This theme included six initial categories for monitoring; however, only those for which there was >75% agreement on its importance among YP, parents, and HCPs were included in the app. These were monitoring (1) symptoms, (2) general well-being, (3) activities, and (4) sleep. Categories that did not achieve this concordance were related to tracking (5) weather and (6) nutrition.

XSL•FO
RenderX

Figure 2. Results of content analysis of transcripts from focus groups for phase I.



All participants identified tracking of joint symptomatology, including stiffness, swelling, and pain as very important, particularly as this varies in intensity, location, and duration over time. They felt that using the app would reduce recall bias and enable both patients and clinicians to monitor symptoms in real time:

> *I want to note the dates of when I have symptoms before I forget them, so that I can show the doctors, and help them better understand my illness and what is really going on.* [Patient 3]

One patient stated:

> *Daily monitoring of pain is better than giving a monthly average at the clinic.* [Patient 2]

Participants indicated that the app should include continuous visual analog scales (VAS) to indicate the intensity of JIA-related symptoms by moving a bar across a horizontal line in preference to categorical selections. One adolescent stated:

> *Instead of choosing just one number, I want to be able to choose a range between the numbers, like between 6 to 7.* [Patient 3]

As requested by YP, each scale was accompanied with a cartoon face that changed facial expressions from extremely positive at the left to extremely negative at the right. Psychometric properties of the VAS have been widely explored with children as young as 8 years of age and have confirmed its test-retest reliability, construct validity, and criterion-related validity [62,63].

Additionally, YP wanted to use a body map to indicate how painful and swollen a specific joint is. Body maps can facilitate children's reports of location of pain [64,65], and electronic versions have been used successfully with children as young as 8 years old [40]. The app will include a simple body map (see Figure 3) with circles at major joint areas in the body, mirroring what is used in routine clinical assessments [66]. YP can click on a joint that is painful or stiff and use a VAS to indicate the intensity of their pain and stiffness, and the joint's color will then change accordingly. This is important to include, as YP often experience different pain intensities from different joints, and their stiffness intensity also vary by location and time of day. Both YP and HCPs thought that recording this will help them better explain their symptoms and assess whether treatment is required (eg, physiotherapy exercises targeting specific joints). Lastly, YP requested an open text area where they can input

XSL•FO
RenderX

additional details regarding their symptoms and add comments about the type of pain, with one of the patients stating the following reason:

> *It's about how you feel and how you describe it instead of just having a number.* [Patient 4]

Being able to monitor general well-being such as mood, fatigue, and functional health were also identified as important. For example, YP indicated:

> *Tracking stressful events and mood and how that impacts your symptoms is most important.* [Patient 4]

Three positive (excited, energetic, happy) and three negative emotions (anxiety/fear, sad, angry) were chosen together with YP and HCPs, which were adapted from the Positive and Negative Affect Schedule for Children [67]. These were assessed using the VAS and the horizontal line ranged from "not at all" on the left to "extremely" on the right. A stress scale was also included, but a verbal rating scale for stress (VRSS) was preferred by YP instead of a VAS. The VRSS was found to be more reliable than the VAS and can be easily understood and completed by children as young as 7 years old [68]. It measures stress on a scale from zero to five with responses describing an increase in stress levels. In addition, YP wanted an electronic version of a JIA-related childhood health-assessment questionnaire (CHAQ) [69], which is a widely used functional health status measure in YP with JIA and is administered during routine clinic consultations. They recognized that completing it once every 3 to 6 months during clinic consultations may provide unreliable responses, as it does not reflect variations over time:

> *I'd like to track more precisely because the answers to those questionnaires can change every day, or even every hour.* [Patient 10]

Digital CHAQs have been developed and tested in previous studies [63] and were found to yield similar outcomes when compared with the paper form of CHAQ.

Recording activities such as participation in school and physical exercise and sports was another important area for YP. One YP said,

> *Exercise usually makes me feel better afterwards and doing sports takes the pain away.* [Patient 8]

Another stated,

> *I want to be able to write down what type of exercise I did, such as swimming.* [Patient 1]

YP suggested that the app should provide visual feedback of their data so that they can see how different factors impacted their condition, and what techniques were helpful in managing their symptoms.

Theme 2: Treatment Adherence

Participants were interested in using the app to improve treatment adherence by setting reminders for when to take regular medications, request repeat prescriptions, complete physiotherapy exercises, and attend hospital appointments for important assessments and tests (eg, regular blood tests). One of the patients said,

> *Sometimes when life gets busy I do forget, so having a reminder would be really useful.* [Patient 4]

However, some participants felt that these reminders should be optional to avoid any disengagement with the app because of annoyance from compulsory reminders. Furthermore, YP were interested in recording when medication doses were missed, reasons for missing doses, side effects, medication change, and whether they took any additional medicines such as analgesics or antibiotics and anti-emetics (see Figure 4). Regarding physiotherapy exercises, YP could select which ones they were assigned or would like to perform from a list of possible exercises recommended by physiotherapists at GOSH and could input the number of sets and repetitions. Additionally, both patients and HCPs wanted a section where patients could input their blood test results, with an indication of whether these results were normal and what the tests measured:

> *I think it's interesting to have your blood results on the app, together with the doctor's comments such as there's nothing to worry about or that something might be a problem.* [Patient 2]

Theme 3: Education and Support

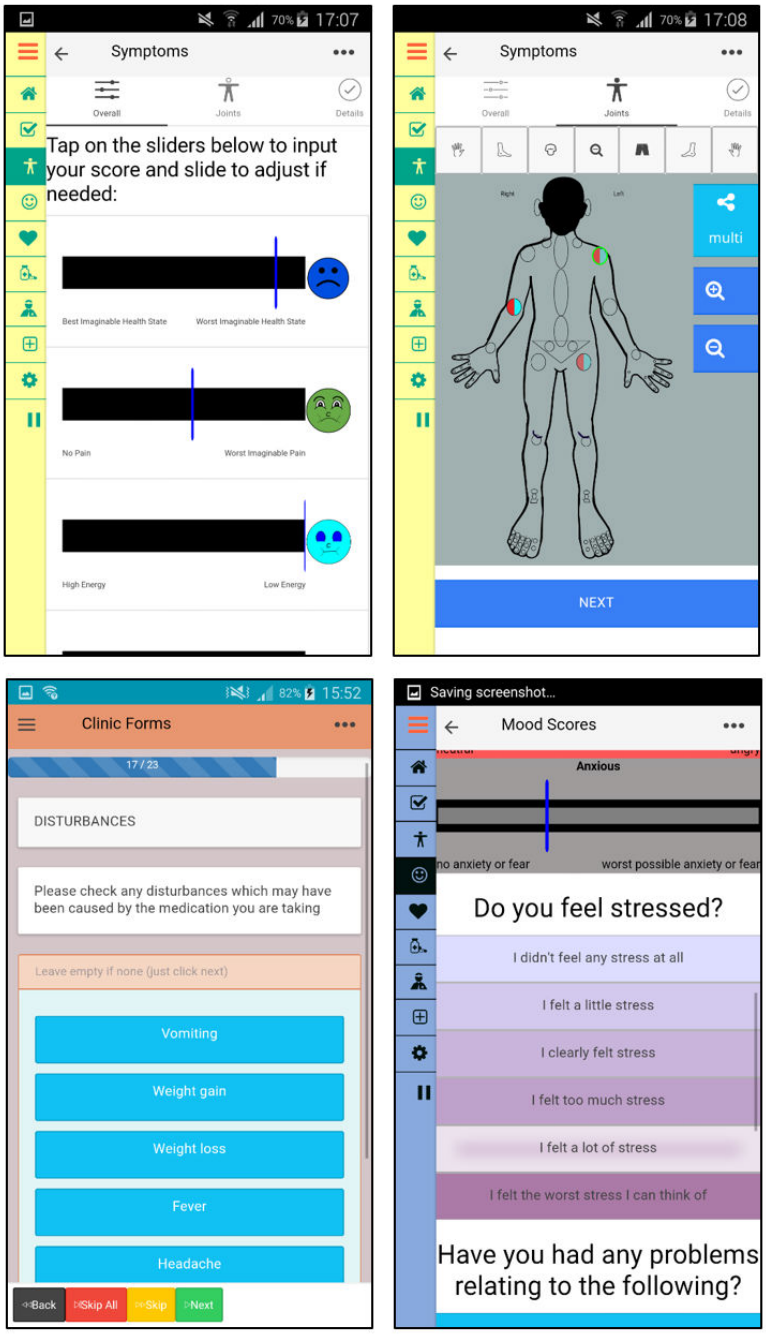
Besides wanting to learn more about their condition, YP wanted to have a discussion forum in the app with other patients. Also, YP from all age groups highlighted the need to have easy access via the app to condition- and treatment-related information, news about research, and social support:

> *The app should have information on arthritis…also news about research, involvement, and being able to talk to other people.* [Patient 2]

Additionally, YP said that learning different ways of managing their stress and pain from peers, and knowing that they are not alone, will make them feel more empowered and confident. Moreover, YP mentioned that their arthritis symptoms tend to worsen during times of stress and exams and would like information on how to better manage stressful situations and pain. Therefore, the app will include general information regarding JIA, medications, physiotherapy exercises, and practical coping advice, with links to more detailed and trusted Web resources (see Figure 4). It will also signpost users to information regarding support groups that are reputable and age-appropriate, as well as possible involvement activities organized by local organizations and charities.

XSL•FO
RenderX

Figure 3. Screenshots of the app for monitoring symptoms, thoughts, & feelings.



XSL•FO
RenderX

Figure 4. Screenshots of the app for adherence and information.

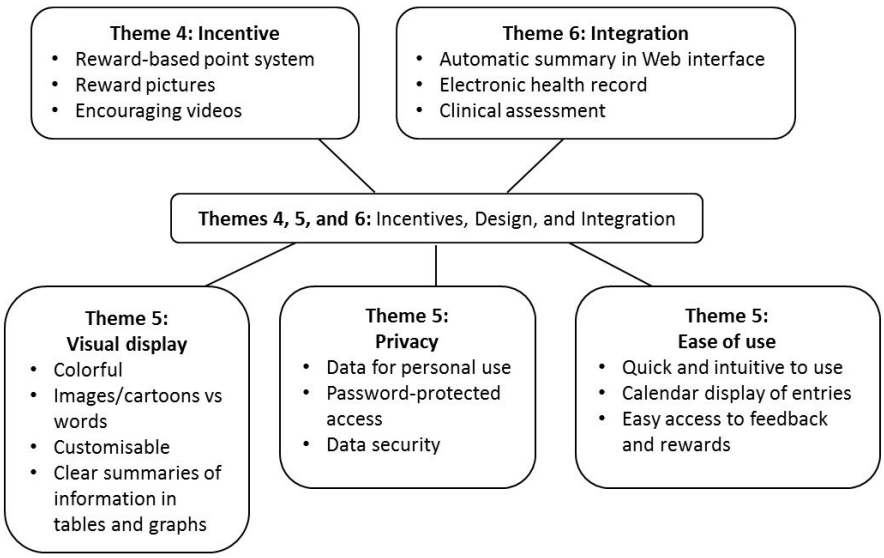
Older patients (18-24 years) requested additional information on how to become more independent as they become an adult and what support is available for higher education and work. The app will provide general information covering these topics, with links to how other YP with JIA overcame challenges of growing up. In addition, there is a diary section where YP can write down any questions or comments that they have and can easily refer to during consultations.

Phase II: Evaluation and Integration Into Clinical Service

During phase II, three more themes emerged (see Figure 5), which related to incentives, design, and integration.

XSL•FO
RenderX

Figure 5. Results of content analysis of transcripts from focus groups for phase II.



Theme 4: Incentives

It was observed that YP wanted to be rewarded for using the app, and they preferred a reward-based rather than a punitive system. One of the patients stated:

> *You should get points to buy rewards for using the app but you shouldn't get points taken away, even if you don't use it for a while.* [Patient 6]

Therefore, the app will incorporate a point-based reward system and include encouraging video messages from celebrities and athletes. However, this was more appealing to the younger group rather than the older group. Older group enjoyed choosing reward pictures (eg, of landscapes, animals, and sports) to use as their wallpaper and motivational quotes that they could display on the home page.

Theme 5: Design of Patient Interface

In terms of designing the user interface, three aspects were important for YP and HCPs: (1) data security, (2) visual appeal/clarity, and (3) convenience. First, as the app will contain patient's personal details, it needs to be password-protected:

> *Arthritis is quite personal so some things you might not want to share with other people.* [Patient 6]

Also, YP wanted the option of diarizing information that would not necessarily be shared with clinicians:

> *Some things might be a bit personal, and just for me to look at in a diary format.* [Patient 13]

Second, YP wanted a colorful visual display, easy-to-read fonts, and pictures and were keen on being able to personalize colors and backgrounds. One participant stated,

> *It's important to have pictures and as little words as possible.* [Patient 3]

Participants also suggested that scales should include faces to represent different pain or fatigue levels:

> *Using faces and cartoon will brighten it up a bit* [Patient 4]

Moreover, YP wanted more space on the home page to display reward videos and pictures on the bottom; therefore, the home page displays a dial (instead of a drop-down ladder) to count down the number of days that are left until the next weekly deadline (see Figure 4).

Lastly, there was a consensus among YP that it should typically take less than 10 min to input all required information:

> *Input must be very quick to do and not feel like a chore.* [Patient 12]

Moreover, YP felt that there needed to be an agreed minimum input to provide clinicians with interpretable data. One adolescent stated:

> *Even if you have no pain, you can just indicate "no pain" so that doctors know that you didn't just forget about the app.* [Patient 10]

It was agreed by both YP and HCPs that the most important data to collect were arthritis-related symptoms, CHAQ, and mood, which should be provided at least once a week. The frequency of completion of other sections should be decided by the YP themselves to give them control over the process.

Ease of use was thus improved in several important ways. First, YP were now able to access the three main monitoring components of the app (symptoms, CHAQ, and mood) directly from the home page rather than from the menu bar. Second, after submitting each input, YP were presented with a pop-up showing how many days ago was their last input for each of the other section, which allowed them to quickly select another section that had not been completed recently. Similarly, hyperlinks were added within the app itself to facilitate access to relevant tips and advice. YP confirmed that these changes made inputting data and finding relevant information faster and more convenient. Lastly, YP requested a calendar display of entries, and the ability to input data retrospectively. One patient said,

XSL•FO
RenderX

> *Sometimes we have a bad day and we're tired so we don't want to the use the app, but we might want to go back and re-enter it later.* [Patient 1]

All the themes and their associated app functionalities were mapped onto the COM-B model (see Table 2).

Theme 6: Clinical Practice Integration

The HCPs discussed how JIApp can be integrated seamlessly into clinical practice and how reports from patients will be received. Clinically relevant data collected from the app will be automatically aggregated and presented in table and graphical summaries (see Figure 4) on a clinician Web interface. The interface was optimized for usability and data are presented in an intuitive format such that any changes over time or associations between various factors (eg, pain and mood) are visually presented to aid comprehension. Figure 6 illustrates the system, where raw data recorded by YP are stored and clinically relevant information can be summarized and displayed on the clinician interface.

The within-application workflow was also aligned with clinician's regular work streams. For example, HCPs can access the Web interface via a patient's existing electronic health records, which can increase the likelihood of integration and access from all multidisciplinary members of the rheumatology team. Information can also be downloaded by HCPs and used to populate clinic letters.

The HCPs also identified possible barriers to integration such as accessibility to all patients and demands for additional resources. For example, HCPs proposed that a desktop version of the app will be beneficial to allow access to YP without a mobile phone. In addition, HCPs were concerned about the additional time needed to monitor remote symptom deteriorations and interpret additional information collected from the app. It was agreed that HCPs are only expected to review the data during or right before a clinic consultation. Therefore, before downloading the app, participants will be informed that data will not be monitored in real time for this stage of the project but will be reviewed during clinic consultations. This is also explicitly explained on the app with a disclosure statement. The HCPs suggested that the effect of app use on HCP-patient interactions and cost-effectiveness of the intervention should be formally assessed in the next stages of the project during a prospective pilot study.

Table 2. Applying the Capability, Opportunity, Motivation-Behavior (COM-B) model to major themes identified for JIApp.

Themes	Capability	Opportunity	Motivation
Theme 1: Self-monitoring	Self-awareness and understanding patterns		Perception of illness and self-efficacy
			Better patient-clinician communication
Theme 2: Treatment adherence		Reminders for treatment adherence	Recording past adherence can stimulate action
Theme 3: Education and support	Knowledge of juvenile idiopathic arthritis and treatments		Confidence in skills and empowerment
	Knowledge of self-management skills and emotion regulation		
Theme 4: Incentives			Expectation of rewards
Theme 5: Design of user interface		Convenient monitoring	
		Convenient access to reliable information and support	

Figure 6. JIApp system.

Phase III: Usability and Acceptability Pilot

All YP endorsed the app and thought that it would be beneficial for them right from the point of initial diagnosis (mean acceptability rating=4.29; standard deviation, SD=0.70). Also, YP were able to input important data within 10 min and found it intuitive to use and easy to navigate (mean usability rating=4.25, SD=0.79).

The usefulness of each of the app functionality was reviewed using qualitative interviews. It was apparent that there were multiple reasons for using the app such as its potential to provide YP with *opportunities* for better self-management, as all YP

XSL•FO
RenderX

discussed the importance of having medication reminders and easy access to information, advice, and support. Furthermore, YP commented on how the app made them feel as if all the information they need is now at their "fingertips," which makes managing their arthritis seem less difficult and increases the likelihood of actually reading the information and engaging in recommended behaviors and self-monitoring. They also appreciated how the information displayed was relevant to their condition and concise. One patient said,

> *I like the way the sections are organized, and how it is not too overwhelming, but how I do have the option of finding out more if I want to.* [Patient 22]

In terms of improving *motivation* to self-manage, YP believed that through self-monitoring, they were able to understand the importance of their own behaviors and how factors under their control can influence disease activity:

> *I never know when I will flare up…so if this can help me see if what I do can "actually" affect my arthritis, then that would be very helpful.* [Patient 29]

Moreover, YP emphasized how reading about other people's successes and having a record of their own past success with managing pain and adhering to treatments can increase their confidence in being able to effectively cope with future challenges.

Also, YP were especially interested in doctors acknowledging their input and using data collected from the app to make better assessments and improve their care. They were enthusiastic about integrating the app into their care, as it will make it easier to discuss issues and symptoms that were recorded beforehand:

> *This app can help me remember what I wanted to say to my doctors, which I often forget by the time I come to the hospital…it will make it easier for me to show them what I mean.* [Patient 18]

Improving communication and patient-clinician relationship can in turn motivate further engagement with health care. Lastly, YP thought the reward system makes it enjoyable to use and is simple, yet engaging:

> *The app itself is already beneficial to the users, but the images as rewards is actually pretty nice and simple. I like it.* [Patient 4]

Most of the YP also fed back on how using the app can improve areas related to self-management *capability*, as reading reliable recommendations regarding pain management techniques will teach them better coping strategies:

> *Sometimes there's nothing you can do except deal with it yourself, and this app can help you do that.* [Patient 21]

Patients also highlighted the fact that recording symptoms the moment they occur using the app enables better understanding of their condition and treatment beyond what can be provided by education alone. One of the patients said,

> *The app is a good way to keep track of my symptoms and should be given by doctors, especially when you're first diagnosed.* [Patient 4]

They believed that each individual is different, and inputting data on a regular basis can help them become aware of their personal triggers and feel more in control of their arthritis:

> *I like visual information and to see how my stress levels interact with my arthritis.* [Patient 26]

Discussion

Principal Findings

This study presents the design and development of a smartphone app to remotely record symptoms and encourage self-management and engagement with health care for YP of all ages (10-24 years) with JIA. We worked closely with YP, parents, and HCPs to identify a catalog of initial requirements for the app system and the necessary changes needed to improve the functionalities of the app and its ease of use. Phase III usability testing demonstrated high rates of overall satisfaction with the app among YP and further affirmed the theoretical validity of the choices made in its design.

Through this app, YP are now able to record numerous additional parameters relating to their physical and psychological well-being, access relevant educational information and social support forums, receive treatment-related reminders, and improve communication with HCPs. More importantly, clinically relevant data is collected by YP themselves in the real-world environment rather than by the HCPs in the hospital environment and is owned by the patient, thus fostering the concept of independent management of health care as the YP mature [70-72]. Moreover, having a clear theoretical basis allows future studies to identify which functions of the app system are driving improvements in self-management.

From the HCP perspective, this app also has the potential to standardize the way reports of clinical symptoms and medication adherence are captured in natural settings and enable fine-grained monitoring of arthritis in a way that is impossible to obtain with occasional clinic visits. Having a more accurate and holistic overview of the symptoms and problems of YP may improve the dynamics of consultation, as time spent on gathering information can be directed to discussing issues that are of concern to the patient. Such a system may also allow systematic collection of a clinically relevant standardized dataset, intended to ensure uniform data collection in the clinical setting [73] that can then inform NHS commissioning processes such as monitoring the efficacy of biologic therapy and the impact of such therapies on patient's functional status. We also wish to prospectively assess in the next stages of the project the feasibility of collecting and processing additional mood- and pain-related data for HCPs, and the impact this has on future resource use (efficient use of consultation time, early identifications of problems, and efficient referrals), treatment effectiveness (whether it facilitates joint decision making with YP and the patient-doctor relationship), and, ultimately, health-related outcomes.

From the health care information technology perspective, the effort was not limited to a standalone app. JIApp is the front-end data-collection and patient-interaction component of a larger platform, involving a substantial back-end architecture, which

XSL•FO
RenderX

includes (1) a flexible database mechanism that synchronizes cached phone data to the data repository, (2) a YP reward service, (3) user feedback and issue tracking system, and (4) secure transport of data from devices to the back-end. Most importantly, the platform also comprises a Web-based clinician's interface to facilitate retrieval and presentation of patient data in a suitable and efficient manner for HCPs.

Moreover, this is not purely a bespoke app for GOSH and UCLH Rheumatology, and the development efforts did not limit the system design to this domain. For example, the software component for the questionnaires could generate any form of questionnaire with a variety of answer options such as optional/blank, single, multiple, free-text, slider-based, and so on. One could reuse such components for rapid development and deployment of similar data-oriented health care apps. The front-end JIApp will be made available via popular app stores (Google Play and Apple).

Comparison With Prior Work

To our knowledge, this study is the first to report a theory-driven and user-centered approach to develop a smartphone-based multifunctional app system for HCPs as well as YP with JIA. Previous JIA-related apps have only focused on monitoring symptoms [47-49] and have not considered how the app can be integrated into clinical practice. It is also the only app developed that addresses all components of the COM-B model for self-management.

Similar self-management apps have been developed for chronic pediatric conditions other than JIA, particularly for YP with diabetes [74,75], asthma [76,77], and cancer [78,79]. Themes that emerged from previous studies were related to what YP with JIA discussed in terms of ease of use, information sharing, rewards, and design. However, medication reminders and tracking missed medications and blood test results were not discussed by YP with other chronic conditions. This shows how certain requirements are unique to JIA patients and highlights the importance of developing condition-specific apps.

In terms of app effectiveness, some asthma and diabetes apps have been shown to benefit YP by increasing asthma control [76] and frequency of blood glucose monitoring [74], whereas others have found no benefits of mobile apps [75,77]. Research in this area is still scarce, and more studies with larger sample sizes and longer follow-ups are needed to evaluate the impact of apps on self-management in YP. However, it is likely for apps that go through iterative development cycles involving both YP and HCPs to be more effective.

Limitations

Our user groups were diverse, even though relatively small, with participating patients having different JIA subtypes at various stages of their treatment and from different ethnical backgrounds. The app's design and evaluation were not affected by the characteristics of YP, but participation bias may have influenced the themes that emerged from our FGDs and interviews. For example, all study participants were computer literate and had easy access to smartphones; whether further refinements of the app are necessary for patients who are less familiar with this technology needs to be established. We considered this limitation by using open frameworks to develop the app so that it can be deployed as a desktop version (running on most modern browsers such as Chrome, Safari, and Firefox) with minimal effort. It is important for future studies to evaluate patient's smartphone usage/ownership and its association with recruitment, app usage, and app satisfaction to ensure that the intervention is suitable and easily accessible for all YP.

Conclusions

In summary, a qualitative user-centered approach was used to develop the first comprehensive smartphone app for YP with JIA. Next steps will involve long-term feasibility testing of the app to establish its effect on patient's self-management skills, disease outcomes, and general well-being, as well as the cost-effectiveness of such an intervention. In addition, patient's app usage will be monitored to assess long-term sustainability and understand reasons for disengagement.

Acknowledgments

We would like to thank the young people, families, and health care professionals who took part in the design and evaluation of the smartphone app. This work was supported by Academic Health Sciences Centre grant, Arthritis Research UK Strategic Award 20164, UCL Proof of Concept grant PoC-13-027, and The Health Foundation Innovating for Improvement program. YI is supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. LRW is supported by National Institute for Health Research Great Ormond Hospital Biomedical Research Centre and Great Ormond Street Children's Charity. RAC, DE, and YI are supported by Arthritis Research UK Grant 20164.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview questions.

[PDF File (Adobe PDF File), 31KB - mhealth_v5i8e121_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2

Usability questionnaire.

[PDF File (Adobe PDF File), 38KB - mhealth_v5i8e121_app2.pdf]

References

1. Prakken B, Albani S, Martini A. Juvenile idiopathic arthritis. The Lancet 2011 Jun;377(9783):2138-2149. [doi: 10.1016/S0140-6736(11)60244-4]
2. Vidqvist K, Malin M, Varjolahti-Lehtinen T, Korpela MM. Disease activity of idiopathic juvenile arthritis continues through adolescence despite the use of biologic therapies. Rheumatology (Oxford) 2013 Nov;52(11):1999-2003 [FREE Full text] [doi: 10.1093/rheumatology/ket256] [Medline: 23893666]
3. Selvaag AM, Aulie HA, Lilleby V, Flatø B. Disease progression into adulthood and predictors of long-term active disease in juvenile idiopathic arthritis. Ann Rheum Dis 2016 Jan;75(1):190-195. [doi: 10.1136/annrheumdis-2014-206034] [Medline: 25362042]
4. World Health Organization. WHO. Geneva, Switzerland: World Health Organization; 2001. The second decade: improving adolescent health and development URL: http://apps.who.int/iris/bitstream/10665/64320/1/WHO_FRH_ADH_98.18_Rev.1.pdf [accessed 2017-08-10] [WebCite Cache ID 6sbe5gCX4]
5. Webb K, Wedderburn LR. Advances in the treatment of polyarticular juvenile idiopathic arthritis. Curr Opin Rheumatol 2015 Sep;27(5):505-510 [FREE Full text] [doi: 10.1097/BOR.0000000000000206] [Medline: 26147756]
6. Oberle EJ, Harris JG, Verbsky JW. Polyarticular juvenile idiopathic arthritis - epidemiology and management approaches. Clin Epidemiol 2014 Oct;6:379-393 [FREE Full text] [doi: 10.2147/CLEP.S53168] [Medline: 25368531]
7. Kim KN. Treatment of juvenile rheumatoid arthritis. Korean J Pediatr 2010 Nov;53(11):936-941 [FREE Full text] [doi: 10.3345/kjp.2010.53.11.936] [Medline: 21218015]
8. Ansell B. Prognosis in juvenile arthritis. Adv Exp Med Biol 1999;455:27-33. [doi: 10.1007/978-1-4615-4857-7_5] [Medline: 10599320]
9. Kasapçopur Ö, Barut K. Treatment in juvenile rheumatoid arthritis and new treatment options. Turk Pediatri Ars 2015 Mar;50(1):1-10 [FREE Full text] [doi: 10.5152/tpa.2015.2229] [Medline: 26078691]
10. Long AR, Rouster-Stevens KA. The role of exercise therapy in the management of juvenile idiopathic arthritis. Curr Opin Rheumatol 2010 Mar;22(2):213-217. [doi: 10.1097/BOR.0b013e328335d1a2] [Medline: 20010296]
11. Haverman L, Grootenhuis MA, van den Berg JM, van VM, Dolman KM, Swart JF, et al. Predictors of health-related quality of life in children and adolescents with juvenile idiopathic arthritis: results from a Web-based survey. Arthritis Care Res (Hoboken) 2012 May;64(5):694-703 [FREE Full text] [doi: 10.1002/acr.21609] [Medline: 22238240]
12. April KT, Feldman DE, Zunzunegui MV, Duffy CM. Association between perceived treatment adherence and health-related quality of life in children with juvenile idiopathic arthritis: perspectives of both parents and children. Patient Prefer Adherence 2008 Feb 02;2:121-128 [FREE Full text] [Medline: 19920952]
13. Len CA, Miotto e Silva VB, Terreri MT. Importance of adherence in the outcome of juvenile idiopathic arthritis. Curr Rheumatol Rep 2014 Apr;16(4):410. [doi: 10.1007/s11926-014-0410-2] [Medline: 24504596]
14. Iversen MD, Hammond A, Betteridge N. Self-management of rheumatic diseases: state of the art and future perspectives. Ann Rheum Dis 2010 Jun;69(6):955-963. [doi: 10.1136/ard.2010.129270] [Medline: 20448289]
15. Feldman DE, De CM, Dobkin PL, Malleson PN, Meshefedjian G, Duffy CM. Effects of adherence to treatment on short-term outcomes in children with juvenile idiopathic arthritis. Arthritis Rheum 2007 Aug 15;57(6):905-912 [FREE Full text] [doi: 10.1002/art.22907] [Medline: 17665485]
16. Shaw KL, Southwood TR, McDonagh JE, British Society of PaediatricAdolescent Rheumatology. Young people's satisfaction of transitional care in adolescent rheumatology in the UK. Child Care Health Dev 2007 Jul;33(4):368-379. [doi: 10.1111/j.1365-2214.2006.00698.x] [Medline: 17584391]
17. Dean AJ, Walters J, Hall A. A systematic review of interventions to enhance medication adherence in children and adolescents with chronic illness. Arch Dis Child 2010 Sep;95(9):717-723. [doi: 10.1136/adc.2009.175125] [Medline: 20522463]
18. Stringer E, Scott R, Mosher D, MacNeill I, Huber AM, Ramsey S, et al. Evaluation of a rheumatology transition clinic. Pediatr Rheumatol Online J 2015 Jun 11;13:22 [FREE Full text] [doi: 10.1186/s12969-015-0016-x] [Medline: 26063057]
19. April KT, Feldman DE, Platt RW, Duffy CM. Comparison between children with juvenile idiopathic arthritis and their parents concerning perceived treatment adherence. Arthritis Rheum 2006 Aug 15;55(4):558-563 [FREE Full text] [doi: 10.1002/art.22104] [Medline: 16874776]
20. Rapoff MA. Management of adherence and chronic rheumatic disease in children and adolescents. Best Pract Res Clin Rheumatol 2006 Apr;20(2):301-314. [doi: 10.1016/j.berh.2005.11.002] [Medline: 16546058]
21. Lawson EF, Hersh AO, Applebaum MA, Yelin EH, Okumura MJ, von SE. Self-management skills in adolescents with chronic rheumatic disease: a cross-sectional survey. Pediatr Rheumatol Online J 2011 Dec 06;9(1):35 [FREE Full text] [doi: 10.1186/1546-0096-9-35] [Medline: 22145642]

XSL•FO
RenderX

22. Hazel E, Zhang X, Duffy CM, Campillo S. High rates of unsuccessful transfer to adult care among young adults with juvenile idiopathic arthritis. Pediatr Rheumatol Online J 2010 Jan 11;8:2 [FREE Full text] [doi: 10.1186/1546-0096-8-2] [Medline: 20148143]
23. Rapoff MA, Lindsley CB. Improving adherence to medical regimens for juvenile rheumatoid arthritis. Pediatr Rheumatol Online J 2007 May 18;5:10 [FREE Full text] [doi: 10.1186/1546-0096-5-10] [Medline: 17550635]
24. Ruperto N, Garcia-Munitis P, Villa L, Pesce M, Aggarwal A, Fasth A, Paediatric Rheumatology International Trials Organisation, Paediatric Rheumatology European Society. PRINTO/PRES international website for families of children with rheumatic diseases: www.pediatric-rheumatology.printo.it. Ann Rheum Dis 2005 Jul;64(7):1101-1106 [FREE Full text] [doi: 10.1136/ard.2004.030221] [Medline: 15760930]
25. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008 Jul 15;59(7):1009-1017 [FREE Full text] [doi: 10.1002/art.23817] [Medline: 18576310]
26. White M, Stinson JN, Lingley-Pottie P, McGrath PJ, Gill N, Vijenthira A. Exploring therapeutic alliance with an internet-based self-management program with brief telephone support for youth with arthritis: a pilot study. Telemed J E Health 2012 May;18(4):271-276. [doi: 10.1089/tmj.2011.0150] [Medline: 22424081]
27. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
28. Jackson C, Eliasson L, Barber N, Weinman J. Applying COM-B to medication adherence: a suggested framework for research and interventions. Eur Heal Psychol 2014;16:7-17.
29. Barker F, Atkins L, de LS. Applying the COM-B behaviour model and behaviour change wheel to develop an intervention to improve hearing-aid use in adult auditory rehabilitation. Int J Audiol 2016 Jul;55(Suppl 3):S90-S98. [doi: 10.3109/14992027.2015.1120894] [Medline: 27420547]
30. Alexander KE, Brijnath B, Mazza D. Barriers and enablers to delivery of the healthy kids check: an analysis informed by the theoretical domains framework and COM-B model. Implement Sci 2014 May 23;9:60 [FREE Full text] [doi: 10.1186/1748-5908-9-60] [Medline: 24886520]
31. Pai AL, McGrady M. Systematic review and meta-analysis of psychological interventions to promote treatment adherence in children, adolescents, and young adults with chronic illness. J Pediatr Psychol 2014 Sep;39(8):918-931 [FREE Full text] [doi: 10.1093/jpepsy/jsu038] [Medline: 24952359]
32. Livingstone S, Haddon L, Vincent J, Mascheroni G, Olafsson K. LSE. London: London School of Economics and Political Science; 2014. Net children go mobile: the UK report URL: https://www.lse.ac.uk/media@lse/research/EUKidsOnline/EU%20Kids%20III/Reports/NCGMUKReportfinal.pdf [accessed 2017-08-10] [WebCite Cache ID 6sboc96Mv]
33. Parmanto B, Pramana G, Yu DX, Fairman AD, Dicianno BE, McCue MP. iMHere: a novel mhealth system for supporting self-care in management of complex and chronic conditions. JMIR Mhealth Uhealth 2013 Jul 11;1(2):e10 [FREE Full text] [doi: 10.2196/mhealth.2391] [Medline: 25100682]
34. 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]
35. Applebaum MA, Lawson EF, von SE. Perception of transition readiness and preferences for use of technology in transition programs: teens' ideas for the future. Int J Adolesc Med Health 2013;25(2):119-125. [doi: 10.1515/ijamh-2013-0019] [Medline: 23740658]
36. de la Vega R, Roset R, Galán S, Miró J. Fibroline: a mobile app for improving the quality of life of young people with fibromyalgia. J Health Psychol 2016 May 24:- Epub ahead of print. [doi: 10.1177/1359105316650509] [Medline: 27225285]
37. 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 Sep;55(1):65-69. [doi: 10.3414/ME14-01-0106] [Medline: 26391694]
38. Palermo TM, Valenzuela D, Stork PP. A randomized trial of electronic versus paper pain diaries in children: impact on compliance, accuracy, and acceptability. Pain 2004;107(3):213-219. [doi: 10.1016/j.pain.2003.10.005]
39. Stinson JN, Jibb LA, Lalloo C, Feldman BM, McGrath PJ, Petroz GC, et al. Comparison of average weekly pain using recalled paper and momentary assessment electronic diary reports in children with arthritis. Clin J Pain 2014 Dec;30(12):1044-1050. [doi: 10.1097/AJP.0000000000000072] [Medline: 24535055]
40. Bromberg MH, Connelly M, Anthony KK, Gil KM, Schanberg LE. Self-reported pain and disease symptoms persist in juvenile idiopathic arthritis despite treatment advances: an electronic diary study. Arthritis Rheumatol 2014 Feb;66(2):462-469 [FREE Full text] [doi: 10.1002/art.38223] [Medline: 24504820]
41. Connelly M, Bromberg MH, Anthony KK, Gil KM, Franks L, Schanberg LE. Emotion regulation predicts pain and functioning in children with juvenile idiopathic arthritis: an electronic diary study. J Pediatr Psychol 2012 Oct;37(1):43-52 [FREE Full text] [doi: 10.1093/jpepsy/jsr088] [Medline: 22037006]
42. Fisher RS, Blum DE, DiVentura B, Vannest J, Hixson JD, Moss R, et al. Seizure diaries for clinical research and practice: limitations and future prospects. Epilepsy Behav 2012 Jul;24(3):304-310. [doi: 10.1016/j.yebeh.2012.04.128] [Medline: 22652423]

XSL•FO
RenderX

43. Connelly M, Anthony KK, Sarniak R, Bromberg MH, Gil KM, Schanberg LE. Parent pain responses as predictors of daily activities and mood in children with juvenile idiopathic arthritis: the utility of electronic diaries. J Pain Symptom Manage 2010 Mar;39(3):579-590 [FREE Full text] [doi: 10.1016/j.jpainsymman.2009.07.013] [Medline: 20303032]
44. 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]
45. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
46. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014 Jul;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
47. Stinson JN, Petroz GC, Tait G, Feldman BM, Streiner D, McGrath PJ, et al. e-Ouch: usability testing of an electronic chronic pain diary for adolescents with arthritis. Clin J Pain 2006;22(3):295-305. [doi: 10.1097/01.ajp.0000173371.54579.31] [Medline: 16514331]
48. Stinson JN, Petroz GC, Stevens BJ, Feldman BM, Streiner D, McGrath PJ, et al. Working out the kinks: testing the feasibility of an electronic pain diary for adolescents with arthritis. Pain Res Manag 2008;13(5):375-382. [doi: 10.1155/2008/326389]
49. Stinson JN, Stevens BJ, Feldman BM, Streiner D, McGrath PJ, Dupuis A, et al. Construct validity of a multidimensional electronic pain diary for adolescents with arthritis. Pain 2008 Jun;136(3):281-292. [doi: 10.1016/j.pain.2007.07.002] [Medline: 17723279]
50. Anderson R. New MRC guidance on evaluating complex interventions. BMJ 2008 Oct 22;337:a1937. [doi: 10.1136/bmj.a1937]
51. Mackenzie M, O'Donnell C, Halliday E, Sridharan S, Platt S. Do health improvement programmes fit with MRC guidance on evaluating complex interventions? BMJ 2010 Feb 01;340:c185. [doi: 10.1136/bmj.c185]
52. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
53. Petty RE, Southwood TR, Manners P, Baum J, Glass DN, Goldenberg J, International League of Associations for Rheumatology. International league of associations for rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. J Rheumatol 2004 Feb;31(2):390-392. [Medline: 14760812]
54. Marshall NJ, Wilson G, Lapworth K, Kay LJ. Patients' perceptions of treatment with anti-TNF therapy for rheumatoid arthritis: a qualitative study. Rheumatology (Oxford) 2004 Aug;43(8):1034-1038 [FREE Full text] [doi: 10.1093/rheumatology/keh237] [Medline: 15150436]
55. Krueger R, Casey M. Focus Groups: A Practical Guide for Applied Research. Thousand Oaks, CA: SAGE Publications; 2000.
56. Kidd PS, Parshall MB. Getting the focus and the group: enhancing analytical rigor in focus group research. Qual Health Res 2000 May 01;10(3):293-308. [doi: 10.1177/104973200129118453]
57. Lund AM. Measuring usability with the USE questionnaire. Usability Interface 2001;8:3-6.
58. Guest G. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 2006 Feb 01;18(1):59-82. [doi: 10.1177/1525822X05279903]
59. 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]
60. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
61. QSR International Pty Ltd. 2012. NVivo qualitative data analysis software version 10 URL: http://www.qsrinternational.com/nvivo-product [accessed 2017-07-03] [WebCite Cache ID 6rgRKrV8G]
62. Huguet A, Stinson JN, McGrath PJ. Measurement of self-reported pain intensity in children and adolescents. J Psychosom Res 2010 Apr;68(4):329-336. [doi: 10.1016/j.jpsychores.2009.06.003] [Medline: 20307699]
63. Geerdink LM, Prince FH, Looman CW, van Suijlekom-Smit LW. Development of a digital childhood health assessment questionnaire for systematic monitoring of disease activity in daily practice. Rheumatology (Oxford) 2009 Aug;48(8):958-963 [FREE Full text] [doi: 10.1093/rheumatology/kep135] [Medline: 19509189]
64. Schanberg LE, Sandstrom MJ. Causes of pain in children with arthritis. Rheum Dis Clin North Am 1999 Feb;25(1):31-53. [doi: 10.1016/S0889-857X(05)70054-5]
65. Schanberg LE, Anthony KK, Gil KM, Maurin EC. Daily pain and symptoms in children with polyarticular arthritis. Arthritis Rheum 2003 May;48(5):1390-1397 [FREE Full text] [doi: 10.1002/art.10986] [Medline: 12746912]
66. Consolaro A, Ruperto N, Bazso A, Pistorio A, Magni-Manzoni S, Filocamo G, Paediatric Rheumatology International Trials Organisation. Development and validation of a composite disease activity score for juvenile idiopathic arthritis. Arthritis Rheum 2009 May 15;61(5):658-666 [FREE Full text] [doi: 10.1002/art.24516] [Medline: 19405003]

XSL•FO
RenderX

67. Laurent J, Catanzaro SJ, Joiner TE, Rudolph KD, Potter KI, Lambert S, et al. A measure of positive and negative affect for children: scale development and preliminary validation. Psychol Assessment 1999;11(3):326-338. [doi: 10.1037//1040-3590.11.3.326]
68. Alfvén G, Nilsson S. Validity and reliability of a new short verbal rating scale for stress for use in clinical practice. Acta Paediatr 2014 Apr;103(4):e173-e175. [doi: 10.1111/apa.12558] [Medline: 24533818]
69. Singh G, Athreya BH, Fries JF, Goldsmith DP. Measurement of health status in children with juvenile rheumatoid arthritis. Arthritis Rheum 1994 Dec;37(12):1761-1769. [Medline: 7986222]
70. Sawyer SM, Aroni RA. Self-management in adolescents with chronic illness. What does it mean and how can it be achieved? Med J Aust 2005 Oct 17;183(8):405-409. [Medline: 16225444]
71. Von Korff M. Collaborative management of chronic illness. Ann Intern Med 1997 Dec 15;127(12):1097. [doi: 10.7326/0003-4819-127-12-199712150-00008]
72. Lorig KR, Holman HR. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01]
73. Armitt G, Cobb J, Shoop S, Rashid A, Sunil S, Douglas S, et al. CAPTURE-JIA Patient Status Dataset: a consensus-derived core dataset to improve clinical care for childrenyoung people with Juvenile Idiopathic Arthritis. Proceedings of the British Society for AdolescentPaediatric Rheumatology Annual Conference. 2015 Presented at: Proceedings of the British Society for Adolescent and Paediatric Rheumatology Annual Conference; September 24-25, 2015; Belfast, UK.
74. 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;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
75. Frøisland DH, Arsand E, Skårderud F. Improving diabetes care for young people with type 1 diabetes through visual learning on mobile phones: mixed-methods study. J Med Internet Res 2012 Aug 06;14(4):e111 [FREE Full text] [doi: 10.2196/jmir.2155] [Medline: 22868871]
76. Burbank AJ, Lewis SD, Hewes M, Schellhase DE, Rettiganti M, Hall-Barrow J, et al. Mobile-based asthma action plans for adolescents. J Asthma 2015 Jan;52(6):583-586. [doi: 10.3109/02770903.2014.995307] [Medline: 25494553]
77. Ryan D, Price D, Musgrave SD, Malhotra S, Lee AJ, Ayansina D, et al. Clinical and cost effectiveness of mobile phone supported self monitoring of asthma: multicentre randomised controlled trial. BMJ 2012;344:e1756 [FREE Full text] [Medline: 22446569]
78. Aldiss S, Taylor R, Soanes L, Maguire R, Sage M, Kearney N, et al. Working in collaboration with young people and health professionals. A staged approach to the implementation of a randomised controlled trial. J Res Nurs 2011 Nov;16(6):561-576. [doi: 10.1177/1744987110380803]
79. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013 Mar;15(3):e51 [FREE Full text] [doi: 10.2196/jmir.2350] [Medline: 23475457]

Abbreviations

CHAQ: Childhood Health Assessment Questionnaire
ERA: enthesitis-related arthritis
FGD: focus group discussion
GOSH: Great Ormond Street Hospital
HCP: health care professional
JIA: juvenile idiopathic arthritis
sc: subcutaneously
SD: standard deviation
UCLH: University College London Hospital
VAS: visual analog scales
VRSS: verbal rating scale for stress
YP: young people

XSL•FO
RenderX

Edited by G Eysenbach; submitted 02.01.17; peer-reviewed by E Lawson, R Grainger; comments to author 02.02.17; revised version received 13.03.17; accepted 14.04.17; published 15.08.17.

Please cite as:
Cai RA, Beste D, Chaplin H, Varakliotis S, Suffield L, Josephs F, Sen D, Wedderburn LR, Ioannou Y, Hailes S, Eleftheriou D
Developing and Evaluating JIApp: Acceptability and Usability of a Smartphone App System to Improve Self-Management in Young People With Juvenile Idiopathic Arthritis
JMIR Mhealth Uhealth 2017;5(8):e121
URL: http://mhealth.jmir.org/2017/8/e121/
doi:10.2196/mhealth.7229
PMID:28811270

©Ran A Cai, Dominik Beste, Hema Chaplin, Socrates Varakliotis, Linda Suffield, Francesca Josephs, Debajit Sen, Lucy R Wedderburn, Yiannakis Ioannou, Stephen Hailes, Despina Eleftheriou. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 15.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Clinical Validation of Heart Rate Apps: Mixed-Methods Evaluation Study

Thijs Vandenberk[1,2], MSc; Jelle Stans[1], MSc; Christophe Mortelmans[3], MD; Ruth Van Haelst[3], MD; Gertjan Van Schelvergem[1], MSc; Caroline Pelckmans[1], MSc; Christophe JP Smeets[1,2], MSc; Dorien Lanssens[1], MSc; Hélène De Cannière[1,2], MSc; Valerie Storms[1], MSc, PhD; Inge M Thijs[1], MSc, PhD; Bert Vaes[3], Prof MD; Pieter M Vandervoort[1,2], Prof MD

[1]Mobile Health Unit, Faculty of Medicine and Life Sciences, Hasselt University, Hasselt, Belgium

[2]Department of Cardiology, Ziekenhuis Oost-Limburg, Genk, Belgium

[3]Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

Corresponding Author:
Thijs Vandenberk, MSc
Mobile Health Unit
Faculty of Medicine and Life Sciences
Hasselt University
Martelarenlaan 42
Hasselt, 3600
Belgium
Phone: 32 11268111
Fax: 32 11268199
Email: thijs.vandenberk@uhasselt.be

Related Article:

This is a corrected version. See correction statement: http://mhealth.jmir.org/2018/3/e19/

Abstract

Background: Photoplethysmography (PPG) is a proven way to measure heart rate (HR). This technology is already available in smartphones, which allows measuring HR only by using the smartphone. Given the widespread availability of smartphones, this creates a scalable way to enable mobile HR monitoring. An essential precondition is that these technologies are as reliable and accurate as the current clinical (gold) standards. At this moment, there is no consensus on a gold standard method for the validation of HR apps. This results in different validation processes that do not always reflect the veracious outcome of comparison.

Objective: The aim of this paper was to investigate and describe the necessary elements in validating and comparing HR apps versus standard technology.

Methods: The FibriCheck (Qompium) app was used in two separate prospective nonrandomized studies. In the first study, the HR of the FibriCheck app was consecutively compared with 2 different Food and Drug Administration (FDA)-cleared HR devices: the Nonin oximeter and the AliveCor Mobile ECG. In the second study, a next step in validation was performed by comparing the beat-to-beat intervals of the FibriCheck app to a synchronized ECG recording.

Results: In the first study, the HR (BPM, beats per minute) of 88 random subjects consecutively measured with the 3 devices showed a correlation coefficient of .834 between FibriCheck and Nonin, .88 between FibriCheck and AliveCor, and .897 between Nonin and AliveCor. A single way analysis of variance (ANOVA; P=.61 was executed to test the hypothesis that there were no significant differences between the HRs as measured by the 3 devices. In the second study, 20,298 (ms) R-R intervals (RRI)–peak-to-peak intervals (PPI) from 229 subjects were analyzed. This resulted in a positive correlation (rs=.993, root mean square deviation [RMSE]=23.04 ms, and normalized root mean square error [NRMSE]=0.012) between the PPI from FibriCheck and the RRI from the wearable ECG. There was no significant difference (P=.92) between these intervals.

Conclusions: Our findings suggest that the most suitable method for the validation of an HR app is a simultaneous measurement of the HR by the smartphone app and an ECG system, compared on the basis of beat-to-beat analysis. This approach could lead to more correct assessments of the accuracy of HR apps.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2017;5(8):e129) doi:10.2196/mhealth.7254

KEYWORDS

heart rate; software validation; remote sensing technology

Introduction

The rapid evolution of technology has brought highly sophisticated electronic devices such as smartphones in our daily lives. The market for these devices is growing at a rapid pace. Globally, there are about 2.6 billion smartphone subscriptions, and by 2020, this number is projected to reach 6.1 billion [1]. Smartphones with multimedia capabilities open new possibilities for app development and service delivery [2]. Recently, smartphones have started to be used for medical purposes to measure numerous vital parameters such as heart rate (HR) and body temperature. This enables the use of a smartphone as a wireless HR monitor [3]. HR is nowadays measured by nurses who have congested schedules and therefore limited time to measure the HR of patients. HR-sensing devices may be a solution for this problem and can be useful in extending the reach of vital signs monitoring in- and outside hospitals, which is typically limited by constraints on human resources [4]. Nowadays, the use of wireless monitors for assessment of HR is a common component of health and fitness programs. Unlike HR apps on smartphones, these HR monitors require a telemetric strap to be worn around the thoracic region or arm to ensure electrocardiography (ECG)-derived HR [5]. The heart is an electromechanical pump with a rhythmic pumping cycle, in which the electrical activity of the heart can be represented in the electrocardiogram by a *P-*, *QRS-*, *and T-wave*. For HR and rhythm analysis, the ECG still remains the gold standard. The contraction of the heart propagates a blood pressure pulse wave through the arterial system that travels to the peripheries. A typical arterial blood pressure waveform comprises a systolic upstroke representing the ventricular ejection. After the systolic contraction, the aortic valve closes, which results in a sudden drop in pressure called the dichroic notch [6]. When the pulse pressure wave is passed, these capillaries relax and eject the excessive blood they accumulated, allowing them to return to their initial state. When the areas with dense capillary beds are studied (ie, fingertips, toes, and earlobes), it is possible to observe this pooling of blood by using optical technologies. This technique is also known as the photoplethysmography (PPG) principle. The relationship between ECG, arterial blood pressure (ABP), and PPG is visualized in Figure 1.

PPG is already used in the clinic to measure oxygen saturation and pulse rate [7]. Additionally, it can also be used to estimate cardiac output [8]. PPG used as signal to measure HR is described as the pulse signal. As such, PPG can be used to measure HR without the need for an ECG device. Furthermore, the HR derived from the PPG signal can be used in a series of calculations to determine the heart rate variability (HRV) [9].

PPG is easy to set up, convenient, simple, and economically efficient. It uses a probe that contains a light source and a photodetector to detect the blood volume pulse. The amount of backscattered light corresponds with the variation in blood volume [10]. Hertzman [11] were the first to find a relationship between the intensity of backscattered light and blood volume in 1938.

Traditional PPG systems typically use a narrow wavelength light source (ie, light-emitting diodes [LEDs] with certain colors such as infrared, red, or green) and a specific photodetector to detect PPG signals through the skin. Interestingly, the smartphone camera in combination with the LED flashlight is able to detect these small variations in skin color caused by the blood flow (Figure 2). The camera uses wide-bandwidth pixel-enabling color detection in the red, green, and blue range (RGB-color).

In 2010, Jonathan and Leahy presented a case study which concluded that HR could indeed be measured through PPG by using a smartphone. This case experiment was confirmed by Gregoski et al in 2011 [5,12]. Currently, numerous smartphone apps exist that measure HR. However, the validity of these apps has not always been confirmed [13]. At this moment, there is no consensus on a gold standard method for the validation of a HR app based on a PPG signal. This results in different validation processes that not always reflect the veracious outcome of comparison. Validation can be done in two ways: (1) by comparing the HR [4,14] or (2) by comparing the ECG-derived R-R intervals (RRI) [15,16] and the PPG-derived peak-to-peak intervals (PPI) [17] as shown in Figure 3. The goal of this paper was to explore which of the two validation approaches is more suited and to investigate and describe the necessary elements in validating and comparing HR apps versus standard technology.

XSL•FO
RenderX

Figure 1. Visualized relation between electrocardiography (ECG), arterial blood pressure (ABP), and photoplethysmography (PPG).

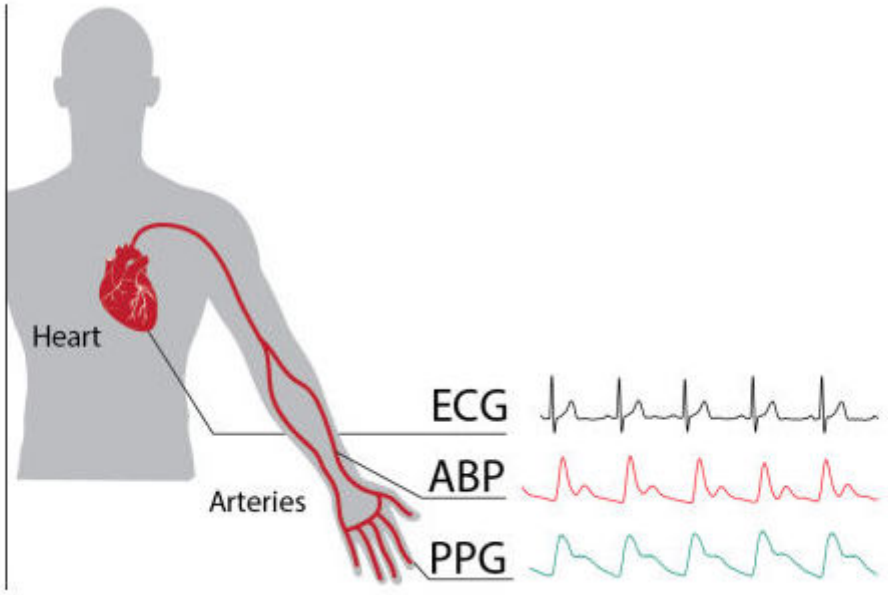
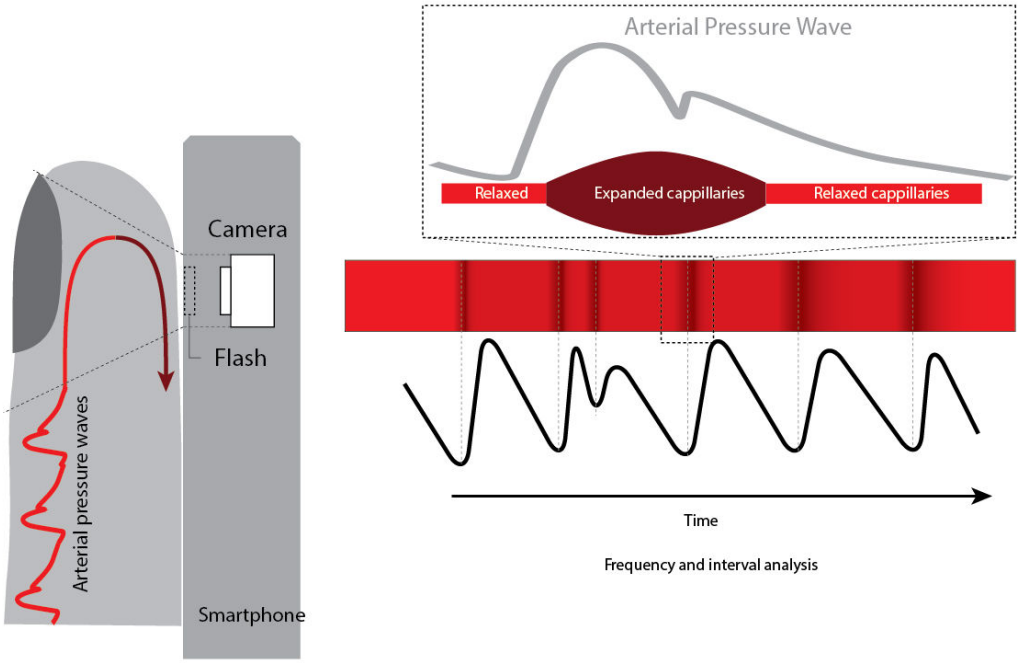
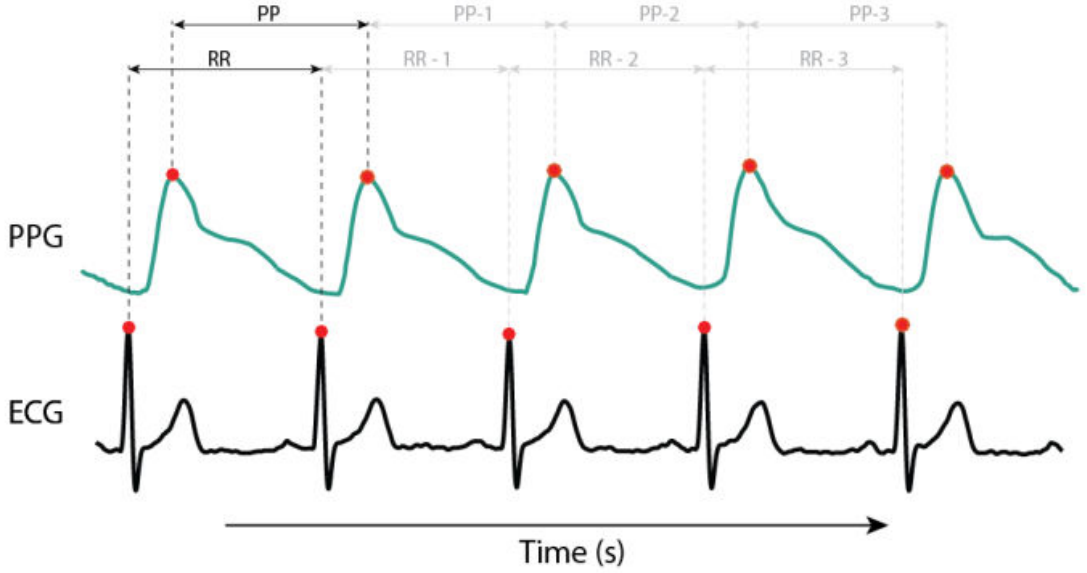


Figure 2. Photoplethysmography (PPG) principle by smartphone.



XSL•FO
RenderX

Figure 3. Beat-to-beat analysis from R-R intervals (RRI) and peak-to-peak intervals (PPI).



Methods

To investigate the correct method that should be used to clinically validate smartphone apps that measure HR, the smartphone app FibriCheck was used as test case. For this purpose, two separate, independent, prospective nonrandomized studies were performed. In the first study, the HR as measured by FibriCheck was compared with the HR measured by 2 sequentially used Food and Drug Administration (FDA)-cleared HR-measuring devices. In the second study, the beat-to-beat (RRI/PPI) accuracy of the FibriCheck app was compared with a raw single-lead ECG that was recorded in a synchronized way. For this, a validated and wearable ECG recorder [18] (Imec Holst Centre,) was used. Both studies comply with the Declaration of Helsinki. The study protocol was approved by the local committee on human research, and all participants provided written informed consent.

Study 1: FibriCheck Compared With FDA-Approved HR Devices

Only 2 FDA-cleared HR measurement devices were used, that is, Nonin oximeter and AliveCor. These 2 devices, which employ different measurement methods, were used to validate a novel smartphone app that measured the participant's HR based on the PPG principle. Nonin uses the transmission PPG method as a stand-alone device, whereas AliveCor uses the ECG as a method measured with a smartphone. The participant's HR was measured 3 times with each measuring device according to the protocol of Terbizan et al [19]. Both FibriCheck and AliveCor were installed on an iPhone 5 (Apple Inc). Participants were recruited in the tertiary care center Ziekenhuis Oost-Limburg (ZOL, Genk, Belgium) in 2015. Inclusion criteria were 18 years or older and able to provide the Dutch written informed consent. Exclusion criteria were failure to obtain valid data with any device or failure to correctly follow the protocol.

A normalization period of 10 min before the first measurement was used to obtain a resting HR. For standardization, all measurements were performed in the same order, that is, FibriCheck app, Nonin oximeter, and AliveCor. The FibriCheck app measures the HR for 10 s by placing the index finger over the rear camera and LED while holding the smartphone in the other hand (Figure 4, left). Nonin and AliveCor measurements were performed according to the manufacturers' guidelines.

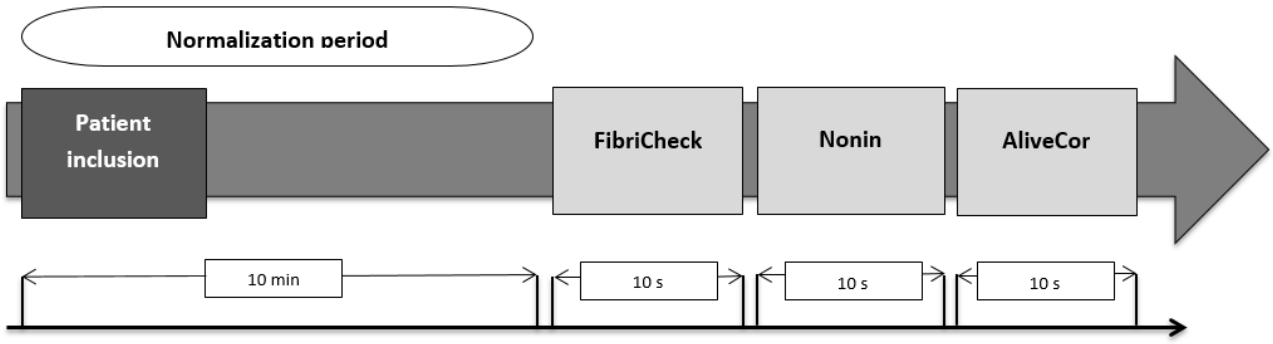
Figure 5 represents a graphical overview of the step-by-step approach of measurement in study 1. In case of the FibriCheck app, the shown HR result value in beats per minutes (BPM) was used, whereas for both the Nonin oximeter and AliveCor app, the minimum and maximum HR during a 10 s measurement were averaged. Subsequently, all results of HRs measured by the different devices were statistically compared with each other.

The Shapiro-Wilk test was performed to test for normality. Different tests were performed to analyze the results. First, a Pearson correlation test of each possible pair of methods was performed to assess correlation. Second, the agreement between methods was assessed by the construction of Bland-Altman plots of the same pairs. Finally, a paired student *t* test and single-way analysis of variance (ANOVA) test were executed to see whether there was a significant difference between the HR as measured by the different methods. Statistical analysis and generation of Bland-Altman plots were performed by using R statistical software (version 3.2.2).

XSL•FO
RenderX

Figure 4. Graphical representation of how measurements are performed using the different devices. Left, FibriCheck application; Middle, Nonin oximeter; Right, AliveCor.

Figure 5. Graphical overview measurement-process study 1.



Study 2: FibriCheck Beat-to-Beat Accuracy Compared With Wearable ECG in Broad Dynamic Range

The beat-to-beat accuracy of the FibriCheck app was verified by comparing it with a wearable ECG patch. To do so, the FibriCheck smartphone app was used and installed on an iPhone 5S. This app also enables synchronization of the PPG signal, with a simultaneously measured ECG signal of a single-lead wearable ECG patch. This wearable device was attached to the upper left corner of the patient's chest with 2 disposable electrodes (Figure 6). This enables comparing the raw data of the 2 devices (ie, FibriCheck and wearable ECG) and measurement principles (ie, PPG and ECG). Inclusion criteria were 18 years or older and able to provide the Dutch written informed consent.

Patients with an active pacemaker rhythm were excluded. Patients were either included by a general practitioner (GP) or by a researcher in ZOL between November 2015 and March 2016.

The GP enrolled male and female patients over the age of 65 years, with or without a history or diagnosis of atrial fibrillation (AF). The researcher included subjects who were diagnosed with AF by a 12-lead ECG system and healthy subjects who underwent a sports session. The study population is heterogeneous since it contains patients with a regular or irregular heart rhythm as well as low and high HRs.

Subjects, included by the GP, were measured in a sitting position and asked to perform three consecutive measurements of 60 s. AF patients, included by the researcher, were measured 3 to 6 times in a lying or sitting position. The sports session involved 5 min cycling at a high pace on a stationary exercise bike to reach a maximum HR. Two measurements before and after the exercise were done.

The FibriCheck app converts 60 Hz video data to raw signals, which were processed with Matlab (Math-Works) to derive the corresponding PPG signal. Time synchronization between ECG and PPG was automatically done by the FibriCheck app. Subsequently, peak detection of the ECG signal and the preconditioned PPG signal was performed by blinded and manual annotation of the identified peaks using Matlab. Finally, it was possible to extract the interpeak distance. An example of the automatic synchronization is shown in Figure 7.

The Kolmogorov-Smirnov test was performed to test for normality. The not normally distributed data are expressed as a median and interquartile range (IQR). A two-sided Wilcoxon signed-rank test was performed to compare two continuous variables for the not-normally distributed data. Correlation between the two continuous variables was calculated by a Spearman correlation test. All analyses were two-sided, and the level of significance was set at a value of .05. Root mean square error (RMSE) and normalized root mean square error (NRMSE) were performed to evaluate the range of errors between predicted and observed values. Data analyses were performed with R statistical software (version 3.2.2). Graphical presentations, such as correlations plots, Bland-Altman plot, and Kernel density plots, were made in RStudio version 0.99.486 (Rstudio Inc).

XSL•FO
RenderX

Figure 6. Measurement setup for simultaneous photoplethysmography (PPG) and electrocardiography (ECG) recording. A wearable ECG sensor to measure 1-lead ECG data; B, FibriCheck app to measure PPG data.

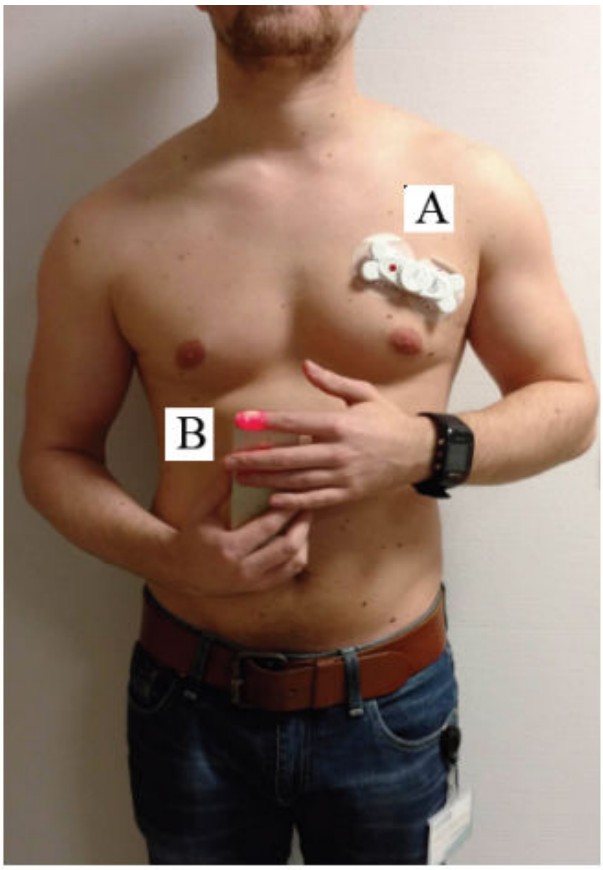
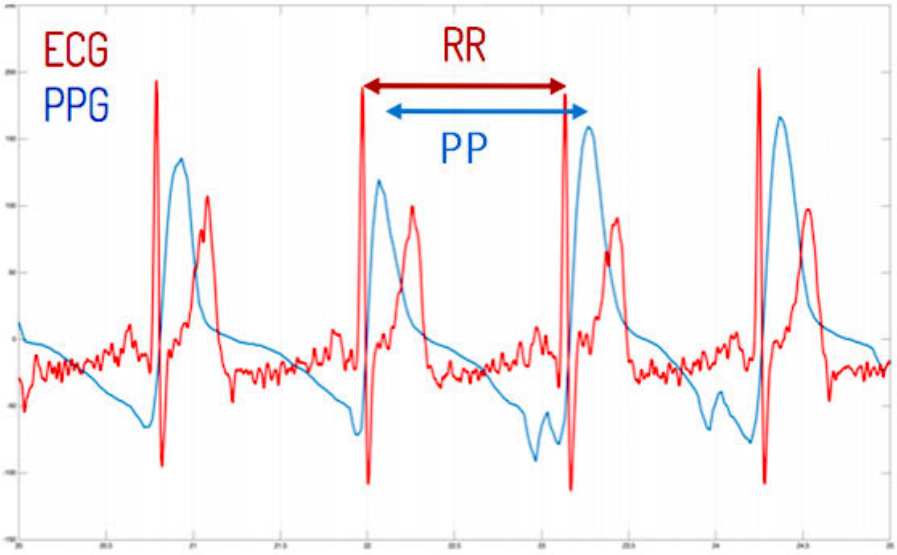


Figure 7. Synchronization of electrocardiography (ECG) and photoplethysmography (PPG) signal.



XSL•FO
RenderX

Results

Study 1: FibriCheck Compared With FDA-Approved HR Devices

In total, 91 persons were included in the study. A total of 3 persons were excluded from analysis because of failure to obtain valid data with 1 or more devices. This resulted in a final study population of 88 subjects. Table 1 shows the characteristics of these patients. Data are expressed as mean (standard deviation [SD]).

The HR measurements as acquired by the three different methods were compared for assessing the ability of the FibriCheck app to correctly measure subjects' HR. First, two-sided Pearson correlation tests were performed to evaluate the correlation between each possible pair of devices. Second, a paired student *t* test was performed. Thereafter, the RMSE and NRMSE were calculated (Table 3).

Table 1. Characteristics of patients in study 1.

Variable	Men (n=50)	Women (n=38)	All
Age in years, mean (SD)	49.34 (17.62)	44.63 (17.69)	47.31 (17.70)
Height, mean (SD)	177.14 (8.11)	165.97 (5.22)	172.26 (8.92)
Weight, mean (SD)	82 (14.12)	66.55 (6.56)	75.25 (13.75)

Table 2.

Measuring device	Heart rate, mean (SD[a])
Nonin, bpm[b]	69 (12)
FibriCheck, bpm	71 (13)
AliveCor, bpm	69 (12)

[a]SD: standard deviation.

[b]BPM: beats per minute.

Table 3. Correlation coefficients, statistical significance, root mean square error, and normalized root mean square error for each pair of devices.

Pair of devices	Correlation coefficient (*r*)	Statistical significance (two-tailed)	Root mean square error (beats per minute)	Normalized root mean square error (beats per minute)
FibriCheck–Nonin	.834	*P*=.36	7.40	0.11
FibriCheck–AliveCor	.88	*P*=.45	6.26	0.09
Nonin–AliveCor	.897	*P*=.87	5.46	0.08

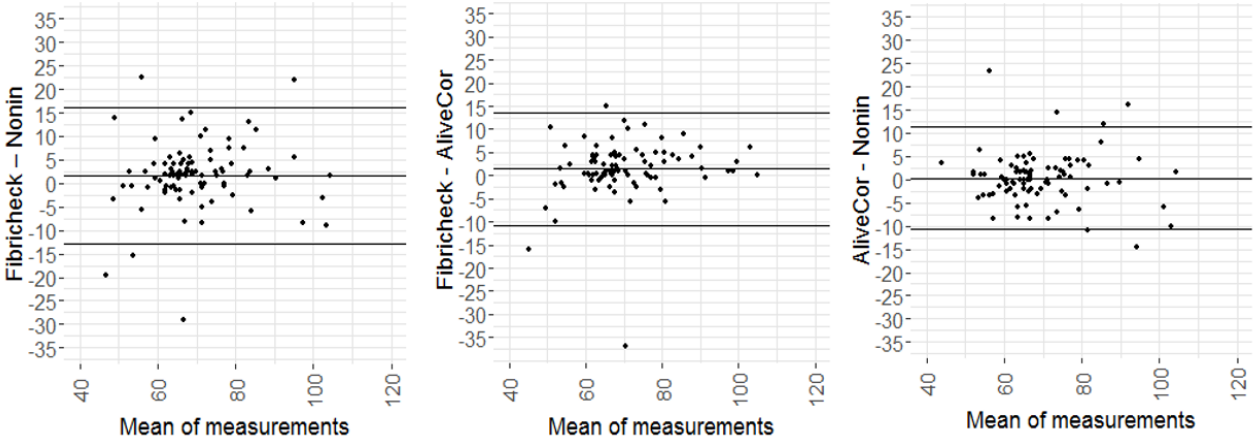
Finally, an ANOVA test was performed to evaluate whether there was a significant difference between the results of the HR measurements of the different devices. The results indicate no significant difference (*P*=.61) between the HRs measured by the 3 different devices.

Results show high correlations without significant differences for all device pairs. However, since correlation does not necessarily imply agreement, Bland-Altman plots were constructed to evaluate agreement between each pair of devices (Figure 8).

The mean bias ranged from 0.29 bpm (Nonin–AliveCor) to 1.42 bpm (FibriCheck–AliveCor) and 1.72 bpm (FibriCheck–Nonin). Some measurements were not situated between the lower limit of agreement (LLA) and the upper limit of agreement (ULA).

XSL•FO
RenderX

Figure 8. Bland-Altman plots for each device pair. The mean difference (bias), 1.96 (lower limit of agreement, LLA) and +1.96 standard deviations (upper limit of agreement, ULA) are plotted as full lines.



Study 2: FibriCheck Beat-to-Beat Accuracy Compared With Wearable ECG in Broad Dynamic Range

A total of 247 subjects were measured with the FibriCheck app in the presence of a GP (n=238) or a researcher in ZOL (n=19). The researcher included both healthy subjects (n=12) and patients who were diagnosed with AF by a 12-lead ECG system (n=7). Around 18 patients from the total study population, all included by the GP, had a pacemaker and were all excluded. Therefore, the final study population included 229 subjects. Table 4 shows the characteristics of these patients.

Table 4. Characteristics of patients in study 2.

Variable	Men (n=105)	Women (n=120)	All[a]
Age in years, mean (SD[b])	73.54 (13.86)	75.51 (14.34)	74.59 (14.12)
Height in centimeters, mean (SD)	173.79 (8.04)	161.22 (7.32)	166.25 (13.79)
Weight in kilograms, mean (SD)	80.3 (13.59)	67.52 (13.83)	73.45 (15.11)
Diabetic, n (%)	20 (19)	24 (20)	44 (20)
Atrial fibrillation, n (%)	50 (47.62)	48 (40)	98 (43.56)
Systolic blood pressure in mm Hg, mean (SD)	128.54 (13.84)	130.6 (20.34)	129.65 (17.13)[c]
Diastolic blood pressure in mm Hg, mean (SD)	74.39 (7.59)	73.55 (11.10)	73.94 (9.62)[c]
CHA_2 DS_2-VASc[d] score, mean (SD)	3.61 (1.75)	4.58 (1.86)	4.13 (1.87)

[a]Demographics of 4 patients were reported as missing data.[b]SD: standard deviation.Systolic and diastolic blood pressure were not included for patients who underwent the sport session.[d]CHA2DS2-VASc calculates the stroke risk for patients with atrial fibrillation.

Table 5 provides the study results. In total, 237 measurements (PPG-ECG pairs) were performed, which resulted in a 20,298 beat-to-beat analysis. An average interval of 758 (RRI) and 758 (PPI) was observed.

Table 5. Overview study results.

Variable	ECG[a]	PPG[b]
Number of Intervals	20,298	20,298
Average interval (ms)	758.4 (351.6)	758.2 (333.3)
Minimum value (ms)	312.5	316.7
Maximum value (ms)	2223.0	2233.0

[a]ECG: electrocardiography.

[b]PPG: photoplethysmography.

The Wilcoxon signed-rank test showed no significant difference between ECG and PPG (*P*=.92). To calculate the correlation and difference between the ECG and PPG measurement, the Spearman rank-order correlation, RMSE, and NRMSE were calculated. A correlation of *rs*=.993 was found, with RMSE=23.04 ms and NRMSE=0.012 ms.

Additionally, a Bland-Altman plot was made, showing the differences between the beat-to-beat intervals of the PPG-ECG

XSL•FO
RenderX

pairs in function of the means. The error distribution and the distribution of the mean duration of the intervals of the PPG-ECG pairs are visualized by kernel density plots (Figure 9). The mean bias is 0.26 (23.045) with a 95% CI from −45.82 to 46.35. The CI (μ±1.96σ) is visualized by the dashed lines. An in-depth analysis was performed to investigate differences within the study results. This detailed analysis was based on two categories: low versus high HR and regular versus irregular intervals.

On the basis of the definition by Laskowski of a resting HR [20] a distinction was made between a resting HR (40-100 BPM) and high HR (100-170 BPM). Figure 10 visualizes and Table 6 describes the study results for this distinction.

Table 6. Summary of intervals divided in low, high, and overall heart rate.

Variable	Interval 40-100 heart rate		Interval 100-170 heart rate	
	ECG[a]	PPG[b]	ECG	PPG
Number of intervals	13,913	13,913	6385	6385
Average interval (ms)	869.5 (203.2)	868.9 (200)	516.4 (70.3)	516.9 (66.6)
Minimum value (ms)	601.6	483.3	312.5	316.7
Maximum value (ms)	2223.0	2233.0	597.7	1000

[a]ECG: electrocardiography.

[b]PPG: photoplethysmography.

No significant difference was observed between both techniques within the interval 40-100 (P=.76) or interval 100-170 (P=.69). Correlation of interval 40-100 between ECG and PPG was strong (rs=.985; RMSE=25.32 ms and NRMSE=0.014). Interval 100-170 was also strongly correlated (rs=.956; RMSE=17.06 ms and NRMSE=0.025). The correlation between both techniques is plotted in Figure 11.

A last step in analysis was performed by investigating the differences between regular and irregular intervals. Figure 12 visualizes the measurements divided into irregular and regular intervals.

Table 7 describes the study results for this category. A total of 2648 intervals were obtained from patients with AF versus 17649 from patients with regular heart rhythms.

Table 7. Summary of intervals divided into regular and irregular beat-to-beats.

Variable	Regular		Irregular	
	ECG[a]	PPG[b]	ECG	PPG
Number of intervals	17,649	17,649	2648	2648
Average interval (ms)	738.9 (204.8)	738.6 (205.3)	888.5 (241.4)	888 (241)
Minimum value (ms)	312.5	316.6	406.3	400
Maximum value (ms)	1835.9	1850	2222.6	2233.3

[a]ECG: electrocardiography.

[b]PPG: photoplethysmography.

No significant difference was observed between both techniques within group regular HR (P=.92) or group AF (P=.93). Correlation for group regular between ECG and PPG was strong (rs=.994; RMSE=20.49 ms and NRMSE=0.013). Group irregular was also strongly correlated (rs=.9832; RMSE=37.62 ms and NRMSE=0.021). The correlation between both techniques is plotted in Figure 13.

XSL•FO
RenderX

Figure 9. Bland-Altman plot comparing the reference R-R intervals (electrocardiography [ECG]) to the peak-to-peak intervals (photoplethysmography, PPG).

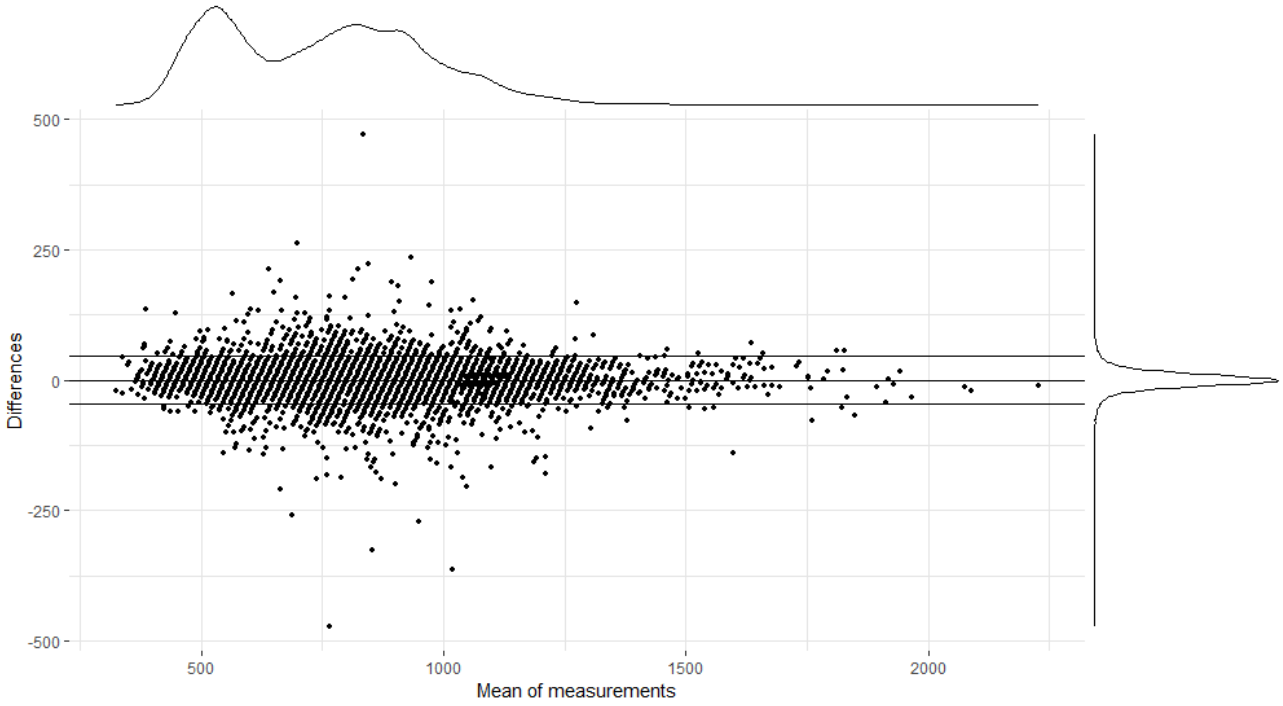
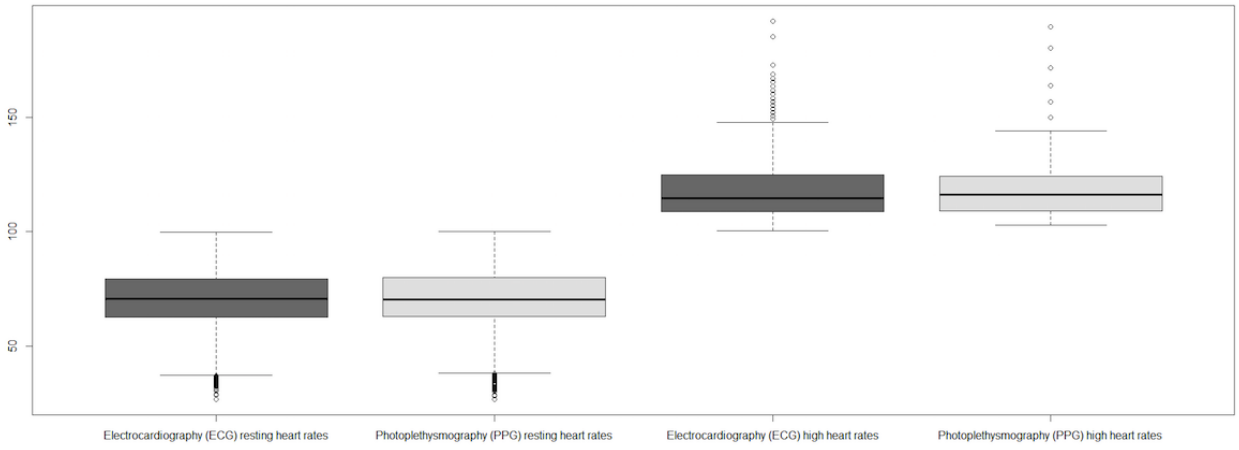


Figure 10. Overview heart rates divided in resting and high heart rates.



XSL•FO
RenderX

Figure 11. Correlation of intervals for both techniques (ie, electrocardiography [ECG] and photoplethysmography [PPG]) in milliseconds. Gray, high heart rates; Black, low heart rates.

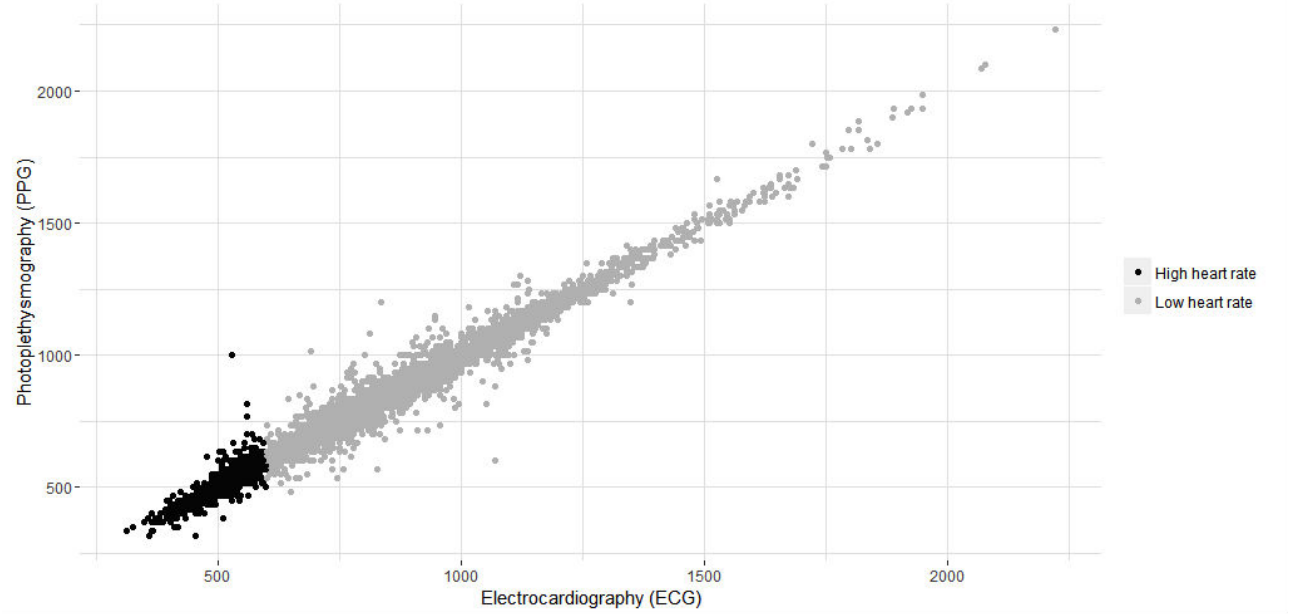
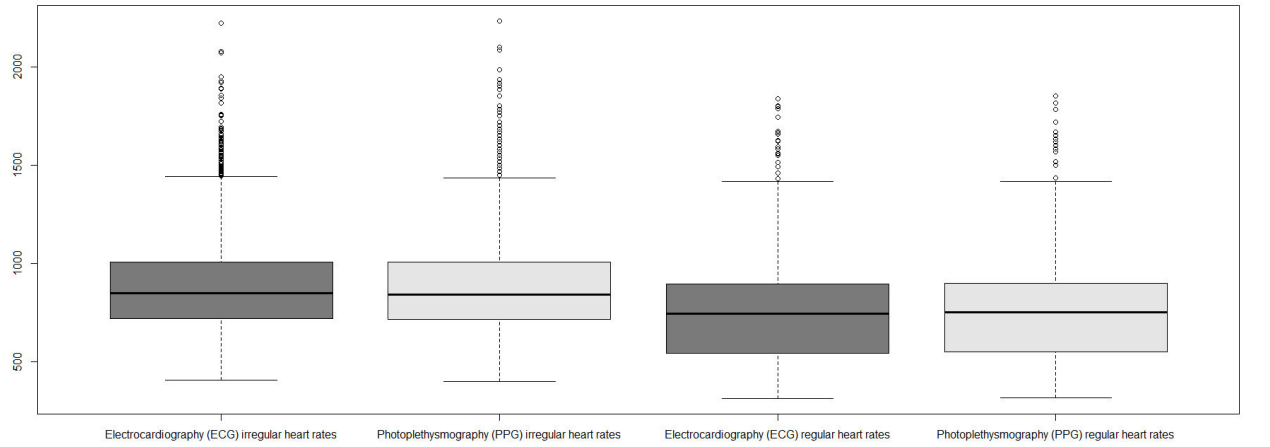
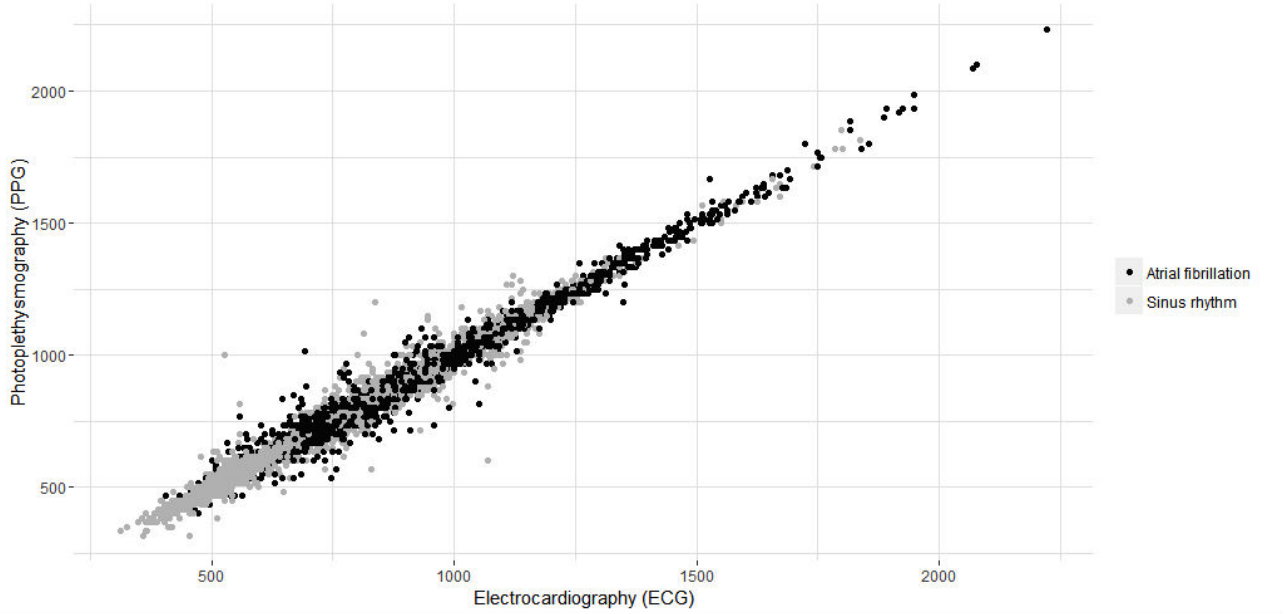


Figure 12. Overview of beat-to-beat intervals divided in irregular and regular beats.



XSL•FO
RenderX

Figure 13. Correlation of intervals for both techniques (ie, electrocardiography [ECG] and photoplethysmography [PPG]) in milliseconds. Gray, irregular intervals; Black, regular intervals.



Discussion

Principal Findings

We sought to determine an approach to validate an HR-measuring app. For this experiment, we set up two different studies for determining the correct approach to answer the research question. The results were interpreted on the criterion validity (demonstrated by statistical test for a high correlation between new tool and the existing standard) and construct validity (refers to the systematic change in results when the input variable is under varying conditions) as described by Franko [21].

Study 1, FibriCheck compared with FDA approved HR devices, compared 3 tools for measuring HRs in a large sample of volunteers. The tools (Nonin and AliveCor) are approved by the FDA and are already used in clinical practice. The third one is the FibriCheck app. The results of the study, for criterion validity, show a correlation coefficient of .834 between FibriCheck and Nonin, .88 between FibriCheck and AliveCor, and .897 between Nonin and AliveCor. A single way ANOVA, P=.61 was executed to construct validity indicating that there is no significant difference between the HRs as measured by the 3 devices.

Study 2, FibriCheck beat-to-beat accuracy compared with wearable ECG, compared the RRI-PPI intervals at the same moment from the FibriCheck app in relation to the data of a wearable ECG. The results of the study show a positive correlation of .993 between RRIs and PPIs. This result supports the validity criteria. For construct validity, no significant difference (P=.92) was shown between the intervals from FibriCheck and the intervals from the wearable ECG.

Terbizan et al [19] suggested a minimum correlation of .9 for heart monitors to be clinically reliable. On the basis of the measured results in study 1, no pair of devices complies with this correlation. Terbizan et al suggest to interpret the device as “not reliable.” This is contradictory because both AliveCor and Nonin have an FDA approval. Bland-Altman plots showed some outliers between the devices. In this study, outliers need to be included in the dataset because of the legitimate character of the observation.

A “not reliable” correlation could have multiple causes. For example, there are device-related (eg, different hardware) causes that could influence the signal of the measurement. Furthermore, algorithms converting the PPG signal into HR measurements differ between manufacturers, including in the way they cope with nonperfect measurements. Therefore, when the captured PPG signal is incomplete, for example, because of vibrations or movement by the finger, resulting HR measurements can differ between HR apps and monitors, even when the raw data are identical.

These differing results can be assessed by running the algorithms on a reference database such as the MIT-BIH arrhythmia database for ECG records [22].

In addition, it is important to consider device specifications when evaluating an HR app on the smartphone. The app therefore needs to be validated on a smartphone with minimal device requirements. Smartphones with lower system specifications than required could result in “‘not reliable” results of the app. It is important for the manufacturer of that HR app to ensure the minimal hardware requirements of hardware. This creates the obligation for manufacturers to evaluate apps on multiple smartphones.

Besides possible hardware and algorithm explanations, there could be time-related causes (eg, measurement on different time) that could result in physiological changes causing a change in HR. This could be eliminated by doing synchronous measurements with these devices.

Another explanation could be that taking average of the minimum and maximum HR during a 10-s interval is not the optimal procedure to obtain a reading from these devices. The

XSL•FO
RenderX

FibriCheck app gives a single result after a 10-s measurement, whereas the Nonin oximeter gives a continuous reading and the AliveCor a minimum and maximum HR result after 10 s. To address this mismatch, the average of the minimum and maximum HR of a 10-s reading was used in case of the Nonin oximeter and AliveCor.

Further research should be conducted to investigate whether there is a stronger correlation between Nonin and AliveCor than the current results suggest; some suggestions are given below.

Related to the possible time- and analytics-related causes of this result, the next step in validation was performed. Experiment 2 for the beat-to-beat detection between the FibriCheck and an ECG device was set up.

Study 2 shows a positive correlation result of .993, an RMSE of 23.04 ms, and an NRMS of 0.012 for the intervals of the FibriCheck app and ECG device. This means that both methods are almost identical. This result suggests that the FibriCheck app could be used as a clinically validated app for measuring HR. The protocol of study 2 confirms the research question of an approach to validate an HR-measuring smartphone app.

Study Limitations

Although the results of this study are encouraging, there are a number of limitations to the study that could be taken into account for further research. First of all, the sample comprised both healthy and unhealthy volunteers who were recruited in a hospital setting and in general practice. However, this means that the sample may not be representative of the general population outside the hospital and general practice that could benefit from the smartphone app.

Measurements of the PPG signal could result in multiple limitations. For example, people with small or calloused fingertips may not be suitable for the detection of a PPG signal measured by the smartphone app. They will have inaccurate HR measurements because of problems with light absorption, on which the PPG principle is based. Additionally, patients with poor blood circulation can also show bad signals. Besides physical factors, environmental factors should also be taken into account. For example, ambient temperature has an influence on the blood circulation in the fingertips.

Study 1 coped with some specific limitations based on the study protocol. Due to the need to use both hands for the AliveCor and FibriCheck app, HRs from the 3 different devices were measured sequentially, leading to time intervals of about 30 s to 1 min between the different measurements. This could cause small changes in HR because of small physiological changes. Another limitation of nonsimultaneous HR measurement could be a learning effect for using a mobile HR app. This learning effect could result in a (small) decrease in HR.

There were a number of measurements that fell outside the LLA and ULA. These deviations could be caused by several factors compromising an optimal reading. For example, it could be possible that pressing too hard on the smartphone's camera impairs the possibility of a good PPG measurement.

For the Nonin oximeter, an incorrect positioning of the device on the finger could hamper a correct reading. Further research could assess whether incorrect usage of these devices can cause deviant HR readings and how to optimally instruct people to avoid these errors.

Conclusions

Smartphones with multimedia capabilities open new possibilities for app development and service delivery [2]. In the last decades, smartphone apps that measure different vital parameters, such as HR, were developed. At this time, apps that measure the HR of a subject can be installed on a variety of smartphones [10]. However, the validity of these apps has not always been confirmed. This paper describes an approach for the clinical validation of an HR app. The current findings suggest that the most suitable method for the validation of an HR app is a simultaneous measurement of the HR by the smartphone app and an ECG system and comparing the obtained intervals. This approach could lead to almost exact accuracy in the clinical setting. Further studies are needed to evaluate the accuracy outside the hospital and in daily life of subjects.

Conflicts of Interest

None declared.

References

1. Techcrunch. Smartphone users globally by 2020, overtaking basic fixed phone subscriptions 2015 URL: https://techcrunch.com/2015/06/02/6-1b-smartphone-users-globally-by-2020-overtaking-basic-fixed-phone-subscriptions/ [WebCite Cache ID 6si3uHw3X]
2. Pelegris P, Banitsas K, Orbach T, Marias K. A novel method to detect heart beat rate using a mobile phone. 2010 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 31, 2010; United States. [doi: 10.1109/IEMBS.2010.5626580]
3. 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]
4. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. J Med Internet Res 2016 Sep 20;18(9):e253 [FREE Full text] [doi: 10.2196/jmir.6025] [Medline: 27651304]

XSL•FO
RenderX

5. Gregoski MJ, Mueller M, Vertegel A, Shaporev A, Jackson BB, Frenzel RM, et al. Development and validation of a smartphone heart rate acquisition application for health promotion and wellness telehealth applications. Int J Telemed Appl 2012;2012:696324 [FREE Full text] [doi: 10.1155/2012/696324] [Medline: 22272197]
6. Politi MT, Ghigo A, Fernández JM, Khelifa I, Gaudric J, Fullana JM, et al. The dicrotic notch analyzed by a numerical model. Comput Biol Med 2016 May 01;72:54-64. [doi: 10.1016/j.compbiomed.2016.03.005] [Medline: 27016670]
7. Friberg L, Engdahl J, Frykman V, Svennberg E, Levin L, Rosenqvist M. Population screening of 75- and 76-year-old men and women for silent atrial fibrillation (STROKESTOP). Europace 2013 Jan;15(1):135-140. [doi: 10.1093/europace/eus217] [Medline: 22791299]
8. Couetil JP. [Studies of mitral valve insufficiency and indications for surgery]. Soins Chir 1987 Apr(74):5-8. [doi: 10.1186/1475-925X-12-19] [Medline: 3649882]
9. Pinheiro N, Couceiro R, Henriques J, Muehlsteff J, Quintal I, Gonçalves L. Can PPG be used for HRV analysis? 2016 Aug Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) Cite this publication; August 2016; United States. [doi: 10.1109/EMBC.2016.7591347]
10. Elgendi M. On the analysis of fingertip photoplethysmogram signals. Curr Cardiol Rev 2012 Feb;8(1):14-25 [FREE Full text] [Medline: 22845812]
11. Hertzman A. Maxpulsecordurohealth. 1938. The blood supply of various skin areas as estimated by the photoelectric plethysmograph URL: http://www.maxpulsecordurohealth.com/files/skinBloodSupply.pdf [accessed 2017-08-21] [WebCite Cache ID 6ssdbyFBg]
12. Jonathan E, Leahy M. Investigating a smartphone imaging unit for photoplethysmography. Physiol Meas 2010 Nov;31(11):N79-N83. [doi: 10.1088/0967-3334/31/11/N01] [Medline: 20871134]
13. Ho C, Fu Y, Lin M, Chan S, Hwang B, Jan S. Smartphone applications (apps) for heart rate measurement in children: comparison with electrocardiography monitor. Pediatr Cardiol 2014 Apr;35(4):726-731. [doi: 10.1007/s00246-013-0844-8] [Medline: 24259012]
14. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med 2016 Sep;15(3):540-547 [FREE Full text] [Medline: 27803634]
15. Giles D, Draper N, Neil W. Validity of the polar V800 heart rate monitor to measure RR intervals at rest. Eur J Appl Physiol 2016 Mar;116(3):563-571 [FREE Full text] [doi: 10.1007/s00421-015-3303-9] [Medline: 26708360]
16. Gamelin FX, Berthoin S, Bosquet L. Validity of the polar S810 heart rate monitor to measure R-R intervals at rest. Med Sci Sports Exerc 2006 May;38(5):887-893. [doi: 10.1249/01.mss.0000218135.79476.9c] [Medline: 16672842]
17. Selvaraj N, Jaryal A, Santhosh J, Deepak KK, Anand S. Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. J Med Eng Technol 2008;32(6):479-484. [doi: 10.1080/03091900701781317] [Medline: 18663635]
18. Torfs T, Smeets CJ, Geng D, Berset T, Van DA, Vandervoort P, et al. Clinical validation of a low-power and wearable ECG patch for long term full-disclosure monitoring. J Electrocardiol 2014;47(6):881-889. [doi: 10.1016/j.jelectrocard.2014.08.012] [Medline: 25201418]
19. Terbizan DJ, Dolezal BA, Albano C. Validity of seven commercially available heart rate monitors. Meas Phys Educ Exerc Sci 2002 Dec;6(4):243-247. [doi: 10.1207/S15327841MPEE0604_3]
20. Laskowski ER. Mayoclinic. 2016. What's a normal resting heart rate? URL: http://www.mayoclinic.org/healthy-lifestyle/fitness/expert-answers/heart-rate/faq-20057979 [accessed 2017-08-14] [WebCite Cache ID 6si60rCBl]
21. Franko. Validate an app: how to design your study and get published. Journal MTM 2012 Jun 01;1(2):1-4. [doi: 10.7309/jmtm.9]
22. Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Eng Med Biol Mag 2001;20(3):45-50. [doi: 10.1109/51.932724] [Medline: 11446209]

Abbreviations

ABP: arterial blood pressure
AF: atrial fibrillation
ANOVA: analysis of variance
BPM: beats per minute
ECG: electrocardiography
FDA: Food and Drug Administration
GP: general practitioner
HR: heart rate
HRV: heart rate variability
IQR: interquartile range
LED: light-emitting diode
LLA: lower limit of agreement
NRMSE: normalized root mean square error

XSL•FO
RenderX

PPI: peak-to-peak interval
PPG: photoplethysmography
RGB: red, green, and blue
RMSE: root mean square error
RRI: R-R intervals
SD: standard deviation
ULA: upper limit of agreement

Edited by G Eysenbach; submitted 03.01.17; peer-reviewed by Y Ma, B Yan; comments to author 04.02.17; revised version received 20.03.17; accepted 21.07.17; published 25.08.17.

Please cite as:
Vandenberk T, Stans J, Mortelmans C, Van Haelst R, Van Schelvergem G, Pelckmans C, Smeets CJP, Lanssens D, De Cannière H, Storms V, Thijs IM, Vaes B, Vandervoort PM
Clinical Validation of Heart Rate Apps: Mixed-Methods Evaluation Study
JMIR Mhealth Uhealth 2017;5(8):e129
URL: http://mhealth.jmir.org/2017/8/e129/
doi:10.2196/mhealth.7254
PMID:

©Thijs Vandenberk, Jelle Stans, Christophe Mortelmans, Ruth Van Haelst, Gertjan Van Schelvergem, Caroline Pelckmans, Christophe JP Smeets, Dorien Lanssens, Hélène De Cannière, Valerie Storms, Inge M Thijs, Bert Vaes, Pieter M Vandervoort. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 25.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

How Do Apps Work? An Analysis of Physical Activity App Users' Perceptions of Behavior Change Mechanisms

Taylor H Hoj[1], BS; Emarie L Covey[1], BS; Allyn C Jones[1], BA; Amanda C Haines[1], BS; P Cougar Hall[1], PhD; Benjamin T Crookston[1], MPH, PhD; Joshua H West[1], MPH, PhD

Computational Health Science Research Group, Department of Health Science, Brigham Young University, Provo, UT, United States

Corresponding Author:
Joshua H West, MPH, PhD
Computational Health Science Research Group
Department of Health Science
Brigham Young University
2139 Life Sciences Building
Provo, UT, 84602
United States
Phone: 1 801 422 3444
Fax: 1 801 422 0273
Email: josh.west@byu.edu

Abstract

Background: Physical activity apps are commonly used to increase levels of activity and health status. To date, the focus of research has been to determine the potential of apps to influence behavior, to ascertain the efficacy of a limited number of apps to change behavior, and to identify the characteristics of apps that users prefer.

Objective: The purpose of this study was to identify the mechanisms by which the use of physical activity apps may influence the users' physical activity behavior.

Methods: This study used a cross-sectional survey of users of health-related physical activity apps during the past 6 months. An electronic survey was created in Qualtrics' Web-based survey software and deployed on Amazon Mechanical Turk. Individuals who had used at least one physical activity app in the past 6 months were eligible to respond. The final sample comprised 207 adults living in the United States. 86.0% (178/207) of respondents were between the ages of 26 and 54 years, with 51.2% (106/207) of respondents being female. Behavior change theory informed the creation of 20 survey items relating to the mechanisms of behavior change. Respondents also reported about engagement with the apps, app likeability, and physical activity behavior.

Results: Respondents reported that using a physical activity app in the past 6 months resulted in a change in their attitudes, beliefs, perceptions, and motivation. Engagement with the app (P<.001), frequency of app use (P=.03), and app price (P=.01) were related to the reported impact of the behavior change theory or mechanisms of change. The mechanisms of change were associated with self-reported physical activity behaviors (P<.001).

Conclusions: The findings from this study provide an overview of the mechanisms by which apps may impact behavior. App developers may wish to incorporate these mechanisms in an effort to increase impact. Practitioners should consider the extent to which behavior change theory is integrated into a particular app when they consider making recommendations to others wishing to increase levels of physical activity.

(JMIR Mhealth Uhealth 2017;5(8):e114) doi:10.2196/mhealth.7206

KEYWORDS

mHealth; mobile apps; health behavior; smartphone

Introduction

Regular physical activity is effective in primary and secondary prevention of chronic diseases, including those relating to both physical and mental health outcomes [1]. It is associated with cardiovascular and metabolic health, reduced body mass, decreased risk of type 2 diabetes, optimal bone health, and strong motor control and physical functioning [2]. Obesity-related conditions, including heart disease, stroke, and type 2 diabetes, remain the leading causes of preventable death

XSL•FO
RenderX

[3]. Consistently engaging in physical activity has also been linked to improved mental health status, as well as increased cognitive functioning and academic performance [4,5].

Despite clear and consistent guidelines establishing standards for physical activity, only half of the adults worldwide meet the recommended levels [6]. Inactivity is much higher in the United States where nearly 80% of adults fail to get the recommended amount of physical activity each week [7]. In light of the low global prevalence of physical activity and the strong connection to health outcomes, the current paradigm could rightly be considered a pandemic [8]. Promoting physical activity will require a sustained effort across multiple domains and disciplines, including behavioral science [8]. Efforts to date have included environment and policy changes [9], school and worksite interventions [10,11], and technology solutions [12-14]. Moving forward, our public health approaches may benefit from increased innovation and creativity [8], including an emphasis on mobile technologies with increased capacities for delivering timely and adapted promotion [15].

Smartphone apps have emerged as a potential tool for individuals seeking to increase levels of physical activity in an effort to improve health status [13-15]. Indeed, tens of thousands of health apps are available across all platforms and have now been downloaded billions of times [16]. The results from studies of these tools appear promising [17,18]. However, little is known about the behavior change techniques included in most physical activity apps [19]. It's one thing to measure and observe that an app may increase levels of physical activity, but it's another to understand the mechanisms to explain such effects [20]. These mechanisms may involve techniques inspired by behavior change theory, an accepted approach for increasing the effectiveness of physical activity interventions [8,21-23].

Despite the wide use of health-related physical activity apps by millions of Americans [24], only a few studies have considered the effectiveness of these apps at impacting the factors known to influence changes in behavior. Health behavior theory can inform behavior change interventions. The health belief model, for example, contains several primary constructs used to predict whether and why people take action to prevent, detect, or control disease outcomes. These constructs include perceived susceptibility, perceived severity, perceived benefits, and barriers to engaging in a behavior. Also addressed are cues to action, meaning internal or external factors that could trigger health behavior [25]. Social cognitive theory elucidates the interaction and impact of environment as it relates to influencing health behaviors. Cognitive influences on behavior (eg, knowledge, self-efficacy, and outcome expectancies) as well as the physical and social environment are core constructs [26]. The theory of planned behavior represents an attempt to predict behavior through impacting behavioral intentions. This is accomplished by changing attitudes, subjective norms, and self-efficacy [26].

The focus of several recent studies has been to analyze the content of physical activity apps and identify the extent to which behavior change theory is integrated [13,15,27-30]. Studies of this nature have provided an initial framework and context by which app potential can be evaluated. However, such studies are limited in that they only evaluate potential rather than actual impact, and there is no indication as to the role of these theory-based mechanisms of behavior change. For example, prominent theories such as the social cognitive theory and the health belief model emphasize the self-efficacy construct. Although attempts to label its inclusion in a physical activity app have been made recently, no empirical effort has been made to identify whether, in an app environment, it can be impacted and whether that can result in a change in physical activity-related behaviors. Constructs from the transtheoretical model [31-33] and the theory of planned behavior [34] also inform mechanisms whose impact should be measured, as each of these have been successfully used previously to influence physical activity. Provided the complexities relating to human behavior and in the context of physical activity, it may be that no one single theory adequately accounts for the most influential determinants. Therefore, a practical approach may involve a combination of distinct constructs and elements from each theory, effectively forming a polytheoretical approach [35]. The purpose of this study was to, first, identify whether and which behavior change mechanisms are impacted by using a physical activity app, and second, whether changes in these mechanisms also relate to physical activity behaviors.

Methods

Design

This study involved a cross-sectional survey designed to analyze the self-reported impact of using a physical activity app on the mechanisms of behavior change. This was done through a survey directed to individuals who had utilized at least one physical activity app within the last 6 months. The survey gathered information regarding mechanisms of behavior change informed by behavior change theory, app engagement and likeability, app usage, and self-reported physical activity behavior.

Recruitment

The study sample comprised respondents who were recruited through Amazon Mechanical Turk (MTurk) and Turk Prime. MTurk is a crowdsourcing Internet marketplace that enables individuals and businesses to coordinate the completion of tasks, of which surveys are a common variety.

The sample was limited to respondents who were 18 years of age or older and residents of the United States. A total of 251 respondents completed the survey. The results for 10 respondents were excluded from the final sample because they reported not having used a health-related physical activity app in the past 6 months. Additionally, only respondents who completed all of the survey items were included in the final sample, which included a total of 207 respondents.

Procedure

An electronic survey constructed through Qualtrics' Web-based survey software was used to collect data through MTurk. The survey was available in MTurk for approximately 2 weeks with a US $1 incentive. The incentive was increased to US $2 after 2 weeks, and the survey was relaunched with Turk Prime to improve the response rate. As part of relaunching the survey using Turk Prime, an authenticator was built into the survey to

XSL•FO
RenderX

prevent repeat respondents. Compensation was entirely based upon completing the survey and not on the quality of the responses.

Measurement

Demographic information was gathered and respondents were asked to report their age, race, ethnicity, sex, highest level of education obtained, and annual household income (Table 1). Additional information relating to app usage was also gathered. This included the number of physical activity apps respondents had used in the past 6 months, how often each app was used, which apps they used most frequently, and the average price of the apps that they used. The average survey completion time was 5 min and 48 sec.

Respondents gave responses to a series of questions designed to measure the app's impact on the mechanisms of behavior change. The questions were based on three prominent behavior change theories: social cognitive theory, the theory of planned behavior, and the health belief model. Items were developed to measure specific constructs within these theories. For example, "My belief that physical inactivity leads to disease" is a reflection of outcome expectancies, a fundamental construct of social cognitive theory. A full list of items and their corresponding theories are displayed in Table 2. These items were adapted from Cowan et al [13]. The adaptations were done to make them applicable for users' perceptions of the apps' impact. The items were then pilot-tested with 36 individuals. The pilot test comprised completing the survey in its entirety and then discussing each item with the pilot test takers to determine that each item was clearly written, understandable, and interpreted in the way it was intended.

A standard, 5-point Likert response scale was used to measure these items, ranging from Strongly disagree (−2) to Strongly agree (+2). A composite theory variable was constructed by summing the values of these 20 items, and the Cronbach alpha of this variable was .931. This variable provided a global assessment of the extent to which the apps impacted constructs believed to be important in influencing behavior change. A polytheoretical measure was determined to be in line with the viewpoint that multifactorial behaviors are too complex for any one single theory and may be best addressed with multiple theories [35]. The composite variable constructed in this study was not normally distributed, and a square root transformation was used to normalize it.

Five items related to the likeability and engagement (actual items displayed in Table 3) of the apps were also assessed using the same Likert scale. A composite variable was constructed to provide an assessment of the extent to which respondents liked and engaged with the apps. The Cronbach alpha for this scaled variable was .890. A square root transformation was used to normalize this scaled variable. Respondents were also asked to report about physical activity behaviors (actual items displayed in Table 4) that changed as a result of using the apps. These items were adapted from a recent study of physical activity apps [36]. The same Likert scale was used to measure physical activity. A composite variable was constructed with these 4 items, and the Cronbach alpha of this variable was .854. Again, a square root transformation was used to normalize this composite variable.

Statistical Analysis

Stata version 14 was used to calculate all statistics. Descriptive statistics were calculated for each of the demographic, app usage, theory, engagement, and behavior variables. Multiple regression analysis was used to identify factors associated with behavior change theory constructs as well as with physical activity behavior, after controlling for potentially confounding variables.

Results

Demographics

The majority of respondents were between the ages of 26 and 54 years, with 45.9% (95/207) reporting their age between 26 and 34 years, and 40.1% (83/207) reporting their ages between 35 and 54 years (Table 1). Concerning race and ethnicity, 82.1% (170/207) of respondents reported being white and 94.2% (195/207) of respondents reported being of a non-Hispanic/Latino ethnicity. Females comprised 51.2 % (106/207) of respondents. Whereas the levels of education varied from less than a high school education to a professional degree, 63.8% (132/207) of respondents had either a 4-year degree or some college (not graduated) education. When asked about the number of physical activity apps used in the past 6 months, 60.9% (126/207) of respondents reported using only 1 physical activity app, whereas 29.0% (60/207) reported using 2 physical activity apps. Regarding frequency of physical activity app use in the past 6 months, 41.0% (85/207) of respondents reported using the apps daily, whereas 48.3% (100/207) of respondents reported using the apps multiple times a week. The most commonly used apps as reported by study respondents were Fitbit and MyFitnessPal, with 22.2% (46/207) of respondents reporting using Fitbit and 17.4 % (36/207) of respondents reporting using MyFitnessPal.

Responses

A majority of respondents (58.0%, 120/207) reported "Strongly agree" that using the apps increased their desire to be healthy (Table 2). Similarly, 56.0% (116/207) strongly agreed that the apps increased their desire to be physically active. Respondents reported similar "Strongly agree" response rates for increased motivation, intention, goal setting desire, and ability to be physically active as a result of app use. A minority of respondents strongly agreed that the apps increased their belief that people important to them want them to be physically active (21.7%, 45/207), their knowledge of the diseases that are caused by physical inactivity (22.2%, 46/207), and their belief that physical inactivity leads to disease (24.6%, 51/207).

XSL•FO
RenderX

Table 1. Demographics (N=207).

Demographics	n (%)
Age, in years	
18-25	17 (8.2)
26-34	95 (45.9)
35-54	83 (40.1)
55-64	11 (5.3)
65 or older	1 (0.5)
Race	
American Indian or Alaska Native	1 (0.5)
Asian	16 (7.7)
Black or African American	18 (8.7)
Native Hawaiian or Other Pacific Islander	2 (1.0)
White	170 (82.1)
Ethnicity	
Hispanic/Latino	12 (5.8)
Non-Hispanic/Latino	195 (94.2)
Gender	
Male	101 (48.8)
Female	106 (51.2)
Education level	
Less than high school	2 (1.0)
High school/General educational development	27 (13.0)
Some college (not graduated)	56 (27.1)
2-year college degree	28 (13.5)
4-year college degree	76 (36.7)
Master’s degree	15 (7.3)
Professional degree (JD, MD, etc)	3 (1.5)
Region of residence in the United States of America	
Northeast	36 (17.4)
Midwest	34 (16.4)
South	93 (44.9)
West	44 (21.7)
Annual household income[a]	
Less than 30,000	47 (22.7)
30,000-39,999	39 (18.8)
40,000-49,999	23 (11.1)
50,000-59,999	28 (13.5)
60,000-69,999	15 (7.3)
70,000-79,999	19 (9.2)
80,000-89,999	12 (5.8)
90,000-99,999	6 (2.9)
100,000 or more	18 (8.7)

XSL•FO
RenderX

[a]All values are in 2016 US dollars.

Table 2. Responses to behavior change constructs (N=207). A composite behavior theory variable was computed by summing these variables, Cronbach alpha=.931.

Construct or mechanism of change[a]	n (%)				
	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
My belief that physical inactivity leads to disease (outcome expectations)[b]	11 (5.3)	39 (18.8)	32 (15.5)	74 (35.8)	51 (24.6)
My belief that diseases related to physical inactivity are harmful (outcome expectancies)[b]	8 (3.9)	26 (12.6)	36 (17.4)	65 (31.4)	72 (34.8)
My belief that being physically active can prevent disease (behavioral belief)[c]	2 (1.0)	14 (6.8)	27 (13.0)	85 (41.1)	85 (41.1)
My belief that physical activity is important in preventing disease (behavioral belief)[c]	2 (1.0)	15 (7.3)	24 (11.6)	86 (41.6)	80 (38.7)
My ability to be physically active (self-efficacy)[b]	1 (0.5)	8 (3.9)	14 (6.8)	80 (38.7)	104 (50.2)
My confidence that I can be physically active (self-efficacy)[b]	1 (0.5)	5 (2.4)	8 (3.7)	104 (50.2)	89 (43.0)
My motivation to be physically active (behavioral attitudes)[c]	0 (0)	3 (1.5)	10 (4.8)	79 (4.8)	115 (55.6)
My desire to be physically active (behavioral attitudes)[c]	0 (0)	1 (0.5)	17 (8.2)	73 (35.3)	116 (56.0)
My intentions to be physically active (behavioral intention)[c]	0 (0)	1 (0.5)	13 (6.3)	81 (39.1)	112 (54.1)
My attitudes about the importance of physical activity in preventing disease (behavioral attitudes)[c]	1 (0.5)	13 (6.3)	25 (12.1)	87 (42.0)	81 (39.1)
My belief that people important to me want me to be physically active (subjective norm)[c]	9 (4.4)	31 (15.0)	50 (24.2)	72 (34.8)	45 (21.7)
My perception that many other people are physically active (situational perception)[b]	8 (3.9)	29 (14.0)	39 (18.8)	71 (34.3)	60 (29.0)
My knowledge of ways in which I can be physically active (knowledge)[b]	2 (1.0)	14 (6.8)	15 (7.3)	95 (45.9)	81 (39.1)
My knowledge of the diseases that are caused by physical inactivity (knowledge)[b]	15 (7.3)	43 (20.8)	36 (17.4)	67 (32.4)	46 (22.2)
My awareness of the benefits of being physically active (perceived benefits)[d]	1 (0.5)	8 (3.9)	22 (10.6)	88 (42.5)	88 (42.5)
My desire to be healthy (behavioral attitudes)[c]	0 (0)	1 (0.5)	12 (5.8)	74 (35.8)	120 (58.0)
The social support I have received for being physically active (reinforcement)[b]	7 (3.4)	35 (16.9)	45 (21.8)	67 (32.4)	53 (35.6)
The positive feedback I have received for being physically active (reinforcement)[b]	7 (3.4)	21 (10.1)	38 (18.4)	80 (38.7)	61 (29.5)
My desire to set goals to be physically active (attitude toward behavior)[b]	0 (0)	1 (0.5)	10 (4.8)	87 (42.0)	109 (52.7)
My ability to achieve my physical activity goals (self-efficacy)[b]	1 (0.5)	3 (1.5)	11 (5.3)	91 (44.0)	101 (48.8)

[a]All theory questions in the survey were preceded by this statement: “Now think about the physical activity/exercise apps that you have used in the past 6 months. Using the apps has increased”:

[b]Social cognitive theory.

[c]Theory of planned behavior.

[d]Health belief model.

XSL•FO
RenderX

Table 3. Responses to likeability and engagement items (N=207). A composite engagement variable was computed by summing these variables, Cronbach alpha=.890.

Item[a]	n (%)				
	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
The app was useful.	0 (0)	1 (0.5)	4 (1.9)	67 (32.4)	135 (65.2)
The app was easy to use.	0 (0)	1 (0.5)	3 (1.5)	64 (30.9)	139 (67.2)
I enjoyed using the app.	0 (0)	2 (1.0)	13 (6.3)	68 (32.9)	124 (59.9)
I liked the app.	0 (0)	0 (0)	7 (3.4)	72 (34.8)	128 (61.8)
I would recommend the app to others.	0 (0)	1 (0.5)	5 (2.4)	72 (34.8)	129 (62.3)

[a]All engagement questions in the survey were preceded by this statement: "Considering the apps that you have used in the past 6 months":

Regarding the app likeability and engagement (Table 3), most respondents (65.2%, 135/207, "Strongly agree" and 32.4%, 67/207, "Somewhat agree") found the physical activity apps "useful." Respondents reported similarly for "The app was easy to use," "I enjoyed using the app," "I liked the app," and "I would recommend the app to others."

More than half of respondents strongly agreed that apps influenced frequency (58.5%, 121/207) and consistency (58.9%, 122/207) of physical activity (Table 4). Whereas 46.4% (96/207) of respondents strongly agreed that the apps helped their actual goal setting to be physically active, 42.0% (87/207) strongly agreed that their intensity of physical activity increased as a result of using the apps.

Regression Analysis

App engagement (P<.001), frequency of app use (P=.03) and app price (P=.01) were all positively associated with health behavior theory (Table 5). Health behavior theory and app engagement were positively associated (P<.001) with physical activity (Table 6).

Table 4. Responses to physical activity behavior items (N=207). A composite behavior change variable was computed by summing these variables, Cronbach alpha=.854.

Item[a]	n (%)				
	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
My actual goal setting to be physically active	1 (0.5)	2 (1.0)	6 (2.9)	102 (49.3)	96 (46.4)
My frequency of physical activity	1 (0.5)	0 (0)	9 (4.4)	76 (36.7)	121 (58.5)
My intensity of physical activity	0 (0)	14 (6.8)	24 (11.6)	82 (39.6)	87 (42.0)
My consistency in being physically active	0 (0)	3 (1.5)	7 (3.4)	75 (36.2)	122 (58.9)

[a]All theory questions in the survey were preceded by this statement: "Now think about the physical activity/exercise apps that you have used in the past 6 months. Using the apps has increased":

Table 5. Regression analysis and behavior change theory (N=207).

Variable	Coefficient (Standard error)	*t*	*P*
App engagement	.23 (0.04)	6.22	<.001
Frequency of app use	.39 (0.17)	2.27	.03
Price	.46 (0.18)	2.54	.01
Age	.05 (0.11)	0.46	.65
Gender	.22 (0.16)	1.31	.19
Education	.01 (0.06)	0.11	.91

XSL•FO
RenderX

Table 6. Regression analysis and physical activity (N=207).

Variable	Coefficient (Standard error)	*t*	*P*
Theory	.21 (0.028)	7.52	<.001
App engagement	.40 (0.074)	5.45	<.001
Frequency of app use	−.01 (0.067)	−0.02	.99
Price	−.01 (0.07)	−0.17	.86
Age	.01 (0.04)	0.27	.79
Gender	−.06 (0.06)	−0.98	.33
Education	.019 (0.023)	0.83	.41

Discussion

Principal Findings

The purpose of this study was to identify mechanisms of behavior change that are impacted by using a physical activity app. Second, we sought to explore the relationship between mechanisms of behavior change and self-reported actual changes in physical activity behaviors. The majority of respondents reported that apps had a favorable impact on their perceptions, attitudes, and beliefs. Physical activity apps certainly resulted in a marked increase in their desire to be healthy and motivation to be physically active. A recent review of app-based interventions reported that the method for increasing motivation to be physically active may include providing prompt and timely feedback to the user [37]. The positive impact on beliefs, attitudes, and perceptions may be useful for prioritizing features in future app-based interventions. These theory-based mechanisms have been shown in nontechnology settings to predict changes in behavior [38], which is corroborated by their impact in this study as well. By contrast, only a few respondents reported an increase in their knowledge of diseases related to physical inactivity as a result of app use. This could be the result of a lack of information available through the apps. This seems somewhat unlikely, provided that the previous research into health and fitness apps has shown that the provision of information is a dominant feature [13,15]. Alternative explanations may include that the respondents already had knowledge regarding the prevalence and effects of these diseases and that using the apps did little to improve upon their preexisting knowledge, or that the respondents were already motivated to be physically active and therefore had little interest in attending to any knowledge-related content. More importantly, it is not clear whether the lack of effect observed in this study is because of the apps truly being ineffective at impacting this aspect of behavior change theory, or whether the apps simply did not address it. A future study could be designed to study this aspect.

An association between the frequency of use and reported impact on the theory-based mechanisms of behavior change was observed. The exact reason for this relationship is not known. Some possible explanations may include the user-friendly nature of the apps that impacted theoretical constructs, increased user motivation, or higher user satisfaction. In a recent study of physical activity app users, respondents valued receiving push prompts and feedback [39], which may be indicators of engagement. Self-reported measures of app engagement and likeability were generally positive in this study, a finding that has also been reported broadly regarding health apps [37]. The relationship between mechanisms of change and self-reported app engagement is a novel finding. It is plausible that engaging apps influenced respondents' perceptions of the overall impact of an app. It may also be the case that engaging with the apps allows for sufficient opportunities for impact. Whatever the reason, it may be of interest to practitioners and developers wishing to use apps to note the importance of engagement. Additionally, strategies to increase the frequency of use of the app should also be considered. Frequency of use continues to be a topic of interest for researchers in this space [40], but it stands to reason that an effective app will be more impactful if it is used more frequently.

Higher priced apps in this study were more likely to have a positive influence on the mechanisms of change, including constructs such as respondents' attitudes, beliefs, and perceptions. A similar finding has also been reported in other studies of health and fitness apps, where apps with a higher cost were evaluated to have greater potential for influencing behavior change [15,41]. There could be several explanations for this correlation. It is possible that paid-for apps provide additional features or have higher quality programming and functionality allowing for greater behavior change theory integration. Additionally, respondents paying for apps may be more dedicated to using the apps because of their financial investment. Respondents who already regularly engage in physical activity may also consider a paid-for app a wise investment, as they will use it often. Similarly, the willingness of respondents to pay for physical activity apps may be representative of a stronger commitment or dedication to engage in physical activity. This finding may have implications for health professionals in that it might be advantageous for practitioners to recommend paid apps, as opposed to those that can be downloaded for free.

This study makes two unique contributions to extant literature. First is the identification of the theory-based mechanisms that are most impacted by using a physical activity app. Second is the connection between the theory-based mechanisms of change and physical activity behavior. The significant findings of the latter provide at least some validation of the major findings of this study. If there had been no association between the mechanisms of change and physical activity behavior, the study findings would be of little practical significance. In light of this finding, future efforts could focus on development-related

XSL•FO
RenderX

questions to determine the most effective way to integrate and impact these theory-based mechanisms of change.

Limitations

The limitations of this study should be considered when interpreting the findings. First, respondents in this study had limited diversity regarding race, ethnicity, education, and income. Respondents in this study were primarily white, educated, and had higher income status. This limitation is likely a reflection of the demographic using mTurk's Web-based surveying system, which also tends to mirror these demographic characteristics [42]. Furthermore, a national survey of health app users revealed that most are young, educated, and have a higher income status [43]. Perhaps a future study could include a more diverse sample. Second, this study would be strengthened by collecting additional respondent data and information. For example, collecting information regarding respondents' body mass index, health status, or motivations for using a physical activity app may have been useful to explore the extent to which such characteristics determine apps' impact. Furthermore, additional validation of the survey is required before using it in new studies. Lastly, respondents' use of apps and the attending impact was not measured longitudinally. Future studies may benefit from qualitative and quantitative designs that explore these relationships in a longitudinal fashion. Despite these limitations, this study represents an initial effort to understand the mechanisms by which physical activity may be impacted and strategies for both future app development and research design.

Conclusions

The purpose of this study was to investigate the mechanisms by which changes in physical activity might occur when using a physical activity app. Findings indicate that increased engagement and use may be related to mechanisms of behavior change informed by behavior change theory. Furthermore, these mechanisms of change appear to be related to physical activity. Those wishing to develop physical activity apps may consider ways to integrate these mechanisms of change. Additionally, practitioners in search of apps for recommendation to improve physical activity behaviors should consider apps with an emphasis on these theory-informed mechanisms with more confidence, as they may be more likely to result in behavior change.

Conflicts of Interest

None declared.

References

1. 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]
2. Surgeon General. Washington, DC: U.S. Department of Health and Human Services; 2015. Step It Up! The Surgeon General's Call to Action to Promote Walking and Walkable Communities URL: https://www.surgeongeneral.gov/library/calls/walking-and-walkable-communities/call-to-action-walking-and-walkable-communites.pdf [accessed 2016-12-21] [WebCite Cache ID 6muomzjvf]
3. Bauer UE, Briss PA, Goodman RA, Bowman BA. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. Lancet 2014 Jul 5;384(9937):45-52. [doi: 10.1016/S0140-6736(14)60648-6] [Medline: 24996589]
4. Jackson WM, Davis N, Sands SA, Whittington RA, Sun LS. Physical activity and cognitive development: a meta-analysis. J Neurosurg Anesthesiol 2016 Oct;28(4):373-380. [doi: 10.1097/ANA.0000000000000349] [Medline: 27768674]
5. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010 May 11;7:40 [FREE Full text] [doi: 10.1186/1479-5868-7-40] [Medline: 20459784]
6. Bauman A, Bull F, Chey T, Craig CL, Ainsworth BE, Sallis JF, The IPS Group. The international prevalence study on physical activity: results from 20 countries. Int J Behav Nutr Phys Act 2009 Mar;6:21 [FREE Full text] [doi: 10.1186/1479-5868-6-21] [Medline: 19335883]
7. Ogden CL, Carroll MD, Fryar CD, Flegal KM. Prevalence of obesity among adults and youth: United States, 2011-2014. NCHS Data Brief 2015 Nov(219):1-8 [FREE Full text] [Medline: 26633046]
8. Kohl 3rd HW, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, Lancet Physical Activity Series Working Group. 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]
9. Sallis JF, Bauman A, Pratt M. Environmental and policy interventions to promote physical activity. Am J Prev Med 1998 Nov;15(4):379-397. [Medline: 9838979]
10. Wechsler H, Devereaux RS, Davis M, Collins J. Using the school environment to promote physical activity and healthy eating. Prev Med 2000 Aug;31(2):S121-S137. [doi: 10.1006/pmed.2000.0649]
11. Dishman RK, Oldenburg B, O'Neal H, Shephard RJ. Worksite physical activity interventions. Am J Prev Med 1998 Nov;15(4):344-361. [Medline: 9838977]

XSL•FO
RenderX

12. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007 Apr 27;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
13. Cowan LT, Van Wagenen SA, Brown BA, Hedin RJ, Seino-Stephan Y, Hall PC, et al. Apps of steel: are exercise apps providing consumers with realistic expectations?: a content analysis of exercise apps for presence of behavior change theory. Health Educ Behav 2013 Apr;40(2):133-139. [doi: 10.1177/1090198112452126] [Medline: 22991048]
14. Doshi A, Patrick K, Sallis JF, Calfas K. Evaluation of physical activity web sites for use of behavior change theories. Ann Behav Med 2003;25(2):105-111. [Medline: 12704012]
15. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012 May 14;14(3):e72 [FREE Full text] [doi: 10.2196/jmir.1977] [Medline: 22584372]
16. Trujillo R. Bizjournals. 2011 Jan. Apple reaches 10 billion apps sold URL: https://www.bizjournals.com/sacramento/news/2011/01/24/apple-reaches-10-billion-apps-sold.html [accessed 2017-07-06] [WebCite Cache ID 6rlDbwTTJ]
17. Turner-McGrievy GM, Beets MW, Moore JB, Kaczynski AT, Barr-Anderson DJ, Tate DF. Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mHealth weight loss program. J Am Med Inform Assoc 2013 May 1;20(3):513-518 [FREE Full text] [doi: 10.1136/amiajnl-2012-001510] [Medline: 23429637]
18. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012 Nov 21;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]
19. Conroy DE, Yang CH, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
20. Michie S, Abraham C. Interventions to change health behaviours: evidence-based or evidence-inspired? Psychol Health 2004 Feb;19(1):29-49. [doi: 10.1080/0887044031000141199]
21. Yabroff KR, Mandelblatt JS. Interventions targeted toward patients to increase mammography use. Cancer Epidemiol Biomarkers Prev 1999 Sep;8(9):749-757 [FREE Full text] [Medline: 10498393]
22. Bartholomew LK, Mullen PD. Five roles for using theory and evidence in the design and testing of behavior change interventions. J Public Health Dent 2011;71(Suppl 1):S20-S33. [Medline: 21656946]
23. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. Health Psychol 2010 Jan;29(1):1-8. [doi: 10.1037/a0016939] [Medline: 20063930]
24. Nielson Company. 2014 Apr 16. Hacking Health: How Consumers Use Smartphones and Wearable Tech to Track Their Health URL: http://www.nielsen.com/us/en/insights/news/2014/hacking-health-how-consumers-use-smartphones-and-wearable-tech-to-track-their-health.html [accessed 2017-07-06] [WebCite Cache ID 6rlDjvbZz]
25. Sharma M. Theoretical Foundations of Health Education and Health Promotion. 3rd edition. Burlington, MA: Jones & Bartlett Learning; Mar 2016.
26. Glanz K, Rimer BK, Viswanath K, editors. Health Behavior: Theory, Research, and Practice, 5th edition. San Francisco, CA: Jossey-Bass; Jul 2015.
27. Morrissey EC, Corbett TK, Walsh JC, Molloy GJ. Behavior change techniques in apps for medication adherence: a content analysis. Am J Prev Med 2016 May;50(5):e143-e146. [doi: 10.1016/j.amepre.2015.09.034] [Medline: 26597504]
28. Yang CH, Maher JP, Conroy DE. Implementation of behavior change techniques in mobile applications for physical activity. Am J Prev Med 2015 Apr;48(4):452-455. [doi: 10.1016/j.amepre.2014.10.010] [Medline: 25576494]
29. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]
30. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014 May 01;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
31. Marshall SJ, Biddle SJ. The transtheoretical model of behavior change: a meta-analysis of applications to physical activity and exercise. Ann Behav Med 2001;23(4):229-246. [Medline: 11761340]
32. Woods C, Mutrie N, Scott M. Physical activity intervention: a transtheoretical model-based intervention designed to help sedentary young adults become active. Health Educ Res 2002 Aug;17(4):451-460 [FREE Full text] [Medline: 12197590]
33. Fahrenwald NL, Walker SN. Application of the transtheoretical model of behavior change to the physical activity behavior of WIC mothers. Public Health Nurs 2003;20(4):307-317. [Medline: 12823791]
34. Armitage CJ. Can the theory of planned behavior predict the maintenance of physical activity? Health Psychol 2005 May;24(3):235-245. [doi: 10.1037/0278-6133.24.3.235] [Medline: 15898858]
35. Achterberg C, Miller C. Is one theory better than another in nutrition education? A viewpoint: more is better. J Nutr Educ Behav 2004;36(1):40-42. [Medline: 14756981]
36. Wang Q, Egelandsdal B, Amdam GV, Almli VL, Oostindjer M. Diet and physical activity apps: perceived effectiveness by app users. JMIR Mhealth Uhealth 2016 Apr 07;4(2):e33 [FREE Full text] [doi: 10.2196/mhealth.5114] [Medline: 27056639]

XSL•FO
RenderX

37. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
38. Ajzen I, Madden TJ. Prediction of goal-directed behavior: attitudes, intentions, and perceived behavioral control. J Exp Soc Psychol 1986 Sep;22(5):453-474. [doi: 10.1016/0022-1031(86)90045-4]
39. Smith JJ, Morgan PJ, Plotnikoff RC, Dally KA, Salmon J, Okely AD, et al. Smart-phone obesity prevention trial for adolescent boys in low-income communities: the ATLAS RCT. Pediatrics 2014 Sep;134(3):e723-e731 [FREE Full text] [doi: 10.1542/peds.2014-1012] [Medline: 25157000]
40. Lewis TL, Wyatt JC. App usage factor: a simple metric to compare the population impact of mobile medical apps. J Med Internet Res 2015 Aug 19;17(8):e200 [FREE Full text] [doi: 10.2196/jmir.4284] [Medline: 26290093]
41. Payne HE, Wilkinson J, West JH, Bernhardt JM. A content analysis of precede-proceed constructs in stress management mobile apps. Mhealth 2016 Feb 29;2:5 [FREE Full text] [doi: 10.3978/j.issn.2306-9740.2016.02.02] [Medline: 28293583]
42. Ross J, Irani L, Silberman MS, Zaldivar A, Tomlinson B. Who are the crowdworkers?: shifting demographics in mechanical turk. In: CHI '10 Extended Abstracts on Human Factors in Computing Systems. 2010 Apr 15 Presented at: CHI EA '10; April 10-15, 2010; Atlanta, Georgia, USA p. 2863-2872. [doi: 10.1145/1753846.1753873]
43. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]

Abbreviations

mTurk: Amazon Mechanical Turk

Edited by G Eysenbach; submitted 21.12.16; peer-reviewed by B Turner-McGrievy, J Apolinário-Hagen, A Paglialonga, Y Liu; comments to author 18.03.17; revised version received 22.04.17; accepted 10.06.17; published 03.08.17.

Please cite as:
Hoj TH, Covey EL, Jones AC, Haines AC, Hall PC, Crookston BT, West JH
How Do Apps Work? An Analysis of Physical Activity App Users' Perceptions of Behavior Change Mechanisms
JMIR Mhealth Uhealth 2017;5(8):e114
URL: http://mhealth.jmir.org/2017/8/e114/
doi:10.2196/mhealth.7206
PMID:28778846

©Taylor H Hoj, Emarie L Covey, Allyn C Jones, Amanda C Haines, P Cougar Hall, Benjamin T Crookston, Joshua H West. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 03.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Improving Patient-Centered Care for Young People in General Practice With a Codesigned Screening App: Mixed Methods Study

Marianne Julie Webb[1], BA, MYHEM; Greg Wadley[2], MSc, PhD; Lena Amanda Sanci[1], MBBS, PhD, FRACGP

[1]Department of General Practice, Melbourne Medical School, Faculty of Medicine, University of Melbourne, Parkville, Australia

[2]School of Computing and Information Systems, University of Melbourne, Parkville, Australia

Corresponding Author:
Marianne Julie Webb, BA, MYHEM
Department of General Practice, Melbourne Medical School
Faculty of Medicine
University of Melbourne
Ground Floor
200 Berkeley Street
Parkville, 3010
Australia
Phone: 61 390355011
Email: webbm@unimelb.edu.au

Abstract

Background: Despite experiencing a high prevalence and co-occurrence of mental health disorders and health-compromising behaviors, young people tend not to seek professional help for these concerns. However, they do regularly attend primary care, making primary care providers ideally situated to identify and discuss mental health and lifestyle issues as part of young people's routine health care.

Objective: The aim was to investigate whether using a codesigned health and lifestyle-screening app, Check Up GP, in general practice influenced young people's assessment of the quality of their care (measures of patient-centered care and youth friendliness), and their disclosure of sensitive issues. In addition, this study aimed to explore young people's acceptance and experience of using a screening app during regular health care.

Methods: This was a mixed methods implementation study of Check Up GP with young people aged 14 to 25 years attending a general practice clinic in urban Melbourne, Australia. A 1-month treatment-as-usual group was compared to a 2-month intervention group in which young people and their general practitioners (GPs) used Check Up GP. Young people in both groups completed an exit survey immediately after their consultation about disclosure, patient-centered and youth-friendly care, and judgment. In addition, participants in the intervention group were surveyed about app acceptability and usability and their willingness to use it again. Semistructured interviews with participants in the intervention group expanded on themes covered in the survey.

Results: The exit survey was completed by 30 young people in the treatment-as-usual group and 85 young people in the intervention group. Young people using Check Up GP reported greater disclosure of health issues (P<.001), and rated their GP higher in patient-centered care: communication and partnership (P=.01), personal relationship (P=.01), health promotion (P=.03), and interest in effect on life (P<.001). No differences were found on core indicators of youth-friendly care: trust, level of comfort, expectations met, and time to ask questions. In all, 86% (73/85) of young people felt the app was a "good idea" and only 1% (1/85) thought it a "bad idea." Thematic analysis of qualitative interviews with 14 participants found that Check Up GP created scope to address unmet health needs and increased sense of preparedness, with use moderated by honesty, motivation, app content and functionality, and app administration.

Conclusions: Integrating a health and lifestyle-screening app into face-to-face care can enrich young people's experience of seeing their GP, create scope to identify and address unmet health needs, and increase patient-centered care. Further research is needed to investigate the effect of using a health and lifestyle-screening app in a diverse range of clinic types and settings, and with a diverse range of GPs and youth.

(JMIR Mhealth Uhealth 2017;5(8):e118) doi:10.2196/mhealth.7816

XSL•FO
RenderX

KEYWORDS

adolescent; needs assessment; general practice; primary prevention; health behavior; health information technology; patient-centered care

Introduction

Adolescence and young adulthood are periods of major transition in physical, cognitive, social, and emotional development along the journey from childhood to adulthood [1]. These are also periods when mental health disorders and health-compromising behaviors emerge [1]. Worldwide, substance use, poor diet, lack of exercise, and mental health disorders are the leading risk factors for years lost due to ill health, disability, or early death for young people [2]. These disorders and behaviors tend to co-occur [3,4] and persist into adulthood [5-7].

Despite experiencing a high prevalence and co-occurrence of mental health disorders and lifestyle issues, young people do not usually seek professional help for these concerns [8]. Yet, they do regularly attend primary care, usually to address physiological concerns [9,10], making primary care providers (PCPs) ideally situated to opportunistically discuss mental health and lifestyle issues, and to provide health promotion and early intervention as needed. Both young people and PCPs report wanting to have these discussions [11,12], yet seldom do so [10,12,13], even during well-child visits in the United States, which are appointments dedicated to screening for health and lifestyle issues and providing preventive health care [14].

Patient-centered care is a theoretical concept proposing that for therapeutic benefits patients need to be actively engaged in their care, with a focus on communication, partnership, and health promotion in the doctor-patient consultation [15]. Patient-centered care has been associated with improved symptom burden, satisfaction, and enablement [16]. Although there is limited research in youth populations, evidence suggests that young people prefer a patient-centered approach, and taking an active partnership role in decision making over a more passive role [17]. Despite this, many young people do not receive patient-centered care, even though it is associated with higher ratings of quality care [18].

Similar to patient-centered care, youth-friendly care is an evidence-based theoretical framework for delivering quality health care. However, youth-friendly care looks specifically at the indicators of quality service delivery for adolescents and young adults. Ambresin et al [19] have developed a theoretical framework for youth-friendly care at both the clinic and individual level: accessibility of health care, staff attitude (trustworthy, supportive, respectful), communication, medical competency, evidence-based guideline-driven care (confidentiality, comprehensive care), age-appropriate environment (privacy, teen-orientated health information), involvement in health care, and health outcomes. Thus, the patient-centered and youth-friendly care theoretical frameworks can provide the foundations for designing and evaluating technological interventions designed to improve young people's engagement with health care.

Regular screening of young people for a range of health and lifestyle issues is recommended by peak professional bodies [20,21] and, along with subsequent intervention, may improve health outcomes [22]. It is known that screening via technology increases disclosure when compared with paper [23,24] and face-to-face [25,26] formats. Furthermore, young people prefer initially disclosing health information via technology rather than face-to-face, even when they know the results will be reviewed by their practitioner [27,28]. Previous studies have validated technology-based health and lifestyle-screening tools for young people [29,30].

Evidence in non-youth populations suggests that technology-based screening improves patient-centered care [31,32]; however, evidence is limited for the impact on young people's patient-centered and youth-friendly care. One study by Gadomski et al [33] undertook an analysis of audiotaped consultations and found that technology-based screening tool use increased doctor engagement and discussion of psychosocial and mental issues, without affecting partnership or rapport. However, this study did not interview young people about their experience of care. Another study by Olson et al [34] found that young people using a technology-based screening tool were more likely to feel that they were listened to carefully, that the discussion was confidential, and to feel more satisfied. Other core elements of youth-friendly care, such as trust and respect, and key elements of patient-centered care, such as health prevention and promotion and shared power, remain unexplored. Both the studies of Gadomski et al and Olson et al were limited to young people aged 19 years or younger and attending well-child visits [33,34]. More evidence is needed about the use of technology-based screening during other occasions of care because less than 50% of young people attend well-child consultations [35,36]; therefore, most are not receiving preventive care. Further, young adults aged 18 to 25 years have a higher prevalence of significant health risks, such as sexually transmitted infections, substance use, and mental health problems [37], and have more unmet health needs [38] compared to younger adolescents. Hence, testing of the utility of technology-based screening is also required in this group.

Guidelines recommend annual preventive screening [20,21]; however, there is a paucity of research on whether young people are willing to use technology-based screening tools on a regular basis and, if so, how often. Although one study in a hospital-based adolescent and young adult clinic found that 84% of young people would be willing to complete a screening tool once a year, participants responded before their consultation [39]. It is possible their responses might be different after the experience of using the tool with their practitioner and at a practice not servicing only youth.

The aim of this study was to investigate how using a codesigned health and lifestyle-screening app, Check Up GP , based on the theoretical principles of patient-centered and youth-friendly care, in an Australian general practice influenced young people's

XSL•FO
RenderX

assessment of care and their engagement through disclosure of sensitive issues. We were also interested in understanding young people's acceptance and experience of using a screening app as part of their regular health care.

This study addresses several aspects of the evidence gap by first conducting an in-depth mixed methods analysis of young people's experience of using a technology-based screening tool incorporated opportunistically into routine health care visits rather than well visits, and second, by including young people aged 14 to 25 years in the study. Furthermore, Australian general practice, although being the most commonly accessed form of primary health care by Australian youth, caters for the population across the life span rather than being a youth-specific service, hence this study also provides evidence relevant to generalist health care settings.

Methods

Study Design and Setting

We conducted a mixed methods implementation study in 2016, comparing a 1-month treatment-as-usual (TAU) phase with a 2-month intervention phase. In the intervention phase, Check Up GP was integrated into the routine care of young people attending a general practice. The length of the study was decided by mutual agreement between the clinic and researchers prior to the commencement. Ethics approval was obtained from the University of Melbourne (Ethics ID #1544281).

One general practice clinic was recruited through the Victorian Primary Care Practice-Based Research Network, managed by the Department of General Practice at the University of Melbourne. The clinic is a large general practitioner (GP)-owned and operated practice located in an area of relative socioeconomic advantage in urban Melbourne, Australia, staffed by 12 GPs, a practice manager, a reception coordinator, and eight receptionists. Open 365 days a year, the practice is funded by a patient copayment on top of the national health care basic amount.

Participants

The four GP principal owners of the practice participated in the study, along with their patients aged 14 to 25 years attending for routine primary care during the study period. Patients were excluded from the study if their GP assessed that their patient was very unwell (eg, vomiting, weak, psychotic), unable to read/speak English, or if they were younger than 18 years of age and not a mature minor [40]. The practice support staff, being the practice manager, reception coordinator, and all eight receptionists, also participated. The practice support staff were responsible for administering Check Up GP to young people.

Check Up GP App and Codesign Process

Based on the design needs identified in codesign workshops held with young people, GPs, practice support staff, and parents (reported previously [41]), we contracted a commercial software developer to build the health and lifestyle-screening tool, Check Up GP . To continue the codesign process, we recruited two reference groups representing the main end users of the tool; one of young people aged 14 to 25 years and another of GPs. These reference groups provided feedback and guidance on the design, content, and implementation of Check Up GP to ensure the final product reflected the requirements identified in the original codesign workshops. This process was also directly guided by the two theoretical frameworks, with each component of patient-centered care and youth-friendly care (described previously) being incorporated into Check Up GP. The resulting key technology and design features of Check Up GP are mapped onto each component of patient-centered care and youth-friendly care theoretical frameworks in Table 1.

Check Up GP consisted of two components: the questionnaire that patients answered and the summary report for GPs. The questionnaire was adapted from the HEEADSSS (Home environment, Education and employment, Eating, peer-related Activities, Drugs, Sexuality, Suicide and depression, Safety from injury and violence) preventive health framework for interviewing adolescents [42,43] recommended by the Royal Australian College of General Practitioners [21]. The framework covers the range of health, social, psychological, and physiological issues and behaviors contributing to the major burdens of disease for young people and suggests that they be raised and explored with young patients. We included validated screening tools in many of the HEEADSSS domains: eating disorders (SCOFF Questionnaire) [44], anxiety (Generalized Anxiety Disorder-2 [GAD-2]) [45], depression (Patient Health Questionnaire-2 [PHQ-2]) [46], and drugs and alcohol (CRAFFT screening test) [47]. At the beginning of the app, we included information about the research study and the purpose of the app, supplemented by a short video. Youth responses to the questions in the app in the format of a clinician summary report were immediately available to the GP via a secure website. The summary highlighted areas of concern, and strengths, along with tips on youth-friendly consultations and suggested actions to take on areas of concern including referral options, information, and resources. Screenshots of the youth and GP interfaces are included in Multimedia Appendix 1.

Study Procedure and Measures

The study had three main phases: a TAU phase that consisted of a GP profile survey conducted at the start of data collection on GPs self-perceived ratings of consulting and communicating with youth, and youth "exit surveys" conducted at the practice of young people postconsultation with their GP; an intervention phase, that involved exit surveys of young people postconsultation after Check Up GP had been implemented; and a postintervention phase employing semistructured interviews with young people. The measures and procedure for each phase are described subsequently and summarized in Table 2.

XSL•FO
RenderX

Table 1. Key technology and design features of Check Up GP according to codesign workshop requirements and theoretical framework components.

Young people's identified requirements	Theoretical framework (component)[a]	GPs' identified requirements[a]	Theoretical framework (component)[a]
Link to Check Up GP sent via SMS at time of making the appointment	YFC (age-appropriate, accessibility)	Secure server	YFC (guideline-driven: confidentiality)
Secure log in	YFC (guideline-driven: confidentiality, accessibility)	Areas of concern/strengths displayed via traffic-light system from low risk (green) through to moderate (yellow) and high risk (red)	YFC (guideline-driven care), PCC (partnership, communication)
Choice of whether to complete on own device prior to attending appointment or on tablet in clinic waiting room immediately prior to appointment	YFC (age-appropriate; accessibility)	High level summary, with ability to expand for more detail	PCC (partnership, communication)
Youth-friendly language, design	YFC (age-appropriate, communication), PCC (communication)	Tips on youth-friendly practice including communication skills, negotiating a management plan	YFC (guideline-driven care, involvement in health care, communication, staff attitude), PCC (partnership, communication)
Ability to skip questions and flag issues for discussion	YFC (involvement in health care), PCC (partnership)	Suggested evidence-based/recommended for actions and health promotion, referrals to appropriate services and self-help apps	YFC (medical competency, guideline-driven, health outcomes), PCC (health promotion, partnership, communication)
		Ability to save as PDF for export to electronic health record	YFC (medical competency, guideline-driven)

[a]PCC: patient-centered care theoretical framework; PDF: portable document format; YFC: youth-friendly care theoretical framework.

Phase 1: Treatment As Usual

The GPs completed a brief paper-based profile survey of demographic information and self-rated their enthusiasm, knowledge, and confidence in consulting and communicating with youth. We then collected exit interviews from young people attending the clinic. During this phase, researchers approached young patients in the waiting room when they arrived and, if they were seeing a participating GP and aged 14 to 25 years, the researcher provided them with information about the study and invited them to participate. If they agreed to participate, the young patient was asked to take a form into their consultation for their GP to assess their eligibility to participate in the study. The patient handed the form back to the researcher on returning to the waiting room. After their consultation, each participant completed the short exit survey on a tablet in the waiting room, with consent provided at the start of the survey. As described in Table 2, this survey asked young people to self-rate levels of disclosure and to rate patient-centered and youth-friendly care and perceived judgment from their GP.

Phase 2: Intervention

Following the TAU phase, clinic staff were introduced to the Check Up GP app. The GPs and practice support staff were trained in administrating and using Check Up GP. This included showing a short video, which demonstrated how to administer and integrate the tool into the consultation with young people. We then met individually with each GP to ensure that they could easily access the online clinician summary on their computer and felt comfortable using it. At this time, two training points were stressed: (1) the importance of acknowledging all issues raised and, if time was not available to address them, to then organize a follow-up consultation and (2) to ask for time alone when discussing personal or sensitive issues if a parent was attending with the young person. These practice points were also available to GPs in the clinician interface of the app. We worked closely with GPs and practice support staff throughout the intervention phase to refine the app and its administration in rapid cycles of continuous quality improvement [51].

XSL•FO
RenderX

Table 2. Measures used in the treatment-as-usual (TAU), intervention, and postintervention phase.

Name of measure	Type, source, and content of measure	Method and phase of administration
GP measures		
GP profile survey	Demographic information survey; age, gender, previous training in youth health	Paper questionnaire, TAU
Knowledge and confidence	Self-rated Likert scales (from 1=not at all to 7=extremely) from Sanci et al [48]; how knowledgeable and confident in consulting with young people aged 14-17, consulting with young people aged 18-25, consulting with male young people, consulting with female young people, communicating with young people, and exploring issues beyond the presenting problem	Paper questionnaire, TAU
Enthusiasm	Self-rated Likert scale (from 1=very unenthusiastic to 11=very enthusiastic) from Sanci et al [48]; enthusiasm for seeing young people	Paper questionnaire, TAU
Young people's measures (completed post consultation)		
Disclosure	Self-rated Likert scales (from 1=no disclosure to 5=full disclosure, or not applicable) adapted[a] from Bradford and Rickwood [25]; disclosure about home life, school/work, bullying/harassment, alcohol and drug use, sexual health and sexuality, safety, mood, addiction, stressful events, diet, exercise, sleep patterns, and risky behavior	Tablet questionnaire, TAU and intervention
Patient-centered care	Self-rated Likert scales (from 1=very strongly agree to 7=very strongly disagree), validated tool by Little et al [16]; a 21-item validated tool consisting of 5 components of patient-centered care: communication and partnership, personal relationship, health promotion, positive approach to diagnosis and prognosis, and interest in the effect on life	Tablet questionnaire, TAU and intervention
Youth-friendly care	Self-rated Likert scales, selected items from tool[b] by Haller et al [49]; trust in the GP (from 1=poor to 5=excellent), whether expectations were met (from 1=strongly disagree to 4=strongly agree), whether there was enough time to ask questions (from 1=strongly disagree to 4=strongly agree), and level of comfort with GP (from 1=not at all to 5=very comfortable)	Tablet questionnaire, TAU and intervention
Fear of judgment	Self-rated Likert scales (from 1=not at all concerned to 5=very concerned) adapted[c] from Bradford and Rickwood [25]; how worried young people were that the GP would think they were a bad person if they talked about everything they had been thinking, feeling, and doing; the GP would think there was something really wrong with them; they would learn things about themselves they didn't want to know; they would lose control of their emotions; and that the GP would judge them	Tablet questionnaire, TAU and intervention
App acceptability	Self-rated Likert scales (from 1=not at all to 5=very much) from Bradford and Rickwood [25]; confidence of Check Up GP providing an accurate picture, comfort in disclosing personal information through Check Up GP, whether questions were difficult to understand, whether questions caused upset, and whether GP addressed issues raised	Tablet questionnaire, intervention
App usability	Self-rated Likert scales selected items from validated tool by Stoyanov et al [50]; how easy Check Up GP was to learn (from 1=very hard to 5=able to use immediately) and how good Check Up GP looked (from 1=no visual appeal to 5=very attractive)	Tablet questionnaire, intervention
Overall opinion and willingness to use again	Self-rated categorical responses; opinion of using Check Up GP (good idea, bad idea, don't know), whether willing to use again (yes, no, and, if so, how often)	Tablet questionnaire, intervention
Youth interview	Semistructured interview; experience of using Check Up GP, how it was administrated and integrated into routine care, if they would like to use again in the future	Phone interview, audiotaped, postintervention

[a]Adaption was replacing "therapist" with "doctor" and "I lied or misrepresented myself" with a "not applicable" option.

[b]English version of tool validated in Bosnian language.

[c]Adaption was replacing "therapist" with "doctor."

Receptionists informed young people or parents about Check Up GP when they called the clinic to make an appointment. Receptionists then flagged the appointment in the clinical software so the reception coordinator would send the patient a short message service (SMS) text message containing the tool link. Patients could choose to complete Check Up GP before they arrived at the clinic or in the waiting room just prior to their appointment. When young patients arrived at the clinic, receptionists asked those aged 14 to 25 years who were patients of participating GPs if they had received the SMS text message and completed the app. Patients who had not completed the app were handed a tablet to complete it in the waiting room before their consultation. The receptionist made a note in the clinical software reminding the GP to review the young person's responses prior to calling them into the consultation.

During the intervention phase, researchers approached young people in the waiting room to participate in the study and complete an exit survey postconsultation, in the same manner as in the TAU phase. This exit survey had the same questions as those in the TAU phase, with additional items related to

XSL•FO
RenderX

Check Up GP as described in Table 2: acceptability, usability, overall opinion, and willingness to use again in the future.

Phase 3: Postintervention

At the end of the exit survey, participants were invited to participate in a follow-up postintervention interview held within 2 weeks of their consultation at a mutually agreed time. The interview guide was informed in part by the themes emerging from the codesign workshops [41]. The aim of these interviews was to explore young people's experience of using Check Up GP, including how it affected the consultation and their relationship with their GP, and whether they would like to use the app again in the future.

Reimbursement

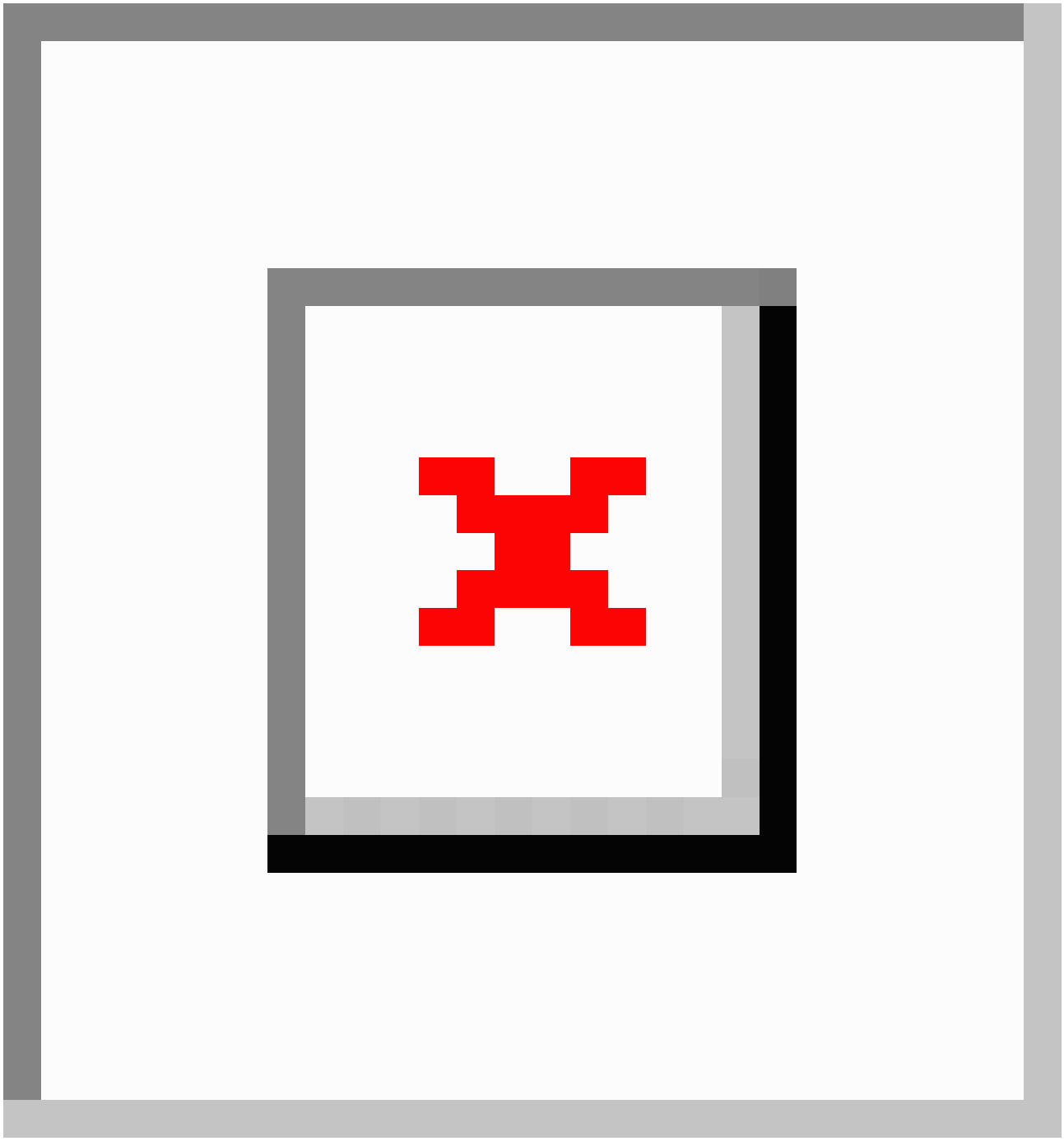
Each patient who completed an exit survey received an Aus $10 gift voucher. Youth participants who were interviewed received an additional Aus $30 gift voucher. The clinic received Aus $2000 for their involvement. In addition, each receptionist received Aus $100 at the end of the study for the additional work required of them to administer the app.

Analysis

Quantitative data from youth exit surveys were analyzed using SPSS version 23. Youth participant characteristics in the TAU and intervention groups were compared using chi-square test for independence. Where assumptions of expected cell frequency were not met, Fisher exact test was used. Data from all scaled variables showed major deviation from the normal distribution. Transformation did not adequately improve normality, so the nonparametric Mann-Whitney *U* analysis was used to compare groups for all variables. The Mann-Whitney *U* test has the added advantage of being suitable for uneven sample sizes. In comparing disclosure ratings between groups, "not applicable" responses were not included in the analysis. Scores for patient-centered care items were combined into five core components of the patient-centered model as described by Little et al [16]: communication and partnership, personal relationship, health promotion, positive and clear approach to problem, and interest in effect on life. Items for disclosure and patient-centered care were also combined to create a total disclosure and patient-centered care score. Internal consistencies of the following were calculated using Cronbach alpha: items in each of the four patient-centered core factors, total patient-centered care, and total disclosure. Effect sizes for all Mann-Whitney *U* analyses were calculated and reported as *r* with magnitudes using Cohen criteria (.1=small, .3=medium, .5=large) [52].

XSL•FO
RenderX

Figure 1. Recruitment breakdown of young people approached in the treatment-as-usual and intervention phases of the study.



Interviews were audio-recorded and transcribed verbatim by a professional transcription service. The first author compared the transcriptions to the audio recordings to ensure accuracy before coding all transcripts using NVivo 11 software. The second and third authors used this coding structure to independently code two transcripts. All authors then met to compare codes, revising the coding structure as required. The first author recoded all transcripts using the updated coding framework before conducting a thematic analysis [53]. These themes were discussed with all authors resulting in final themes.

Results

Participant Characteristics: Quantitative Phase

All four GPs were male and reported having had previous training in adolescent health. As shown in Table 3, each GP rated their enthusiasm for seeing a young person highly, with a mean rating of 9.5 (SD 1.3) out of 11. The participating GPs also highly rated their knowledge and confidence in consulting and communicating with young people, and exploring issues beyond the presenting problem.

Figure 1 shows the breakdown of young people who were approached and participated in the TAU and intervention phases of the study. In the TAU phase, 45 young patients were

XSL•FO
RenderX

approached, of whom 30 (67%) agreed to participate. In the intervention phase, 209 patients were invited to participate, of whom 85 (40.7%) agreed. Of the 124 patients (59.3%) who declined to participate, 15 had completed Check Up GP, but did not have time to complete the exit survey.

Table 3. General practitioner (GP) self-rated enthusiasm, and knowledge and confidence of consulting and communicating with young people.

GP characteristics	GP 1	GP 2	GP 3	GP 4	Mean (SD)
Age group (year range)	45-54	45-54	45-54	55-64	
Enthusiasm for seeing young people (max=11)	10	8	9	11	9.5 (1.3)
Consulting with young people (age 14-17) (max=7)					
Knowledgeable	5	5	6	7	5.8 (1.0)
Confident	6	4	5	7	5.5 (1.3)
Consulting with young people (age 18-25) (max=7)					
Knowledgeable	5	5	6	7	5.8 (1.0)
Confident	6	5	6	7	6.0 (.8)
Consulting with male young people (max=7)					
Knowledgeable	5	5	6	7	5.8 (1.0)
Confident	6	5	5	7	5.8 (1.0)
Consulting with female young people (max=7)					
Knowledgeable	5	5	5	7	5.5 (1.0)
Confident	6	4	5	7	5.5 (1.3)
Communicating with young people (max=7)					
Knowledgeable	5	5	5	7	5.5 (1.0)
Confident	6	4	5	7	5.5 (1.3)
Exploring issues beyond presenting problem (max=7)					
Knowledgeable	6	4	6	7	5.8 (1.3)
Confident	6	3	4	7	5.0 (1.8)

XSL•FO
RenderX

Table 4. Characteristics of youth participants in the treatment-as-usual (TAU) (n=30) and intervention (n=85) phase.

Youth characteristics	TAU (n=30)	Intervention (n=85)	Group comparison[a]	
			χ^2_1	*P*
Gender, n (%)			0.8	.38
Male	11 (37)	39 (46)		
Female	19 (63)	46 (54)		
Age (years), mean (SD)	19.13 (2.62)	19.93 (3.32)		
Age range (years)	14-23	14-25	0.2	.66
14-17	9 (30)	22 (26)		
18-25	21 (70)	63 (74)		
Sexuality, n (%)				
Heterosexual	27 (90)	83 (97)		.20
Bisexual		1 (1)		
Questioning	2 (7)	2 (2)		
Other	1 (3)			
Activities, n (%)				.19
Full-time work	3 (10)	16 (19)		
Part-time work		10 (12)		
Unemployed		2 (2)		
Home duties		1 (1)		
Have job, but not there due to illness		1 (1)		
Student attending school	9 (30)	22 (26)		
Student attending university	18 (60)	33 (33)		
First time at clinic, n (%)				.28
Yes	2 (7)	2 (2)		
No	28 (93)	83 (98)		
First time with doctor, n (%)				.56
Yes	6 (20)	12 (14)		
No	24 (80)	73 (86)		
Attending with parent, n (%)				.20
Yes, went into consult with young person	12 (40)	27 (31)		
Yes, but remained in waiting room		8 (9)		
No	18 (60)	51 (59)		

[a]Those comparisons that were not chi-square values were calculated with Fisher exact test.

Demographic characteristics of youth participants who completed the quantitative exit survey are presented in Table 4. There were no differences between the characteristics of patients in the TAU group and those in the intervention group.

Participant Characteristics: Qualitative Interviews

Fourteen youth participants from the intervention phase participated in qualitative interviews. Of those interviewed, nine were female and five were male; three participants were aged 14 to 17 years and 11 were aged 18 to 25 years (mean 20.9, SD 3.4 years). Eight participants were university or college students, two were at high school, three in full-time employment, and one was unemployed. Five of the 14 attended the clinic with a parent, with two of these five young people reporting their parent went into the consultation with them.

Quantitative Findings

Young people's median ratings of disclosure of health and lifestyle issues are shown in Table 5, with higher ratings indicating greater disclosure. Compared to the TAU group, young people using Check Up GP reported higher ratings of disclosure on all health and lifestyle issues, except for physical health. Median ratings of disclosure for four issues (alcohol/other drugs, sexual health, sexuality, and hurt

XSL•FO
RenderX

self/others) improved from the minimum rating (1/5) in TAU to the maximum rating (5/5) in the intervention group. Medium effect sizes were found for all domains, except for sexual activity, sexuality, and diet that each had small effect sizes.

The items for each of the five core components of patient-centered care had high internal consistency. The Cronbach alpha coefficients were .99 for communication and partnership, .97 for personal relationship, .96 for health promotion, .97 for positive and clear approach to problem, and .89 for interest in effect on life.

Young people's median ratings of patient-centered care are shown in Table 6, with lower ratings indicating greater patient-centered care. Compared to those in the TAU group, young people in the intervention group rated patient-centered care better (lower median ratings) in four of the five core factors of patient-centered care: communication and partnership, personal relationship, health promotion, and interest in effect on life. Effect sizes were small for each, except for "interest in effect on life," which had a medium effect size. There was no difference between TAU and intervention groups for positive and clear approach to the problem.

There were no overall differences between TAU and intervention groups for patients' ratings of youth-friendly care: how comfortable they felt, the level of trust in the GP, the extent to which their service met their expectations, and whether they had enough time to ask the GP everything they wanted to. Young people in the intervention group had lower ratings for "I would lose control of my emotions" (z=–2.087, *P*=.04) compared to those in the TAU group. The effect size was small (*r*=.19). There were no other differences in rating of judgment.

Table 5. Youth disclosure ratings of health and lifestyle issues in the treatment-as-usual (TAU) and intervention phase.

Health domain	TAU		Intervention		*U*	z	*P*	*r*
	Median (IQR)	n	Median (IQR)	n				
Physical health	5.0 (2.0)	30	5.0 (2.0)	80	1042.50	–1.259	.21	.12
Home life	2.0 (1.0)	25	4.0 (3.0)	76	487.50	–3.789	<.001	.38
School/work	2.0 (2.0)	27	4.0 (2.0)	81	523.50	–4.188	<.001	.41
Bullying	1.0 (0)	22	3.5 (4.0)	56	312.50	–3.663	<.001	.41
Alcohol/other drugs	1.0 (0)	24	5.0 (3.0)	68	411.00	–3.839	<.001	.40
Sexual health	1.0 (2.0)	25	5.0 (4.0)	69	627.00	–2.168	.03	.22
Sexuality	1.0 (1.0)	24	5.0 (4.0)	66	548.00	–2.463	.01	.23
Hurt self/others	1.0 (0)	21	5.0 (4.0)	63	337.00	–3.717	<.001	.41
Mood	2.0 (1.0)	27	4.0 (2.0)	78	517.50	–4.142	<.001	.40
Addiction	1.0 (0)	19	4.5 (4.0)	58	256.50	–3.756	<.001	.43
Stressful events	2.0 (2.0)	28	4.0 (3.0)	80	611.50	–3.732	<.001	.36
Diet	3.0 (2.0)	27	4.0 (4.0)	75	706.50	–2.417	.02	.24
Exercise	2.5 (3.0)	26	5.0 (4.0)	75	607.50	–3.044	.002	.30
Sleep	3.0 (2.0)	27	4.0 (2.0)	81	681.50	–3.069	.002	.31
Risky behavior	1.0 (0)	23	2.5 (4.0)	66	475.00	–2.923	.003	.31
Safety	1.0 (0)	22	3.0 (4.0)	65	422.00	–3.129	.002	.34
Total[a]	27.5 (11.0)	30	44.5 (45.0)	84	648.50	–3.938	<.001	.37

[a]Cronbach alpha=.968.

Table 6. Reported ratings of patient-centered care by patients in treatment-as-usual (TAU) (n=30) and intervention groups (n=85).

Patient-centered care components	TAU, median (IQR)	Intervention, median (IQR)	*U*	z	*P*	*r*
Communication and partnership	26.5 (44.5)	12.0 (14.0)	849.00	–2.794	.01	.26
Personal relationship	9.5 (9.5)	4.0 (7.5)	880.50	–2.605	.01	.24
Health promotion	7.5 (5.8)	4.0 (6.0)	944.50	–2.175	.03	.20
Positive and clear approach to problem	10.5 (12.3)	5.0 (9.0)	991.00	–1.851	.06	.23
Interest in effect on life	7.0 (5.3)	3.0 (4.5)	734.50	–3.550	<.001	.33
Total[a]	54.5 (69.0)	29.0 (35)	801.50	–3.030	.002	.28

[a]Cronbach alpha=.991.

XSL•FO
RenderX

Table 7. Youth experience and acceptability of using Check Up GP.

Youth experience and acceptability	Not at all, n (%)	Only a little bit, n (%)	Somewhat, n (%)	Quite a bit, n (%)	Very much, n (%)
How confident are you that Check Up GP was able to provide an accurate picture of yourself to your GP?	2 (2)	4 (5)	31 (37)	37 (44)	11 (13)
How comfortable were you disclosing personal information through Check Up GP?	0	10 (12)	23 (27)	24 (28)	28 (33)
Were the questions in Check Up GP difficult to understand?	55 (65)	15 (18)	11 (13)	4 (5)	0
Did any of the questions in Check Up GP cause you to become upset?	66 (78)	10 (12)	6 (7)	1 (1)	2 (2)
To what extent did you feel your GP address the issues raised in CUGP?	3 (4)	7 (8)	17 (28)	32 (38)	26 (31)

Check Up GP was highly acceptable to patients. In all, 86% (73/85) thought the app was a "good idea," 13% (11/85) felt unsure ("don't know"), and only 1% (1/85) thought it was a "bad idea." Most young people (63/85, 74%) wanted to use Check Up GP again, and only 6% (5/85) did not. Of those who wanted to use Check Up GP again, and responded to the question, more than half (36/62, 58%) wanted to use it twice a year or more, 15% (9/62) wanted to use it once a year, and a further 23% (14/62) wanted to use the app every time they saw their GP.

Table 7 shows the experience and acceptance by patients using Check Up GP. The majority of patients (48/85, 56%) were "very much" or "quite a bit" confident that the app provided an accurate picture of them to their GP. The majority of patients (52/85, 61%) were also either "very much" or "quite a bit" comfortable about disclosing personal information about themselves through Check Up GP. Most patients (58/85, 68%) reported that their GP addressed issues raised by Check Up GP "quite a bit" or "very much." Almost all (76/85, 89%) reported that the questions caused them to become upset either "not at all" or "only a little bit." In all, 82% of patients (70/85) rated the questions as being difficult to understand either "not at all" or "only a little bit."

The usability and design of Check Up GP was also highly rated by patients. In all, 95% (81/85) rated Check Up GP as either being "able to use immediately" or "easy to learn"' and 75% (64/85) rated Check Up GP as having either a "high level of visual appeal" or being "very attractive, memorable."

Qualitative Findings

Our analysis of interviews with participants found four main themes: identifying unmet needs, creating scope, moderating factors for use, and future use.

Identifying Unmet Health Needs

Many participants reported that using Check Up GP helped them to identify and disclose issues that they had not previously discussed, or thought would be relevant to discuss, with their GP:

> *It definitely brought things up that wouldn't normally have been discussed in a consult...I feel like even the fact that those questions are being asked about your sleeping habits, your alcohol consumption, your drug use; the fact that they're being asked in a health context at all could be beneficial.* [Male, 24]

In particular, participants reported that using Check Up GP made it easier to disclose sensitive issues, even though they knew the responses would be available to their GP. These young people "wouldn't really like to tell the doctor in person...I've just always felt more comfortable with [technology]" (female, 19 years). There was a recognition that once the difficulty of disclosure could be overcome, problems could then be addressed:

> *I've wanted to bring up [mental health issues] with my doctor before and haven't because I hadn't really had the platform and felt a bit nervous about it.* [Male, 23]

In addition to facilitating disclosure, it was felt that using Check Up GP enabled participants to go into the consultation feeling more prepared and in control, and knowing what topics were going to be discussed. Check Up GP allowed them to "indicate to your doctor that this is what I'm going to talk to you about and this is where I'm heading" (female, 18 years). Thus, a shared understanding of what would be focused on in the consultation was achieved through the app in a timely manner:

> *You and the GP are obviously on a similar understanding, and then you can get right into whatever your concerns are.* [Female, 17]

Creating Scope for Addressing Unmet Health Needs

As well as facilitating the disclosure of unmet health needs, using Check Up GP created scope for addressing these needs within the consultation:

> *[The GP] took my responses...and gave a good reason to keep that [behavior] in check. Which I thought was good. It wasn't just tick, tick, tick, it's like he used it to explore further about my lifestyle choices and he did ask probing questions.* [Male, 24]

Although participants were generally very satisfied with how their GP addressed their concerns, a few felt they did not have sufficient time because their GP was either running well behind schedule or their responses were only addressed at the very end of their consultation. However, these participants also reported that their GP invited issues to be discussed at a future time:

> *We didn't really address any of the problems in detail because it was towards the end of the consultation and he was going over time. But he said at the end, "If you do ever want to talk about any of these, you're more than welcome to."* [Female, 23]

XSL•FO
RenderX

A positive consequence of discussing unmet needs was that many participants felt that their GP had improved their understanding of the young person's life.

> *When you go in sometimes you just focus on one question. So I think it was good to have a bit more of a general understanding.* [Female, 24]

Indeed, the process of unpacking responses led to some participants feeling a greater sense of connection and rapport with the GP. These young people felt that their GP showed a genuine interest and "actually care about what's happening" (female, 24 years) in their lives:

> *They're not just saying "how's it going," they're sitting there and asking you about things that might be concerning you.* [Male, 23]

In addition to addressing unmet needs, the two participants who went into the consultation with a parent reported that the GP asked for time alone. One participant reported that this was the first time that time without parents had been requested: "I've never had that before, so that was good" (female, 15 years).

Only one participant reported having a negative experience of using Check Up GP. This young person felt their GP did not adequately attempt to explore the issues raised due to a lack of time:

> *He just said "I don't think I need to lecture you on drugs and alcohol"...I just thought well what's the point of me putting it in there if I'm going to get a bit of a judgy comment about it rather than in an endeavor to understand a bit more to see if it's worth talking about...if he'd had more time then he would have gone into that a bit more.* [Male, 23]

Moderating Factors of Use

We discovered three factors that moderated use of the tool: young people's motivation and honesty, the content and functionality of the tool, and administration.

All participants recognized the potential value of the tool, even those who had no current issues or had a good relationship with their doctor:

> *Perhaps it doesn't necessarily apply to me, but I can see it working really well for other people, or at least it didn't apply to me that time. Who knows in the future?* [Male, 19]

Four of the five males interviewed, along with one of the nine females, volunteered that Check Up GP gave them something useful to do in the waiting room:

> *I get bored very easily, especially in the waiting room, and if I have something to do that is actually relevant, it definitely will be a big help.* [Male, 16]

Almost all participants reported they answered questions honestly because they understood that disclosing was the aim of the app: "I wouldn't lie because I think I'm only hurting myself if I'm lying to the GP because then he's not going to be able to help me" (male, 24 years).

In contrast, five young people explained that there may be occasions when they would lie or fail to disclose the full extent of a behavior: if they were uncomfortable discussing an issue (such as sexting) or if they believed it did not affect their health:

> *Anything that I felt comfortable telling him, I definitely told the truth about. But there were maybe one or two things that I probably didn't want to discuss with him that I just chose not to.* [Female, 23]

Another moderating factor to using Check Up GP was its content and functionality. Although participants felt there was a "good range of questions to get a general idea on someone" (female, 18 years), the language of the questions could be also perceived as "very direct and people might not want to answer that truthfully" (female, 24 years). Also, for a few participants, the length was "a little bit long because if the doctor was running on time, you probably wouldn't have enough time to finish it" (female, 25 years).

We found that the context of using Check Up GP could be an important factor influencing young people's engagement with the app. Although all young people felt they had sufficient privacy when completing it in the waiting room, a few participants also expressed the concern that their responses would be influenced if a parent or family member was sitting next to them:

> *If I'd been sitting next to a parent or family member, and they'd been looking over my shoulder—people curious [about] what it is—I think that would influence a lot of people's answers, my own included.* [Male, 19]

Many participants felt they had received insufficient explanation about Check Up GP by receptionists. Instead, most relied on the information provided at the start of the tool or by the researcher to understand the app's purpose and how the information would be used by GPs:

> *Reception didn't really say much about it, which was a bit weird. But on the tool the front page gave it a pretty decent, pretty lengthy explanation (male, 19 years).*

Future Use

All participants reported that they wanted to use Check Up GP regularly, when they see their GP one or two times per year. They felt that this would enable new issues that emerge to be identified and discussed:

> *[Twice a year] seems like the right space of time where things can change, life events happen, you know, relationships, or family members going through difficulties, yourself going through difficulties, employment changing.* [Male, 19]

Participants felt regular use of Check Up GP would also ensure that issues that had emerged previously could be revisited to "see how I'm tracking" (female, 23 years).

Finally, participants suggested a number of ways that the Check Up GP app could be developed. One suggestion was adding the ability to view previous answers. Another was the ability to

XSL•FO
RenderX

receive automated, personalized health information, such as on healthy diet or exercise.

Discussion

Principal Results

This mixed methods study aimed to investigate whether the implementation of a health and lifestyle-screening app, Check Up GP, codesigned by young people and GP staff and grounded in the theories of patient-centered and youth-friendly care, improved disclosure of health risks and care experiences of young patients attending general practice. We were also interested in understanding young people's acceptance of and experience with using a screening app as part of their routine health care with their GP.

Overall, we found using Check Up GP improved disclosure across all health and lifestyle domains, including sexual health, alcohol and other drugs, and risky behavior, except for physical health where disclosure was already high. This supports previous research showing that the use of technology improves young people's disclosure of sensitive health issues [24,25]. Our qualitative interviews suggested that using Check Up GP not only led to identification of unmet needs, but created scope for discussing and addressing these needs in a way that does not usually occur for young people during routine care. It is encouraging that the majority of young people (75/85, 88%) felt the issues raised were at least somewhat addressed, particularly given previous research has found that GPs ignored certain domains identified as issues by young people [54]. It is possible that the functionality of Check Up GP for the GPs (eg, traffic-light summary report with suggested actions and referral options), in addition to the emphasis in GPs' training on the need to follow up on all raised issues, facilitated GPs in our study to more consistently address raised issues. It is also possible that the GPs in this study felt particularly engaged and skilled to address issues, and willing to spend extra time doing this.

Young people who used Check Up GP felt more prepared to discuss the issues raised in their consultation. In our quantitative findings, young people using the app were less concerned that they would lose control of their emotions than those in the TAU group. This finding was reflected in our interviews with young people, with participants appreciating that the tool provided them a chance to prepare and reflect on their responses before the consultation. In doing so, young people felt more emotionally prepared to have a conversation about sensitive issues than they otherwise would have been. Although similar results were found with young people using a screening tool with a nurse in a school setting [55], this is the first study to our knowledge that explores young people's experience of the tool in the general practice setting, outside of well-child visits.

Using Check Up GP had a profound impact on young people's perception of patient-centered care. These findings support and extend previous research that had only measured and found improvements in one element of communication [34]. As well as communication and partnership, we found that using Check Up GP improved personal relationships, health promotion, and interest in the effect of the problem on life. In contrast, Gadomski et al [33] only found improvements in doctor engagement and no change in partnership or rapport. This difference may be due to measuring patient-centered care by analyzing audio recordings of consultations, rather than patient's own ratings. Another recently published study found no changes in communication and experience of care, or in health information [56]. However, both of these studies were conducted with young people at well visits, which are not well-attended by key subgroups of young people such as low income and uninsured [14]. Thus, our findings suggest that the use of a technology-based screening tool has the potential to transform young people's experience of care.

In contrast to patient-centered care, using Check Up GP did not improve core indicators of youth-friendly care. We found no difference between TAU and intervention groups for young people's assessment of trust, whether they had enough time to ask the GP everything they wanted to, their level of comfort, nor the extent to which the service met their expectations. There was also no change on all but one measure of judgmental reactions. The lack of improvement may be due to a ceiling effect, with these variables being rated very highly in the TAU group. Indeed, a previous randomized controlled trial of a paper-based screening tool also found youth participants rated their level of trust in their GP highly in both control and intervention groups [57]. Perhaps practices volunteering for these studies are already interested in meeting the needs of young people and therefore have high baseline scores. That there was almost no negative effect on youth-friendly variables in our study is a positive finding and suggests that this technology can be integrated into routine care and can direct focus away from the presenting issue without necessarily undermining the delivery of quality youth-friendly care. However, further research is needed to determine whether factors other than the ceiling effect or participating GP characteristics are responsible for the nonsignificant differences in youth-friendly care.

There are a number of established theories that may explain the underlying causal mechanisms of our findings. The Culture of Situated Cognition theory proposes that cultural mindsets influence what feels fluent and what is engaged with [58]. Drawing on this theory, the patient decision aid theoretical framework developed by Alden and colleagues [59] proposes that patient screening and decision aids that are culturally targeted to specific cultural groups, using appropriate imagery, evidential information, linguistics, and values relevant to that group, will enhance feelings of congruency with the material, thereby increasing "processing fluency." Improved processing fluency in turn predicts enhanced preparedness (eg, confidence and openness to participating) leading to improved outcomes. In addition, in the context of technology and using the Stimulus-Organism-Response theory, Jiang and colleagues [60] demonstrated that interactive features or cues on websites trigger a greater cognitive and emotional involvement of users, which in turn elicit desired behaviors. Thus, these theories would predict that an interactive technology-based screening tool that is designed specifically for a youth subculture is likely to enhance processing fluency and increase engagement, resulting

XSL•FO
RenderX

in increased disclosure and perceptions of patient-centered care. Further research is needed to further explore the underlying causal mechanisms with this technology.

A particularly interesting finding from the qualitative interviews was the dynamic nature of honesty in participants' responses on Check Up GP. For these young people, honesty was not a binary but was fluid, considered and controlled by the young person. Indeed, a substantial minority (31/85, 37%) reported that Check Up GP only "somewhat" provided an accurate picture of them. Given this was the first time participants had used such a tool, it may be that, understandably, young people wanted to "test the water" with their GP, to ensure any disclosure would be addressed with sensitivity. Alternatively, it may be that some participants were not yet ready to acknowledge they had a problem or to change their behavior. Perhaps using an app, such as Check Up GP, has the potential to be a catalyst for young people to reflect on their health and lifestyle and to begin the process of behavior change.

A positive finding of this study was the high proportion of young people who not only assessed Check Up GP as a "good idea," but who also wanted to continue to use it at least once a year as part of their routine care. Although best practice guidelines recommend this type of screening to be done annually [20,21], and previous studies of technology-based [39] and paper-based [57] screening have also found a high acceptance rate, there has been no research on whether young people themselves want to be screened regularly for health and lifestyle issues. Our results suggest that young people see value in this kind of technology being integrated into their routine care with their GP.

Finally, our findings suggest that administration of the app may influence young people's engagement with and use of Check Up GP. Most young people we interviewed felt that they did not receive adequate explanation from the clinic about Check Up GP. We also found that 42% (52/124) of young people who were not eligible to be included in the study were excluded because they did not have time to finish Check Up GP in the waiting room before their consultation. The introduction of technology into general practice needs to ensure all young people have the opportunity to use and benefit from it. There is now a solid body of research to suggest that the implementation of technology in health settings such as general practice is determined by interrelated individual, organizational, and social factors [61]. Further research is needed to understand how a screening app for young people can be successfully integrated within a general practice setting; another component of this study will address this theme in future work.

Limitations

This study had a number of limitations. Because it was conducted at one clinic in a relatively socioeconomically advantaged area of one city, the generalizability of results may be limited. The four participating GPs were all male and of a similar age, all had previous adolescent health training, and all reported being highly confident and knowledgeable in consulting with young people. It is possible that GPs with less experience and enthusiasm for young people may not use and engage with Check Up GP in the same way as the GPs in this study, which in turn may influence young people's experience of using it within their routine care. Young people in other socioeconomic areas may experience Check Up GP differently to those in this study. Another limitation of our study is that it did not have an experimental design and so did not have a randomized control group. Although we found no significant difference between the TAU and intervention groups in the quantitative data in key characteristics such as gender, age, and sexuality, it is possible that there were underlying and unidentified differences between the two groups.

Future Research

This implementation study of a health and lifestyle-screening app for young people was conducted in one general practice clinic. Further research is needed to investigate the effect of using a health and lifestyle-screening app in a diverse range of clinic types and settings, and with a diverse range of GPs and youth participants. Future research could also investigate whether improvements found in this study are sustained over time and influence young people's health and lifestyle. Finally, there is a need to explore the experience of PCPs and practice support staff in integrating Check Up GP into their clinical and administrative work, respectively; we will report on that phase of the study in the future.

Conclusions

Previous research has found that using a technology-based health and lifestyle-screening tool as part of routine care can improve disclosure and communication with health professionals, which aids in early identification of issues and delivery of preventive health care. This study extends this research by providing new insights about the use and experience of using such an app by young people as part of their routine care in general practice. The results suggest that integrating a health and lifestyle-screening app into face-to-face regular care can enrich young people's experience of seeing their GP, create scope to address unmet health needs, and become integrated into their regular health care.

Acknowledgments

This research is supported by the Young and Well Cooperative Research Centre, an Australian-based international research center that unites young people with researchers, practitioners, innovators, and policy makers from over 70 partner organizations. We would like to thank the young patients and their parents, and the general practitioners, practice manager, reception coordinator, and receptionists from the clinic where Check Up GP was piloted; without their participation this research would not have been possible. We would also like to thank research assistants Oshara de Silva and Shaunagh O'Sullivan for their role in collecting data.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1

App screen grabs of Check Up GP youth and general practitioner interface.

[PDF File (Adobe PDF File), 380KB - mhealth_v5i8e118_app1.pdf]

References

1. Sawyer SM, Afifi RA, Bearinger LH, Blakemore S, Dick B, Ezeh AC, et al. Adolescence: a foundation for future health. Lancet 2012 Apr 28;379(9826):1630-1640. [doi: 10.1016/S0140-6736(12)60072-5] [Medline: 22538178]
2. Gore FM, Bloem PJN, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10-24 years: a systematic analysis. Lancet 2011 Jun 18;377(9783):2093-2102. [doi: 10.1016/S0140-6736(11)60512-6] [Medline: 21652063]
3. Spring B, Moller AC, Coons MJ. Multiple health behaviours: overview and implications. J Public Health (Oxf) 2012 Mar;34 Suppl 1:i3-i10 [FREE Full text] [doi: 10.1093/pubmed/fdr111] [Medline: 22363028]
4. Martin J, Kauer SD, Sanci L. Road safety risks in young people attending general practice: a cross-sectional study of road risks and associated health risks. Aust Fam Physician 2016 Sep;45(9):666-672 [FREE Full text] [Medline: 27606371]
5. Oesterle S, Hill KG, Hawkins JD, Guo J, Catalano RF, Abbott RD. Adolescent heavy episodic drinking trajectories and health in young adulthood. J Stud Alcohol 2004 Mar;65(2):204-212 [FREE Full text] [Medline: 15151351]
6. Patton GC, Coffey C, Carlin JB, Sawyer SM, Williams J, Olsson CA, et al. Overweight and obesity between adolescence and young adulthood: a 10-year prospective cohort study. J Adolesc Health 2011 Mar;48(3):275-280. [doi: 10.1016/j.jadohealth.2010.06.019] [Medline: 21338899]
7. Pine DS, Cohen P, Gurley D, Brook J, Ma Y. The risk for early-adulthood anxiety and depressive disorders in adolescents with anxiety and depressive disorders. Arch Gen Psychiatry 1998 Jan;55(1):56-64. [Medline: 9435761]
8. Bailey V, Baker A, Cave L, Fildes J, Perrens B, Plummer J. Mission Australia. 2016. Mission Australia 2016 Youth Survey Report URL: https://www.missionaustralia.com.au/publications/research/young-people [WebCite Cache ID 6p8nKIohj]
9. Tylee A, Haller DM, Graham T, Churchill R, Sanci LA. Youth-friendly primary-care services: how are we doing and what more needs to be done? Lancet 2007 May 05;369(9572):1565-1573. [doi: 10.1016/S0140-6736(07)60371-7] [Medline: 17482988]
10. Australian Bureau of Statistics. 2013. Australian Health Survey: health service usage and health related actions, 2011-12 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4364.0.55.002Main+Features12011-12?OpenDocument [accessed 2017-03-22] [WebCite Cache ID 6p8neqhFN]
11. Stein RE, Horwitz SM, Storfer-Isser A, Heneghan A, Olson L, Hoagwood KE. Do pediatricians think they are responsible for identification and management of child mental health problems? Results of the AAP periodic survey. Ambul Pediatr 2008;8(1):11-17. [doi: 10.1016/j.ambp.2007.10.006] [Medline: 18191776]
12. Klein JD, Wilson KM. Delivering quality care: adolescents' discussion of health risks with their providers. J Adolesc Health 2002 Mar;30(3):190-195. [Medline: 11869926]
13. Ma J, Wang Y, Stafford R. US adolescents receive suboptimal preventive counseling during ambulatory care. J Adolescent Health 2005 May;36(5):441.e1-441.e7 [FREE Full text] [doi: 10.1016/j.jadohealth.2004.08.024] [Medline: 15841517]
14. Irwin CE, Adams SH, Park MJ, Newacheck PW. Preventive care for adolescents: few get visits and fewer get services. Pediatrics 2009 Apr;123(4):e565-e572. [doi: 10.1542/peds.2008-2601] [Medline: 19336348]
15. Little P, Everitt H, Williamson I, Warner G, Moore M, Gould C, et al. Preferences of patients for patient centred approach to consultation in primary care: observational study. BMJ 2001 Feb 24;322(7284):468-472 [FREE Full text] [Medline: 11222423]
16. Little P, Everitt H, Williamson I, Warner G, Moore M, Gould C, et al. Observational study of effect of patient centredness and positive approach on outcomes of general practice consultations. BMJ 2001 Oct 20;323(7318):908-911 [FREE Full text] [Medline: 11668137]
17. Alden DL, Merz MY, Akashi J. Young adult preferences for physician decision-making style in Japan and the United States. Asia Pac J Public Health 2012 Jan;24(1):173-184. [doi: 10.1177/1010539510365098] [Medline: 20460277]
18. Toomey SL, Elliott MN, Schwebel DC, Tortolero SR, Cuccaro PM, Davies SL, et al. Relationship between adolescent report of patient-centered care and of quality of primary care. Acad Pediatr 2016;16(8):770-776. [doi: 10.1016/j.acap.2016.01.006] [Medline: 26802684]
19. Ambresin A, Bennett K, Patton GC, Sanci LA, Sawyer SM. Assessment of youth-friendly health care: a systematic review of indicators drawn from young people's perspectives. J Adolesc Health 2013 Jun;52(6):670-681. [doi: 10.1016/j.jadohealth.2012.12.014] [Medline: 23701887]
20. American Academy of Pediatrics. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, 4th Ed. Elk Grove Village, IL: American Academy of Pediatrics; 2008. URL: http://reader.aappublications.org/

XSL•FO
RenderX

bright-futures-guidelines-for-health-supervision-of-infants-children-and-adolescents-4th-ed/1 [accessed 2017-03-22] [WebCite Cache ID 6sHmn8Sc7]
21. Royal Australian College of General Practitioners. Guidelines for Preventive Activities in General Practice. 8th ed. East Melbourne: Royal Australiasian College of General Practitioners; 2012. URL: http://www.racgp.org.au/download/Documents/Guidelines/Redbook9/17048-Red-Book-9th-Edition.pdf [accessed 2017-07-28] [WebCite Cache ID 6rwCMaYre]
22. Webb MJ, Kauer SD, Ozer EM, Haller DM, Sanci LA. Does screening for and intervening with multiple health compromising behaviours and mental health disorders amongst young people attending primary care improve health outcomes? A systematic review. BMC Fam Pract 2016 Aug 04;17:104 [FREE Full text] [doi: 10.1186/s12875-016-0504-1] [Medline: 27488823]
23. Spark S, Lewis D, Vaisey A, Smyth E, Wood A, Temple-Smith M, et al. Using computer-assisted survey instruments instead of paper and pencil increased completeness of self-administered sexual behavior questionnaires. J Clin Epidemiol 2015 Jan;68(1):94-101. [doi: 10.1016/j.jclinepi.2014.09.011] [Medline: 25304502]
24. Gnambs T, Kaspar K. Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. Behav Res Methods 2015 Dec;47(4):1237-1259. [doi: 10.3758/s13428-014-0533-4] [Medline: 25410404]
25. Bradford S, Rickwood D. Acceptability and utility of an electronic psychosocial assessment (myAssessment) to increase self-disclosure in youth mental healthcare: a quasi-experimental study. BMC Psychiatry 2015 Dec 01;15:305 [FREE Full text] [doi: 10.1186/s12888-015-0694-4] [Medline: 26627041]
26. Nguyen M, Bin YS, Campbell A. Comparing online and offline self-disclosure: a systematic review. Cyberpsychol Behav Soc Netw 2012 Feb;15(2):103-111. [doi: 10.1089/cyber.2011.0277] [Medline: 22032794]
27. Jasik CB, Berna M, Martin M, Ozer EM. Teen preferences for clinic-based behavior screens: who, where, when, and how? J Adolesc Health 2016 Dec;59(6):722-724. [doi: 10.1016/j.jadohealth.2016.08.009] [Medline: 27884300]
28. Coker TR, Sareen HG, Chung PJ, Kennedy DP, Weidmer BA, Schuster MA. Improving access to and utilization of adolescent preventive health care: the perspectives of adolescents and parents. J Adolesc Health 2010 Aug;47(2):133-142. [doi: 10.1016/j.jadohealth.2010.01.005] [Medline: 20638005]
29. Harrison PA, Beebe TJ, Park E. The Adolescent Health Review: a brief, multidimensional screening instrument. J Adolesc Health 2001 Aug;29(2):131-139. [Medline: 11472872]
30. Salerno J, Marshall VD, Picken EB. Validity and reliability of the rapid assessment for adolescent preventive services adolescent health risk assessment. J Adolesc Health 2012 Jun;50(6):595-599. [doi: 10.1016/j.jadohealth.2011.10.015] [Medline: 22626486]
31. Krist AH, Glasgow RE, Heurtin-Roberts S, Sabo RT, Roby DH, Gorin SNS, et al. The impact of behavioral and mental health risk assessments on goal setting in primary care. Transl Behav Med 2016 Jun;6(2):212-219 [FREE Full text] [doi: 10.1007/s13142-015-0384-2] [Medline: 27356991]
32. Fothergill KE, Gadomski A, Solomon BS, Olson AL, Gaffney CA, Dosreis S, et al. Assessing the impact of a web-based comprehensive somatic and mental health screening tool in pediatric primary care. Acad Pediatr 2013;13(4):340-347 [FREE Full text] [doi: 10.1016/j.acap.2013.04.005] [Medline: 23830020]
33. Gadomski AM, Fothergill KE, Larson S, Wissow LS, Winegrad H, Nagykaldi ZJ, et al. Integrating mental health into adolescent annual visits: impact of previsit comprehensive screening on within-visit processes. J Adolesc Health 2015 Mar;56(3):267-273 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.11.011] [Medline: 25586231]
34. Olson AL, Gaffney CA, Hedberg VA, Gladstone GR. Use of inexpensive technology to enhance adolescent health screening and counseling. Arch Pediatr Adolesc Med 2009 Feb;163(2):172-177. [doi: 10.1001/archpediatrics.2008.533] [Medline: 19188650]
35. Nordin JD, Solberg LI, Parker ED. Adolescent primary care visit patterns. Ann Fam Med 2010;8(6):511-516 [FREE Full text] [doi: 10.1370/afm.1188] [Medline: 21060121]
36. Tsai Y, Zhou F, Wortley P, Shefer A, Stokley S. Trends and characteristics of preventive care visits among commercially insured adolescents, 2003-2010. J Pediatr 2014 Mar;164(3):625-630. [doi: 10.1016/j.jpeds.2013.10.042] [Medline: 24286572]
37. Neinstein LS. The New Adolescents: An Analysis of Conditions, Behaviors, Risks, and Access to Services Among Emerging Young Adults. 2013. URL: https://eshc.usc.edu/thenewadolescents/doc/TheNewAdolescents_Final_Locked.pdf [accessed 2017-07-28] [WebCite Cache ID 6sHnaP7a3]
38. Marshall EG. Do young adults have unmet healthcare needs? J Adolesc Health 2011 Nov;49(5):490-497. [doi: 10.1016/j.jadohealth.2011.03.005] [Medline: 22018563]
39. Hassan A, Blood EA, Pikcilingis A, Krull EG, McNickles L, Marmon G, et al. Youths' health-related social problems: concerns often overlooked during the medical visit. J Adolesc Health 2013 Aug;53(2):265-271. [doi: 10.1016/j.jadohealth.2013.02.024] [Medline: 23643339]
40. Sanci LA, Sawyer SM, Weller PJ, Bond LM, Patton GC. Youth health research ethics: time for a mature-minor clause? Med J Aust 2004 Apr 05;180(7):336-338. [Medline: 15059054]
41. Webb M, Sanci L, Kauer S, Wadley G. Designing a health screening tool to help young people communicate with their general practitioner. Proceedings of the Annual Meeteeting Austalian Special Interest Grourp Computer Human Interaction Melbourne; 2015 Presented at: OzCHI '15 Annual Meeting of the Australian Special Interest Group for Computer Human Interaction; Dec 7-10, 2015; Melbourne, Australia. [doi: 10.1145/2838739.2838756]
42. Goldenring JM, Rosen DS. Getting into adolescent heads: an essential update. Contemp Pediatr 2004;21(1):64-74.

XSL•FO
RenderX

43. Klein DA, Goldenring JM, Adelman WP. Probing for scars: how to ask the essential questions. Contemp Pediatr 2014;31(1):16-28.
44. Morgan JF, Reid F, Lacey JH. The SCOFF questionnaire: assessment of a new screening tool for eating disorders. BMJ 1999 Dec 4;319(7223):1467-1468 [FREE Full text] [Medline: 10582927]
45. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 06;146(5):317-325. [Medline: 17339617]
46. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
47. Knight JR, Sherritt L, Shrier LA, Harris SK, Chang G. Validity of the CRAFFT substance abuse screening test among adolescent clinic patients. Arch Pediatr Adolesc Med 2002 Jun;156(6):607-614. [Medline: 12038895]
48. Sanci L, Grabsch B, Chondros P, Shiell A, Pirkis J, Sawyer S, et al. The prevention access and risk taking in young people (PARTY) project protocol: a cluster randomised controlled trial of health risk screening and motivational interviewing for young people presenting to general practice. BMC Public Health 2012 Jun 06;12:400 [FREE Full text] [doi: 10.1186/1471-2458-12-400] [Medline: 22672481]
49. Haller DM, Meynard A, Pejic D, Sredic A, Huseinagic S, Courvoisier DS, et al. YFHS-WHO+ Questionnaire: validation of a measure of youth-friendly primary care services. J Adolesc Health 2012 Nov;51(5):422-430. [doi: 10.1016/j.jadohealth.2012.01.019] [Medline: 23084162]
50. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
51. Langley G. The Improvement Guide: A Practical Approach to Enhancing Organizational Performance. San Francisco: Jossey-Bass; 1996.
52. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: L Erlbaum Associates; 1988.
53. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
54. Bradford S, Rickwood D. Psychosocial assessments for young people: a systematic review examining acceptability, disclosure and engagement, and predictive utility. Adolesc Health Med Ther 2012;3:111-125 [FREE Full text] [doi: 10.2147/AHMT.S38442] [Medline: 24600292]
55. Golsäter M, Sidenvall B, Lingfors H, Enskär K. Adolescents' and school nurses' perceptions of using a health and lifestyle tool in health dialogues. J Clin Nurs 2011 Sep;20(17-18):2573-2583. [doi: 10.1111/j.1365-2702.2011.03816.x] [Medline: 21752132]
56. Thompson LA, Wegman M, Muller K, Eddleton KZ, Muszynski M, Rathore M, et al. Improving adolescent health risk assessment: a multi-method pilot study. Matern Child Health J 2016 Dec;20(12):2483-2493 [FREE Full text] [doi: 10.1007/s10995-016-2070-5] [Medline: 27406154]
57. Sanci L, Chondros P, Sawyer S, Pirkis J, Ozer E, Hegarty K, et al. Responding to young people's health risks in primary care: a cluster randomised trial of training clinicians in screening and motivational interviewing. PLoS One 2015;10(9):e0137581 [FREE Full text] [doi: 10.1371/journal.pone.0137581] [Medline: 26422235]
58. Oyserman D. Culture as situated cognition: cultural mindsets, cultural fluency, and meaning making. Eur Rev Soc Psychol 2011 Nov 15;22(1):164-214. [doi: 10.1080/10463283.2011.627187]
59. Alden DL, Friend J, Schapira M, Stiggelbout A. Cultural targeting and tailoring of shared decision making technology: a theoretical framework for improving the effectiveness of patient decision aids in culturally diverse groups. Soc Sci Med 2014 Mar;105:1-8. [doi: 10.1016/j.socscimed.2014.01.002] [Medline: 24606791]
60. Jiang Z, Chan J, Tan BCY, Chua WS. Effects of interactivity on website involvement and purchase intention. J Assoc Inf Syst 2010;11(1):34-59.
61. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]

Abbreviations

GP: general practitioner
PCP: primary care provider
SMS: short message service
TAU: treatment as usual

XSL•FO
RenderX

Edited by G Eysenbach; submitted 02.04.17; peer-reviewed by S Bradford, D Alden; comments to author 11.05.17; revised version received 03.06.17; accepted 12.07.17; published 11.08.17.

Please cite as:
Webb MJ, Wadley G, Sanci LA
Improving Patient-Centered Care for Young People in General Practice With a Codesigned Screening App: Mixed Methods Study
JMIR Mhealth Uhealth 2017;5(8):e118
URL: http://mhealth.jmir.org/2017/8/e118/
doi:10.2196/mhealth.7816
PMID:28801302

©Marianne Julie Webb, Greg Wadley, Lena Amanda Sanci. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 11.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Feasibility and Acceptability of a Text Message-Based Smoking Cessation Program for Young Adults in Lima, Peru: Pilot Study

Dora Blitchtein-Winicki[1,2], MD, MPH, DrPH; Karine Zevallos[1,3], MD, MPH; M Reuven Samolski[1], MD(Psych); David Requena[1], MSc, MScBMI; Chaska Velarde[1], BSocSci; Patricia Briceño[1], BComm; Marina Piazza[1,4], MPH, ScD; Michele L Ybarra[5], MPH, PhD

[1]Mental Health, Alcohol and Drug Unit, Public Health Department, Universidad Peruana Cayetano Heredia, Lima, Peru

[2]Executive Office of Research, Peruvian National Institute of Health, Lima, Peru

[3]Centro de Investigación en Enfermedades Tropicales "Maxime Kuczynski", Peruvian National Institute of Health, Loreto, Peru

[4]Peruvian National Institute of Health, Lima, Peru

[5]Center for Innovative Public Health Research, San Clemente, CA, United States

Corresponding Author:
Dora Blitchtein-Winicki, MD, MPH, DrPH
Executive Office of Research
Peruvian National Institute of Health
Cápac Yupanqui 1400
Jesus María
Lima, CP 11
Peru
Phone: 51 999 090917
Fax: 51 748 1111
Email: dblit2007@gmail.com

Abstract

Background: In Peru's urban communities, tobacco smoking generally starts during adolescence and smoking prevalence is highest among young adults. Each year, many attempt to quit, but access to smoking cessation programs is limited. Evidence-based text messaging smoking cessation programs are an alternative that has been successfully implemented in high-income countries, but not yet in middle- and low-income countries with limited tobacco control policies.

Objective: The objective was to assess the feasibility and acceptability of an short message service (SMS) text message-based cognitive behavioral smoking cessation program for young adults in Lima, Peru.

Methods: Recruitment included using flyers and social media ads to direct young adults interested in quitting smoking to a website where interested participants completed a Google Drive survey. Inclusion criteria were being between ages 18 and 25 years, smoking at least four cigarettes per day at least 6 days per week, willing to quit in the next 30 days, owning a mobile phone, using SMS text messaging at least once in past year, and residing in Lima. Participants joined one of three phases: (1) focus groups and in-depth interviews whose feedback was used to develop the SMS text messages, (2) validating the SMS text messages, and (3) a pilot of the SMS text message-based smoking cessation program to test its feasibility and acceptability among young adults in Lima. The outcome measures included adherence to the SMS text message-based program, acceptability of content, and smoking abstinence self-report on days 2, 7, and 30 after quitting.

Results: Of 639 participants who completed initial online surveys, 42 met the inclusion criteria and 35 agreed to participate (focus groups and interviews: n=12; validate SMS text messages: n=8; program pilot: n=15). Common quit practices and beliefs emerged from participants in the focus groups and interviews informed the content, tone, and delivery schedule of the messages used in the SMS text message smoking cessation program. A small randomized controlled pilot trial was performed to test the program's feasibility and acceptability; nine smokers were assigned to the SMS text message smoking cessation program and six to a SMS text message nutrition program. Participant retention was high: 93% (14/15) remained until day 30 after quit day. In all, 56% of participants (5/9) in the SMS text message smoking cessation program reported remaining smoke-free until day 30 after quit day and 17% of participants (1/6) in the SMS text message nutrition program reported remaining smoke-free during the entire program. The 14 participants who completed the pilot reported that they received valuable health information and approved the delivery schedule of the SMS text messages.

XSL•FO
RenderX

Conclusions: This study provides initial evidence that a SMS text message smoking cessation program is feasible and acceptable for young adults residing in Lima.

(JMIR Mhealth Uhealth 2017;5(8):e116) doi:10.2196/mhealth.7532

KEYWORDS

Pilot Projects, Text Messaging, Smoking Cessation, Young Adult, Cognitive Therapy, Feasibility Studies, Latinos

Introduction

Smoking tobacco and exposure to particulate matter and carcinogenic agents in tobacco smoke have detrimental effects on health [1,2] leading to a greater incidence of cancer; cardiac, cerebrovascular, and respiratory failure; eye and skin damage; and fertility problems [1]. Smoking is highly addictive due to pharmacological [3] and psychological factors, making it difficult to quit [4,5]. Studies have demonstrated that interventions for smoking cessation can be effective [6,7] at reducing morbidity and mortality related to tobacco use [6,8-10].

Increased awareness of tobacco-related morbidity and mortality has led many countries to implement policies, control measures, and interventions for smoking cessation [5,7], which have helped to reduce the prevalence of smoking in developed countries [11]. However, developing countries such as Chile, Costa Rica, Jamaica, Suriname, and Peru [12] have maintained a high prevalence of tobacco smoking: approximately 20% to 30% in men and 10% to 20% in women [13]. In urban areas of Peru, tobacco smoking begins between 12 and 18 years. Young adults have a higher prevalence rate. In 2013, 3 in 10 individuals aged between 19 and 28 years in Lima smoked at least one cigarette during the year compared to 1 in 10 individuals aged between 12 and 65 years [14]. Between 6 and 7 of 10 adolescents in urban areas of Peru who smoke tobacco have tried to quit, most of whom were not successful [15].

Mobile phones can be effective public health interventions due to their ubiquity and relative ease of use. Delivering text messages is program-initiated (one-way) and proactive, reaching the user wherever he or she happens to be, with minimal barriers (the mobile phone notifies the user when a short message service [SMS] text message is received and SMS text messages may automatically appear on the screen) [16-19]. In addition, using cell phones avoids the high cost of in-person interactions with health personnel and facilitates the efficient collection and processing of information. Messages can be adapted based on the user's characteristics and delivered at any time of day [20,21]. It is also cost-effective and can reach a large number of people in a large geographical area [22].

Most people in Peru own mobile phones. In 2014, there were approximately 104 mobile phones for every 100 inhabitants according to the Ministry of Communications and Transportation [23]. In particular, young people are more prone to carry mobile phones and use text messages. In Peru, the use of mobile technology has demonstrated great potential as a cost-effective health strategy to improve access to health care [24].

There is evidence that SMS text messaging programs using cognitive behavioral therapy (CBT) for smoking cessation are effective, increasing its likelihood of success by 35% in comparison with control groups [17,25-27]. Its success rate is similar to other interventions, such as helpline services to quit smoking, and face-to-face cognitive behavioral and cognitive therapy and counseling [25,28], with the additional benefit of being delivered at a lower cost [17]. Compared with other treatment methods, there is evidence that SMS text message-only interventions affect 6-month cessation outcomes similarly to other effective treatments [26,29-31]. Likewise, a meta-analysis that included 20 manuscripts with 22 interventions showed that SMS text message cognitive behavioral smoking cessation programs can be an effective strategy to help individuals stop smoking [27].

Intervention strategies for smoking cessation should consider its target population [30,32]. Attitudes and behaviors related to tobacco use are often dependent on factors such as gender, race, and age, among others. For example, most Latinos prefer nonpharmacological methods to quit smoking (eg, replacement of nicotine, bupropion, varenicline) [33,34], whereas a lower proportion of young people decide to use these therapies compared with other age groups [35].

Theoretical Frameworks

The theoretical frameworks that guided the development of the SMS text message cognitive behavioral smoking cessation program were the transtheoretical model, CBT, increasing self-efficacy, and relapse prevention. As defined by the transtheoretical model, behavior change is a process over time and involves progression through a series of stages: precontemplation, contemplation, preparation, action, and maintenance [36,37]. Program content was based on CBT and increasing self-efficacy, and program components identified in smoking cessation telephone-based counseling (quitlines) [4,30,38-41]. CBT focuses on cognitive restructuring, education, self-monitoring, and practical coping strategies aimed at successful smoking cessation. CBT implies that mental rules guide behavior, such that applying illogical beliefs (cognitive processes) lead to dysfunctional behavior (behavioral actions). To acquire the capability to quit smoking, individuals need to recognize, understand, and change their behavior patterns. Increasing self-efficacy was also considered a key element in developing the SMS text message cognitive behavioral program content [42-44] because it includes constructs such as perceived temptations and self-regulatory strategies increases relapse prevention [45] and low self-efficacy may undermine the ability to maintain or initiate efforts to cope with high-risk situations [46-48]. Relapse prevention is a cognitive behavioral approach to relapse with the goal of identifying and preventing high-risk situations; in the case of smoking cessation, preventing smoking after a period of abstinence [49,50]. Relapse prevention includes adapting coping responses to life stressors and risky situations by minimizing exposure to cues. For example, understanding

XSL•FO
RenderX

that discomfort will reach a peak and then subside, encouraging patience to wait out the urge, using nicotine replacement treatment (NRT), and seeking health professional advice about other pharmacological evidence-based treatments [51-53].

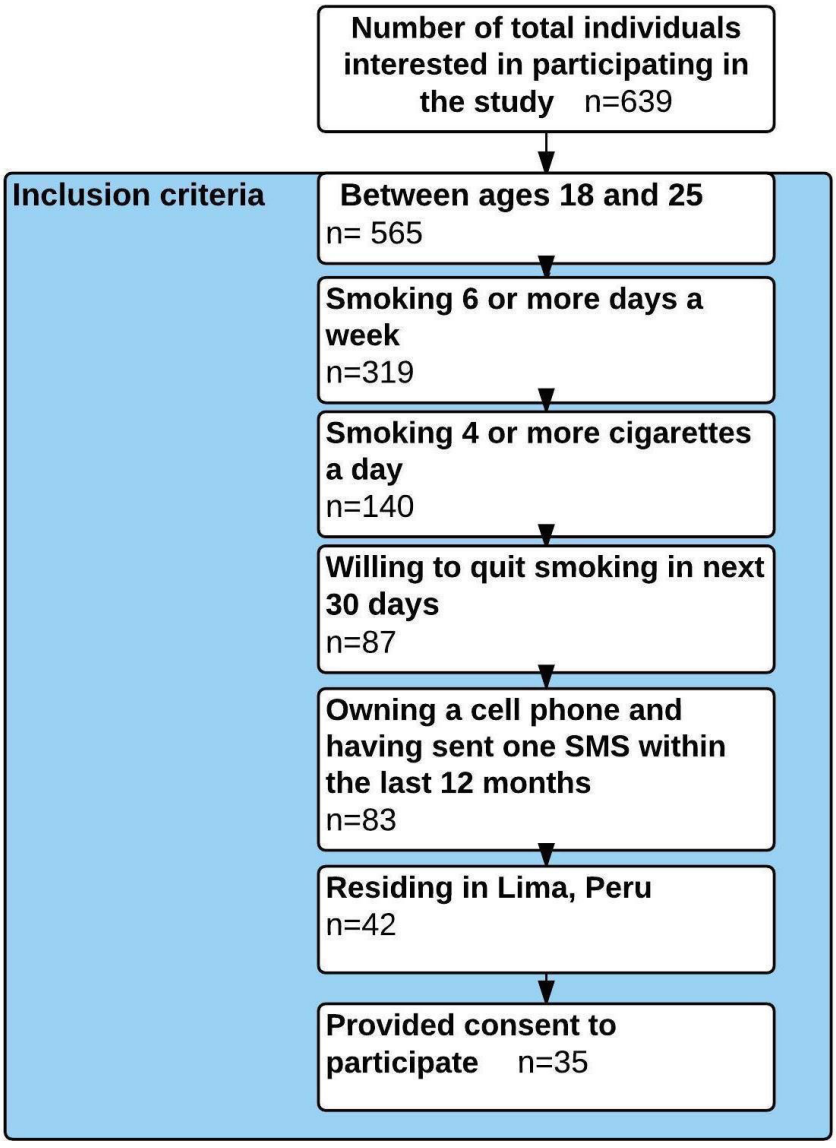
The objective of this study is to evaluate the feasibility and acceptability of an SMS text message cognitive behavioral program for smoking cessation among young people aged 18 to 25 years who live in Lima, Peru. In this study, we report on feasibility and acceptability of an SMS text message cognitive behavioral program for smoking cessation and develop initial content and methodology in preparation for a randomized controlled trial (RCT).

Methods

Recruitment

Participants were recruited through advertisements on Facebook and Google, a Facebook fan page, and flyers in locations where young people frequent (eg, cafeterias, stores, pubs, technical schools, universities, near public transportation), which directed them to a website with smoking cessation information. Participants had to be interested in quitting smoking and seeking help to quit. Those interested in participating completed an online eligibility questionnaire on the website, which was hosted on Google Drive. Eligible participants were residents of Lima, between ages 18 and 25 years, smoked four or more cigarettes daily, had their own cell phone, and sent and received at least one SMS text message in the previous 12 months. We measured intention to quit by asking: "Are you seriously thinking of quitting smoking?" Response options were (1) "No, not thinking about quitting," (2) "Sometime, but not within the next 6 months," (3) "Yes, within the next 6 months," and (4) "Yes, within the next 30 days." Only those who selected the fourth option were eligible. Those who met the eligibility criteria were contacted by a member of the research team who provided a description of the study, answered questions, and asked participants to provide informed consent. Recruitment continued during the entirety of the study. Figure 1 shows the eligibility criteria used to evaluate individuals interested in participating in the study. Content development and technical development are discussed in further detail subsequently.

Figure 1. Flowchart of individuals included in the study based on inclusion criteria.



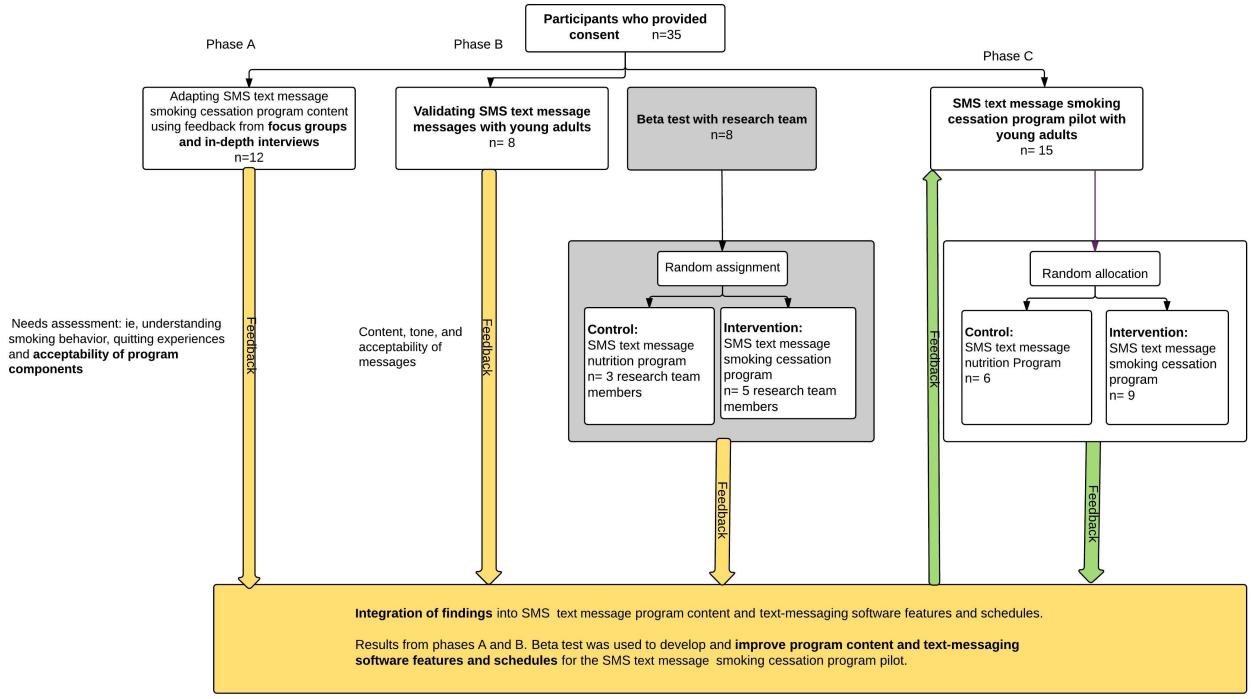
XSL•FO
RenderX

All participants who met the inclusion criteria were included in only one of the following study phases, which were conducted consecutively:

1. Phase A: adapting SMS smoking cessation and nutrition program content with focus groups and in-depth interviews;
2. Phase B: individual sessions to validate SMS text messages with young adults; and
3. Phase C: SMS text message smoking cessation program pilot where participants were randomly assigned to the SMS text message smoking cessation program (intervention) or the SMS text message nutritional program (control).

Figure 2 describes the main components of the study and the number of participants in each phase. All participants provided informed consent. Prior to phase C, the smoking cessation program was piloted with members of the research team. The pilot study was conducted between July 2013 and November 2014. Recruitment lasted 10 months.

Figure 2. Program outline and participation in different phases of the study.



Content Development

General Characteristics of the SMS Text Message Smoking Cessation Program (Intervention)

The smoking cessation SMS text message program content consisted of tailored messages and pathways based on individual assessment [54]. Program content was tailored to create individualized intervention strategies according to gender, stage of quitting (prequit, quit, relapse), and smoking status at different stages of the study. Likewise, participants received bidirectional messages, which allowed them to provide self-reported smoking cessation data on days 2, 7, and 30 after quit date. Participants received tailored sets of messages depending on their responses, which were sent automatically using algorithms.

Content was developed based on the theoretical frameworks described previously, and participant feedback through focus groups and in-depth interviews to validate the SMS text messages with the target population. Examples of the type of information included in the SMS text message smoking cessation program are shown in Figure 3.

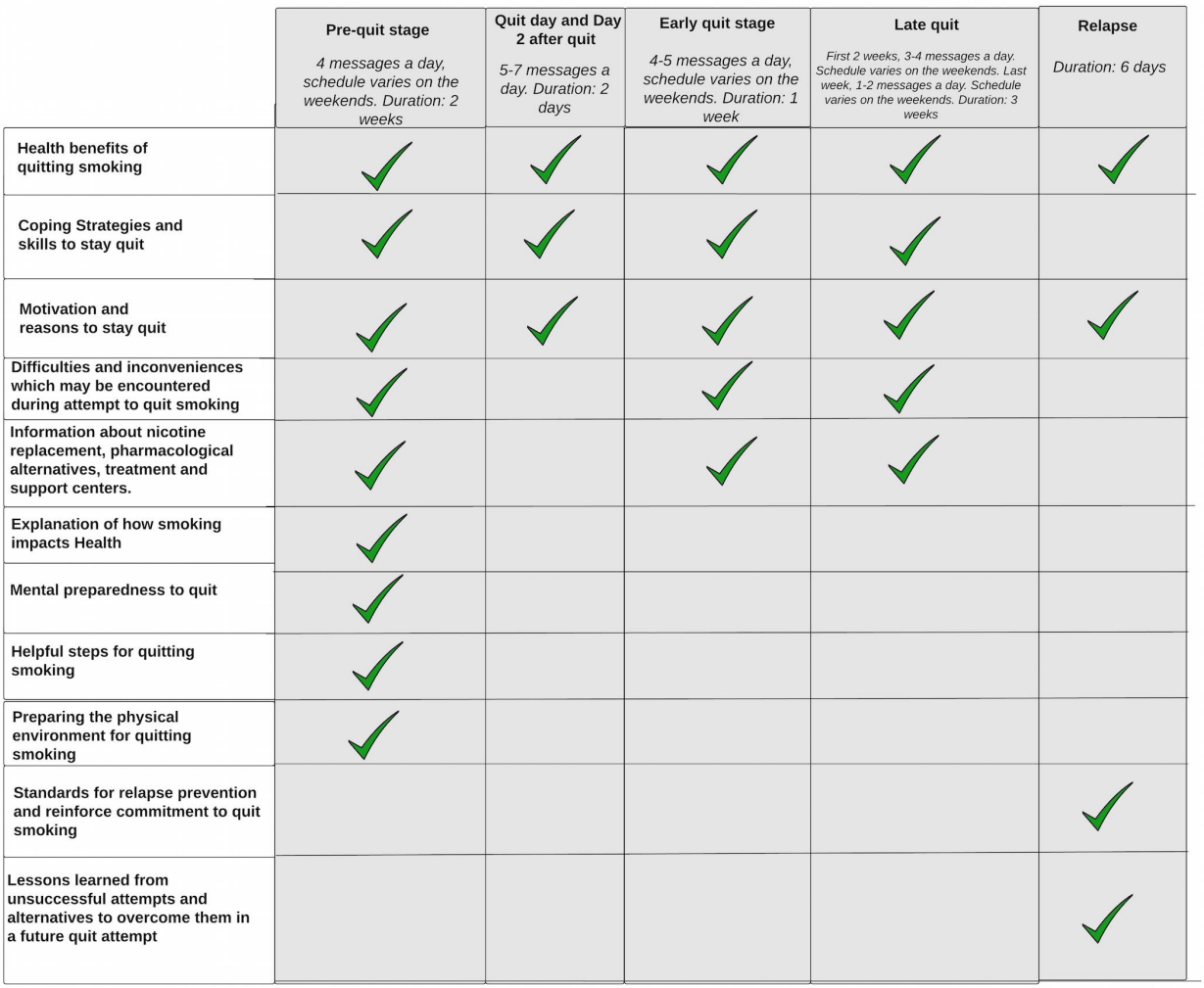
Examples of Types of Information Included in Different Stages of the SMS Text Message Cognitive Behavioral Smoking Cessation Program

Prequit Stage

1. Organizing a plan to quit smoking: setting a date; informing family and friends; asking for their support; identifying situations, triggers, and risks (eg weekends, friends, and social activities), and alternatives to cope with or avoid them; identifying social support resources; planning to reward oneself for achieving short-term smoking cessation goals; and educating oneself about how to avoid nicotine withdrawal symptoms.
2. Preparing the physical environment for quitting smoking: removing cigarettes and objects related to smoking.
3. Preparing mentally to quit smoking: for example, debunking myths such as “smoking is ok as long as I engage in other healthy behaviors” or “e-cigarettes are the most effective method to quit smoking.” Messages that promote self-confidence and identifies benefits and reasons to quit and stay tobacco free.
4. Motivational content that reinforces commitment to quit.

XSL•FO
RenderX

Figure 3. Types of information included in each stage of the SMS text message smoking cessation program.



Quit Day and Day 2 After Quit

1. Coping strategies and skills to stay quit: activities and suggestions that help to stay quit and meet short-term quit goals.
2. Motivation: celebrate successes and reinforce commitment to quit.

Early Quit Stage

1. Coping strategies and skills to stay quit: reminders of reasons for quitting; engaging in other enjoyable activities; how to cope with withdrawal side effects; helpful tips, such as keeping hydrated; and capitalizing on prior successful actions and strategies to stay quit.
2. Difficulties and inconveniences: for instance, how to cope with common withdrawal symptoms, what to expect and how to face them, and when they are expected to appear.

Late Quit

1. Coping strategies and skills to stay quit: address concerns such as gaining weight and how to manage stress without smoking, how to identify and handle relapse triggers, and how to deal with cravings.
2. Motivation to stay quit: planning activities to celebrate each milestone achieved, preparing to continue, perceiving self as a nonsmoker, and continuing setting short-term goals.
3. Difficulties and inconveniences: for instance, recognizing gains in health and information about withdrawal symptoms and timeline.

Relapse

1. Standards for relapse prevention and commitment to quit smoking: positive reminders of achievements; information about how successfully quitting smoking may require several attempts, but is achievable; reinforcing messages about benefits of quitting and motivations and steps for continuing to try quitting.
2. Lessons learned from an unsuccessful attempt and alternatives to overcome them in a future quit attempt: identify reasons for relapse and employ prevention strategies in next quit attempt, and encouragement to try quitting smoking again within a short period of time (one month).

General Characteristics of the SMS Text Message Nutrition Program (Control)

The nutrition text messages were designed to motivate healthy eating and physical activity in order to achieve health benefits.

XSL•FO
RenderX

Messages included increasing vegetable intake, decreasing sweetened beverage consumption, and improving eating behaviors and physical activity.

Focus Groups and In-Depth Interviews

We held two focus groups and four in-depth interviews with 12 participants with the aim to provide insight into young adult smoking behavior, SMS text message-related behavior and preferences, experiences with quitting smoking, and perceived reasons for prior unsuccessful smoking attempts. The themes explored included participants' initial smoking experiences, smoking practices, smoking-related perceptions, reasons for continuing smoking, previous attempts to quit smoking, reasons for deciding to quit, the influence of close friends and family on quitting, perceived consequences of smoking, strategies and activities to stay quit, knowledge and prior use of alternative tobacco products, use of mobile phones for text messaging, acceptability of using SMS text messages to quit smoking, and preferences regarding the type, frequency, and content of SMS text messages they would like to receive. Their answers were audiotaped, transcribed, and analyzed and considered in the adaptation and development of the program content.

Validating SMS Text Messages With Young Adults

Once we had a final pool of messages, eight participants were recruited to review a set of messages from the proposed SMS text message smoking cessation program. A member of the research team met for 1.5 to 2 hours with each participant individually and explained the purpose of the SMS text message program, and the role and importance of using messages in the program. Each participant was given a list of 60 messages to review between the initial meeting and a follow-up meeting 7 to 10 days later, using a guide provided by the researchers. During the follow-up meeting, they were requested to provide feedback on the messages. At least two participants were asked to review each SMS text message to ensure that all the messages had been reviewed. Each participant received messages from all the stages in the program (prequit, quit day, early quit, late quit, and relapse). Participants were asked to provide feedback on the following: was the message understandable, how did they perceive the tone of the message, was the message phrased in a way that young adults could relate, and did the messages provide relevant information according to the participant's smoking status. Four research team members modified the SMS text messages based on participant feedback.

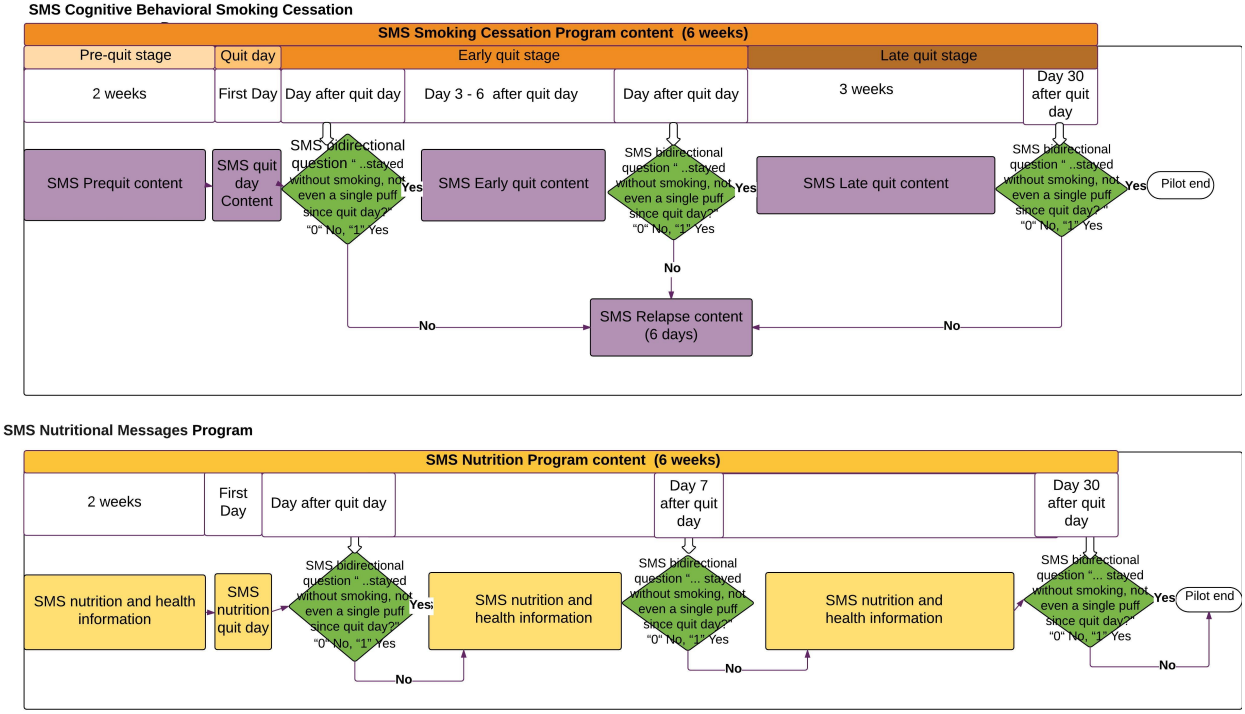
Participants received a code of prepaid text message credit equivalent to US $6 after each face-to-face meeting (two meetings per participant).

Technological Development

The automated text messaging system was developed using the following components: list of SMS text messages for both programs, list of participants' cellphone numbers, and randomizing methodology and tailoring criteria for the SMS text message smoking cessation program and the SMS text message nutrition program. The system used a script written in Python 2.7 in the app SMS Scheduler [55] to manage, schedule, and send SMS messages. The system was configured on a Samsung Galaxy S5 mini with Android 4.4 (KitKat).

On days 2, 7, and 30 after the participant's quit day, a message was sent to participants asking them about their ability to remain smoke-free. Participants answered 1 ("yes") or 0 ("no"), which the system registered and used to automatically assign the next set of messages to be received by the participant in the SMS text message smoking cessation program. In the SMS text message nutrition program, participants received the same set of messages regardless of their response, as shown in Figure 4.

Figure 4. Flowchart of the SMS text message smoking cessation program and the SMS text message nutrition program pathways.



XSL•FO
RenderX

Beta Test of the SMS Text Message Smoking Cessation Program With the Research Team

Before piloting the SMS text message smoking cessation program with the participants, it was tested by members of the research team over 6 weeks to troubleshoot any technological difficulties. Three research team members received the SMS text message nutrition program (control) and five members the SMS smoking cessation program (intervention). During the beta test, various difficulties were addressed: the software was corrected to include letters and symbols used in the Spanish language; difficulties with scheduled SMS text message delivery were solved by sending SMS text messages separately and restricting simultaneous messages; due to constant changes in cell phone ownership, the research team frequently updated the participants' contact information and encouraged participants to take advantage of mobile numerical portability (MNP), which allows users to retain their mobile telephone numbers independent of changes in cell phone or mobile network carrier.

Piloting the SMS Text Message Smoking Cessation Program with Young Adults

Fifteen participants were recruited to test the feasibility and acceptability of the SMS text message smoking cessation program. The participants were randomly assigned to the SMS text message smoking cessation program (nine participants) or the nutrition program (six participants). Only two women were recruited for the pilot; one woman was assigned to each arm through randomized assignment. Every week, participants sent feedback about the content and timing of the SMS text messages received. Feedback received was used to measure acceptability of the program among participants.

The outcome—successfully quitting on days 2, 7, and 30 after quit day—was measured using self-reports of smoking cessation for both arms of the study. We sent an SMS bidirectional question on days 2, 7, and 30 after the participant's quit day, asking "Have you stayed tobacco free (not even a single puff) since the day you quit?" If yes, the participant answered "1" and if no, the answer was "0." The answer allowed the program to assign the next set of messages that the participant received as shown in Figure 4. The research assistant called to remind the participant to answer the SMS text question the same day the bidirectional question was sent. This approach was effective for almost all the participants, except for one who left the country during the study.

The participants were asked to provide feedback weekly to minimize recall bias. They sent their comments through SMS text message and they received the equivalent of US $3 in credit for mobile calls and SMS text message through a code sent to them by SMS text message to compensate for the cost of sending SMS text messages related to the program.

To assess the feasibility we considered issues related to ease of implementation of the SMS text message pilot study, and any technical problems and issues which arose from taking part in the intervention or control arm as other SMS text message studies have previously measured [16,20,56,57].

We included the following measures:

1. Recruitment rate: ability to recruit our target sample size of 75 participants from Peru (of which 40 were residents of Lima) in the time allotted for the project, which was 10 months.
2. Retention rates: measures included (1) within the SMS text message smoking cessation program, whether participants responded to queries about their smoking status on days 2, 7, and 30 after quit day, and (2) in the research, to be able to achieve the follow-up rates necessary to conduct a RCT at day 30 after quit day (93% of 15 participants).
3. Performance of the software that delivers the program's text messages (eg, successful delivery of messages to all designed mobile phone carriers).

To assess the acceptability of the program, we considered consent and response rates during the entirety of the study, and comments and feedback about their experiences with the study [16,57,58].

We included the following measures: participant responses to the text messages, feedback from participants, the number of participants who asked to be removed from the program during the intervention phase, and a series of questions about the program's characteristics and likability, administered one month after the quit day.

This pilot study was approved by the institutional review boards of Michigan State University and Universidad Peruana Cayetano Heredia.

Statistical Analysis

All data during the SMS text message smoking cessation program pilot was entered into a Microsoft Excel 2011 database. Data was cleaned and checked for disparities before and after the data entry. Descriptive statistics (percentage, mean, median, standard deviation) were used to describe demographic, smoking-related characteristics, SMS text message use, nicotine dependence, and smoking cessation self-report outcome of respondents. Descriptive analyses were performed using STATA (12.0 version). Nicotine dependence was classified using the Fagerström Test of Nicotine Dependence (FTND) score and used the following categories: very low/none, low, medium, and high [3].

Results

Participants

A total of 639 individuals answered our online screening survey; 565 met the inclusion age criteria (18-25 years), 319 reported having smoked 6 to 7 days a week (frequency), 140 smoked four or more cigarettes a day (intensity), 87 had decided to quit smoking in the next 30 days, 46 lived in Lima, and two did not have a mobile phone (Table 1). Forty-two participants had sent at least one SMS text message in the previous 12 months of which 35 consented to participate in the pilot study: focus groups and in-depth interviews (n=12), young adult participants to validate messages (n=8), and the validation of the SMS text message smoking cessation program (n=15). We recruited during the entirely of the study using paid Facebook ads to increase recruitment rates, recruiting 70 individuals with the

XSL•FO
RenderX

inclusion criteria at the national level, of which 42 were Lima residents.

XSL•FO
RenderX

Table 1. Characteristics of participants in the SMS text message smoking cessation program pilot.

Characteristics of participants	Smoking cessation (intervention), n (%) (n=9)	Nutrition (control), n (%) (n=6)	Total, n (%) (N=15)
Age			
18-19	1 (11)	2 (33)	3 (20)
20-21	4 (44)	3 (50)	7 (47)
22-23	4 (44)	1 (17)	5 (33)
Gender			
Female	1 (11)	1 (17)	2 (13)
Male	8 (89)	5 (83)	13 (87)
Education level			
Secondary education completed	1 (11)	0 (0)	1 (7)
Technical studies not completed	1 (11)	2 (33)	3 (20)
University education not completed	6 (67)	4 (67)	10 (67)
University completed	1 (11)	0 (0)	1 (7)
Carrier			
Claro	7 (78)	4 (67)	11 (73)
Movistar	2 (22)	1 (17)	3 (20)
Entel	0 (0)	1 (17)	1 (7)
Service plan			
Prepaid	2 (22)	0 (0)	2 (13)
Postpaid	7 (78)	6 (100)	13 (87)
Average number of SMS text messages received per day at the beginning of the study			
≤5	5 (56)	3 (50)	8 (53)
6-10	2 (22)	2 (33)	4 (27)
≥11	2 (22)	1 (17)	3 (20)
When do you check your SMS text messages?			
As soon as it arrives	8 (89)	5 (83)	13 (87)
During the afternoon	0 (0)	1 (17)	1 (7)
At night	1 (11)	0 (0)	1 (7)
Dependency level (FTND scale)			
Very low	5 (56)	4 (67)	9 (60)
Low	2 (22)	0 (0)	2 (13)
Mild	1 (11)	2 (33)	3 (20)
High	1 (11)	0 (0)	1 (7)
Do you live with other people who smoke?			
No	5 (56)	2 (33)	7 (47)
Yes	4 (44)	4 (67)	8 (53)
Have you had problems with alcohol and other drugs?			
No	5 (56)	4 (67)	9 (60)
Yes	4 (44)	2 (33)	6 (40)
Do you have a medical problem?			
No	7 (78)	4 (67)	11 (73)
Yes	2 (22)	2 (33)	4 (27)

XSL•FO
RenderX

Characteristics of participants	Smoking cessation (intervention), n (%) (n=9)	Nutrition (control), n (%) (n=6)	Total, n (%) (N=15)
Medical problem			
Asthma	0 (0)	1 (17)	1 (7)
Back pain	1 (11)	0 (0)	1 (7)
Prediabetes	1 (11)	0 (0)	1 (7)
Renal inflammation nephritis	0 (0)	1 (17)	1 (7)
Have you tried to quit smoking before? (yes)	9 (100)	6 (100)	15 (100)
Number of previous quit attempts			
1	1 (11)	1 (17)	2 (13)
2	2 (22)	2 (33)	4 (27)
3	4 (44)	1 (17)	5 (33)
4	2 (22)	2 (33)	4 (27)
Have you used some kind of aid when you tried to quit?			
No	5 (56)	5 (83)	10 (67)
Yes	4 (44)	1 (17)	5 (33)
What type of aid?			
Chewing gum	1 (11)	0 (0)	1 (7)
E-cigarette	2 (22)	1 (17)	3 (20)
Book	1 (11)	0 (0)	1 (7)
What's the longest time you've been smoke-free (not even a puff), from any prior quit attempt?			
1 day	0 (0)	1 (17)	1 (7)
3 days	3 (33)	0 (0)	3 (20)
1 week	1 (11)	2 (33)	3 (20)
2 weeks	0 (0)	1 (17)	1 (7)
3 weeks	1 (11)	0 (0)	1 (7)
1 month	1 (11)	0 (0)	1 (7)
2 months	0 (0)	1 (17)	1 (7)
3 months	3 (33)	1 (17)	4 (27)
Quit report "not even a puff" on day 7 after quit[a]			
No (relapse)	4 (44)	2 (33)	6 (40)
Yes	5 (56)	4 (67)	9 (60)
Quit report "not even a puff" on day 7 after quit[a]			
No (relapse)	0 (0)	3 (50)	3 (20)
Yes	5 (56)	1 (17)	6 (40)
Quit report "not even a puff" on day 30 after quit day[a]			
No (relapse)	0 (0)	0 (0)	0 (0)
Yes	5 (56)	1 (17)	6 (40)

[a] Participants answered 0="no" or 1="yes." Each participant could only answer "no" once. On answering no, they were sent messages from the relapse stage for 6 days.

XSL•FO
RenderX

Adapting SMS Text Message Smoking Cessation and Nutrition Program Content with Focus Groups and in-Depth Interviews

Findings from the focus group discussions and in-depth interviews demonstrated that the majority of the participants (n=12) who wanted to quit smoking did not have a plan. Most had tried previously but were unsuccessful. What they least liked about smoking was the smell, increased probability of having wrinkles and yellow teeth, and the shortness of breath during regular activities such as climbing stairs. They were unfamiliar with evidence-based alternatives and were not be willing to take medication. They were willing to try nicotine patches, but they did not know where to get them; cost was also a barrier. Many considered the e-cigarette an easily accessible alternative to cigarettes and knew acquaintances who had used it to quit.

The biggest obstacle to quitting was smoking triggers during the weekends: meeting friends who smoke, being at bars and parties, and being around people who smoke. One of the most common reasons for individuals to not quit smoking was the fear of gaining weight.

From the results of the focus groups, messages were added or modified to address participant concerns and triggers (see Multimedia Appendix 1). For instance, one theme that emerged was the idea that healthy behaviors compensated for the negative health effects of smoking. Another main concern about quitting was weight gain. We included messages related to the following: including exercise in their plan to quit, consuming low-calorie foods, drinking lots of water, and focusing all their efforts on quitting (only after they successfully quit smoking should they focus on other goals).

Validating SMS Text Messages with Young Adults

Once we had a final pool of messages, we recruited eight participants to validate the messages. The group read and provided feedback on each text message. Recommendations included rewording the messages, expressions and tone as if a peer had sent the messages, sending different messages for men and women, personalizing certain messages by including their name, and sending messages during certain hours or days of the week (eg, sending messages on the weekends about alternative enjoyable activities instead of going to a bar with friends). At least two participants reviewed each message. Suggestions were reviewed by four research team members and were used to improve the program messages and delivery schedule. Examples of the messages are found in Multimedia Appendix 2.

Beta Test With Research Team

Overall the software program and messaging infrastructure (eg, mobile phone and server) were reliable. Text messages were sent to eight research team members during the beta test. Certain issues arose such as the initial inability to use accented letters and Latin characters, inability to send multiple messages simultaneously, and frequent changes in mobile network operators. We identified strategies to overcome these barriers, such as finding an alternative to include accented letters and Latin characters, and asked participants to take advantage of MNP when their mobile phone was lost or stolen. MNP allows an individual to retain the same phone number regardless of changing from one mobile network carrier to another was an advantage.

SMS Text Message Smoking Cessation Program Pilot

Of the 15 young adults who provided consent to participate in the SMS text message smoking cessation program, more than 87% (13/15) were male, and all reported having attempted to quit smoking in the past on at least two previous occasions (Table 1). Three young adults (20%) used an e-cigarette as an aid to quit smoking in previous attempts; 53% (8/15) lived with other smokers and 40% (6/15) reported previous consumption of alcohol and/or drugs. The level of nicotine dependence was measured with the FTND. Based on the FTND score, nine (60%) had very low dependence, two (13%) had low, three (20%) had mild, and one (7%) had high.

The intervention retention rate was high: 14 of 15 participants (93%) remained until day 30 after quit day. One participant traveled out of the country and could not continue. Fourteen participants were contacted weekly during their participation and on day 30 after quit day. We called all the participants at the end of the same day the question was sent. If they did not answer the question, we reminded them to answer the text question, on which the participant was assigned to the corresponding pathway.

All nine individuals assigned to the intervention to quit smoking completed the program. Length of participation in the program was dependent on their answers regarding quitting smoking; for instance, if the participant responded “no” to “Have you stayed tobacco free (not even a puff) since the day you quit?” then they were assigned to receive messages with content from the relapse stage for 6 days. The control arm was completed by five of six (83%) of the participants.

All participants in the intervention group (n=9) stated that they received valuable information about health issues, treatment resources, NRT, and strategies to avoid smoking. Furthermore, 78% (7/9) reported that the messages were motivating and easy to understand, 89% (8/9) reported that they would recommend the program to others, 78% (7/9) agreed with the number of messages sent during the duration of the program, and 100% (9/9) agreed with the number and schedule for delivering messages. In all, 33% (3/9) suggested increasing the number of messages in the late quit stage and 11% (1/9) suggested including a partner to share messages when necessary.

Of the nine participants in the SMS text message smoking cessation program, 56% (5/9) reported having remained smoke-free (not even a puff) during the entire program (until day 30 after quit day); 44% (4/9) reported relapsing on day 2 after quit day. However, of six participants receiving nutritional messages, one (17%) reported having remained smoke-free during the entire program; 33% (2/6) reported relapsing on day 2 after quit day and 50% (3/6) reported relapsing on day 7 after quit day.

XSL•FO
RenderX

Discussion

This pilot study suggests that a SMS text message cognitive behavioral smoking cessation program would be used by young adults in Lima, Peru. Main findings supporting this conclusion include high recruitment rate and intervention retention rate during the pilot of 93% (14/15), content and technological acceptability, and high self-reported abstinence rate at day 30 after quit day (40%, 6/15; 56%, 5/9 in smoking cessation SMS text message intervention arm).

This pilot demonstrated that it is possible to use low-cost methods (website and social media) to identify and recruit specific populations who may otherwise be difficult to recruit at an on-site location (young adults between ages 18 and 25 who smoke at least four cigarettes daily, etc) [59]. Web-based recruitment strategies are suggested to be associated with better abstinence outcomes than other recruitment strategies [27], probably due to the fact participants are more familiar with technology and technological interventions. Facebook ads may be less cost-effective in Peru than in other countries [60]; however, it may be more effective for recruitment at the national level. Compared with alternatives for pilot studies on smoking cessation, the number of participants recruited was sufficient [59].

Evidence shows that content adaptation is critical for the successful development of mobile behavioral interventions [61]. Relevant information about past experiences and behaviors that contribute to quit characteristics of young adults and Latino populations [33,62-66] was included into the SMS text message cognitive behavioral smoking cessation program content. Participant suggestions were used in designing and modifying the content, syntax, and tone of the SMS text message smoking cessation program content. Participants expressed preferences for messages that contained accurate spelling and grammar (eg, Latin characters), used positive versus negative framing, and were designed to help participants achieve personal goals. These findings coincide with results from other studies [61,67-69]. In addition, adapting the SMS text message cognitive behavioral smoking cessation program content to young adult and Latino populations [33], who have expressed resistance to the use of pharmacotherapy, was an important factor in the development and adaptation of program content to achieve behavioral change [70]. An advantage in Peru for SMS text message health program development and implementation is that there is no cost for receiving SMS text messages unlike other countries, which makes health strategies using SMS text messaging quite accessible.

Some of the preliminary findings from the focus groups and in-depth interviews are similar to findings from systematic reviews and studies in other countries [71,72]: participants perceive the use of e-cigarettes as an accessible, convenient aid in smoking cessation, and less harmful than cigarettes. This alternative can be considered as a resource if evidence (RCTs and population-based studies with precise exposure and safety measures) show that e-cigarettes are at least as effective as NRT in helping young adults smokers to quit [71].

Limitations

If we want to scale up the pilot, we need to consider strategies to continue and strengthen the high adherence shown in this preliminary results because it may be related to the provision of weekly credit for text messages which allowed participants to provide program feedback (93% overall participation for the entirety of the study; 100% in the intervention arm and 83% in the control arm). Interaction may have increased adherence and response [21]. Additional research is needed to measure adherence without a prepaid card: how much money are participants willing to spend and will they make an effort to continue in the program? Future studies may also help identify whether prepaid cards can help to improve adherence and/or results in long-term studies.

Self-reporting [16,73] (answering yes to “have not smoked even one puff since the quit day”) was used to measure the intervention’s effectiveness; nonetheless, this measure is not the most accurate way to measure abstinence and may produce a high level of negative self-reports among some populations [73,74]. Thus, it would be important to consider biochemical verification of smoking status as an outcome measure in future studies.

In future studies we may consider more accurate forms of measuring abstinence. One approach would be to ask individuals to approach a physical location to collect samples demonstrating abstinence. Another alternative would be to use strips to detect cotinine in the saliva, which can either be mailed by courier or have a photo of the result taken by the participant and sent by mobile phone or email to be read as abstinence confirmation. Portable meters that allow carbon monoxide measurement [75] have already been used in more limited geographical contexts such as universities [76]. This would enable us to request measurement at any time of the day in real time to prevent dose handling. All these measures should be considered for an RCT. Other forms of measuring abstinence include measurement of the saliva cotinine level, carbon monoxide confirmation, and cotinine analysis [73,74]. In addition, in future studies we want to know outcomes in terms of true smoking cessation, which requires increasing follow-up periods to at least 6 months [17,77].

This preliminary data show that a SMS text message cognitive behavioral smoking cessation program can be implemented and should be acceptable for young adults in Lima, Peru, in addition to contributing knowledge about the current evidence gap about appropriateness of a SMS text message smoking cessation program for young adults in a middle-income country with an inactive tobacco control policy [26,27].

The results from our pilot could broaden access to this type of health intervention and be adapted to help other groups quit smoking, such as younger populations, other health conditions (mental, chronic, cardiovascular, neoplastic and infectious diseases), other cultures, and rural areas.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank Manuel Catacora for his contribution during the beta trial and smoking cessation pilot and for helping develop the text messages. We also want to thank Manuel Le Noir for helping with recruitment through social media, Raquel Huamani for logistical support, Michelle Chan for helping with editing and valuable suggestions of data presentation, and Yomtov Levy for his technical suggestions during the recruitment and initial phases of the pilot. Financial support was from NIH Fogarty and NIDA D43TW00589 (DB,KZ,RS,MY) and K05DA015799 (JCA). This original contribution has not been published elsewhere. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse, the National Institutes of Health, Universidad Peruana Cayetano Heredia, or Michigan State University.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Some of the focus group and in-depth interviews analysis results taken in consideration to adapt characteristics of the smoking cessation SMS text message program.

[JPEG File, 1MB - mhealth_v5i8e116_app1.jpeg]

Multimedia Appendix 2

Examples of SMS text messages.

[JPEG File, 2MB - mhealth_v5i8e116_app2.jpeg]

References

1. Colombo G, Clerici M, Giustarini D, Portinaro NM, Aldini G, Rossi R, et al. Pathophysiology of tobacco smoke exposure: recent insights from comparative and redox proteomics. Mass Spectrom Rev 2014 Nov;33(3):183-218. [doi: 10.1002/mas.21392] [Medline: 24272816]
2. Mossina A, Lukas C, Merl-Pham J, Uhl FE, Mutze K, Schamberger A, et al. Cigarette smoke alters the secretome of lung epithelial cells. Proteomics 2017 Jan;17(1-2):1. [doi: 10.1002/pmic.201600243] [Medline: 27891773]
3. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [Medline: 1932883]
4. Perkins K, Conklin C, Levine M. In: Group TF, editor. Cognitive-Behavioral Therapy for Smoking Cessation: A Practical Guidebook to the Most Effective Treatments. New York: New York Routledge; 2008:1119-1127.
5. Perk J, De BG, Gohlke H, Graham I, Reiner Z, Verschuren M, Scholte op Reimer Wilma J M, European Association for Cardiovascular Prevention & Rehabilitation (EACPR), ESC Committee for Practice Guidelines (CPG). European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). Eur Heart J 2012 Jul;33(13):1635-1701 [FREE Full text] [doi: 10.1093/eurheartj/ehs092] [Medline: 22555213]
6. Anderson JE, Jorenby DE, Scott WJ, Fiore MC. Treating tobacco use and dependence. Chest 2002 Mar;121(3):932-941. [doi: 10.1378/chest.121.3.932]
7. Jha P, Peto R. Global effects of smoking, of quitting, and of taxing tobacco. N Engl J Med 2014 Jan 02;370(1):60-68. [doi: 10.1056/NEJMra1308383] [Medline: 24382066]
8. Stead LF, Hartmann-Boyce J, Perera R, Lancaster T. Telephone counselling for smoking cessation. Cochrane Database Syst Rev 2013;8:CD002850. [doi: 10.1002/14651858.CD002850.pub3] [Medline: 23934971]
9. Webb M, Rodríguez-Esquivel D, Baker E. Smoking cessation interventions among Hispanics in the United States: a systematic review and mini meta-analysis. Am J Health Promot 2010;25(2):109-118. [doi: 10.4278/ajhp.090123-LIT-25] [Medline: 21039291]
10. Mehring M, Haag M, Linde K, Wagenpfeil S, Schneider A. Effects of a guided web-based smoking cessation program with telephone counseling: a cluster randomized controlled trial. J Med Internet Res 2014 Sep 24;16(9):e218 [FREE Full text] [doi: 10.2196/jmir.3536] [Medline: 25253539]
11. Ng M, Freeman MK, Fleming TD, Robinson M, Dwyer-Lindgren L, Thomson B, et al. Smoking prevalence and cigarette consumption in 187 countries, 1980-2012. JAMA 2014 Jan 08;311(2):183-192. [doi: 10.1001/jama.2013.284692] [Medline: 24399557]
12. World Health Organization. WHO Report on the Global Tobacco Epidemic, 2013: Enforcing Bans on Tobacco Advertising, Promotion and Sponsorship. Geneva, Switzerland: World Health Organization; 2013.

XSL•FO
RenderX

13. Islami F, Stoklosa M, Drope J, Jemal A. Global and regional patterns of tobacco smoking and tobacco control policies. Eur Urol Focus 2015 Aug;1(1):3-16. [doi: 10.1016/j.euf.2014.10.001]
14. Comision Nacional para el Desarrollo y Vida sin Drogas, DEVIDA. I Estudio sobre prevencion y consumo de drogas en la poblacion general de Lima Metropolitana y el Callao – 2013. Lima, Peru: Observatorio Peruano de Drogas; 2014. URL: http://www.devida.gob.pe/wp-content/uploads/2015/02/I_Estudio_Lima_Callao_completo.pdf [accessed 2017-07-26] [WebCite Cache ID 6sFMHL8fy]
15. Zavaleta A, Salas M, Peruga A, Hallal AL, Warren CW, Jones NR, et al. Prevalence of smoking and other smoking related behaviors reported by the Global Youth Tobacco Survey (GYTS) in four Peruvian cities. BMC Public Health 2008 Dec 15;8 Suppl 1:S2 [FREE Full text] [doi: 10.1186/1471-2458-8-S1-S2] [Medline: 19091045]
16. Ybarra ML, Holtrop JS, Bağci BA, Bilir N, Korchmaros JD, Salih EA. Feasibility and acceptability of a text messaging-based smoking cessation program in ankara, Turkey. J Health Commun 2013 Aug;18(8):960-973. [doi: 10.1080/10810730.2012.757399] [Medline: 23627304]
17. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: a meta-analysis. J Subst Abuse Treat 2015 Sep;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]
18. Crutchfield T, Kistler C. Getting patients in the door: medical appointment reminder preferences. Patient Prefer Adherence 2017;11:141-150 [FREE Full text] [doi: 10.2147/PPA.S117396] [Medline: 28182131]
19. Serrano K, Yu M, Riley W, Patel V, Hughes P, Marchesini K, et al. Willingness to exchange health information via mobile devices: findings from a population-based survey. Ann Fam Med 2016;14(1):34-40 [FREE Full text] [doi: 10.1370/afm.1888] [Medline: 26755781]
20. Britto MT, Munafo JK, Schoettker PJ, Vockell AB, Wimberg JA, Yi MS. Pilot and feasibility test of adolescent-controlled text messaging reminders. Clin Pediatr (Phila) 2012 Feb;51(2):114-121. [doi: 10.1177/0009922811412950] [Medline: 22019792]
21. Douglas N, Free C. 'Someone batting in my corner': experiences of smoking-cessation support via text message. Br J Gen Pract 2013 Nov;63(616):e768-e776 [FREE Full text] [doi: 10.3399/bjgp13X674459] [Medline: 24267860]
22. Gonzales R, Ang A, Murphy DA, Glik DC, Anglin MD. Substance use recovery outcomes among a cohort of youth participating in a mobile-based texting aftercare pilot program. J Subst Abuse Treat 2014 Jul;47(1):20-26 [FREE Full text] [doi: 10.1016/j.jsat.2014.01.010] [Medline: 24629885]
23. OSIPTEL. Telecommunications in Peru, Statistics (Option: Densidad por departamento). 2012. Sector Telecomunicaciones, Información Estadística de Telecomunicaciones, Indicadores: Estadísticos. 2. Indicadores del Servicio Móvil URL: http://www.osiptel.gob.pe/articulo/22-densidad-por-departamento [accessed 2017-07-08] [WebCite Cache ID 6ro8YaIPx]
24. Ruiz E, Proaño Á, Ponce O, Curioso W. [Mobile health for public health in Peru: lessons learned]. Rev Peru Med Exp Salud Publica 2015;32(2):364-372 [FREE Full text] [Medline: 26338384]
25. Head KJ, Noar SM, Iannarino NT, Grant HN. 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]
26. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
27. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
28. Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2012;11:CD006611. [doi: 10.1002/14651858.CD006611.pub3] [Medline: 23152238]
29. Kong G, Ells DM, Camenga DR, Krishnan-Sarin S. Text messaging-based smoking cessation intervention: a narrative review. Addict Behav 2014 May;39(5):907-917 [FREE Full text] [doi: 10.1016/j.addbeh.2013.11.024] [Medline: 24462528]
30. Ybarra ML, Holtrop JS, Bağci BA, Emri S. Design considerations in developing a text messaging program aimed at smoking cessation. J Med Internet Res 2012;14(4):e103 [FREE Full text] [doi: 10.2196/jmir.2061] [Medline: 22832182]
31. Buhi ER, Trudnak TE, Martinasek MP, Oberne AB, Fuhrmann HJ, McDermott RJ. Mobile phone-based behavioural interventions for health: a systematic review. Health Educ J 2012 Jul 10;72(5):564-583. [doi: 10.1177/0017896912452071]
32. Ybarra ML, Prescott TL, Holtrop JS. Steps in tailoring a text messaging-based smoking cessation program for young adults. J Health Commun 2014 Dec;19(12):1393-1407. [doi: 10.1080/10810730.2014.901441] [Medline: 24766267]
33. Levinson AH, Borrayo EA, Espinoza P, Flores ET, Pérez-Stable EJ. An exploration of Latino smokers and the use of pharmaceutical aids. Am J Prev Med 2006 Aug;31(2):167-171. [doi: 10.1016/j.amepre.2006.03.022] [Medline: 16829334]
34. Anguiano B, Brown-Johnson C, Rosas LG, Pechmann C, Prochaska JJ. Latino adults' perspectives on treating tobacco use via social media. JMIR Mhealth Uhealth 2017 Feb 08;5(2):e12 [FREE Full text] [doi: 10.2196/mhealth.6684] [Medline: 28179217]
35. Curry SJ, Sporer AK, Pugach O, Campbell RT, Emery S. Use of tobacco cessation treatments among young adult smokers: 2005 National Health Interview Survey. Am J Public Health 2007 Aug;97(8):1464-1469. [doi: 10.2105/AJPH.2006.103788] [Medline: 17600243]

XSL•FO
RenderX

36. Norcross JC, Krebs PM, Prochaska JO. Stages of change. J Clin Psychol 2011 Feb;67(2):143-154. [doi: 10.1002/jclp.20758] [Medline: 21157930]
37. Schulz D, Kremers S, de Vries H. Are the stages of change relevant for the development and implementation of a web-based tailored alcohol intervention? A cross-sectional study. BMC Public Health 2012 May 17;12:360 [FREE Full text] [doi: 10.1186/1471-2458-12-360] [Medline: 22594949]
38. Clinical Practice Guideline Treating Tobacco Use and Dependence 2008 Update Panel, Liaisons, and Staff. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A U.S. Public Health Service report. Am J Prev Med 2008 Aug;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]
39. Stead LF, Perera R, Lancaster T. Telephone counselling for smoking cessation. Cochrane Database Syst Rev 2006 Jul 19(3):CD002850. [doi: 10.1002/14651858.CD002850.pub2] [Medline: 16855992]
40. Zhou X, Nonnemaker J, Sherrill B, Gilsenan AW, Coste F, West R. Attempts to quit smoking and relapse: factors associated with success or failure from the ATTEMPT cohort study. Addict Behav 2009 Apr;34(4):365-373. [doi: 10.1016/j.addbeh.2008.11.013] [Medline: 19097706]
41. Zhu SH, Stretch V, Balabanis M, Rosbrook B, Sadler G, Pierce JP. Telephone counseling for smoking cessation: effects of single-session and multiple-session interventions. J Consult Clin Psychol 1996 Feb;64(1):202-211. [Medline: 8907100]
42. Wilson DK, Wallston KA, King JE. Effects of contract framing, motivation to quit, and self-efficacy on smoking reduction. J Appl Social Pyschol 1990 Apr;20(7):531-547. [doi: 10.1111/j.1559-1816.1990.tb00426.x]
43. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [Medline: 847061]
44. Wood W, Neal DT. A new look at habits and the habit-goal interface. Psychol Rev 2007 Oct;114(4):843-863. [doi: 10.1037/0033-295X.114.4.843] [Medline: 17907866]
45. DiClemente CC, Prochaska JO, Fairhurst SK, Velicer WF, Velasquez MM, Rossi JS. The process of smoking cessation: an analysis of precontemplation, contemplation, and preparation stages of change. J Consult Clin Psychol 1991 Apr;59(2):295-304. [Medline: 2030191]
46. Gwaltney CJ, Metrik J, Kahler CW, Shiffman S. Self-efficacy and smoking cessation: a meta-analysis. Psychol Addict Behav 2009 Mar;23(1):56-66 [FREE Full text] [doi: 10.1037/a0013529] [Medline: 19290690]
47. Reese ED, Veilleux JC. Relationships between craving beliefs and abstinence self-efficacy are mediated by smoking motives and moderated by nicotine dependence. Nicotine Tob Res 2016 Jan;18(1):48-55. [doi: 10.1093/ntr/ntv054] [Medline: 25744968]
48. Borrelli B, Mermelstein R. Goal setting and behavior change in a smoking cessation program. Cogn Ther Res 1994 Feb;18(1):69-83. [doi: 10.1007/BF02359396]
49. Donovan D, Witkiewitz K. Relapse prevention: from radical idea to common practice. Addict Res Theory 2012 Jan 12;20(3):204-217. [doi: 10.3109/16066359.2011.647133]
50. Marlatt GA, George WH. Relapse prevention: introduction and overview of the model. Br J Addict 1984 Sep;79(3):261-273. [Medline: 6595020]
51. Bowen S, Marlatt A. Surfing the urge: brief mindfulness-based intervention for college student smokers. Psychol Addict Behav 2009 Dec;23(4):666-671. [doi: 10.1037/a0017127] [Medline: 20025372]
52. O'Connell KA, Hosein VL, Schwartz JE, Leibowitz RQ. How does coping help people resist lapses during smoking cessation? Health Psychol 2007 Jan;26(1):77-84. [doi: 10.1037/0278-6133.26.1.77] [Medline: 17209700]
53. Ferguson SG, Shiffman S. The relevance and treatment of cue-induced cravings in tobacco dependence. J Subst Abuse Treat 2009 Apr;36(3):235-243. [doi: 10.1016/j.jsat.2008.06.005] [Medline: 18715743]
54. Kreuter MW, Skinner CS. Tailoring: what's in a name? Health Educ Res 2000 Feb;15(1):1-4 [FREE Full text] [Medline: 10788196]
55. lylynx.com. SMS Scheduler. URL: https://play.google.com/store/apps/details?id=com.lylynx.smsscheduler&hl=es_419 [accessed 2017-04-03] [WebCite Cache ID 6pR2Kzdgh]
56. Bock B, Heron K, Jennings E, Morrow K, Cobb V, Magee J, et al. A text message delivered smoking cessation intervention: the initial trial of TXT-2-Quit: randomized controlled trial. JMIR Mhealth Uhealth 2013;1(2):e17 [FREE Full text] [doi: 10.2196/mhealth.2522] [Medline: 25098502]
57. Richmond SJ, Keding A, Hover M, Gabe R, Cross B, Torgerson D, et al. Feasibility, acceptability and validity of SMS text messaging for measuring change in depression during a randomised controlled trial. BMC Psychiatry 2015 Apr 03;15:68 [FREE Full text] [doi: 10.1186/s12888-015-0456-3] [Medline: 25886377]
58. Balmford J, Borland R, Benda P, Howard S. Factors associated with use of automated smoking cessation interventions: findings from the eQuit study. Health Educ Res 2013 Apr;28(2):288-299 [FREE Full text] [doi: 10.1093/her/cys104] [Medline: 23107931]
59. Marcano Belisario JS, Bruggeling MN, Gunn LH, Brusamento S, Car J. Interventions for recruiting smokers into cessation programmes. Cochrane Database Syst Rev 2012;12:CD009187. [doi: 10.1002/14651858.CD009187.pub2] [Medline: 23235672]
60. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]

XSL•FO
RenderX

61. Muench F, van SK, Morgenstern J, Kuerbis AN, Markle K. Understanding messaging preferences to inform development of mobile goal-directed behavioral interventions. J Med Internet Res 2014 Feb;16(2):e14 [FREE Full text] [doi: 10.2196/jmir.2945] [Medline: 24500775]
62. Trinidad DR, Pérez-Stable EJ, White MM, Emery SL, Messer K. A nationwide analysis of US racial/ethnic disparities in smoking behaviors, smoking cessation, and cessation-related factors. Am J Public Health 2011 Apr;101(4):699-706. [doi: 10.2105/ajph.2010.191668]
63. Sedjo RL, Li Y, Levinson AH. Smoking-cessation treatment: use trends among non-Hispanic white and English-speaking Hispanic/Latino smokers, Colorado 2001-2012. Am J Prev Med 2016 Aug;51(2):232-239. [doi: 10.1016/j.amepre.2016.02.015] [Medline: 27020316]
64. Cox LS, Okuyemi K, Choi WS, Ahluwalia JS. A review of tobacco use treatments in US ethnic minority populations. Am J Health Promot 2011;25(5 Suppl):S11-S30 [FREE Full text] [doi: 10.4278/ajhp.100610-LIT-177] [Medline: 21510783]
65. Fu SS, Kodl MM, Joseph AM, Hatsukami DK, Johnson EO, Breslau N, et al. Racial/Ethnic disparities in the use of nicotine replacement therapy and quit ratios in lifetime smokers ages 25 to 44 years. Cancer Epidem Biomar 2008 Jul 01;17(7):1640-1647. [doi: 10.1158/1055-9965.epi-07-2726]
66. Bock BC, Niaura RS, Neighbors CJ, Carmona-Barros R, Azam M. Differences between Latino and non-Latino White smokers in cognitive and behavioral characteristics relevant to smoking cessation. Addict Behav 2005 May;30(4):711-724. [doi: 10.1016/j.addbeh.2004.08.017] [Medline: 15833576]
67. Romanowich P, Lamb RJ. The effect of framing incentives as either losses or gains with contingency management for smoking cessation. Addict Behav 2013 Apr;38(4):2084-2088 [FREE Full text] [doi: 10.1016/j.addbeh.2013.01.007] [Medline: 23403276]
68. Cornacchione J, Smith SW. The effects of message framing within the stages of change on smoking cessation intentions and behaviors. Health Commun 2012 Aug;27(6):612-622. [doi: 10.1080/10410236.2011.619252] [Medline: 22292861]
69. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]
70. 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]
71. Glasser AM, Collins L, Pearson JL, Abudayyeh H, Niaura RS, Abrams DB, et al. Overview of electronic nicotine delivery systems: a systematic review. Am J Prev Med 2017 Feb;52(2):e33-e66. [doi: 10.1016/j.amepre.2016.10.036] [Medline: 27914771]
72. Adkison SE, O'Connor RJ, Bansal-Travers M, Hyland A, Borland R, Yong H, et al. Electronic nicotine delivery systems: international tobacco control four-country survey. Am J Prev Med 2013 Mar;44(3):207-215 [FREE Full text] [doi: 10.1016/j.amepre.2012.10.018] [Medline: 23415116]
73. Noonan D, Jiang Y, Duffy SA. Utility of biochemical verification of tobacco cessation in the Department of Veterans Affairs. Addict Behav 2013 Mar;38(3):1792-1795 [FREE Full text] [doi: 10.1016/j.addbeh.2012.11.006] [Medline: 23261494]
74. Connor GS, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. Nicotine Tob Res 2009 Jan;11(1):12-24. [doi: 10.1093/ntr/ntn010] [Medline: 19246437]
75. Meredith SE, Robinson A, Erb P, Spieler CA, Klugman N, Dutta P, et al. A mobile-phone-based breath carbon monoxide meter to detect cigarette smoking. Nicotine Tob Res 2014 Jun;16(6):766-773 [FREE Full text] [doi: 10.1093/ntr/ntt275] [Medline: 24470633]
76. Dallery J, Raiff BR, Kim SJ, Marsch LA, Stitzer M, Grabinski MJ. Nationwide access to an internet-based contingency management intervention to promote smoking cessation: a randomized controlled trial. Addiction 2016 Dec 06;112(5):875-883. [doi: 10.1111/add.13715] [Medline: 27923264]
77. Stanton A, Grimshaw G. Tobacco cessation interventions for young people. Cochrane Database Syst Rev 2013;8:CD003289. [doi: 10.1002/14651858.CD003289.pub5] [Medline: 23975659]

Abbreviations

CBT: cognitive behavioral therapy
FTND: Fagerström Test of Nicotine Dependence
MNP: mobile number portability
NRT: nicotine replacement treatment
RCT: randomized controlled trial
SMS: short message service

XSL•FO
RenderX

Edited by G Eysenbach; submitted 19.02.17; peer-reviewed by J Apolinário-Hagen, I Montagni; comments to author 26.03.17; revised version received 22.05.17; accepted 13.06.17; published 04.08.17.

Please cite as:
Blitchtein-Winicki D, Zevallos K, Samolski MR, Requena D, Velarde C, Briceño P, Piazza M, Ybarra ML
Feasibility and Acceptability of a Text Message-Based Smoking Cessation Program for Young Adults in Lima, Peru: Pilot Study
JMIR Mhealth Uhealth 2017;5(8):e116
URL: http://mhealth.jmir.org/2017/8/e116/
doi:10.2196/mhealth.7532
PMID:28778850

©Dora Blitchtein-Winicki, Karine Zevallos, M Reuven Samolski, David Requena, Chaska Velarde, Patricia Briceño, Marina Piazza, Michele L Ybarra. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 04.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Smartphone App for Families With Preschool-Aged Children in a Public Nutrition Program: Prototype Development and Beta-Testing

Pamela Hull[1], PhD; Janice S Emerson[2], PhD; Meghan E Quirk[2,3], PhD; Juan R Canedo[4,5], DHSc; Jessica L Jones[2], MS; Violetta Vylegzhanina[6], MS; Douglas C Schmidt[6], PhD; Shelagh A Mulvaney[7], PhD; Bettina M Beech[8], MPH, DrPH; Chiquita Briley[9], PhD; Calvin Harris[2], MS; Baqar A Husaini[2], PhD

[1]Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

[2]Center for Prevention Research, Tennessee State University, Nashville, TN, United States

[3]The National Academies of Sciences, Engineering, and Medicine, Washington, DC, United States

[4]Progreso Community Center, Nashville, TN, United States

[5]Meharry-Vanderbilt Alliance, Meharry Medical College, Nashville, TN, United States

[6]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, United States

[7]School of Nursing, Vanderbilt University, Nashville, TN, United States

[8]Departments of Pediatrics and Family Medicine, University of Mississippi Medical Center, Jackson, MS, United States

[9]Department of Family and Consumer Sciences, Tennessee State University, Nashville, TN, United States

Corresponding Author:
Pamela Hull, PhD
Division of Epidemiology
Department of Medicine
Vanderbilt University Medical Center
2525 West End Avenue, Suite 800
Nashville, TN,
United States
Phone: 1 615 936 3241
Fax: 1 615 343 5938
Email: pam.hull@vanderbilt.edu

Abstract

Background: The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) in the United States provides free supplemental food and nutrition education to low-income mothers and children under age 5 years. Childhood obesity prevalence is higher among preschool children in the WIC program compared to other children, and WIC improves dietary quality among low-income children. The Children Eating Well (CHEW) smartphone app was developed in English and Spanish for WIC-participating families with preschool-aged children as a home-based intervention to reinforce WIC nutrition education and help prevent childhood obesity.

Objective: This paper describes the development and beta-testing of the CHEW smartphone app. The objective of beta-testing was to test the CHEW app prototype with target users, focusing on usage, usability, and perceived barriers and benefits of the app.

Methods: The goals of the CHEW app were to make the WIC shopping experience easier, maximize WIC benefit redemption, and improve parent snack feeding practices. The CHEW app prototype consisted of WIC Shopping Tools, including a barcode scanner and calculator tools for the cash value voucher for purchasing fruits and vegetables, and nutrition education focused on healthy snacks and beverages, including a Yummy Snack Gallery and Healthy Snacking Tips. Mothers of 63 black and Hispanic WIC-participating children ages 2 to 4 years tested the CHEW app prototype for 3 months and completed follow-up interviews.

Results: Study participants testing the app for 3 months used the app on average once a week for approximately 4 and a half minutes per session, although substantial variation was observed. Usage of specific features averaged at 1 to 2 times per month for shopping-related activities and 2 to 4 times per month for the snack gallery. Mothers classified as users rated the app's WIC Shopping Tools relatively high on usability and benefits, although variation in scores and qualitative feedback highlighted several

XSL•FO
RenderX

barriers that need to be addressed. The Yummy Snack Gallery and Healthy Snacking Tips scored higher on usability than benefits, suggesting that the nutrition education components may have been appealing but too limited in scope and exposure. Qualitative feedback from mothers classified as non-users pointed to several important barriers that could preclude some WIC participants from using the app at all.

Conclusions: The prototype study successfully demonstrated the feasibility of using the CHEW app prototype with mothers of WIC-enrolled black and Hispanic preschool-aged children, with moderate levels of app usage and moderate to high usability and benefits. Future versions with enhanced shopping tools and expanded nutrition content should be implemented in WIC clinics to evaluate adoption and behavioral outcomes. This study adds to the growing body of research focused on the application of technology-based interventions in the WIC program to promote program retention and childhood obesity prevention.

(JMIR Mhealth Uhealth 2017;5(8):e102) doi:10.2196/mhealth.7477

KEYWORDS

pediatric obesity; health education; public health informatics; mobile apps

Introduction

The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) serves 9.3 million low-income, nutritionally at-risk mothers, infants, and children under age 5 years in the United States by providing free supplemental healthy foods, nutrition education, breastfeeding support, and health care referrals. Prevalence of childhood overweight and obesity has increased rapidly in recent decades, and risk for obesity increases sharply after age 5 years [1,2]. Prevalence is greater among black and Hispanic children compared to white children and greater among low-income families served by WIC compared to the general population [3-5]. The WIC program improves diet quality among low-income children [6], but participation in the program declines as children age, particularly among 2- to 4-year-olds [5].

Studies have found that the vast majority of WIC participants are satisfied with the nutrition education provided by WIC every 3 months, predominantly delivered in person in WIC clinics. However, repetition of information, long wait times, and accompanying children have been identified as barriers and predictors of program attrition [7,8]. Compared to the nutrition education component, WIC participants tend to have lower satisfaction with the process of redeeming their WIC benefits in the grocery store [8,9].

WIC benefits consist of approved food packages assigned to each WIC participant in the family, which list the specific types, brands, sizes, and quantities of products that can be purchased. In addition, a cash value voucher (CVV) provides a flat dollar amount per WIC participant that can be spent on fruits or vegetables, currently $8 for children and $10 for mothers. As of May 2017, the majority of states (30/50, 60%), including Tennessee, have not yet implemented electronic benefits transfer (EBT) statewide and continue to provide benefits in the form of paper vouchers; the majority of these remaining states are scheduled to transition to EBT in 2018 (including Tennessee) or 2019, with the deadline of 2020. Challenges with the WIC shopping experience can act as a barrier for participants to redeem all of their benefits and can lead to program attrition, particularly with paper vouchers [8-12]. An expert review panel recommended that in-clinic nutrition education be complemented with reinforcement education using innovative, multilevel strategies to make it easier for participants to purchase all of the nutritionally beneficial items included in their food packages and to consume them at home [13].

Very limited research has been conducted to date on the use of technology-based interventions in the WIC program [14-16]. Technology-supported interventions such as smartphone apps can impact a substantially larger number of people more frequently and usually at a lower cost than in-person education. Nearly two-thirds of all adults and 85% of young adults in the United States own a smartphone, with higher levels of ownership among racial/ethnic minorities compared to whites [17-19]. Mobile phone–based interventions have been effective with minority and low-income populations to improve nutrition and other health-promoting behaviors [20-23]. A handful of smartphone apps are commercially available that include some WIC shopping features and limited nutrition information or recipes. However, no published studies to date have reported on the development or testing of a smartphone app for WIC participants that combines nutrition education and WIC shopping tools.

This paper describes the development and beta-testing of the Children Eating Well (CHEW) smartphone app. This app was developed for WIC-participating families with preschool-aged children as a home-based intervention to reinforce WIC nutrition education and improve the WIC shopping experience. The objective of beta-testing was to test the CHEW app prototype with target users, focusing on usage, usability, and perceived barriers and benefits of the app.

Methods

Needs Identified by Stakeholders

As part of the larger Nashville CHEW for Health project, the CHEW Community Advisory Board (CAB) was formed in 2011, comprising 8 to 10 WIC participants and several nonprofit organizations focused on food security. The CAB provided input on the project's objective to develop a home-based nutrition education intervention for WIC-participating families with preschool children ages 2 to 4 years with a particular focus on black and Hispanic families, given their elevated risk for obesity. CAB members indicated that they were largely satisfied with availability of nutritious foods available through the WIC program but would like for the cash value voucher for fruits and vegetables to be increased in value.

XSL•FO
RenderX

CAB members expressed that their main concern with the WIC program, similar to the findings of previous studies [8-11], was that the WIC shopping experience was confusing and included the following challenges: each family member having a different paper voucher; specific limits on brands, quantities, and sizes; allowed choices and combinations within categories; the CVV requiring math skills; embarrassment at the register when items did not match up to their approved benefits; and not always getting all of their allowed items due to these complexities. They strongly desired that something be done to make the process easier. In terms of specific topics for nutrition education, CAB members indicated that they would be interested in receiving quick and simple recipes with WIC items that their children would eat and practical advice on parent feeding strategies to help them more easily provide healthy food to their children.

The CHEW team also engaged in a series of conversations with state and local WIC program staff to identify program needs and feasible intervention strategies that would enhance the existing WIC program. WIC program staff expressed interest in opportunities to provide participants with tools and resources to complement the nutrition education they received during WIC clinic visits. The program staff had two main concerns: many participants not redeeming all of their monthly WIC benefits and lower program retention as children age past 2 years. They agreed with the feedback from the CAB that frustrations with the shopping experience could contribute to both of these problems. The combined feedback of the CAB and WIC program staff inspired the idea of creating an app with WIC shopping tools plus nutrition education to include simple recipes and practical advice.

In 2012, the team conducted the CHEW Nutrition Survey in a multiethnic sample of 150 families with WIC-participating children ages 2 to 4 years [24,25]. Analysis of the survey data on the dietary needs of the children informed the selection of dietary targets for the nutrition education to be developed. Compared to the US dietary guidelines that were in effect in 2012 [26], the children in the survey consumed less than the recommended amounts of vegetables, whole grains, dairy, and fiber [27]. Intake of whole grains and water was lowest among Hispanic children, while intake of fat, added sugars, and whole milk was highest among black non-Hispanic children, even though WIC did not provide whole milk to participants older than 2 years at the time. Commonly reported barriers to eating more fruits and vegetables included that the cost was more than the value of the CVV, parents or other family members did not like them, and parents did not have time to prepare fresh fruits and vegetables (unpublished findings). Finally, the team identified some cultural differences in food preferences for black and Hispanic children compared to white children, based on which fruits and vegetables the parents chose to purchase with the preschool child's CVV.

Theoretical Framework and Conceptual Model

Based on the input from the CAB and WIC program, we established the goal of developing a smartphone app for use as a home-based nutrition education intervention to complement and reinforce WIC's in-clinic nutrition education with a specific focus on black and Hispanic children ages 2 to 4 years. The theoretical framework that guided the development of the CHEW app was the socioecological model [28-30]. The CHEW app interplays with the following 5 levels of influence on child dietary intake (Figure 1):

1. The *policy level* represents the state WIC program that determines the approved food packages compliant with federal regulations for each family and provides periodic nutrition education and other services.
2. The *grocery store food environment* is where clients will use the app to facilitate WIC shopping and ideally maximize purchasing all of their eligible items.
3. The *home food environment* is where the user will interact with the app most of the time, bringing home more healthy food, planning shopping, and viewing content.
4. The *interpersonal level* is where the parent will interact with the child, implementing tips learned through the app for feeding the child. Parents are the primary agents of change within the family, especially for younger children; thus they are the primary audience for the app.
5. The *individual level* represents the child, where all of the influences will converge to impact the child's consumption of food and beverages.

The concept is for the app to improve the WIC shopping experience, which should lead to maximizing benefit redemption and an improved home environment as well as increased satisfaction and program retention. Moreover, the nutrition education provided through the app will improve the home food environment and improve parent-feeding practices, which will lead to improved child dietary intake.

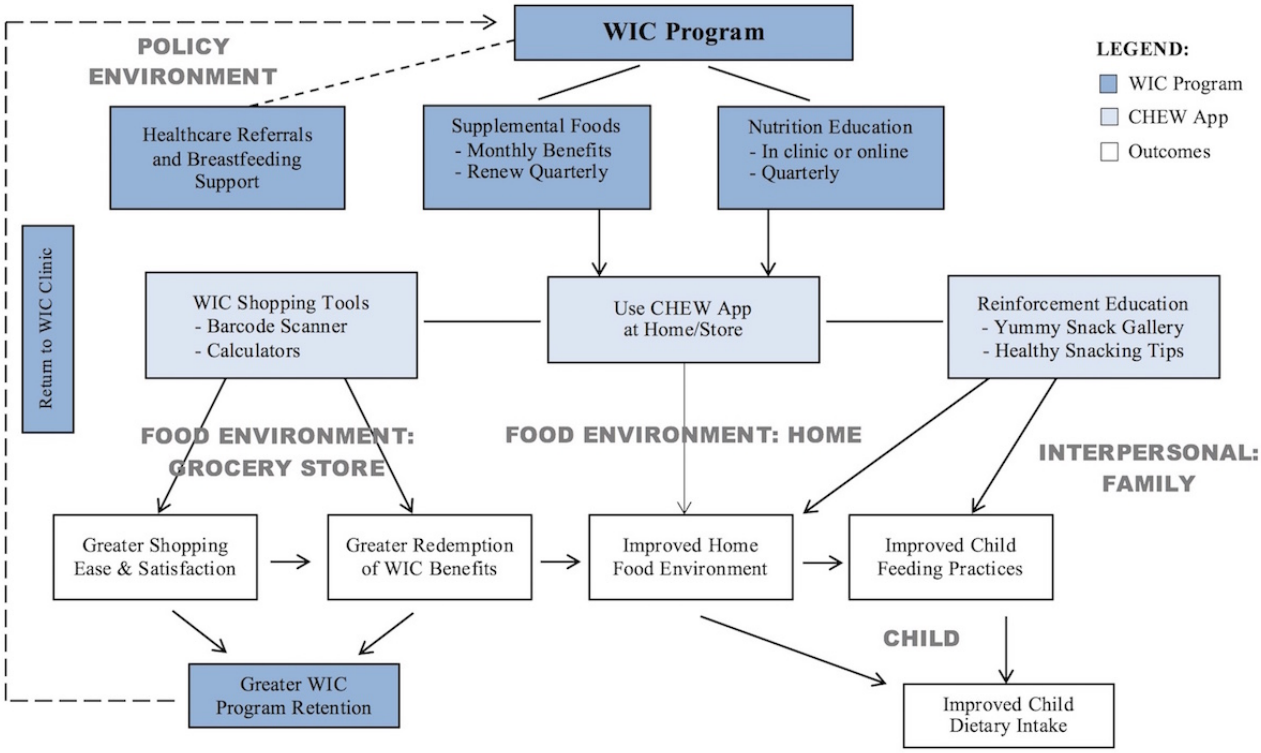
Development of the Children Eating Well Smartphone App Prototype

Overview

The CHEW team carried out an iterative [31], user-centered [32] design process to develop the CHEW app prototype with periodic input on concepts, content, and functional prototypes from the CAB and the WIC program as the target end-users. Computer science engineering graduate and undergraduate students, under supervision of engineering faculty, programmed the prototype version for the Android operating system. The goals of the app are to improve the WIC shopping experience, increase benefit redemption, and improve parent snack feeding practices. The app consists of 2 components: WIC shopping tools to make the WIC shopping process easier and nutrition education for parents of 2- to 4-year-olds [33,34]. See screenshots in Figure 2. All app content was prepared in both English and Spanish with the assistance of one of our key partners, a Hispanic community organization. Users have the option to switch between English and Spanish versions of the app.

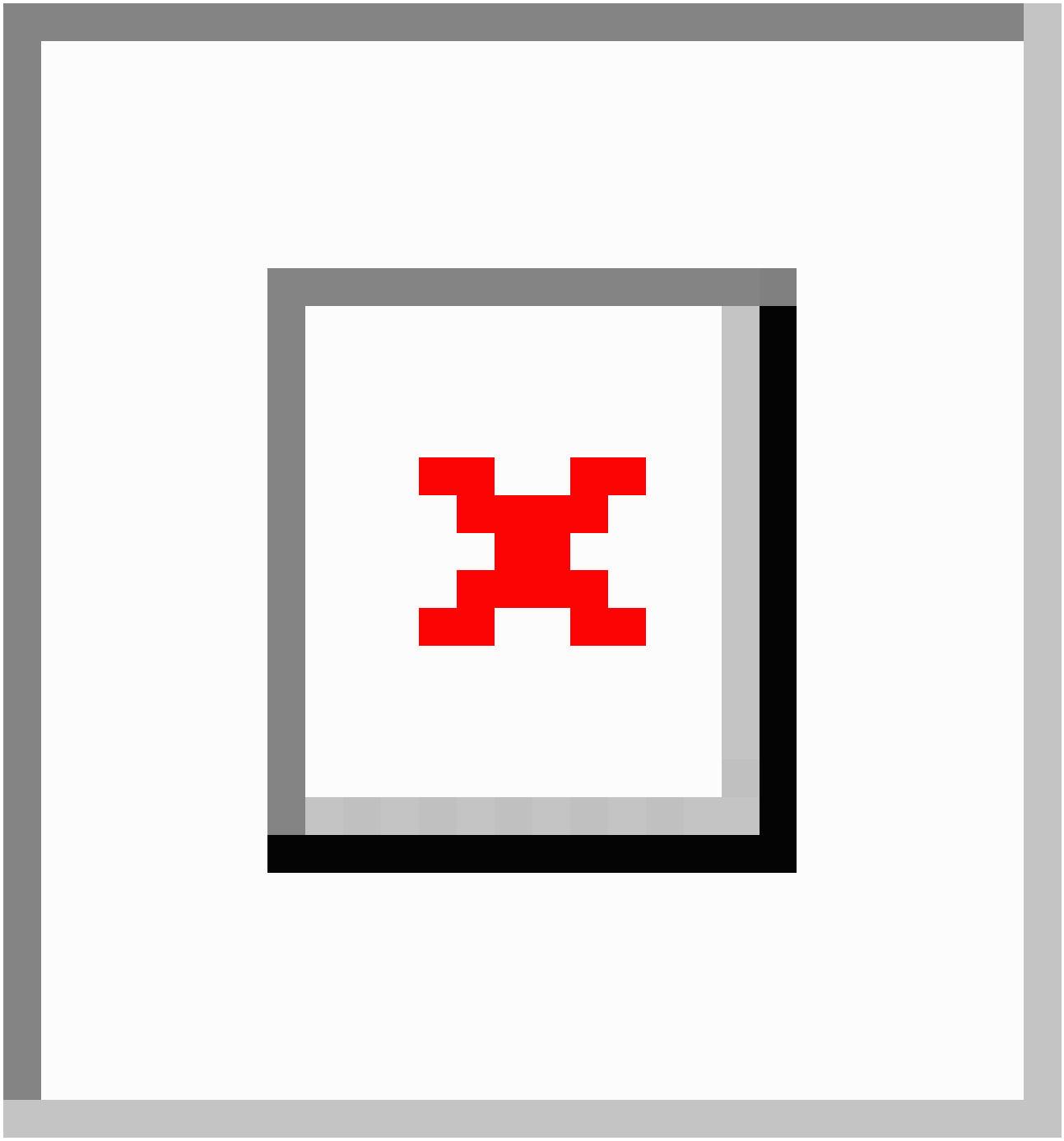
XSL•FO
RenderX

Figure 1. Conceptual model for Children Eating Well smartphone app based on socioecological framework.



XSL•FO
RenderX

Figure 2. Screenshots and features of Children Eating Well smartphone app prototype.



WIC Shopping Tools

Several app features were developed to assist users with making selections while in the grocery store. Two large grocery chains where the majority of WIC benefits are redeemed in Tennessee provided the research team with lists of all the WIC-authorized items carried by each store and the corresponding universal product codes (UPC). The research team combined the databases and coded approximately 1000 items to indicate to which regular vouchers each corresponded as an allowed item along with over 12,000 fruit and vegetable items that could be purchased with CVVs. In addition, the programmers constructed logical arguments to represent which items, combinations of items, and item limits were included on each regular voucher type.

For the prototype version, only standard food package types were included, excluding food packages for special dietary needs. The programmers built a barcode scanner feature that enables the user to scan the UPC barcode of any item in the store to see if the item is (1) approved by WIC and (2) on one of the family members' vouchers. Users can then select if they want to purchase the item, for which voucher, and the quantity, if applicable. The app notifies users when they reach the limits for specific categories. At any time during the shopping trip, users can view the products and quantities allowed for each voucher and a running list of the items they have selected for each voucher.

For the CVV, the programmers built calculator tools to assist the shopper with using the specified dollar amount for each

XSL•FO
RenderX

CVV. The cost of fruits and vegetables is challenging for participants to keep track of given the various ways that they are priced in grocery stores. Thus, 4 calculators are available depending on which type of fruits or vegetables the user wants to purchase: packaged fresh—fixed price (eg, bag of carrots, package of strawberries), packaged frozen—fixed price (eg, box of frozen spinach, bag of frozen peaches), loose produce—price per pound (eg, apples or green beans priced per pound), and loose produce—price per item (eg, individual cucumbers or watermelons priced per unit). Since prices in each store vary day to day, the app does not include a price database, so the user is prompted to enter the package or unit price and quantity. The cost for the total quantity is then calculated. The app keeps a running tab of the amount spent on each CVV and alerts the user when maximum values are reached.

Nutrition Education

Based on feedback from the CAB and WIC program and the survey findings for the prototype version, the team chose to focus the nutrition education content on age-appropriate snacks and beverages for preschoolers. The two nutrition education features in the prototype version are a Yummy Snack Gallery and Healthy Snacking Tips. The objective of these features is to give parents practical skills and strategies to improve the following dietary targets for children: increased intake of fruits, vegetables, fiber, low-fat milk, and water and decreased intake of sugar-sweetened beverages.

In the prototype version, the Yummy Snack Gallery includes 35 snack and beverage recipes developed by the CHEW team to be quick and easy to prepare, have few ingredients, include WIC-approved items, require no cooking skills, and be appealing to 2- to 4-year-olds. Each recipe includes images for visual step-by-step instructions. The recipes were selected to target the dietary preferences of black and Hispanic WIC participants based on findings from the CHEW Nutrition Survey of the most frequent fruits and vegetables purchased by each group. Examples of recipes include fast collard greens, peanut butter and fruit smash-wich, avocado toast, chayote salad, veggie boats, and peachy banana smoothie. Users can browse the snack gallery, view the details of each recipe, flag specific recipes as favorites, and add the ingredients to a snack shopping list.

In the prototype version, the Healthy Snacking Tips feature includes a database of 21 short educational messages and 21 reminders to use the various app features. The educational messages are based on evidence from the scientific literature about parent feeding strategies to provide parents with advice on practical strategies they can employ with their preschool children to improve the targeted dietary behaviors. The main topic areas cover beverages (eg, “Whole milk is for toddlers under two. For older kids and adults, gradually switch to 2% then 1% and then skim. Try mixing them for an easier switch!”), kid-friendly snacking (eg, “Let healthy foods fill the gaps! If your child misses fruit at lunch, offer it as an afternoon snack.”), introducing new foods (eg, “It may take more than 12 tries for your child to like a new food. Stay positive and keep motivating them to try a bite!”), parents as role models (eg, “Be a great role model! Let your child see you enjoying fruits, vegetables, water, and low-fat dairy every day.”), and health benefits (eg, “Fiber helps the tummy do its work and helps you go regularly! Get it from: Apples * Carrots * Broccoli * Pears * Spinach”). These topics were selected with input from the CHEW CAB. Examples of reminders include, “Remember to use your WIC shopping tools when you're ready to buy WIC items at the store,” and “Just a few minutes until snack time? Check out the Yummy Snack Gallery for quick and tasty ideas!”

The app is programmed to make the messages appear on the phone via push notifications per a predetermined schedule. In the prototype version, the schedule was set to vary the number of notifications sent each week, with each messages being sent twice across the 3-month testing period. The default time to send the notification is 10:00 AM on the designated day, and the user has the option to change the time.

Study Sample and Recruitment

The prototype app was beta-tested with a sample of 80 mothers of WIC-participating children ages 2 to 4 years, stratified by race/ethnicity of the child with 40 black non-Hispanic children and 40 Hispanic children of any race. Study participants were enrolled from August 2014 through January 2016. To recruit families actively enrolled in the WIC program, potentially eligible families were randomly sampled from periodic lists of black and Hispanic participants 2 to 4.5 years of age who made a WIC clinic visit in the previous 3 months. The upper age cutoff was set at 4.5 (4 years and 6 months) to allow time to recruit families and complete the testing period before the child turned 5 and would no longer be eligible for WIC benefits. The study team then contacted the adult contact person to invite the mother to be screened and participate in the study if eligible. The mother was targeted for inclusion in the study since mothers are typically the parent primarily responsible for feeding young children in families. They were first contacted by mail and given the opportunity to opt out of further contact, then study staff followed up with phone calls.

Eligibility screening was completed over the phone in English or Spanish, per the preference of the respondent. Inclusion criteria were as follows: child is between 2 and 4.5 years of age, either non-Hispanic black/African American or Hispanic of any race (as identified by a parent or guardian), currently receives benefits from WIC, will receive benefits for the 3 months following enrollment, and receives one of the standard food packages (not special dietary needs); mother is 18 years of age or older (or if under 18 years of age, her guardian provides consent to participate), currently owns and regularly uses an Android smartphone, and uses the family’s WIC vouchers at one of the two major grocery chain stores included in the UPC database. The following exclusion criteria were applied: mother does not speak and read English or Spanish, is currently pregnant, currently has a child under the age of 6 months, is unable to provide consent, or someone other than the mother regularly does the WIC shopping for the child participant or family. The study inclusion and exclusion criteria were extensive and strict in order to help ensure that we enrolled people who would most likely be able to use the app and complete the study; it was not intended as a model for how the app will be disseminated in practice settings in the future.

XSL•FO
RenderX

Out of all the families on the sampled lists, 22% were not reachable due to incorrect contact information, 20% were nonresponsive, 10% opted out of further contact or declined to be screened, and the remaining 48% were successfully screened for eligibility. Among those screened, approximately two-thirds (69%) did not qualify based on the inclusion and exclusion criteria listed above. The most common reasons for failing screening were as follows: not owning an Android smartphone (39%), having an infant under 6 months of age (15%), not shopping regularly at one of the two targeted chain grocery stores (14%), currently being pregnant (10%), and speaking a language other than English or Spanish (7%). Potential participants may have owned smartphones with other operating systems (eg, iPhone iOS, Windows), but for the purposes of testing the prototype programmed on the Android platform, they were not eligible to participate in this study. Among those screened, 14% were deemed eligible but did not follow through with the baseline interview appointment or subsequently declined and 17% were deemed eligible and enrolled in the study (ie, 55% of those who were screened as eligible decided to enroll). This study was approved by the Institutional Review Board of the lead institution.

Study Design and Procedures

Beta-testing of the CHEW app prototype was conducted using an observational design based on data collected after the testing period. Interested mothers who were screened as eligible scheduled an in-home visit with trained interviewers to complete enrollment. After completing the informed consent process, enrolled mothers completed a baseline questionnaire in English or Spanish. The interviewers then installed the CHEW app prototype on their Android smartphones, loaded the family's current WIC voucher information into the app, gave the mother an overview of how to use the app, and asked her to test the app during the next 3 months. The app was successfully installed for 74 of the 80 enrolled mothers (92%); partially installed for 3/80 mothers (4%), meaning that the main app worked, but the scanning function was not fully operable; and was not successfully installed for 3/80 mothers (4%) due to incompatibility of their model of Android phone.

A second in-home visit was scheduled for 3 months after the baseline visit, during which the mothers completed a follow-up questionnaire that included a series of questions about their experiences with the app prototype. At the follow-up visit, the interviewers also attempted to manually download an app usage log from the participant's phone that tracked the frequency of using specific features in the app. Follow-up visits were successfully completed with 63/80 mothers (79% retention), with 7/80 (9%) declining follow-up and 10/80 (13%) lost to follow-up. Among the 63 who completed the follow-up visit, usage logs were successfully downloaded from 42/63 participants' phones (67%). Interviewers were unable download logs from the remaining participants: 3/63 (5%) whose installation failed at baseline due to phone incompatibility, 11/63 (18%) who had replaced their phones since the first interview and no longer had the app, 3/63 (5%) whose phone was currently broken or lost, and 4/63 (6%) in which the interviewer was unable to locate and retrieve the log file from the phone. Participants received a $25 for each completed in-home visit.

Measures

App Usage Log

Overview

The app was programmed with a simple logging function to record time-stamped notations when certain activities or events were performed by the user in the app, such as opening the app or one of the features, into an archived file. The log could then be manually downloaded from the phone in the form of a raw text file. Next our team converted each text file into a spreadsheet with usable data fields. Then we aggregated all of the event rows up to the level of the individual user, creating computed variables using arithmetic functions such as sum or mean. Finally the data points for each user were merged into a cumulative log database with one row per participant.

Frequency and Duration of App Sessions

Individual app events were sorted by date and time to identify discrete user sessions. A new session was operationalized as any new event (eg, opening the home page) occurring at least one hour after the previous event. When aggregating the data to the individual level, we hand-counted the number of discrete app sessions per user. At the same time, we calculated the duration of each session in minutes, then upon aggregation calculated the average duration across all of the individual user's sessions.

Frequency of Using App Features

Within each participant's usage log, we counted the frequency of several specific logged events as indicators for using a specific app feature. For using the WIC shopping tools, we counted the number of times the user selected a grocery store (one of the two stores included in the app) to initiate a shopping session as well as the number of times the user opened any of the produce calculators. For use of the snack gallery, we counted the number of times the user opened one of the snack recipes. It was not possible to track receipt or reading of the healthy snacking tips messages using the log since these messages were delivered on a predefined schedule to the user through the Android phone's native push notification function.

Questionnaire

At baseline, mothers provided sociodemographic information about themselves, the preschool-aged child, and the family. At follow-up, mothers responded to a series of items about their experiences with the CHEW app prototype during the testing period, described below.

Self-Reported Usage

The mother reported whether or not she had used each of the main features of the app at least once (Yes or No) during the past 3 months: WIC shopping tools, snack gallery, and healthy snacking tips. A "yes" response on any of these 3 items was then coded in a new variable to indicate using any of the features in the past 3 months.

Usability

Concepts from an existing smartphone app usability model [35] were used to construct a series of usability items specific to the features of the CHEW app in terms of ease of use, helpfulness,

XSL•FO
RenderX

usefulness, and satisfaction. The response choices for all of the items consisted of 5-point Likert scales ranging from 1=strongly disagree to 5=strongly agree, with higher values indicating greater usability. Some examples of items include, "The Snack Gallery recipes were easy to follow" (ease of use), "The fruit and vegetable calculator was helpful" (helpfulness), "The Healthy Snacking Tips gave me new information" (usefulness), and "I would recommend the WIC Shopping Tool to other WIC families" (satisfaction).

Perceived Benefits

A series of items was created specific to the CHEW app prototype to assess the extent to which the user felt the app helped the mother perform the targeted behaviors. For each app component (shopping tools, snack gallery, snacking tips), the items asked whether this component helped the user perform specific WIC shopping tasks or recommended parent feeding strategies. The response choices for all of the items consisted of a 5-point Likert scale ranging from 1=strongly disagree to 5=strongly agree, with higher values indicating greater benefits. Examples of items include, "The WIC Shopping Tool made checking out easier with my cash value vouchers for fruits and vegetables," "The Snack Gallery helped me introduce new vegetables to my child," "The Healthy Snacking Tips helped me give my child water instead of sugary drinks," and "The Healthy Snacking Tips helped me with my picky eater."

Qualitative User Feedback

In addition to the structured, quantitative questionnaire items described above, the interviewers asked the participants several open-ended questions to elicit qualitative responses of user feedback regarding the key app features. For all of these questions, the mother could answer however she wanted, and the interviewer typed her response into the data collection form as close to verbatim as possible.

Reasons for nonusers: If the participant indicated in the usage questions above that she did not use or view one of the 3 components (shopping tools, snack gallery, healthy snacking tips) at all during the past 3 months, the interviewer asked why she did not use the feature, as an open-ended question.

Barriers and benefits for users: At the end of each set of usability and benefits items for each of the 3 main app features, the interviewer asked the following 2 open-ended questions regarding each feature: "Is there anything you would like to see changed about the [fill in] feature in the CHEW app?" and "Do you have anything else you'd like to say about the [fill in] feature in the CHEW app?"

Data Analysis

Analyses were performed on the data from the 63 mothers who completed the posttesting follow-up questionnaires and the subset of 42 mothers with app usage logs that were downloaded at follow-up. Sample characteristics were described using cross-tabulations with chi-square tests to indicate associations with race/ethnicity of the child, since the sample was stratified by African Americans and Hispanics. Descriptive statistics were used to report the other quantitative variables, using frequencies and percentages for categorical variables and means with standard deviations for continuous and ordinal variables.

The qualitative responses to the open-ended questions were pooled together, reviewed to identify emergent common themes where applicable, and assigned categories corresponding to the themes. The themes were then summarized in a table with counts to provide a general idea of how often they were mentioned by participants. The responses of nonusers for each feature (ie, reasons for not using a feature at all) were summarized separately from the responses of mothers who had at least some experience using a feature. For these mothers, the themes emerging from the 2 open-ended questions were divided into barriers and benefits for using the respective app feature.

For purposes of informing development of future versions of the app, the team chose to interpret mean scores of 4.0 or higher as successful in usability or benefits, mean scores of 3.0 to 3.99 as a moderately positive responses that indicate areas needing focus for improvements in usability or benefits, and mean scores below 3.0 as not meeting minimum criteria for usability or benefits and needing major changes.

Results

Sample Characteristics

Table 1 reports the characteristics of the sample by racial/ethnic group of the child. The racial/ethnic group of the mother matched that of the child for all but 2 women, one in each group (results not shown). The majority of children were 2 or 3 years old, given that the upper age cut off was 4.5 years. The sample had slightly more male than female children, all of whom were born in the United States. For mothers, the largest age group was 25 to 34 years. Roughly equal percentages of mothers were either married or single/never married. All but one of the mothers of Hispanic children were born outside of the United States. Over half of the mothers had only one child receiving WIC, approximately one-third had 2 family members receiving WIC, and nearly one-tenth had three or more family members. More than 4 out of 5 families also received benefits from the federal Supplemental Nutrition Assistance Program (SNAP) in addition to WIC. Race/ethnicity of the child was significantly associated with mothers' education (P<.001), employment status (P=.01), country of birth (P<.001), and food insecurity (P=.001). We compared the baseline demographic characteristics in Table 1 for those who did and did not complete the 3-month follow-up interview. The only significant difference was that families also receiving SNAP benefits were more likely to complete follow-up (results not shown).

XSL•FO
RenderX

Table 1. Sample characteristics from the Children Eating Well smartphone app prototype study (note: some column percentages do not add up to 100% due to rounding).

Variable	Black children N=33 n (%)	Hispanic children N=30 n (%)	*P* value
Child—age			
2 years	18 (54)	13 (45)	.09
3 years	14 (42)	10 (35)	
4 years	1 (3)	6 (21)	
Child—gender			
Male	17 (52)	17 (57)	.68
Female	16 (49)	13 (43)	
Child born in United States			
No	0	0	1.00
Yes	33 (100)	30 (100)	
Mother—age			
18-24 years	5 (17)	1 (3)	.21
25-34 years	14 (47)	18 (60)	
35-44 years	11 (37)	11 (37)	
Mother—marital status			
Married	10 (30)	17 (57)	.10
Single, never married	20 (61)	12 (40)	
Single, divorced	3 (9)	1 (3)	
Mother—education			
<High school degree	2 (6)	15 (52)	<.001
High school degree	8 (24)	10 (35)	
Any college/technical	22 (67)	4 (13)	
Mother—employment			
Not employed	12 (36)	22 (73)	.01
Employed full-time	14 (42)	5 (17)	
Employed part-time	7 (21)	3 (10)	
Mother born in United States			
No	4 (12)	29 (97)	<.001
Yes	29 (88)	1 (3)	
Family members on WIC[a]			
1	19 (57)	18 (60)	.73
2	11 (33)	9 (30)	
3 or more	3 (9)	2 (9)	
Family receives SNAP[b]			
No	4 (12)	5 (17)	.61
Yes	29 (88)	25 (83)	
Run out of money for food			
Never	9 (27)	1 (3)	.001
Sometimes	11 (33)	5 (17)	

XSL•FO
RenderX

Variable	Black children N=33 n (%)	Hispanic children N=30 n (%)	*P* value
Often	2 (6)	12 (41)	
Nearly every month	11 (33)	11 (38)	

[a]WIC: Special Supplemental Nutrition Program for Women, Infants, and Children.

[b]SNAP: Supplemental Nutrition Assistance Program.

Usage

Table 2 presents usage of the main app components. According to the available app logs, on average mothers engaged in 12.57 discrete sessions, which roughly translates to once a week over a 3-month period. More than one-third of mothers (15/42, 36%) used it more than 12 times, or more than once a week. However, frequency of usage varied substantially, with a range from 0 to 70 and fairly even distribution across the various categories. Over 20% (9/42) only used the app once or not at all, while over half (22/42, 52%) used it more than 6 times, or more than twice a month. The average duration of sessions also varied substantially, with a mean of 4.67 minutes and a range from 0 to 22 minutes, which was spread fairly evenly across 2-minute intervals up to 6+ minutes.

Mothers engaged in a shopping session in the app by selecting a grocery store on average 5.14 times, with a range of 0 to 19. Over half (22/42, 52%) selected a store more than 3 times, or at least once a month. Users opened the produce calculators on average 3.21 times, or approximately once a month, ranging from 0 to 17 times. Over one-third opened the calculator more often than 3 times (15/42, 36%). The snack gallery had the highest average usage at 12.76 times, or roughly once a week. However, the wide range of 0 to 111 times and SD 24.67 suggest a right-tail skewed distribution influenced by a few outliers. Nevertheless, 40% (17/42) of participants opened a recipe 6 or more times, or more than twice a month. Each feature had no logged record of use for nearly a third of mothers.

In the self-reported questionnaire items, 4 out of 5 mothers (51/63, 81%) reported that they used the WIC shopping tools, over two-thirds (43/63, 68%) said they used the snack gallery, and three-fourths (47/63, 75%) indicated that they received or read the healthy snacking tips. Combined, 9 out of 10 mothers (57/63, 91%) reported using at least 1 of the 3 components at some point during the 3 months. This overall percentage coincides with the app log data indicating that 90% (38/42) opened at least 1 app session, and the 2 estimates for using the snack gallery at least once were similar. However, mothers self-reported using the shopping tools at least once more often than the logged count of users selecting a store or using the produce calculators at least once.

Qualitative Feedback on Barriers and Benefits

Table 3 includes a summary of the reasons given by mothers who indicated they had not used a specific app feature at all during the 3-month test period. Many in this group experienced technical barriers that precluded their ability to use a specific feature or the entire app, such as a broken phone, unsuccessful installation, and problems with features working properly on their phone. A few pointed to the app not being easy enough to use, lack of interest in the content, forgetting to use it, or not noticing alerts.

Also included in the table are themes that represent barriers and benefits mentioned by mothers in the open-ended questions. Notably, the overwhelming majority of users did not mention any barriers to using each of the features when asked what they would want to change about them. Nevertheless, the users identified some important issues with the shopping tools that may have interfered with some users taking full advantage of them. The most common problem was the barcode scanner not functioning well on certain Android phones. Some noticed potential errors in the database of WIC-approved items provided by the 2 grocery store chains as well as a delay in our team updating the database when WIC implemented some food package changes in October 2014. A few users experienced some challenges in using the shopping tools. At the same time, even though the open-ended questions did not specifically ask mothers to comment on what they liked about the app, many were enthusiastic about how they liked the shopping tools and found them to be helpful.

Many mothers also expressed enthusiasm and satisfaction with the snack gallery, saying they liked it, it was helpful, their children loved it, and it was easy and affordable. The common barrier noted for the snack gallery was a demand for it to be expanded with more recipes. One mother also said she would prefer to use it on a different device besides her phone.

Regarding the healthy snacking tips, a couple users experienced technical issues in receiving the alerts, some did not like the delivery schedule, and some were not interested in the information or reminders. Roughly as many mothers pointed out that they liked the tips or wanted more. A few said they were helpful, specifically for helping them buy or eat more fruits and vegetables.

XSL•FO
RenderX

Table 2. Children Eating Well smartphone app prototype usage (note: some column percentages do not add up to 100% due to rounding).

Variable	n (%)
Logged app usage (n=42)	
App sessions (mean 12.57, SD 14.98)	
None	4 (9.5)
1	5 (11.9)
2-3	6 (14.3)
4-6	5 (11.9)
7-9	4 (9.5)
10-12	3 (7.1)
13-15	1 (2.4)
16-18	4 (9.5)
19 or more	10 (23.8)
Average minutes per session (mean 4.67, SD 4.29)	
Less than 2	11 (28.9)
2-3.9	8 (21.1)
4-5.9	8 (21.1)
6 or more	11 (28.9)
Times selected store for shopping (mean 5.14, SD 5.43)	
None	13 (31.0)
1-3	7 (16.7)
4-6	10 (23.8)
7-9	2 (4.8)
10-12	4 (9.5)
13 or more	6 (14.2)
Times opened produce calculators (mean 3.21, SD 4.25)	
None	14 (33.3)
1-3	13 (31.0)
4-6	8 (19.0)
7-9	4 (9.5)
10 or more	3 (7.1)
Times opened a snack recipe (mean 12.76, SD 24.67)	
None	12 (28.6)
1-5	13 (31.0)
6-10	5 (11.9)
11-20	4 (9.5)
21-30	3 (7.1)
31 or more	5 (11.9)
Self-reported usage in follow-up interview (n=63)	
Used any feature at least once	57 (90.5)
Used WIC[a]shopping tools at least once	51 (81.0)
Used snack gallery at least once	43 (68.3)
Viewed healthy snacking tip alerts at least once	47 (74.6)

XSL•FO
RenderX

[a]WIC: Special Supplemental Nutrition Program for Women, Infants, and Children.

Usability and Perceived Benefits

Table 4 reports means and SD for the usability and perceived benefits items for each of the app prototype components. None of the items scored below a 3 on usability or perceived benefits dimensions, which would have indicated a need for major changes.

For usability, the WIC shopping tools overall, the barcode scanner, and produce calculators scored just under 4 for ease of use (3.81-3.98), indicating a need for some improvement. The barcode scanner also scored just under 4 for helpfulness (3.90) and usefulness/correct information (3.96), while the calculators and overall tools scored above 4 on these dimensions (4.16-4.40). The WIC shopping tools scored above 4 (4.33) on satisfaction in terms of willingness to recommend them to other WIC participants. In terms of perceived benefits of the shopping tools, 6 of the 7 items scored above 4 (4.10 to 4.51), with 1 item scoring 3.57 for the tools helping to spend more of the CVV, indicating a need for some improvement to facilitate that target behavior.

Mothers who used the Yummy Snack Gallery in the previous 3 months scored it high on all of the indicators of ease of use, helpfulness, usefulness, and satisfaction, with averages ranging from 4.63 to 4.95. For the 2 perceived benefits of the Yummy Snack Gallery, introducing new fruits scored just under 4 (3.93) and introducing new vegetables scored just above 4 (4.05).

Among mothers who viewed the Healthy Snacking Tips in the previous 3 months, on average they also scored it above 4 on all of the indicators, ranging from 4.34 to 4.68. For perceived benefits, 6 of the 8 items scored below 4, ranging from 3.20 to 3.96, with the lowest for switching to low-fat/skim milk. The 2 items scoring above 4 were reducing sugary drinks and offering more fruits.

Discussion

Principal Findings

Overall, our beta-testing successfully demonstrated the feasibility of using the CHEW app prototype with mothers of WIC-enrolled black and Hispanic preschool-aged children. Study participants testing the app for 3 months used the app on average once a week for approximately 4 and a half minutes per session, according to app usage logs. However, substantial variation was observed, with very low to no use among 20% of mothers while another half of mothers used it more than twice a month. Usage of specific features averaged at 1 to 2 times per month for shopping-related activities and 2 to 4 times per month for the snack gallery. Over two-thirds of users were tracked using the shopping tools or the snack gallery at least once over the 3-month period. This moderate level of engagement with the prototype app suggests that the app generated enough interest and perceived benefits for most participants to try it and many to continue using it periodically. In order to further expand initial uptake and ongoing engagement with the next version of the app, it will be crucial to maintain high usability and appeal to generate perceived benefits among users.

Mothers classified as users rated the app's WIC Shopping Tools relatively high on usability and benefits, although variation in scores and qualitative feedback highlighted several barriers that need to be addressed: improving the functionality of the barcode scanner, making the shopping tools and produce calculators easier to use, and establishing a process for resolving occasional errors or changes in the database of WIC-approved items. Explicit attention should be paid to planning a mechanism for a future version of the app to communicate seamlessly with WIC's new EBT system in the future.

The Yummy Snack Gallery and Healthy Snacking Tips scored higher on usability than benefits, suggesting that the nutrition education components may have been appealing but too limited in scope and exposure. Qualitative feedback revealed that some users had technical problems with push notifications, some wanted access to more recipes and tips, and individual preferences varied regarding the timing and frequency of Healthy Snacking Tips. In addition to expanding the quantity and frequency of nutrition content, users may benefit from individual tailoring of content to enhance perceived benefits [36,37].

XSL•FO
RenderX

Table 3. Summary of qualitative feedback from users on Children Eating Well smartphone app prototype features.

Feature	Response	Reason	Number
WIC[a] Shopping Tools			
	Reasons for nonusers (n=12)[b]		
		Lost/broken phone	4
		Initial installation failed	3
		Problems with scanner	3
		Inconvenient	1
		Confusing	1
		Hard in store with children	1
	Barriers for users (n=51)[c]		
		None	34
		Scanner slow/inconsistent	8
		Incorrect after WIC change	3
		Items say non-WIC approved	3
		Confusing	3
		Needed more instructions	2
		Forgot to use	1
		Limited time to shop	1
		Would like to include prices	1
	Benefits for users (n=51)		
		Liked it/great/worked well	4
		Shopping tools were helpful	2
		Scanner was helpful	1
		Helpful for using CVV[d]	1
		Should give to everyone	1
Yummy Snack Gallery			
	Reasons for nonusers (n=20)		
		Forgot to use	5
		Lost/broken phone	3
		Unsuccessful installation	3
		No reason given	3
		Problems with other features	2
		Not interested in recipes	2
		Limited variety in recipes	1
		Did not know how to use	1
	Barriers for users (n=43)		
		None	33
		Wanted more recipes	9
		Wanted on another device	1
	Benefits for users (n=43)		
		Liked it/great/fun	6
		Helpful	5

XSL•FO
RenderX

Feature	Response	Reason	Number
		Kids loved it	1
		Easy	1
		Affordable	1
		Good for everyone (non-WIC)	1
Healthy Snacking Tips			
	Reasons for nonusers (n=16)		
		No reason given	8
		Unsuccessful installation	3
		Lost/broken phone	3
		Did not notice the alerts	2
	Barriers for users (n=47)		
		None	38
		Stopped receiving alerts	2
		Messages too frequent	2
		Child already ate healthy	2
		Not interested in reminders	1
		Repeated messages	1
		Time of message delivery	1
	Benefits for users (n=43)		
		Great info/loved it	3
		Helpful to buy/eat more F/V[e]	2
		Wanted more tips	1
		Reminders were helpful	1
		Worked well	1

[a]WIC: Special Supplemental Nutrition Program for Women, Infants, and Children.

[b]One person provided 2 reasons.

[c]Five people provided 2 reasons.

[d]CVV: cash value voucher.

[e]F/V: fruits and vegetables.

XSL•FO
RenderX

Table 4. Usability and perceived benefits of Children Eating Well smartphone app prototype features (note: response choices for each item ranged from 1=strongly disagree to 5=strongly agree. Interpretation: successful—4.0 and higher; needs improvement—3.0 to 3.99; needs major changes—below 3.0).

Feature	Mean (SD)
WIC[a] Shopping Tools (n=51)	
Usability	
Ease of use	
Overall	3.98 (1.12)
Barcode scanner	3.81 (1.32)
Produce calculators	3.95 (1.16)
Helpful	
Overall	4.40 (0.90)
Barcode scanner	3.90 (1.42)
Produce calculators	4.16 (1.00)
Useful: correct information	
Overall	4.29 (1.15)
Barcode scanner	3.96 (1.52)
Produce calculators	4.39 (0.97)
Satisfaction: recommend to others	4.33 (1.05)
Perceived Benefits	
Keep track of family vouchers	4.45 (1.08)
WIC shopping easier	4.10 (1.13)
Show foods allowed to buy	4.51 (0.83)
Regular checkout easier	4.10 (1.14)
CVV[b]checkout easier	4.30 (1.09)
Buy all regular voucher items	4.37 (1.06)
Spend more of CVV	3.57 (1.35)
Yummy Snack Gallery (n=43)	
Usability	
Ease of use: simple	4.81 (0.50)
Ease of use: easy to follow	4.84 (0.43)
Helpful: photos and steps	4.95 (0.21)
Useful: age appropriate	4.77 (0.43)
Useful: fruits that I buy	4.88 (0.39)
Useful: vegetables that I buy	4.63 (0.73)
Satisfaction: recommend to others	4.84 (0.37)
Perceived benefits	
Introduce new fruits	3.93 (1.24)
Introduce new vegetables	4.05 (1.17)
Healthy Snacking Tips (n=47)	
Usability	
Helpful	4.68 (0.59)
Useful: new information	4.40 (0.83)
Satisfaction: enjoy receiving	4.34 (0.94)

XSL•FO
RenderX

Feature	Mean (SD)
Satisfaction: recommend to others	4.60 (0.77)
Perceived benefits	
Increase child's water	3.77 (1.32)
Reduce sugary drinks	4.07 (1.29)
Replace sugary drinks with water	3.96 (1.35)
Switch to low-fat/skim milk	3.20 (1.58)
Offer more fruits	4.11 (1.12)
Offer more vegetables	3.94 (1.24)
Help with picky eater	3.68 (1.31)
Improve my own diet	3.98 (1.27)

[a]WIC: Special Supplemental Nutrition Program for Women, Infants, and Children.

[b]CVV: cash value voucher.

Qualitative feedback from mothers classified as nonusers pointed to several fundamental barriers that could preclude some WIC participants from using the app at all. These included technical problems with their phone (eg, lost or broken), the app not working properly on their phone, challenges in understanding how to use the app, forgetting that the app is on their phone, and lack of interest. One challenge of programming apps on the Android platform is variation in the compatibility of the app across the multiple manufacturers of Android smartphones to use features such as a barcode scanner and push notifications. The multiple manufacturer situation does not apply to the iOS platform, so it my have fewer compatibility issues in the future.

This study adds to the small but growing body of research focused on the application of technology-based interventions in the WIC program to promote program retention and healthy weight gain during childhood. A report on WIC research agenda recommendations identified the need for interventions to increase benefit redemption and consumption of WIC foods, effective approaches to reinforce WIC education between quarterly visits, and effective use of technology in WIC nutrition education [38]. A study in New York found that the two most common reasons for participants to drop out of WIC were negative shopping experiences and low perceived value of the WIC food package [39]. The CHEW app addresses these needs as a technology-based reinforcement intervention designed to increase benefit redemption through WIC shopping tools and consumption of WIC foods through built-in nutrition education tools. In addition, the app could potentially have a positive effect on retention in the WIC program in preschool ages by improving the shopping experience for parents and caregivers and providing greater incentive to remaining in the program.

A study in Michigan found that online WIC education modules were as effective as traditional WIC education for increasing fruit and vegetable consumption [14]. In a 2011 survey of WIC clients/parents in western US states, the majority among ages 20 to 31 years reported interest in the following options for WIC-related services: checking WIC EBT balance online, accessing WIC-authorized food guides online, a smartphone UPC barcode scanning app to check WIC-authorized items, and online recipes [15]. In a recent focus group study, WIC participants expressed interest in the idea of receiving a comprehensive WIC app with multiple features such as recipes, shopping lists, how-to videos, and tools for tracking health behaviors [16].

Limitations

As a common limitation among most studies using self-report data, the survey responses of participants could have been influenced by social desirability bias [40]. Thus, we complemented the subjective self-report data with objective app usage logs to confirm patterns of usage. However, another limitation was not having a mechanism to automatically download extensive app usage data directly from all of the participants due to budget constraints; thus, we had to rely on manual downloads of app usage data among the participants who completed follow-up. Our analyses indicated that participants receiving SNAP benefits were more likely to complete follow-up, but we did not find other demographic differences between completers and noncompleters. However, it is possible that the attrition of noncompleters could have introduced bias in the app usage data. Budget limitations prevented us from continuously modifying and upgrading the app prototype based on ongoing feedback while it was being beta-tested. We plan to incorporate automatically synced app analytics and additional iterative design cycles to develop the next version of the app. Finally, as a pilot study of a prototype app, our sample size was relatively small, particularly for the app usage logs that could not be collected on all participants. However, the findings have generated very useful insights that will inform our future work with this app on a larger scale.

Conclusions

Our team plans to leverage the findings and lessons learned from the CHEW prototype development and testing process to guide our next phase of research, in which we will undergo a new phase of iterative development and usability testing to upgrade to CHEW version 2.0. We plan to create both Android and iOS versions that will be adapted to communicate directly with Tennessee's forthcoming EBT system, overcoming many of the technical challenges we faced in the prototype phase associated with the paper voucher system, multiple participants

XSL•FO
RenderX

per family, and special dietary needs. Subsequently we will assess the impact of the app on improving food purchasing behaviors, parent feeding strategies, and child dietary intake in a randomized trial. We will also explore how certain educational components of the app could be adapted for a Web-based program and print resources to increase accessibility beyond those who do not use smartphones.

With further development, the CHEW app offers the potential to be disseminated eventually to WIC programs in states across the country to improve nutrition and reduce obesity risk among preschoolers. Future versions of the CHEW app will need an efficient mechanism to update the app with periodic revisions to the WIC food packages, changes to the WIC-approved food lists, and changes in the inventory of WIC-authorized products from a wide range of vendors. Since WIC is a federal program that is administered by the states and includes several options for states to customize the program, considerable variation exists across states regarding which foods are approved. Thus, in order to disseminate the app to states beyond Tennessee, future iterations would need to be adapted to specific state-level rules and approved food databases.

The WIC program represents an ideal setting to disseminate technology-based behavioral interventions targeting parents of young children when early intervention is needed to reduce the risk of childhood obesity. Parents and the home environment play a crucial role in primary prevention of childhood obesity [41-43]. Smartphone apps offer the potential to make a substantial impact on population health outcomes, even if the magnitude of impact on each individual is relatively small, given the potential for smartphone apps to reach large numbers of people [44]. More applied research is needed to develop and evaluate innovative interventions implemented in the WIC context such as the CHEW app.

Acknowledgments

PH led the project to develop and evaluate the app prototype. BH obtained the funding for and led the overall Nashville Children Eating Well for Health grant under which this project was funded. PH, JE, MQ, VV, DS, SM, JJ, CH, and BB contributed to various aspects of the development of the app concept, design, and content. VV and DS oversaw and carried out the computer programming for the app. PH, JE, MQ, SM, JJ, and BB contributed to the study design and questionnaire instrument. JE, JJ, and CH supervised data collection, quality control, and data management. PH conducted data analysis and drafted the paper. All authors were involved in writing and revising the paper and approved the final paper as submitted.

This material is based upon work that was supported by the National Institute of Food and Agriculture, US Department of Agriculture (USDA), under award numbers 2011-68001-30113 and 2017-68001-26352. The research was also partially supported by the National Institutes of Health (NIH), award numbers 5U54CA163072, 5U54CA163066, 5U54CA163069, UL1 RR024975, and UL1 TR000445. The content is solely the responsibility of the authors and does not necessarily represent the official views of the USDA or NIH.

The authors would like to thank the Tennessee WIC Program and the Nashville/Davidson County WIC program for their support and guidance in planning and implementing the Nashville Children Eating Well (CHEW) for Health project. The authors would also like to thank the organizations and members of the CHEW Community Advisory Board, including Second Harvest Food Bank for hosting Community Advisory Board meetings and Progreso Community Center for assisting with content development, Spanish translations, recruitment, and data collection with Hispanic participants. The authors are extremely grateful to the faculty, staff, and students who assisted with aspects of the smartphone app development (Dr Jules White, Chad Ryan, Pankaj Chand, Michael Walker) or recruitment and data collection (Dr Barbara Canada, Van Cain, Chizoba Anyimukwu, Nerissa Aquino, Arielle Arzu, Kelsey Burton, Eboné Colclough, Anjelica Crawford, Jacquelyn Favours, Jolynn Franklin, Valentino Gray, Danielle Haik, Melissa Jones, Abimbola Oginni, Diana Ramirez, Erica Ray, Rafael Reyes, Shari Scott, Erika Stanley, Rodney Suttles II, Darnell Towns).

Conflicts of Interest

None declared.

References

1. Ogden CL, Carroll MD, Lawman HG, Fryar CD, Kruszon-Moran D, Kit BK, et al. Trends in obesity prevalence among children and adolescents in the United States, 1988-1994 through 2013-2014. JAMA 2016 Jun 07;315(21):2292-2299. [doi: 10.1001/jama.2016.6361] [Medline: 27272581]
2. Berenson GS, Bogalusa Heart Study Group. Health consequences of obesity. Pediatr Blood Cancer 2012 Jan;58(1):117-121. [doi: 10.1002/pbc.23373] [Medline: 22076834]
3. Centers for Disease Control and Prevention. Nutrition, Physical Activity and Obesity Data, Trends and Maps. 2017. URL: https://nccd.cdc.gov/dnpao_dtm/rdPage.aspx?rdReport=DNPAO_DTM.ExploreByTopic&islClass=OWS&islTopic=OWS1&go=GO [accessed 2017-05-31]
4. Pan L, Freedman DS, Sharma AJ, Castellanos-Brown K, Park S, Smith RB, et al. Trends in obesity among participants aged 2-4 years in the Special Supplemental Nutrition Program for Women, Infants, and Children—United States, 2000-2014.

XSL•FO
RenderX

MMWR Morb Mortal Wkly Rep 2016 Nov 18;65(45):1256-1260 [FREE Full text] [doi: 10.15585/mmwr.mm6545a2] [Medline: 27855143]
5. Thorn B, Tadler C, Huret N, Ayo E, Trippe C. WIC participant and program characteristics 2014. Alexandria: US Department of Agriculture; 2015. URL: http://www.fns.usda.gov/sites/default/files/ops/WICPC2014.pdf [accessed 2017-07-16] [WebCite Cache ID 6o7tRqmaX]
6. Rose D, Habicht JP, Devaney B. Household participation in the food stamp and WIC programs increases the nutrient intakes of preschool children. J Nutr 1998 Mar;128(3):548-555 [FREE Full text] [Medline: 9482762]
7. Nestor B, McKenzie J, Hasan N, AbuSabha R, Achterberg C. Client satisfaction with the nutrition education component of the California WIC program. J Nutr Educ 2001;33(2):83-94. [Medline: 12031188]
8. Woelfel ML, Abusabha R, Pruzek R, Stratton H, Chen SG, Edmunds LS. Barriers to the use of WIC services. J Am Diet Assoc 2004 May;104(5):736-743. [doi: 10.1016/j.jada.2004.02.028] [Medline: 15127057]
9. Phillips D, Bell L, Morgan R, Pooler J. Transition to EBT in WIC: review of impact and examination of participant redemption patterns: final report. Portland: Altarum Institute; 2014. URL: http://altarum.org/sites/default/files/uploaded-publication-files/Altarum_Transition%20to%20WIC%20EBT_Final%20Report_071614.pdf [accessed 2017-07-16] [WebCite Cache ID 6o7tnwfmv]
10. Ammerman A, De Marco M, Harding M, Hartman T, Rhode J. Buying wisely and well: managing WIC food costs while improving the WIC customer's shopping experience. Durham: Duke-UNC USDA Center for Behavioral Economics and Healthy Food Choice Research (BECR); 2015. URL: https://becr.sanford.duke.edu/wp-content/uploads/2016/05/BECR-WIC-Research-Brief.pdf [accessed 2017-02-08] [WebCite Cache ID 6o7tRqmb0]
11. Bertmann FMW, Barroso C, Ohri-Vachaspati P, Hampl JS, Sell K, Wharton CM. Women, infants, and children cash value voucher (CVV) use in Arizona: a qualitative exploration of barriers and strategies related to fruit and vegetable purchases. J Nutr Educ Behav 2014;46(3 Suppl):S53-S58 [FREE Full text] [doi: 10.1016/j.jneb.2014.02.003] [Medline: 24809997]
12. Jacknowitz A, Tiehen L. Transitions into and out of the WIC program: a cause for concern? Soc Serv Rev 2009 Jun;83(2):151-183. [doi: 10.1086/600111]
13. National Academies of Sciences, Engineering, and Medicine, Committee to Review WIC Food Packages. Review of WIC Food Packages: Improving Balance and Choice: Final Report. Washington: National Academies Press; 2017.
14. Bensley RJ, Anderson JV, Brusk JJ, Mercer N, Rivas J. Impact of internet vs traditional Special Supplemental Nutrition Program for Women, Infants, and Children nutrition education on fruit and vegetable intake. J Am Diet Assoc 2011 May;111(5):749-755. [doi: 10.1016/j.jada.2011.02.010] [Medline: 21515124]
15. Bensley RJ, Hovis A, Horton KD, Loyo JJ, Bensley KM, Phillips D, et al. Accessibility and preferred use of online Web applications among WIC participants with Internet access. J Nutr Educ Behav 2014;46(3 Suppl):S87-S92. [doi: 10.1016/j.jneb.2014.02.007] [Medline: 24810002]
16. Biediger-Friedman L, Crixell SH, Silva M, Markides BR, Smith KS. User-centered design of a Texas WIC app: a focus group investigation. Am J Health Behav 2016 Jul;40(4):461-471. [doi: 10.5993/AJHB.40.4.8] [Medline: 27338993]
17. Duggan M, Rainie L. Cell phone activities 2012. Washington: Pew Internet and American Life Project; 2012. URL: http://www.pewinternet.org/2012/11/25/cell-phone-activities-2012/ [accessed 2017-02-08] [WebCite Cache ID 6o7tRqmb5]
18. Mobile technology fact sheet. Washington: Pew Internet and American Life Project; 2014. URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2017-02-08] [WebCite Cache ID 6o7tRqmbA]
19. Smith A. US smartphone use in 2015. Washington: Pew Internet and American Life Project; 2015 Apr. URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2017-07-16] [WebCite Cache ID 6o7tRqmbF]
20. Mojica CM, Parra-Medina D, Yin Z, Akopian D, Esparza LA. Assessing media access and use among Latina adolescents to inform development of a physical activity promotion intervention incorporating text messaging. Health Promot Pract 2014 Jul;15(4):548-555 [FREE Full text] [doi: 10.1177/1524839913514441] [Medline: 24357863]
21. Osborn CY, Mulvaney SA. Development and feasibility of a text messaging and interactive voice response intervention for low-income, diverse adults with type 2 diabetes mellitus. J Diabetes Sci Technol 2013 May;7(3):612-622 [FREE Full text] [Medline: 23759393]
22. Sharifi M, Dryden EM, Horan CM, Price S, Marshall R, Hacker K, et al. Leveraging text messaging and mobile technology to support pediatric obesity-related behavior change: a qualitative study using parent focus groups and interviews. J Med Internet Res 2013;15(12):e272 [FREE Full text] [doi: 10.2196/jmir.2780] [Medline: 24317406]
23. Nollen NL, Mayo MS, Carlson SE, Rapoff MA, Goggin KJ, Ellerbeck EF. Mobile technology for obesity prevention: a randomized pilot study in racial- and ethnic-minority girls. Am J Prev Med 2014 Apr;46(4):404-408 [FREE Full text] [doi: 10.1016/j.amepre.2013.12.011] [Medline: 24650843]
24. Emerson JS, Towns DR, Jones JL, Cain VA, Hull PC. Racial/ethnic and weight status differences in food preparation among WIC participants. J Health Care Poor Underserved 2015 May;26(2):335-344. [doi: 10.1353/hpu.2015.0044] [Medline: 25913333]
25. Zhou YE, Emerson JS, Husaini BA, Hull PC. Association of infant feeding with adiposity in early childhood in a WIC sample. J Health Care Poor Underserved 2014 Nov;25(4):1542-1551. [doi: 10.1353/hpu.2014.0181] [Medline: 25418226]

XSL•FO
RenderX

26. US Department of Agriculture. Dietary Guidelines for Americans, 7th Edition. Washington: US Government Printing Office; 2010. URL: https://health.gov/dietaryguidelines/dga2010/dietaryguidelines2010.pdf [accessed 2017-07-17] [WebCite Cache ID 6s0oGK4Ta]
27. Emerson JS, Husaini BA, Hull PC, Levine RS. Nashville Children Eating Well (CHEW) for Health. J Nutr Educ Behav 2013;45(4):S81.
28. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [Medline: 3068205]
29. Rosenkranz RR, Dzewaltowski DA. Model of the home food environment pertaining to childhood obesity. Nutr Rev 2008 Mar;66(3):123-140. [doi: 10.1111/j.1753-4887.2008.00017.x] [Medline: 18289177]
30. Sallis JF, Glanz K. Physical activity and food environments: solutions to the obesity epidemic. Milbank Q 2009 Mar;87(1):123-154 [FREE Full text] [doi: 10.1111/j.1468-0009.2009.00550.x] [Medline: 19298418]
31. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. Ann Behav Med 2005 Aug;30(1):65-73. [doi: 10.1207/s15324796abm3001_8] [Medline: 16097907]
32. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
33. Vylegzhanina V, Schmidt D, Hull P, Emerson J, Quirk M, Mulvaney S. Helping children eat well via mobile software technologies. 2014 Presented at: Second International Workshop on Mobile Development Lifecycle; 2014; Portland. [doi: 10.1145/2688412.2688413]
34. Emerson J, Hull P. Community care: Nashville Children Eating Well (CHEW) for Health. Int Innov 2015;182:44-46.
35. Choi H, Choi YJ, Kim KM. The understanding of building trust model on smartphone application: focusing on users' motivation. 2012 Presented at: International Conference on IT Convergence and Security; 2011; Amsterdam.
36. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
37. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of Web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
38. Institute of Medicine. Planning a WIC Research Agenda: Workshop Summary. Washington: National Academies Press; 2010.
39. New York State Department of Health. NY WIC Retention Promotion Study: Keep, Reconnect, Thrive. Albany: New York State Department of Health; 2016. URL: https://www.health.ny.gov/statistics/prevention/nutrition/wic/2016_nys_wic_retention_summary_report.pdf [accessed 2017-07-16] [WebCite Cache ID 6o7x0glMW]
40. Crutzen R, Göritz AS. Social desirability and self-reported health risk behaviors in web-based research: three longitudinal studies. BMC Public Health 2010 Nov 23;10:720 [FREE Full text] [doi: 10.1186/1471-2458-10-720] [Medline: 21092267]
41. Curtis KE, Lahiri S, Brown KE. Targeting parents for childhood weight management: development of a theory-driven and user-centered healthy eating app. JMIR mHealth uHealth 2015;3(2):e69 [FREE Full text] [doi: 10.2196/mhealth.3857] [Medline: 26088692]
42. Pimpin L, Ambrosini GL, Llewellyn CH, Johnson L, van Jaarsveld CHM, Jebb SA, et al. Dietary intake of young twins: nature or nurture? Am J Clin Nutr 2013 Nov;98(5):1326-1334 [FREE Full text] [doi: 10.3945/ajcn.113.065250] [Medline: 24047917]
43. Golan M. Parents as agents of change in childhood obesity—from research to practice. Int J Pediatr Obes 2006;1(2):66-76. [Medline: 17907317]
44. Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. Eval Health Prof 2006 Mar;29(1):126-153. [doi: 10.1177/0163278705284445] [Medline: 16510882]

Abbreviations

CAB: Community Advisory Board
CHEW: Children Eating Well
CVV: cash value voucher
EBT: electronic benefits transfer
NIH: National Institutes of Health
SNAP: Supplemental Nutrition Assistance Program
UPC: universal product code
USDA: US Department of Agriculture
WIC: Special Supplemental Nutrition Program for Women, Infants, and Children

XSL•FO
RenderX

Edited by C Dias; submitted 09.02.17; peer-reviewed by R Bensley, B Nezami, S Iyer; comments to author 03.03.17; revised version received 04.06.17; accepted 19.06.17; published 02.08.17.

Please cite as:
Hull P, Emerson JS, Quirk ME, Canedo JR, Jones JL, Vylegzhanina V, Schmidt DC, Mulvaney SA, Beech BM, Briley C, Harris C, Husaini BA
A Smartphone App for Families With Preschool-Aged Children in a Public Nutrition Program: Prototype Development and Beta-Testing
JMIR Mhealth Uhealth 2017;5(8):e102
URL: http://mhealth.jmir.org/2017/8/e102/
doi:10.2196/mhealth.7477
PMID:28768611

©Pamela Hull, Janice S Emerson, Meghan E Quirk, Juan R Canedo, Jessica L Jones, Violetta Vylegzhanina, Douglas C Schmidt, Shelagh A Mulvaney, Bettina M Beech, Chiquita Briley, Calvin Harris, Baqar A Husaini. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 02.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Analysis of Requirements for Developing an mHealth-Based Health Management Platform

Hehua Zhang[1*], MA; Han Zhang[2*], PhD; Xiaoning Wang[2*], PhD; Zuosen Yang[3*], MPH; Yuhong Zhao[4*], PhD

[1]Shengjing Hospital of China Medical University, China Medical University, Shenyang, Liaoning, China

[2]Department of Medical Informatics, China Medical University, Shenyang, Liaoning, China

[3]Health Information Center of Liaoning Province, Shenyang, Liaoning, China

[4]Shengjing Hospital of China Medical University, China Medical University, Liaoning, Shenyang, China

[*]all authors contributed equally

Corresponding Author:
Yuhong Zhao, PhD
Shengjing Hospital of China Medical University
China Medical University
Heping District, Sanhao
Street No. 36
Liaoning, Shenyang,
China
Phone: 86 18900910777
Email: zhaoyh@sj-hospital.org

Abstract

Background: Studies have consistently shown that mobile and Web-based apps have positive impacts on people's daily lifestyles, health management, and disease treatment. As the development of medical and health informatization in China has evolved, different kinds of mobile-based apps for individuals and hospitals have been developed by software vendors. However, doubts and challenges posed by the media have prevented these apps from having a stable and substantial user base. Analyses of user requirements have not typically been performed prior to the design of such mobile apps. The health information government authority in Liaoning Province, China, was planning to establish a mobile health (mHealth)-based health management platform, aiming to alleviate the difficulties citizens have in seeking hospital services.

Objective: The goal of this study was to determine the actual health and medical needs of citizens that may be addressed by medical information technologies. The results may contribute to the functional design and development of health management and appointed treatment-oriented mobile apps.

Methods: In this study, a semi-structured questionnaire on mHealth requirements was designed and tested, and 240 questionnaires were given to the outpatients of the First Hospital of the China Medical University in Shenyang, Liaoning Province, China; of these, 228 valid responses were collected, for a response rate of 95%. We discussed the current development of mHealth with 50 related experts and engineers from health authorities and a medical information company. SPSS 13.0 was used for statistical analyses.

Results: After detailed analyses of the questionnaire data, several findings were evident: first, most citizens and patients were unclear about their health conditions (64.5%, 147/228) and were interested in receiving a mobile app as a tool to manage their health and medical needs (71.1%, 162/228). Patients in different outpatient departments had different opinions regarding online registration. Conversely, the main problems for outpatients were long waiting times (66.4%, 148/223) and difficulties in making appointments (46.5%, 106/228), and they also worried about payments and Internet problems when using a mobile app for appointment reservations. Furthermore, as the main service target of mHealth is the health management of the general population, we first need to solve the associated interoperability and data security problems associated with such apps.

Conclusions: This study provides insight into the health and medical requirements of smartphone apps, and draws attention to some of the challenges and opportunities of mHealth. We suggest several value-added features and characteristics that app developers should take into consideration when developing health and medical-related apps. The findings also highlight some major challenges that require further consideration and research to ensure that these apps meet the core needs of patients and aid the development of the health information system in Liaoning Province, China.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2017;5(8):e117) doi:10.2196/mhealth.5890

KEYWORDS

online information-seeking; health management; mHealth

Introduction

Development and App of mHealth Apps

The concept of mobile health (mHealth) appeared in the early 21st century with the development of mobile phones and wireless technology [1], and since that time research has consistently shown the deployment of mHealth will benefit patients with chronic disease [2], making it easier to implement primary health care [3] and reduce health care costs [4]. Different kinds of mobile phone or tablet-based apps on Android and/or iOS operating systems have been exploited for use inside [5] and outside of hospitals for the purposes of nursing support, outpatient management, chronic disease management, online medical consulting, daily health, and public service [4,6,7]. Research on evidence-based practices consistently show the positive effects of using mHealth-based apps [8-10]. In China, mHealth is in a stage of rapid growth, and many mHealth apps that are not specifically used by health care providers are familiar to (and used by) patients outside of hospitals, including online doctor services that are available at any time, health data tracking, and real-time analyses using wearable devices and self-help checking. Some hospitals have even developed their own hospital mobile apps in cooperation with online payment companies. Governments in China are taking positive measures to encourage mHealth development, and venture capital funds are investing in mHealth apps to promote the prosperity of the health market [11].

mHealth-Based Health Management Platform in Liaoning Province, China

The entire health information structure of Liaoning Province is at the forefront of development in China. The system includes five parts known as *46312*: a four-level health information platform integrating county-city-province to nation; six key business apps (medications, medical service, health care, health insurance, fertility service, and integrated apps); three-database construction (electronic medical records, electronic health records, and population); one virtual private network connecting all platforms and apps; and a two-standard hierarchy (health information and security). The ultimate aim of this structure is to provide more convenient services for patients using current information technologies, which is the underlying reason for genesis of this project involving an mHealth-based health management platform in Liaoning Province (referred to as *the platform* in the sections below). To alleviate the difficulties that regional citizens experience while seeking medical and health services [12], the platform will be integrated with large regional hospitals by adopting data exchange and mHealth-related technologies, which would solve the main problems that patients confront by providing multiple ways of accessing medical and health services (including mobile phone apps, websites, and Wireless App Protocol-based apps) [13].

Purpose of This Study

Various kinds of mHealth-based apps for patients are presently available on the Internet, yet none can maintain a stable and large-scale user base to support a range of province-based health requirements. We believe the underlying reason for this problem lies in the fact that existing apps do not meet the fundamental health requirements of the general public [14]. Compared with foreign research (eg, on online information-seeking, appointment making, and patient-provider interactions [15,16]), we have not found any comprehensive requirement studies of patients or experts on mHealth for the purpose of a system designed in China, except for some small Internet surveys that are not credible for platform development [17]. In this study, we aimed to acquire and analyze patients' requirements of medical, health, and related mHealth apps in Liaoning Province, which will become the foundation for the functional design and development of a province-wide innovation platform in China.

Methods

Questionnaire Development

There were few preexisting questionnaires that were suitable for the investigation of mHealth requirements in this study, so we designed a new questionnaire that includes two parts: *Part A* for citizens and patients (30 questions), and *Part B* for health information-related persons (10 questions). Part A is divided into three sections: basic information questions, medical and health-related questions, and smartphone app questions (see the detailed contents in Multimedia Appendix 1). Part B concerns the current and future development (and current problems) of mHealth, as relayed by health information-related professionals. We invited four experts (three from the Medical Informatics Department of China Medical University and one from the Ministry of Health in Liaoning Province) to review the content and structure validity of the questionnaire. Subsequently, we distributed the questionnaire to 300 volunteers and, using SPSS13.0 to collect the data and assess its reliability, modified the questionnaire to make it workable for a large-scale investigation. The final reliabilities were 0.802 and 0.848 for Parts A and B, respectively.

Participants

For questionnaire Part A, we randomly selected patients and related citizens (consisting of family or friends accompanying the patients) in six outpatient departments (medicine, dermatology, gynecology, ophthalmology, otorhinolaryngology, and mental health) of the First Hospital of China Medical University, which is the largest comprehensive hospital in Northern China with an outpatient capacity of over 2,630,000 per year. Patients with different kinds of disease conditions come from different cities in Liaoning Province because they prefer to visit this hospital first. Thus, the opinions of participants from the hospital outpatient areas represent a substantial proportion of the viewpoints of citizens in Liaoning

XSL•FO
RenderX

Province. For Part B, we randomly chose engineers, clerks, and managers from three information centers in a hospital, Liaoning Province Ministry of Health, and an information technology (IT) company to complete the questionnaire.

Survey Deployment

Three investigators visited the waiting areas of the six outpatient departments during the periods of December 2-4 and December 9-11 in 2013, when outpatient visits are typically at the highest level. Investigators individually explained the questionnaire to participants and asked the patients or citizens to fill it in, and retrieved each completed questionnaire. For Part B, we investigated health-related persons from three institutions that are the main participants in the health informatization process (a hospital information center, an information center of the health board, and an IT company). At the beginning of each questionnaire, we explained that the survey was part of a platform project that was proposed by China Medical University and that the information collected from the participants was only for research purposes. We did not give out the questionnaires until the patients understood this and were willing to complete the survey.

Analysis

We used SPSS13.0 to process and analyze the questionnaire data, which included the following three parts: *Define Variables*, *Data Entry*, and *Data Statistics*.

Define Variables

In the variable view, each question item of the questionnaire was treated as a variable; we defined the name, data type, width, label, values, and other variables for each.

Data Entry

The types of questions were choice questions or multiple-choice questions; we entered all options selected by our respondents in the questionnaires into SPSS according to the defined variables.

Data Statistics

After data entry, we used the frequency, crosstabs, and tables in the descriptions function, defined multiple response sets, and conducted multiple response analyses. Cross tab analysis is a Chi-square test method to measure relevance between every two variables in SPSS13.0. Multiple response analysis is a frequency analyzed when there can be more than one response to a survey question per participant, and allows the set of responses to be combined and collectively analyzed.

Results

Overview

Two hundred and forty patients and citizens responded to questionnaire A, of whom 12 were excluded due to incomplete answers (more than two missing answers). At total of 228 valid respondents were included in our analysis (see Table 1). Many studies have pointed out that age and sex heavily affect online information-seeking behavior. We hypothesized that visit type (whether a person is visiting for the first time [first] or second time [subsequent], or is a companion [citizen] with patients visiting a hospital during the past 2 months) was a key factor for patients acquiring health information. Patients tend to go to large hospitals instead of general practitioners to seek medical treatment once they get sick. In the current health system of China, the more times that patients go to a hospital, the more they learn about their health conditions. Of the 228 respondents, almost half were women (111/228, 48.7%), and the average age was 31.5 years, with a high rate of young adults (43.9%, 100/228); 42.5% (97/228) were first-time visitors to the hospital, and 35.5% (81/228) were second-time visitors to the hospital.

Nine of 60 respondents to Part B of the questionnaire were excluded due to careless and repeated answers. These respondents covered all health information departments in Liaoning Province; different position levels may have influenced their opinions about mHealth development (see Table 2).

Table 1. Participant demographics of the Part A questionnaire.

Characteristic		n (%)
Total (N)		228 (100.0%)
Age (years)	18-29	100 (43.9%)
	30-40	87 (38.2%)
	41-69	41 (18.0%)
Sex	Male	117 (51.3%)
	Female	111 (48.7%)
Visit type	First	97 (42.5%)
	Subsequent	81 (35.5%)
	Citizen	50 (21.9%)

XSL•FO
RenderX

Table 2. Participant characteristics of the Part B questionnaire.

Department	Position			
	Clerk	Chief Information Officer	Engineer	Researcher
Government health information center, n (%)	0 (0.0%)	8 (34.8%)	15 (65.2%)	0 (0.0%)
Hospital, n (%)	8 (57.1%)	1 (7.1%)	1 (7.1%)	4 (28.6%)
Information technology company, n (%)	0 (0.0%)	1 (7.7%)	12 (92.3%)	0 (0.0%)

Evaluation Outcomes

Questionnaire A

Questionnaire A is divided into three parts: *health information-seeking*, *appointment making and medical treatment*, and *smartphone usage*. Comprehensive analyses are detailed in Multimedia Appendix 2.

Health Information-Seeking

Regardless of age, sex, or visit type, most of the respondents were unaware of their health conditions (62.3%, 142/228) and unclear about former preventive care records (84.6%, 193/228) or hospital visit records (65.4%, 149/228). Most patients (73.7%, 168/228) were willing to check their health, visit, or test records if it was convenient, with most respondents (64.5%, 147/228) obtaining disease information mostly and/or only from their doctors. Patients (citizens) preferred to get health information from a health expert or hospital visits (65.4%, 149/228) compared to Internet searches (21.1%, 48/228); 55.2% (126/228) indicated that they would likely and/or certainly use the Internet or a mobile app to get some information before seeking a doctor. Additionally, 67.1% (153/228) of respondents answered that they might follow some health suggestions from a prestigious health website; 71.5% (163/228) said they wanted to learn medical-related knowledge if it was convenient. Among all of the above, some significant findings include: women were more likely than men to know partially about their former preventive records (χ^2=10.855, P<0.01); subsequent patients knew more about their health status than first-visit patients (χ^2=21.075, P<0.01); patients were more aware of their former hospital visit records than nonpatients (χ^2=9.553, P<0.05); males were more likely than females to choose Internet searches as the way to obtain health information (χ^2=10.592, P<0.05); first-visit hospital patients would rather choose Internet information-seeking than others (χ^2=15.612, P<0.05); and young adults (18-29 years old) would be more likely than others to use the Internet or a mobile app to obtain some health information before seeking a doctor.

Multiple response analysis showed the most wanted health information included a physical examination, disease awareness, and daily health; computer and smartphone apps were the most favorable ways to search the Internet.

Appointment Making and Medical Treatment

Multiple response analysis showed that the three most common problems during medical treatment were long waiting times (66.4%, 148/223), difficulty making appointments (46.5%, 106/228), and unclear treatment results (40.3%, 87/216). Table analysis and Chi square tests showed the most desirable way to make an appointment with a doctor was online (49.6%, 113/228) regardless of age or sex, especially in outpatient departments of internal medicine, otorhinolaryngology, and mental health (χ^2= 213.077, P<0.01). However, the results varied by visit type (χ^2=25.782, P<0.01), which meant citizens were more likely to choose online appointment-making than others. Respondents showed similar degrees of concern regarding online appointment-making, which included: tedious processes (37.3%, 85/228), registering with the wrong department (33.6%, 75/223), inability to find the most wanted doctor (35.0%, 78/223), and more difficult hospital processes (30.9%, 69/223).

Compared with hospital payment (30.7%, 69/225) and uncertain answers (22.7%, 51/225), most respondents indicated that they wanted to pay the appointment-making fee online (46.7%, 105/225) regardless of age, sex, or visit type, especially in outpatient departments of internal medicine, otorhinolaryngology, and mental health (χ^2= 28.843, P<0.05). After online appointment-making, 81.9% (86/105) of the respondents wanted an alert to show the time and process of seeking a doctor in the hospital. If the appointed doctor could not be available on the day (doctors' schedules are often changed in the hospital), 58.9% (66/112) of respondents chose to make a decision later based on the factual situation, while subsequent visit patients were less likely to continue their hospital visit (χ^2=10.911, P<0.05).

In total, 49.1% (112/228) of participants preferred to choose a time to go to the waiting room and wait for the doctor rather than staying in the waiting room after appointment-making, especially in outpatient departments of internal medicine, otorhinolaryngology, and mental health χ^2= 27.543, P<0.01), regardless of age or visit type. However, females showed more willingness to wait than males (χ^2=9.968, P<0.01). Most respondents only showed partial satisfaction with their medical treatment outcome (54.6%, 112/205). Reasons for dissatisfaction with the doctor-seeking experience included: *still unclear about my condition* (29.1%, 62/213), *have unanswered questions* (49.3%, 105/213), and *doctors' poor attitude* (21.6%, 46/213).

Thirty-six respondents wrote down other reasons and suggestions about unsatisfactory hospital outpatient experiences, including: five pointed out a lack of clarity of hospital processes and/or ways to choose the correct doctor because of unclear doctor interfaces; three were unclear about the doctors' prescriptions; three complained about the high medical treatment price, which mainly included medication and physical tests; five respondents did not trust their doctors' diagnosis; three complained of doctors' and nurses' poor service attitudes; five pointed out the prolonged waiting time of outpatients, and

XSL•FO
RenderX

waiting for test results after a difficult appointment-making process in the hospital; four respondents were still unclear about all of their conditions or test results; two reported poor navigation and disorder in the outpatient section in the hospital; and four expressed their opinions about the online appointment-making process (Internet problems causing failure, functions that should be customized, mobile payments should also be acceptable, and the process should record the entire view of a patient's health care).

Smartphone and App Usage

We defined a smartphone as a phone that has a standalone operating system, running memory, and connection to wireless Internet and mobile communication Internet (so that users can install any third-party software, such as games, navigation, and so on). Data analysis showed that 96.4% (216/224) of our respondents had smartphones, but department difference (χ^2=14.18, P<0.05) indicated that elderly patients from the ophthalmology department may not use smartphones. Pearson Chi-square test results showed daily time spent on a smartphone significantly declined as the users' age increased (χ^2=20.593, P<0.01), and differences were noted in different departments (χ^2=31.613, P<0.01). Among the three age groups (18-29, 30-40, >41), the percentages of daily time spent on a smartphone >3 hours were 45.9% (45/98), 28.2% (24/85), and 22.5% (9/40), respectively. Most respondents (69.8%, 150/215) chose "YES" when asked about receiving pushed messages from health and medical-related apps.

Multiple response analysis results showed several trends: the two most used functions of smartphones were apps (85.5%, 189/221) and phone calls (49.3%, 109/221); the two most used kinds of smartphone-based apps were instant messengers (84.2%, 186/221) and news (59.3%, 131/221); the three most desired kinds of apps were health related (63.9%, 138/216), daily life service related (52.3%, 113/216), and medical treatment related (48.6%, 105/216); respondents comprehensively showed high health-related app expectations regarding health management (47%, 101/215), health learning (59.5%, 128/216), and health record queries (52.1%, 112/216); medical-related app expectations included appointment-making (65.1%, 142/218), treatment results queries (59.2%, 129/218), and patient-provider interactions (53.7%, 117/218). The most accepted kinds of push messages from smartphone apps were test results (58.3%, 127/218) and appointment alerts (51.8%, 113/218).

Questionnaire B

Crosstab analysis showed that 68% of respondents (34/50) thought mHealth would benefit general citizens more than other institutions, regardless of working units or positions. Multiple response analysis showed: information technology development (71%, 35/49), the health care industry (43%, 21/49), and citizens' health needs (39%, 19/49) together gave rise to mHealth development; the roles that smartphones play in the mHealth industry mainly involved the information display platform (78%, 39/50) and data collection (or transmission) tools (46%, 23/50); the most used smartphone functions in the health industry were the Internet (70%, 35/50), apps (58%, 29/50), messaging (52%, 26/50), and sensors (42%, 21/50). Based on current health informatization achievements, mHealth was thought to benefit citizens by enhancing health information access (63.3%, 31/49) and improving the hospital visiting process (63.3%, 31/49). Respondents indicated that mHealth usage could benefit the health industry by optimizing health care process management (60%, 30/50) and embodying the value of the whole health informatization structure (48%, 24/50). Data security (49%, 24/49), interoperability (41%, 20/49) and legality (47%, 23/49) were nominated as the main concerns that mHealth is facing in the current situation. The health department (72%, 36/50), hospitals (78%, 39/50), insurance entities (68%, 34/50), researchers (44%, 22/50), and patients (68%, 34/50) were all thought to be important participants in mHealth development. Participants also indicated that data sharing between providers and patients (62%, 31/50), health services (eg, health education and family health; 74%, 37/50) and chronic disease management (56%, 28/50) could be improved by mHealth in the current health environment.

Discussion

Principal Results

Health Information-Seeking

In this study, we found women were more likely than men to be concerned about health information. Patients were more likely to be familiar with their historical health records than nonpatients, which is in accordance with one of the common Chinese health scenarios, which is that people usually do not care about their health until they get sick. Women are usually more conservative than men regarding new technologies, which explains why males were more likely than females to choose an Internet search as a way to obtain health information. This finding is similar to German research conducted by Bidmon and Terlutter [18], who suggested that women were more engaged in searching for health information, while men were more likely to adapt to a virtual patient-physician relationship, such as personal appointment-making, referral to other doctors, writing prescriptions, discussions of normal test results, and doctor's notes/certificates of health. First-time hospital-visiting patients would rather choose Internet information-seeking than subsequent patients; young adults (18-29 years old) would be more likely than others to use the Internet or a mobile app to obtain some health information before seeking a doctor, which provides evidence for the conclusion that health communication efforts utilizing mHealth will have a strong impact when the target population is a younger generation [19]. The Internet is increasingly becoming one of the major health information media formats [20]. Patients were favorably disposed to learning health-related information through the Internet or mobile phones. The results above suggest that we should provide functions for checking historical health records and a medical knowledge database for self-help learning when designing an mHealth-based health management platform in Liaoning Province.

XSL•FO
RenderX

Hospital Visiting

Significant results include: citizens tended to choose online appointment-making more than first and/or subsequent patients, which is similar to the result for online information-seeking and indicates that people would like to search for health information before seeking a doctor. These responses may be due to limited health care access [21], especially in China. Subsequent patients do not seem to like to continue with hospital visits when their appointed doctor changes, indicating that they want to choose the doctor themselves. This trend indicates another problem in the Chinese health system: patients require a constant primary health care provider instead of seeking unknown hospital doctors [22]. The long waiting time after an outpatient appointment has been made is a major problem during hospital visits; thus, most respondents want to choose an appropriate time to come to the hospital waiting area. However, women are more willing than men to wait until the doctor's office calls them, which reflects the fact that men in China are more open-thinking and more readily adapt to new things whereas women are more conservative, which is the opposite to the situation in many foreign countries in which women are more likely to seek and accept online health information [18,23].

The findings from this part of the survey have great significance for our platform design and development. In general, we found appointment-making and long waiting times were the most commonly confronted problems in hospital outpatient visiting, and most patients and citizens were willing to make appointments online, which means that designing an online appointment-making function to alleviate the current situation is necessary. However, some patients worried that they might possibly choose the wrong department or doctor, or encounter a more complicated hospital process when making an appointment online, so we need to take these factors into consideration when developing mobile apps. To integrate with the hospital management system, hospitals need to make some changes to the outpatient process to accommodate online appointment-making to make it more convenient. As we perceived, patients would like to pay the appointment fee online, as this is more convenient; after successful payment, most patients want to receive alert messages to remind them of the hospital visiting time, address, and other instructions. Most citizens, and especially patients, would hesitate if the appointed doctor is changed, and when a patient goes to the hospital outpatient department on the appointed day, they do not like to wait for a long time until it is their turn.

According to the results above, we should take every possible expected (or unexpected) circumstance into careful consideration when designing the function frame and workflows of the platform, keeping users fully informed of every change in their appointments. Online payment is a key point for more convenient appointment-making for hospital outpatient visits. However, there are some special challenges that cannot be solved solely by mHealth platform development, such as the increasingly intense relationship between patients and providers, high medical costs and medication prices, and the deficiency of high quality medical resources.

Smartphone Usage

The results found in this part are in accordance with current mobile technology development. Almost all respondents had a smartphone, and over half were familiar with smartphone-based apps. Time spent on the Internet on smartphones was negatively correlated with age because most of our respondents were young adults who display online health information-seeking behaviors [20,24]. Most of the respondents would like to try new health apps, such as health management and medical treatment-related apps. In addition to online appointment-making and health history recording, respondents want to use apps to query test results, learn about their conditions, and talk with their doctors. Future research should pay more attention to doctor/patient communication.

mHealth Development

We predict the future direction of mHealth is health management and health improvement, targeted at meeting the major requirements of general citizens, such as health management apps, health monitoring, and telemedicine [10,25]. Smartphone apps could be connected with wearable devices through technologies (eg, Bluetooth, Wi-Fi, and wireless sensors) to realize real-time data collection and analytical functions on regional health information platforms, which would provide citizens with online services such as health information storage, information querying, personal health assessments, and chronic disease management [26]. To achieve the final goal of providing more convenient health services for general citizens, mHealth development should be supported by the government, medical institutions, medical insurance, and patients themselves. Governments should set performance standards to evaluate the safety, validity, and practicality, together with hospitals and patients [9,27]. The most common problems we need to solve for mHealth development are safety and interoperability [28]. Due to the explosive increases in health and medical data, data safety is a major problem in the medical field, which is exacerbated by the implementation of mHealth-based app usage [29]. Overall, respondents were all positive about mHealth usage, development, and significance, which is in accordance with current health policy suggestions.

General Suggestions for mHealth-Based Health Management Platform Development

Platform development needs to focus more on basic functional designs. Our requirements analysis showed that most citizens were unclear about their historical health records, so the query function design should use this as a foundation. This approach is in accordance with other studies showing that we should provide better means of navigating online health information by improving systems supporting patients' health information-seeking activities [30,31].

We also need to pay more attention to small details, which is the key to the success of mHealth apps. Study results indicated patients usually encounter many problems during their hospital visits, such as unfamiliarity with hospital outpatient doctors, difficulty knowing which doctor to choose, inability to find a particular location, and frustration from waiting times that are

XSL•FO
RenderX

too long during their hospital visits. mHealth app development should try to solve these particular problems.

We also need to address interoperability. Referring to medical treatment functions (eg, appointment making and online payment), hospitals and the mHealth platform should exchange some key parameters, and the hospital should also make some changes to simplify the process of patients visiting their doctors after they have made online appointments, to make it easier for patients to make an online appointment to encourage app usage. Regarding sustainable development, mHealth development needs to continue to devote funds, manpower, and devices to maintain a sustainable operation, which is a key point [32]. Apart from providing free basic services, some customized value-added services are needed and necessary, and more research is needed to investigate what kinds of services would be acceptable.

Novelty

To our knowledge, this is the first comprehensive study analyzing the requirements of an mHealth platform before proceeding to system design and development in Liaoning Province. We designed and verified a set of questionnaires directed at discovering Internet health information-seeking behavior, online appointment making, mHealth-based app usage, and mHealth development. Based on the current achievements of health informatization, such as the comprehensive regional health platform connecting large hospitals and the unified payment platform, the Health Ministry encourages the development of mHealth-related apps to provide patients with more convenient tools for health management and medical treatment. We believe this study provides many meaningful ideas for functional design and program development of mHealth related apps.

Limitations

There is the possibility of selection bias among respondents, since we only chose patients (or citizens) who were willing to fill in the questionnaire for us. Young adults showed more interest and understood our questions better, which caused a disproportionate age distribution, and a few elder respondents missed one or two questions, which may be due to their limited health literacy or our incomprehensive design of the questionnaire. There may be more health information-seeking problems or requirements that will need to be explored in the future. It is very common in China that elders use fewer computer or smartphone functions than young adults, even when they have access to the Internet, and we also saw this trend in our results. We believe that the design of the mHealth app should mainly be oriented to young adults, who will help elder family members when seeking a doctor in a hospital online, and will also help elder people improve their online health information-seeking behavior [33]. Furthermore, our study is simply an analysis of the requirements for a large-scale information system development process; the effectiveness of our study should be evaluated according to the implementation of the mHealth-based platform and other similar projects.

Acknowledgments

We would like to thank Dr. Zhang and Dr. Wang (China Medical University) for their help with the design of the questionnaire. We would also like to thank the leaders at the Health Information Center of Liaoning Province and the First Hospital of China Medical University. This study was supported by the People's Republic of China Ministry of Science and Technology Project (Northeast Area Natural Population Cohort Study; project number 2017YFC090740)

Conflicts of Interest

None declared.

Multimedia Appendix 1

mHealth requirement survey.

[PDF File (Adobe PDF File), 61KB - mhealth_v5i8e117_app1.pdf]

Multimedia Appendix 2

Health information seeking characteristics.

[PDF File (Adobe PDF File), 63KB - mhealth_v5i8e117_app2.pdf]

References

1. 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]
2. Jaglal SB, Haroun VA, Salbach NM, Hawker G, Voth J, Lou W, et al. Increasing access to chronic disease self-management programs in rural and remote communities using telehealth. Telemed J E Health 2013 Jun;19(6):467-473 [FREE Full text] [doi: 10.1089/tmj.2012.0197] [Medline: 23570277]

XSL•FO
RenderX

3. Leventhal T, Taliaferro JP, Wong K, Hughes C, Mun S. The patient-centered medical home and health information technology. Telemed J E Health 2012 Mar;18(2):145-149. [doi: 10.1089/tmj.2011.0130] [Medline: 22304440]
4. Keh H, Shih C, Chou K. Integrating unified communications and Internet of m-health things with micro wireless physiological sensors. J App Sci Eng 2014;17(3):319-328 EI source with no PMCID. [doi: 10.6180/jase.2014.17.3.12]
5. Park JY, Lee G, Shin SY, Kim JH, Han HW, Kwon TW, et al. Lessons learned from the development of health applications in a tertiary hospital. Telemed J E Health 2014 Mar;20(3):215-222 [FREE Full text] [doi: 10.1089/tmj.2013.0192] [Medline: 23909863]
6. Williams V, Rutter H, Christy T, Tarassenko L, Farmer A. Exploring patients' perspectives of an mHealth application: a qualitative study as part of EDGE COPD. Int J Integr Care 2013 Nov 20;13(7):a. [doi: 10.5334/ijic.1447]
7. Corpman DW. Mobile health in China: a review of research and programs in medical care, health education, and public health. J Health Commun 2013;18(11):1345-1367. [doi: 10.1080/10810730.2013.778370] [Medline: 23915211]
8. Lane ND, Mohammod M, Lin M, Yang X, Lu H, Campbell AT. BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing. 2011 Jan. URL: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=FC947F03E1801D39ADF6C1537546D075?doi=10.1.1.298.2259&rep=rep1&type=pdf [accessed 2017-07-28] [WebCite Cache ID 6sIpO85cC]
9. Rajput ZA, Mbugua S, Amadi D, Chepngeno V, Saleem JJ, Anokwa Y, et al. Evaluation of an Android-based mHealth system for population surveillance in developing countries. J Am Med Inform Assoc 2012;19(4):655-659 [FREE Full text] [doi: 10.1136/amiajnl-2011-000476] [Medline: 22366295]
10. Martínez-Pérez B, De La Torre-Diez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
11. iiMedia Research. 2013. China mobile health market research annual report (2012-2013) URL: http://www.iimedia.cn/36664.html [accessed 2017-07-20] [WebCite Cache ID 6s70K3SNy]
12. Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey. J Med Internet Res 2015 Apr 29;17(4):e106 [FREE Full text] [doi: 10.2196/jmir.4126] [Medline: 25925943]
13. Tang S, Brixi H, Bekedam H. Advancing universal coverage of healthcare in China: translating political will into policy and practice. Int J Health Plann Manage 2014;29(2):160-174. [doi: 10.1002/hpm.2207] [Medline: 23877959]
14. Ronda MC, Dijkhorst-Oei LT, Rutten GE. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014 Nov 25;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
15. Lu SH, Dear BF, Johnston L, Wootton BM, Titov N. An internet survey of emotional health, treatment seeking and barriers to accessing mental health treatment among Chinese-speaking international students in Australia. Couns Psychol Q 2013 Aug 07;27(1):96-108. [doi: 10.1080/09515070.2013.824408]
16. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de Rooij S, et al. Health information-seeking behavior of seniors who use the Internet: a survey. J Med Internet Res 2015 Jan 08;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]
17. Liu J, Liu Z, Zhang Z, Dong S, Zhen Z, Man L, et al. Internet usage for health information by patients with epilepsy in China. Seizure 2013 Nov;22(9):787-790 [FREE Full text] [doi: 10.1016/j.seizure.2013.06.007] [Medline: 23831148]
18. Bidmon S, Terlutter R. Gender differences in searching for health information on the Internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. J Med Internet Res 2015 Jun 22;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
19. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
20. Beck F, Richard JB, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014 May 13;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
21. Bhandari N, Shi Y, Jung K. Seeking health information online: does limited healthcare access matter? J Am Med Inform Assoc 2014;21(6):1113-1117 [FREE Full text] [doi: 10.1136/amiajnl-2013-002350] [Medline: 24948558]
22. Yang H, Huang X, Zhou Z, Wang HH, Tong X, Wang Z, et al. Determinants of initial utilization of community healthcare services among patients with major non-communicable chronic diseases in South China. PLoS One 2014;9(12):e116051 [FREE Full text] [doi: 10.1371/journal.pone.0116051] [Medline: 25545636]
23. Choi N. Relationship between health service use and health information technology use among older adults: analysis of the US National Health Interview Survey. J Med Internet Res 2011 Apr 20;13(2):e33 [FREE Full text] [doi: 10.2196/jmir.1753] [Medline: 21752784]
24. Lee YJ, Boden-Albala B, Larson E, Wilcox A, Bakken S. Online health information seeking behaviors of Hispanics in New York City: a community-based cross-sectional study. J Med Internet Res 2014 Jul 22;16(7):e176 [FREE Full text] [doi: 10.2196/jmir.3499] [Medline: 25092120]

XSL•FO
RenderX

25. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
26. Kartsakli E, Lalos AS, Antonopoulos A, Tennina S, Renzo MD, Alonso L, et al. A survey on M2M systems for mHealth: a wireless communications perspective. Sensors (Basel) 2014 Sep 26;14(10):18009-18052 [FREE Full text] [doi: 10.3390/s141018009] [Medline: 25264958]
27. Sun Y, Wang N, Guo X, Peng Z. J Electron Commer Re. 2013. Understanding the Acceptance of Mobile Health Services: A Comparison and Integration of Alternative Models URL: http://web.csulb.edu/journals/jecr/issues/20132/paper4.pdf [accessed 2017-07-28] [WebCite Cache ID 6sIqQCla7]
28. Payne JD. mHealth Alliance. 2013. Standards and interoperability for mHealth among low- and middle-income countries URL: https://hingx.org/ [accessed 2017-07-21] [WebCite Cache ID 6s77or79b]
29. Luxton DD, Kayl RA, Mishkind MC. mHealth data security: the need for HIPAA-compliant standardization. Telemed J E Health 2012 May;18(4):284-288. [doi: 10.1089/tmj.2011.0180] [Medline: 22400974]
30. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014 Dec 02;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
31. Bratucu R, Gheorghe IR, Purcarea RM, Gheorghe CM, Popa VO, Purcarea VL. Cause and effect: the linkage between the health information seeking behavior and the online environment--a review. J Med Life 2014 Sep 15;7(3):310-316 [FREE Full text] [Medline: 25408746]
32. Mechael P. mHealth Alliance. 2013. Sustainable financing for mobile health (mHealth): options and opportunities for mHealth financial models in low and middle income countries URL: http://www.fphighimpactpractices.org/briefs/mhealth/ [accessed 2017-07-21] [WebCite Cache ID 6s78YkZ5X]
33. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015 Jan 13;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]

Abbreviations

IT: information technology
mHealth: mobile health

Edited by G Eysenbach; submitted 07.07.16; peer-reviewed by H AlSaleh, K Blondon, A Martinez-Millana, TR Soron, WC Su; comments to author 24.10.16; revised version received 25.01.17; accepted 25.03.17; published 03.08.17.

Please cite as:
Zhang H, Zhang H, Wang X, Yang Z, Zhao Y
Analysis of Requirements for Developing an mHealth-Based Health Management Platform
JMIR Mhealth Uhealth 2017;5(8):e117
URL: http://mhealth.jmir.org/2017/8/e117/
doi:10.2196/mhealth.5890
PMID:28778840

©Hehua Zhang, Han Zhang, Xiaoning Wang, Zuosen Yang, Yuhong Zhao. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 03.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Opportunities of mHealth in Preconception Care: Preferences and Experiences of Patients and Health Care Providers and Other Involved Professionals

Matthijs R Van Dijk[1], MD; Maria PH Koster[1], MD, PhD; Ageeth N Rosman[1*], PhD; Regine PM Steegers-Theunissen[1,2*], MD, PhD

[1]Erasmus MC, Obstetrics and Gynecology, Rotterdam, Netherlands

[2]Erasmus MC, Pediatrics, Division of Neonatology, Rotterdam, Netherlands

[*]these authors contributed equally

Corresponding Author:
Regine PM Steegers-Theunissen, MD, PhD
Erasmus MC
Obstetrics and Gynecology
Dr Molewaterplein 50
Rotterdam, PO Box 2040
Netherlands
Phone: 31 107038254
Fax: 31 107036815
Email: r.steegers@erasmusmc.nl

Abstract

Background: The importance of the preconception period and preconception care (PCC) are broadly acknowledged and the potential benefits regarding health promotion have been studied extensively. PCC provides the opportunity to identify, prevent, and treat modifiable and nonmodifiable risk factors to optimize the health of couples trying to become pregnant. The prevalence of modifiable and nonmodifiable risk factors in these couples is high, but the uptake of PCC remains low.

Objective: The aim of this study is to identify the preferences and experiences of women and men (patients) trying to become pregnant and of health care providers and other involved professionals regarding mobile health (mHealth), in particular the coaching platform Smarter Pregnancy, and its potential role in PCC.

Methods: Patients who participated in the Smarter Pregnancy randomized controlled trial (RCT) and health care providers and professionals also involved in PCC were invited to participate in a qualitative study. The barriers, benefits, and opportunities of big data collection by mHealth were discussed in focus group sessions, prompted with statements regarding PCC.

Results: We composed five focus groups, consisting of 27 patients in total (23 women and 4 men), who participated in the RCT, and nine health care providers and other professionals. Of the patients, 67% (18/27) were familiar with the concept of PCC, but only 15% (4/27) received any form of PCC. A majority of 56% (combined percentages of statements 1 [n=18], 2 [n=11], and 3 [n=16]) of the patients believed in the benefit of receiving PCC, and all agreed that men should be involved in PCC as well. Patients did not have a problem using anonymized data obtained from mHealth tools for scientific purposes. Patients and health care providers and other professionals both acknowledged the lack of awareness regarding the importance of PCC and stated that mHealth provides several opportunities to support clinical PCC.

Conclusions: Our findings substantiate previous studies addressing the low uptake of PCC due to unawareness or lack of perception of its relevance by couples who are trying to become pregnant. The positive judgment and experiences with mHealth, in particular Smarter Pregnancy, will stimulate future research and further development of effective and cost-effective personalized mHealth apps for patients, health care providers, and other professionals as an add-on to clinical PCC.

(JMIR Mhealth Uhealth 2017;5(8):e123) doi:10.2196/mhealth.7834

KEYWORDS

Focus group; Lifestyle; mHealth; Nutrition; Personalized medicine; Preconception care

XSL•FO
RenderX

Introduction

Since the recommendation of preconceptional folic acid supplement use for the prevention of neural tube defects in the early 1990s, the importance of the preconceptional period in the physiology and pathophysiology of pregnancy outcome and preconception care (PCC) is broadly acknowledged. The potential benefits of health promotion and interventions during this period of at least 14 weeks before conception has been extensively studied [1-3]. PCC can be used to identify, prevent, and treat modifiable and nonmodifiable risk factors and it optimizes the health of couples trying to conceive and, ultimately, the pregnancy outcome [4]. In the Netherlands, PCC is only delivered to a select group of women, mainly those who have a fertility problem or a high risk for adverse pregnancy outcome due to a known genetic or medical condition or a previous poor pregnancy outcome. However, at their own request, couples can receive PCC from a health care professional, but so far only a very small proportion of the general population takes advantage of this. The low uptake of PCC, combined with the high prevalence of unhealthy nutrition and lifestyle behaviors, illustrates the lack of awareness regarding the importance of PCC in couples who are trying to conceive [5-7].

Currently, rapid developments in the field of telemedicine by means of electronic health (eHealth) and mobile health (mHealth) are opening doors to new opportunities to empower patients and health care providers and professionals and to fill the gaps in patient care [8,9]. In 2010, more than 200 million health-related online apps were downloaded, suggesting that mHealth indeed has the potential to reach, inform, and educate a large population [10]. Inherent to such mHealth apps, programs, or services, an enormous amount of data, referred to as "big data," can be obtained and stored by integrating data of online questionnaires, biofeedback, and diagnostic and monitoring tools. Consequently, big data can be used to study specific populations of interest and is therefore considered to be of great medical and scientific importance in the future [11]. Because nearly all women and men of reproductive age have Internet access and/or own a mobile phone, we believe that mHealth can play a role in providing information that can induce awareness and eventually support the implementation of PCC. Although there are many pregnancy-related mHealth solutions, mHealth solutions focusing on PCC are scarce [12]. Therefore, we consider this study regarding "Smarter Pregnancy" as a pioneer in the field of PCC using mHealth.

The aim of this qualitative study was to explore the preferences and experiences of women and men regarding mHealth, including big data, and its potential role in PCC. Moreover, we discussed these preferences and experiences with health care providers and other involved professionals in the field of PCC.

Methods

Participants and Recruitment

All participants (hereafter referred to as "patients" to improve readability) of the Smarter Pregnancy randomized controlled trial (RCT) (ie, fertile and subfertile couples trying to conceive) who completed the first six months of the program or resigned prematurely were invited to participate. The details of the Smarter Pregnancy mHealth platform and the RCT design have previously been published [13,14]. In short, during the Smarter Pregnancy RCT, the intervention group received individual coaching consisting of a baseline screening and a follow-up screening at 6, 12, 18, and 24 weeks regarding nutrition and lifestyle behavior. Coaching also included a maximum of three interventions per week, which consisted of short message service (SMS) text messages and email messages containing tips, recommendations, vouchers, seasonal recipes, and additional questions addressing gender, behavior, first day of last menstrual period, pregnancy status, body mass index, and adequacy of the diet. The control group did not receive the weekly personal coaching after the baseline screening and only received a minimum of feedback on the screening questionnaires at baseline and at 12 and 24 weeks.

For this qualitative study, all potential participants received an email that included an invitation to participate in a focus group session. In this email, we stated that we were interested in their feedback on our mHealth coaching platform and their views on the general concept of mHealth and big data by means of a semistructured interview, prompted with statements about PCC (Table 1).

Data Collection Procedure

To compose homogeneous focus groups and consequently lower the barrier to participate and increase the response rate, we chose to stratify the groups according to gender, known fertility status, and RCT study group (ie, intervention or control group). We aimed to recruit 6 to 10 patients per group. One week before the meeting, patients received a list of the statements that were going to be discussed during the focus group. At the start of a focus group, patients were asked to fill out a questionnaire regarding their personal information, medical information, and experiences and knowledge on PCC in general.

Table 1. Statements used during the focus groups with patients.

Statement	Topic
1. I consider the background information and coaching received by the mHealth program Smarter Pregnancy as useful.	Preconception care
2. Personal coaching by email and text messages is a valuable additive.	mHealth
3. Smarter Pregnancy has a pleasant way of communicating.	mHealth
4. Mobile health is a right method to give preconception care.	mHealth
5. Data obtained from Smarter Pregnancy can be (anonymously) used for other (non)commercial purposes.	Big data

XSL•FO
RenderX

Every focus group meeting took place at the Erasmus MC, University Medical Centre, Rotterdam (the Netherlands), and was preceded by an individual introduction of each patient and a short presentation by a researcher (MRvD) to repeat the aim of the meeting and to ensure confidentiality. During the 2 to 2.5 hour focus group session, a professional moderator (ANR) guided the discussion. The involved researcher (MRvD) took minutes and ensured optimal audio recording.

Health Care Providers and Professionals

After the focus group sessions with the patients, we also invited health care providers and professionals involved in the fields of reproductive medicine, obstetrics or PCC, policy makers, and representatives of a health care insurance company. Because all focus group sessions with the patients had already been processed and analyzed, health care providers and other professionals were not only asked to discuss their own views regarding PCC, mHealth, and big data, but also to reflect on the patients' input on these topics.

Theoretical Framework and Data Analysis

This study is based on a framework described by Fleuren et al [15], which identifies four main stages in innovation processes: dissemination, adoption, implementation, and continuation. These processes can be considered as potential failure points in which the transition from one stage to another is determined by both positive and negative factors (determinants). The framework considers characteristics of the organization, the innovation itself, the end user, and the sociopolitical environment. By using statements prompted during the focus groups, determinants regarding patients' preferences and experiences were derived. The same was done within the focus group of the professionals; however, specific information from the patient's focus groups was added and discussed.

All recorded audio was transcribed verbatim, using the minutes as guidance. To perform thematic analysis, a set of preliminary codes was developed from the notes and transcripts and discussed between the researchers involved. Subsequently, the codes were structured to the concepts of determinants as previously described. Two researchers (MRvD and MPHK) coded one transcript independently and then compared the coding to reach consensus. Thereafter, the remaining scripts were coded by MRvD. All coding took place using NVivo version 10 (QSR International, Cambridge, MA, USA).

Ethical Considerations

All data were anonymously processed. This qualitative study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving patients were approved by the Medical Ethical and Institutional Review Board of the Erasmus MC, University Medical Centre, Rotterdam, in the Netherlands. Informed consent was obtained from all participants to use anonymized data for analysis.

Results

Study Population

A total of 509 patients received an invitation, of which 23 women and 4 men accepted the invitation and were able to participate in four focus groups. Patients who had an indication to receive fertility treatment by means of an in vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI) were labeled as the IVF-ICSI population, whereas patients who did not receive this treatment were labeled as the general population. Groups were composed as follows:

1. Women, general population (n=5);
2. Women, IVF-ICSI population, intervention group (n=9);
3. Women, IVF-ICSI population, control group (n=9); and
4. Men, IVF-ICSI population (n=4).

Overall, baseline characteristics of these women and men, such as age, ethnicity, educational level, and lifestyle were comparable between patients of the four focus group sessions (Table 2).

The focus group session with health care providers and professionals consisted of nine attendants (ie, a gynecologist, a midwife, a general practitioner, a fertility doctor, a preventive health care center physician, a dietician, a medical advisor from a health insurance company, a representative of the municipality of Rotterdam, and a representative of the Dutch association of parent and patient organizations).

Preconception Care: Beliefs and Perception

A summary of the patients' answers on the additional questionnaire at the start of the focus group session is shown in Table 3, illustrating their perceptions and beliefs regarding PCC. Despite the observation that only 67% (18/27) of patients were familiar with the current concept of PCC (ie, a consultation with a health care professional) and only 15% (4/27) received any form of PCC (Table 2), a majority of 56% (combined percentages of statements 1 [n=18], 2 [n=11], and 3 [n=16] in Table 3) indicated the benefits of receiving PCC and adopting a healthy lifestyle when trying to conceive. Whether they believe that if they become pregnant, their child benefits from received PCC remains questionable because only 32% (combined percentages of statements 4 [n=7] and 5 [n=10] of Table 3) agreed with this statement.

XSL•FO
RenderX

Table 2. Baseline characteristics of all patients, based on the additional questionnaire (N=27).

Baseline characteristics	General population (n=5)	IVF-intervention group (n=9)	IVF-control group (n=9)	Men (n=4)
Focus group, n	1	2	3	4
Age (years), mean (SD)	33.0 (5.1)	33.7 (5.1)	35.2 (4.3)	43.3 (17.5)
Marital status, n (%)				
Single	—	—	1 (11)	—
Married or living together, without children	1 (20)	4 (44)	3 (33)	4 (100)
Married or living together, with children	4 (80)	5 (56)	5 (56)	—
Ethnicity, n (%)				
Dutch or Western	5 (100)	8 (89)	9 (100)	4 (100)
Non-Dutch, non-Western	—	1 (11)	—	—
Education, n (%)				
None	—	—	—	—
Low	—	—	1 (11)	—
Middle	1 (20)	3 (33)	1 (11)	—
High	4 (80)	6 (67)	7 (78)	4 (100)
Smoking (yes), n (%)	—	—	1 (11)	1 (25)
Alcohol consumption (yes), n (%)	1 (20)	1 (11)	4 (44)	4 (100)
Drug use, n (%)	—	—	—	—
Vitamin use, n (%)	1 (20)	7 (78)	7 (78)	1 (25)
Medication use, n (%)	1 (20)	—	1 (11)	1 (25)
Comorbidity, n (%)	3 (60)	2 (22)	3 (33)	2 (50)
Mode of conception, n (%)				
Spontaneous	3 (60)	1 (11)	1 (11)	—
Hormonal treatment	—	—	—	—
IVF or ICSI	—	5 (56)	4 (44)	2 (50)
Nulliparous	1 (20)	6 (67)	5 (56)	4 (100)
Familiar with preconception care, n (%)	2 (40)	8 (89)	4 (44)	4 (100)
Received preconception care, n (%)	2 (40)	—	2 (22)	—

XSL•FO
RenderX

Table 3. Patients perceptions and beliefs regarding PCC, prior to the focus group (N=27).

Patients perceptions and beliefs regarding PCC	Focus group, n				Overall, %
	1	2	3	4	
Preconception care: beliefs and perception					
1. PCC will make me adopt a healthy lifestyle.					
Strongly disagree	0	0	0	0	0
Disagree	0	0	2	1	11
Neither agree nor disagree	1	1	3	1	22
Agree	4	8	4	2	67
Strongly agree	0	0	0	0	0
2. Through PCC, I know whether it's wise to become pregnant.					
Strongly disagree	0	0	0	0	0
Disagree	0	2	2	1	19
Neither agree nor disagree	1	4	4	2	41
Agree	4	3	3	0	37
Strongly agree	0	0	0	1	4
3. Through PCC, I'm better prepared to become pregnant.					
Strongly disagree	0	0	0	0	0
Disagree	0	0	1	1	7
Neither agree nor disagree	1	4	3	1	33
Agree	4	5	4	2	56
Strongly agree	0	0	1	0	4
4. PCC reduces the risk of complications during pregnancy or labor.					
Strongly disagree	0	0	0	0	0
Disagree	0	1	0	0	4
Neither agree nor disagree	3	6	5	4	67
Agree	2	2	3	0	26
Strongly agree	0	0	0	0	0
5. PCC makes my baby more healthy.					
Strongly disagree	0	0	0	0	0
Disagree	0	1	2	0	11
Neither agree nor disagree	3	6	2	3	52
Agree	2	2	4	1	33
Strongly agree	0	0	1	0	4
Preconception care: logistics					
6. PCC should be obligated					
Yes	1	1	1	1	15
No	4	8	8	3	85
7. PCC should be given to:					
Women only	0	0	0	0	0
Men only	0	0	0	0	0
Women and men	5	9	9	4	100
8. When should PCC be reimbursed by an insurance company					
Only if a woman has a high-risk (medical) condition	0	5	3	2	37

XSL•FO
RenderX

Patients perceptions and beliefs regarding PCC	Focus group, n				Overall, %
	1	2	3	4	
Always	5	4	6	2	63
9. For whom should PCC be reimbursed					
Women only	2	7	4	2	56
Couples only	0	1	1	1	11
No opinion	3	1	4	1	33
Preconception care: conditions and content					
10. Would you prefer anonymous PCC over personal					
Yes	0	0	1	0	4
No	5	9	8	4	96
11. PCC should consist of one consultation					
Yes	3	2	1	1	26
No	2	7	7	3	74
12. PCC by mobile health can be useful					
Yes	4	7	9	4	93
No	1	1	0	0	7
13. PCC can be used unconditionally regarding treatment					
Yes	4	5	8	3	74
No	1	4	1	1	26

Preconception Care: Logistics

All patients acknowledged that men should be involved in PCC. On the contrary, more than half (15/27, 56%) stated that only the costs of PCC received by women should be reimbursed by the insurance company. Despite the agreement on the importance of PCC, 85% (23/27) stated that it should not be mandatory for couples trying to conceive.

Preconception Care: Conditions and Content

Most patients (26/27) would not prefer anonymous PCC. Despite previous findings showing a majority stating PCC should not be obligatory, 74% (20/27) stated that PCC should be mandatory as a part of fertility treatment (Table 3).

mHealth

In general, patients feel comfortable using mobile apps. They believe that using mobile devices in health care is a good development and a modern approach to provide patients with information and background. Most male patients acknowledged that mobile health can be used to substitute for certain parts of regular consultations, especially during fertility treatment, but women emphasized the importance of face-to-face contact and nonverbal communication and stated that mHealth should only be used as an additive to routine clinical care:

> *It is not necessary to have a face-to-face consultation, if I need to discuss something, I'll find my way to contact a health care professional.* [Man, group 4]

> *If they ask me over the phone through an app, how am I doing, I'll just say "I'm doing good," but if they ask me during a consultation, they can see me and notice I'm not doing okay.* [Woman, group 1]

Awareness

The most frequently discussed topic during most focus group sessions was "awareness." Some female patients specifically mentioned the visual feedback, as provided by the Smarter Pregnancy platform, as a trigger and motivator to improve behavior. Knowing they would perform better on the next monitoring questionnaire gave them perceived control, but a high frequency of coaching and incorporated positive feedback is needed to secure adherence to the program and to truly improve awareness:

> *It makes you more aware of what you're eating, so when you're tired you won't eat an unhealthy snack because you want to reach the best score on the questionnaire.* [Woman, group 2]

> *If you truly want to change someone's behavior, one reminder per week is nice, but not enough to provide sufficient information.* [Woman, group 3]

> *Besides all the coaching on what to improve, it's also nice to hear you're doing a good job.* [Woman, group 2]

Education

Patients agreed that background information on the importance of healthy nutrition and lifestyle as a part of PCC improves awareness, but only if they believe it is trustworthy and preferably evidence-based. The Smarter Pregnancy coaching platform is supported by multiple health care organizations of

XSL•FO
RenderX

which the logos are displayed at the website. This was highly appreciated by the patients:

Using received information as an online reference book, which was always accessible on demand, was suggested to be of great value. Also adding visual content by means of images and videos was considered valuable.

Personalized Mobile Health

In addition to awareness, participants agreed on the fact that mHealth needs to be highly personalized to be really effective. Impersonal messages or general messages were considered not effective or even countereffective. Men and women both suggested that the psychological aspect of trying to conceive should be integrated in mHealth as well as the functionality of asking online questions to a health care professional:

> *I guess it would really work if patients can use an accessible "chat function" or "email service" to ask questions to a health care professional.* [Woman, group 3]

Big Data

Scientific and Commercial Use

All patients were asked their opinion on data obtained through a mHealth platform being used either anonymously or nonanonymously, for scientific purposes. Patients were very willing to support prepregnancy- and pregnancy-related research in general in this way, provided it was anonymous, because they considered it to be helpful for other patients and future parents. Some patients approved of nonanonymous scientific use of their data. Also, "medical-related" companies and organizations, such as hospitals and pharmaceutical companies, were considered to be relevant purposes. On the contrary, most patients did not allow usage of nonanonymous or anonymous data for commercial purposes. There was a general perception that companies or organizations, and health care insurance companies in particular, should not benefit from this data, although one participant mentioned this could be an opportunity to develop profitable PCC:

> *I am willing to help science, but not willing to help a company sell more of its products.* [Woman, group 1]
>
> *I really value that I decide what to share with the outside world, and with whom.* [Woman, group 1]

Safety and Monitoring

In general, patients were not worried about data leakage due to limited safety of storage by mHealth devices. It was believed that every medical institute itself, or together with governmental support, could guarantee data safety and monitoring.

Health Care Providers and Professionals

Preconception Care

Participating health care providers and professionals also agreed on the general lack of awareness for PCC. To create awareness, the importance of evidence-based information and education was emphasized. For example, consistent online and offline information can educate patients and health care providers and professionals and consequently increase the intrinsic motivation to change certain unhealthy behavior that are not often addressed in health care and PCC. It was suggested that awareness in general can be increased by offering PCC through employers or, even better, through secondary schools integrated within biology or sex education lessons. The health care providers and professionals recognize the fact that patients are familiar with the broad and inconsistent spectrum of online information regarding PCC and notice that especially higher-educated patients use the Internet to obtain information regarding fertility and pregnancy, whereas more lower-educated patients with limited health literacy and the highest health risks prefer to visit a professional first:

> *It feels like selling something to someone who doesn't want to buy it. You are trying to convince them of something they don't believe it's important.* [Gynecologist]
>
> *With the existence of online communities, patients are "educated" by peers instead of professionals. That is their preconception care.* [Fertility doctor]

mHealth

Health care providers and professionals were familiar with mHealth, especially apps to monitor menstrual cycles, fetal development during pregnancy, and for online questionnaires to identify risk factors for certain conditions (eg, depression and anxiety). In addition to monitoring, they were concerned whether mHealth can reach and educate those who need it the most, for example due to a language barrier. Therefore, it is suggested that developing apps in multiple languages will overcome this. If so, it is believed that mHealth can be used to substitute for certain elements of routine consultations (eg, nutrition and lifestyle recommendations). Replacing consultations by alternative techniques, such as video chat, is believed to be an upcoming development, but is currently not appreciated due to the lack of technical support and security.

All health care providers and professionals were very enthusiastic about the concept of using an online questionnaire, such as the one incorporated in the Smarter Pregnancy platform, including a link between the given answers and the patient's medical record. The prospect of having all the results before a face-to-face PCC consultation was considered very useful and timesaving. Moreover, providing questionnaires to patients was in itself already thought to increase awareness:

> *Monitoring over time is very useful, because knowing whether a patient is improving gives the opportunity to give them a compliment, which can be very helpful.* [Gynecologist]

Big Data

The health care providers and professionals unanimously agreed that big data can be of great medical and scientific importance. By obtaining detailed information on target groups and populations, interventions can be designed and clinical care can be tailored at specific behaviors, needs, or risk factors of specific patient groups. Although the health care providers and professionals were aware of the perception of patients toward

XSL•FO
RenderX

the use of big data, they believed that commercial use could also be beneficial in creating large-scale awareness.

Discussion

Principal Findings

This qualitative study addressed the preferences and experiences of patients and health care providers and other professionals regarding PCC in general and mHealth in particular. Based on the four focus group sessions with patients the following can be summarized:

Patients are familiar with PCC in general and confirm that there is a lack of awareness regarding the importance.

Patients believe that mHealth can play a role in PCC, especially regarding awareness and providing evidence-based information, but mainly as an additive to standard care with face-to-face contact with a health care professional.

Patients also believe that mHealth should be personalized, customized, and tailored to their needs, risks, and behaviors to reach its full potential and become truly effective.

Patients approve that data obtained from mHealth, referred to as big data, can be used anonymously for scientific purposes.

The health care providers and other professionals agree on the potential role of mHealth in PCC, especially as an effective tool to inform and educate couples to improve awareness of the importance of PCC care in general. They are optimistic about the concept of mHealth integrated into the patients' medical records, but emphasize that the current situation is not suitable for this innovation due to the lack of technical support.

Comparison With Literature

Our findings correspond with existing literature, in which low uptake of PCC due to unawareness or a lack of perception of relevancy by couples trying to conceive have been described [16-18]. Concerning mHealth and its role in PCC, previous studies have suggested that tailored interventions may improve the uptake of PCC, especially when added to standard care [19,20]. Currently, the development and uptake of commercial and non-evidence-based apps continues, whereas there is an ongoing debate on the efficacy of mHealth in general, because the scientific merit is questionable due to the absence of robust evidence [13,20-23]. Therefore, many studies are conducted to provide scientific evidence on the effectiveness of mHealth interventions in general [14,24-28]. To our knowledge, the perception of patients regarding the use of big data for scientific purposes has not been described before.

Regarding the preferences and experiences of patients using mHealth interventions in general, our findings are in line with previous studies. The personalized character and credibility of mHealth interventions have recently been described to be important to enhance adherence to therapy and nutrition and lifestyle recommendations [29-31].

Strengths and Limitations

Patient involvement during the designing phase of an intervention is essential, followed by end user participation and evaluation of an intervention to further improve customization [32,33]. Consequently, the main strength of this study is the involvement of several end users of our mHealth platform (ie, participants of the Smarter Pregnancy RCT), including the participation of men. Also, we included multiple health care providers and other professionals, representing various organizations and professions in the field of PCC, which is an important strength. These professionals were able to state their own opinion, substantiated by the policy of the organization or profession they represent. Due to the careful stratification and composition of the focus groups, we created a safe environment for the patients in which the structured discussion took place. Furthermore, with the presence of a professional moderator, we were able to give all participants the opportunity to express and interactively discuss their opinions, experiences, and feelings equally and without any consequences.

The most important limitation to address is the low acceptance rate resulting in a very small number of patients in total and per focus group, although this can be considered as confirmation of the main underlying problem: the lack of knowledge and awareness regarding PCC. This, together with the involvement of end users, may also introduce selection bias; the patients involved in this study are generally highly educated and probably more engaged with the topic because they already participated in a previous study regarding mHealth and PCC. Although this bias is hard to overcome when conducting qualitative studies in general, and especially in this field of research with a population of interest that is very hard to reach, it needs to be addressed because it could influence the external validity of the results.

Conclusions

Overall, we conclude that patients and health care providers and professionals believe that mHealth has several unique opportunities for PCC. Our findings imply that future research should focus on the development of mHealth apps as an add-on to standard care, preferably integrated or connected to patients' medical records, allowing health care providers and other professionals to become involved and support their patients. The first step to increase awareness would be to provide evidence-based information, followed by providing apps or programs containing this information, but also tailored to individual conditions. Therefore, patient involvement and end user participation will be indispensable in designing effective interventions.

Acknowledgments

We would like to acknowledge Bart Looman, moderator and project manager at Pharos, a center of expertise regarding health inequities, for moderating the focus group with the health care providers and other professionals. This research was funded by

XSL•FO
RenderX

the Department of Obstetrics and Gynecology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands and through grants from Erasmus MC Mrace program “Health Care Efficiency Research” and ZonMW Health Care Efficiency Research.

Authors' Contributions

MRvD invited all participants; acquired, analyzed, and interpreted data; and drafted the manuscript. MPHK contributed to data analysis, inference, and critically reviewed all versions of the manuscript. ANR moderated the focus groups, contributed to the inference, and critically reviewed all versions of the manuscript. RPMST initiated and developed the “Smarter Pregnancy” program, critically reviewed the manuscript, and was responsible for all aspects of this study.

Conflicts of Interest

Since 2016, RPMST has been CEO of eHealth Care Solutions and CSO of Slimmere Zorg BV. All other authors declare no conflicts of interest.

References

1. Johnson K, Posner SF, Biermann J, Cordero JF, Atrash HK, Parker CS, CDC/ATSDR Preconception Care Work Group, Select Panel on Preconception Care. Recommendations to improve preconception health and health care--United States. A report of the CDC/ATSDR Preconception Care Work Group and the Select Panel on Preconception Care. MMWR Recomm Rep 2006 Apr 21;55(RR-6):1-23 [FREE Full text] [Medline: 16617292]
2. American College of ObstetriciansGynecologists. ACOG Committee Opinion number 313, September 2005. The importance of preconception care in the continuum of women's health care. Obstet Gynecol 2005 Sep;106(3):665-666. [Medline: 16135611]
3. Locksmith GJ, Duff P. Preventing neural tube defects: the importance of periconceptional folic acid supplements. Obstet Gynecol 1998 Jun;91(6):1027-1034. [Medline: 9611019]
4. Temel S, van Voorst SF, de Jong-Potjer LC, Waelput AJ, Cornel MC, Rombout de Weerd S, et al. The Dutch national summit on preconception care: a summary of definitions, evidence and recommendations. J Community Genet 2015;6(1):107-115. [doi: 10.1007/s12687-014-0204-2]
5. Inskip H, Crozier SR, Godfrey KM, Borland SE, Cooper C, Robinson SM, Southampton Women's Survey Study Group. Women's compliance with nutrition and lifestyle recommendations before pregnancy: general population cohort study. BMJ 2009 Feb 12;338:b481 [FREE Full text] [Medline: 19213768]
6. Hosli EJ, Elsinga J, Buitendijk SE, Assendelft WJ, van der Pal-de Bruin KM. Women's motives for not participating in preconception counseling: qualitative study. Community Genet 2008;11(3):166-170. [doi: 10.1159/000113879] [Medline: 18376113]
7. Gormack AA, Peek JC, Derraik JG, Gluckman PD, Young NL, Cutfield WS. Many women undergoing fertility treatment make poor lifestyle choices that may affect treatment outcome. Hum Reprod 2015 Jul;30(7):1617-1624. [doi: 10.1093/humrep/dev094] [Medline: 25924654]
8. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
9. 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. [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
10. Silva B, 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 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
11. de la Torre Diez I, Cosgaya HM, Garcia-Zapirain B, López-Coronado M. Big data in health: a literature review from the year 2005. J Med Syst 2016 Sep;40(9):209. [doi: 10.1007/s10916-016-0565-7] [Medline: 27520614]
12. Gardiner P, Hempstead MB, Ring L, Bickmore T, Yinusa-Nyahkoon L, Tran H, et al. Reaching women through health information technology: the Gabby preconception care system. Am J Health Promot 2013;27(3 Suppl):eS11-eS20 [FREE Full text] [doi: 10.4278/ajhp.1200113-QUAN-18] [Medline: 23286652]
13. Van Dijk MR, Huijgen NA, Willemsen SP, Laven JS, Steegers EA, Steegers-Theunissen RP. Impact of an mHealth platform for pregnancy on nutrition and lifestyle of the reproductive population: a survey. JMIR Mhealth Uhealth 2016 May 27;4(2):e53 [FREE Full text] [doi: 10.2196/mhealth.5197] [Medline: 27234926]
14. van Dijk MR, Oostingh EC, Koster MP, Willemsen SP, Laven JS, Steegers-Theunissen RP. The use of the mHealth program Smarter Pregnancy in preconception care: rationale, study design and data collection of a randomized controlled trial. BMC Pregnancy Childbirth 2017 Jan 26;17(1):46 [FREE Full text] [doi: 10.1186/s12884-017-1228-5] [Medline: 28125970]
15. Fleuren M, Wiefferink K, Paulussen T. Determinants of innovation within health care organizations: literature review and Delphi study. Int J Qual Health Care 2004 Apr;16(2):107-123. [doi: 10.1093/intqhc/mzh030] [Medline: 15051705]
16. Poels M, Koster MP, Boeije HR, Franx A, van Stel HF. Why do women not use preconception care? A systematic review on barriers and facilitators. Obstet Gynecol Surv 2016 Oct;71(10):603-612. [doi: 10.1097/OGX.0000000000000360] [Medline: 27770130]

XSL•FO
RenderX

17. Bortolus R, Oprandi NC, Rech MF, Marchetto L, Filippini F, Agricola E, et al. Why women do not ask for information on preconception health? A qualitative study. BMC Pregnancy Childbirth 2017 Jan 05;17(1):5 [FREE Full text] [doi: 10.1186/s12884-016-1198-z] [Medline: 28056865]
18. M'hamdi HI, van Voorst SF, Pinxten W, Hilhorst MT, Steegers EA. Barriers in the uptake and delivery of preconception care: exploring the views of care providers. Matern Child Health J 2017 Jan;21(1):21-28 [FREE Full text] [doi: 10.1007/s10995-016-2089-7] [Medline: 27423236]
19. Agricola E, Pandolfi E, Gonfiantini MV, Gesualdo F, Romano M, Carloni E, et al. A cohort study of a tailored web intervention for preconception care. BMC Med Inform Decis Mak 2014 Apr 15;14:33 [FREE Full text] [doi: 10.1186/1472-6947-14-33] [Medline: 24731520]
20. Nwolise C, Carey N, Shawe J. Preconception care education for women with diabetes: a systematic review of conventional and digital health interventions. J Med Internet Res 2016 Nov 08;18(11):e291 [FREE Full text] [doi: 10.2196/jmir.5615] [Medline: 27826131]
21. Noordam AC, Kuepper BM, Stekelenburg J, Milen A. Improvement of maternal health services through the use of mobile phones. Trop Med Int Health 2011 May;16(5):622-626 [FREE Full text] [doi: 10.1111/j.1365-3156.2011.02747.x] [Medline: 21342374]
22. Heminger CL, Schindler-Ruwisch JM, Abroms LC. Smoking cessation support for pregnant women: role of mobile technology. Subst Abuse Rehabil 2016;7:15-26 [FREE Full text] [doi: 10.2147/SAR.S84239] [Medline: 27110146]
23. Smith C, Gold J, Ngo TD, Sumpter C, Free C. Mobile phone-based interventions for improving contraception use. Cochrane Database Syst Rev 2015 Jun 26(6):CD011159. [doi: 10.1002/14651858.CD011159.pub2] [Medline: 26115146]
24. Mummah S, King AC, Gardner CD, Sutton S. Iterative development of Vegethon: a theory-based mobile app intervention to increase vegetable consumption. Int J Behav Nutr Phys Act 2016 Aug 08;13:90 [FREE Full text] [doi: 10.1186/s12966-016-0400-z] [Medline: 27501724]
25. Ebert DD, Heber E, Berking M, Riper H, Cuijpers P, Funk B, et al. Self-guided internet-based and mobile-based stress management for employees: results of a randomised controlled trial. Occup Environ Med 2016 May;73(5):315-323. [doi: 10.1136/oemed-2015-103269] [Medline: 26884049]
26. Partridge SR, McGeechan K, Bauman A, Phongsavan P, Allman-Farinelli M. Improved eating behaviours mediate weight gain prevention of young adults: moderation and mediation results of a randomised controlled trial of TXT2BFiT, mHealth program. Int J Behav Nutr Phys Act 2016 Apr 02;13:44 [FREE Full text] [doi: 10.1186/s12966-016-0368-8] [Medline: 27039178]
27. Dobson R, Whittaker R, Jiang Y, Shepherd M, Maddison R, Carter K, et al. Text message-based diabetes self-management support (SMS4BG): study protocol for a randomised controlled trial. Trials 2016 Apr 02;17:179 [FREE Full text] [doi: 10.1186/s13063-016-1305-5] [Medline: 27039300]
28. Poirier J, Bennett WL, Jerome GJ, Shah NG, Lazo M, Yeh HC, et al. Effectiveness of an activity tracker- and internet-based adaptive walking program for adults: a randomized controlled trial. J Med Internet Res 2016 Feb 09;18(2):e34 [FREE Full text] [doi: 10.2196/jmir.5295] [Medline: 26860434]
29. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. Designing health apps to support dietetic professional practice and their patients: qualitative results from an international survey. JMIR Mhealth Uhealth 2017 Mar 31;5(3):e40 [FREE Full text] [doi: 10.2196/mhealth.6945] [Medline: 28363882]
30. Drewes C, Kirkovits T, Schiltz D, Schinkoethe T, Haidinger R, Goldmann-Posch U, et al. EHealth acceptance and new media preferences for therapy assistance among breast cancer patients. JMIR Cancer 2016 Sep 14;2(2):e13 [FREE Full text] [doi: 10.2196/cancer.5711] [Medline: 28410189]
31. Hartzler A, BlueSpruce J, Catz SL, McClure JB. Prioritizing the mHealth design space: a mixed-methods analysis of smokers' perspectives. JMIR Mhealth Uhealth 2016 Aug 05;4(3):e95 [FREE Full text] [doi: 10.2196/mhealth.5742] [Medline: 27496593]
32. Noergaard B, Sandvei M, Rottmann N, Johannessen H, Wiil U, Schmidt T, et al. Development of a web-based health care intervention for patients with heart disease: lessons learned from a participatory design study. JMIR Res Protoc 2017 May 17;6(5):e75 [FREE Full text] [doi: 10.2196/resprot.7084] [Medline: 28526674]
33. Kushniruk A, Nøhr C. Participatory design, user involvement and health IT evaluation. Stud Health Technol Inform 2016;222:139-151. [Medline: 27198099]

Abbreviations

eHealth: electronic health
ICSI: intracytoplasmic sperm injection
IVF: in vitro fertilization
mHealth: mobile health
PCC: preconception care
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by C Dias; submitted 07.04.17; peer-reviewed by N Macklon, M Ashford; comments to author 16.05.17; revised version received 08.06.17; accepted 21.06.17; published 17.08.17.

Please cite as:
Van Dijk MR, Koster MPH, Rosman AN, Steegers-Theunissen RPM
Opportunities of mHealth in Preconception Care: Preferences and Experiences of Patients and Health Care Providers and Other Involved Professionals
JMIR Mhealth Uhealth 2017;5(8):e123
URL: http://mhealth.jmir.org/2017/8/e123/
doi:10.2196/mhealth.7834
PMID:28818820

©Matthijs R Van Dijk, Maria PH Koster, Ageeth N Rosman, Regine PM Steegers-Theunissen. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 17.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Potential of Mobile Apps for Improving Asthma Self-Management: A Review of Publicly Available and Well-Adopted Asthma Apps

Peter Tinschert[1], MSc; Robert Jakob[2], MSc; Filipe Barata[3], MSc; Jan-Niklas Kramer[1], MSc; Tobias Kowatsch[1], PhD

[1]Center for Digital Health Interventions, Institute of Technology Management (ITEM-HSG), University of St. Gallen, St. Gallen, Switzerland

[2]Operations & Supply Chain Management, School of Management, Technical University of Munich, Munich, Germany

[3]Center for Digital Health Interventions, Department for Management, Technology and Economics, ETH Zurich, Zurich, Switzerland

Corresponding Author:
Tobias Kowatsch, PhD
Center for Digital Health Interventions
Institute of Technology Management (ITEM-HSG)
University of St. Gallen
Central Institute Buidling, 1st Fl.
Dufourstrasse 40a
St. Gallen, 9000
Switzerland
Phone: 41 71 224 ext 7244
Fax: 41 71 224 73 01
Email: tobias.kowatsch@unisg.ch

Abstract

Background: Effective disease self-management lowers asthma's burden of disease for both individual patients and health care systems. In principle, mobile health (mHealth) apps could enable effective asthma self-management interventions that improve a patient's quality of life while simultaneously reducing the overall treatment costs for health care systems. However, prior reviews in this field have found that mHealth apps for asthma lack clinical evaluation and are often not based on medical guidelines. Yet, beyond the missing evidence for clinical efficacy, little is known about the potential apps might have for improving asthma self-management.

Objective: The aim of this study was to assess the potential of publicly available and well-adopted mHealth apps for improving asthma self-management.

Methods: The Apple App store and Google Play store were systematically searched for asthma apps. In total, 523 apps were identified, of which 38 apps matched the selection criteria to be included in the review. Four requirements of app potential were investigated: app functions, potential to change behavior (by means of a behavior change technique taxonomy), potential to promote app use (by means of a gamification components taxonomy), and app quality (by means of the Mobile Application Rating Scale [MARS]).

Results: The most commonly implemented functions in the 38 reviewed asthma apps were tracking (30/38, 79%) and information (26/38, 68%) functions, followed by assessment (20/38, 53%) and notification (18/38, 47%) functions. On average, the reviewed apps applied 7.12 of 26 available behavior change techniques (standard deviation [SD]=4.46) and 4.89 of 31 available gamification components (SD=4.21). Average app quality was acceptable (mean=3.17/5, SD=0.58), whereas subjective app quality lied between poor and acceptable (mean=2.65/5, SD=0.87). Additionally, the sum scores of all review frameworks were significantly correlated (lowest correlation: r_{36}=.33, P=.04 between number of functions and gamification components; highest correlation: r_{36}=.80, P<.001 between number of behavior change techniques and gamification components), which suggests that an app's potential tends to be consistent across review frameworks.

Conclusions: Several apps were identified that performed consistently well across all applied review frameworks, thus indicating the potential mHealth apps offer for improving asthma self-management. However, many apps suffer from low quality. Therefore, app reviews should be considered as a decision support tool before deciding which app to integrate into a patient's asthma self-management. Furthermore, several research-practice gaps were identified that app developers should consider addressing in future asthma apps.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2017;5(8):e113) doi:10.2196/mhealth.7177

KEYWORDS

asthma; self care; disease management; mobile applications; smartphone; mHealth; eHealth; mobile health; behavior and behavior mechanisms; review

Introduction

Asthma, a chronic airway disease characterized by respiratory symptoms, affects an estimated 235 million [1] to 334 million [2] people worldwide. In the United States alone, 24 million people suffer from asthma (ie, 8.6% of children and 7.4% of adults; [3]). Due to its high prevalence and ongoing treatment throughout the lifetime of most patients, asthma costs the US health care system around US 56 billion dollars annually [4].

Health information technology can reduce the burden of chronic diseases such as asthma for patients and health care systems [5,6]. Mobile health (mHealth) apps in particular could enable low-cost and clinically efficacious interventions for asthma [7]. With virtually unlimited scalability and availability, apps could empower patients in their everyday asthma self-management by delivering evidence-based interventions. Research has shown that such interventions (eg, dissemination of educational materials and symptom monitoring tools) improve a patient's quality of life and limit excess health care utilization [8]. Moreover, data obtained through mobile phone sensors and connected medical devices (eg, smart inhalers) can be used to deliver self-management interventions tailored to the specific needs of patients, thus increasing the intervention's efficacy [9,10].

However, whereas first studies have indicated that asthma apps can indeed be effective tools for supporting patients in their self-management [11,12], reviews of asthma apps come to either ambiguous [13] or rather disappointing conclusions regarding the value of apps for improving asthma self-management [14,15]. These reviews have based their conclusions mainly on a lack of methodologically sound studies [13] and the limited application of medical guidelines [14,15].

Although an evidence base is highly desirable for health apps in general, it is unrealistic to expect app developers to conduct clinical studies or to hire clinical experts as consultants. Reasons for this are, among others, the high costs and difficulty of clinical research [16]. Additionally, app stores do not require developers to indicate the evidence base of an app (eg, Apple App store [17]). mHealth apps are also often not comprehensively regulated by public authorities (eg, in the United States, the Food and Drug Administration only regulates a subset of the available apps [18]). Even though patients tend to consider the evidence base as an important selection criterion for mHealth apps [19], they are often not able to make an informed judgment regarding such characteristics of an app [20]. All these factors contribute to the fact that patients keep using publicly available asthma apps regardless of missing validation studies or an app's limited application of medical guidelines.

Therefore, it is necessary for the clinical practice to develop a more comprehensive understanding of the value that publicly available asthma apps might offer for asthmatics. Research pursuing this objective would need to address the limitations of the aforementioned reviews, which have focused exclusively on peer-reviewed papers and not publicly available apps [13] or on qualitative analysis of app content without applying review frameworks with quantifiable results [14,15].

To overcome these shortcomings, the present review applies review frameworks (eg, behavior change technique and gamification taxonomies [21,22]) and thereby, quantifies the characteristics of asthma apps. By making use of such frameworks, this review investigates the potential of well-adopted asthma apps for supporting asthmatics, even if clinical evidence is not available. In the next section, a functional definition of the term "app potential" is provided. The definition consists of specific requirements that an asthma app should fulfill to enable a positive effect on a patient's asthma self-management.

In conclusion, this app review assesses the potential of publicly available and well-adopted asthma apps for supporting patients in their disease management. Furthermore, research-practice gaps will be identified in hopes of stimulating the development of new and more sophisticated asthma apps.

Methods

Overview

In this review, an app's potential to support patients in their disease management is defined as the degree to which an app fulfills the following four requirements: (R1) an app has to contain some kind of "active ingredient," that is, functions supporting patients in their disease management; (R2) similar to traditional asthma self-management programs offered by health care entities [23], an app should be able to change or guide behavior relevant for effective asthma self-management; (R3) an app should motivate the patient to use the app to deliver the active ingredients; and (R4) an app has to be of adequate quality (eg, functionality and aesthetics of an app) to guarantee a sufficient implementation quality of the other postulated requirements.

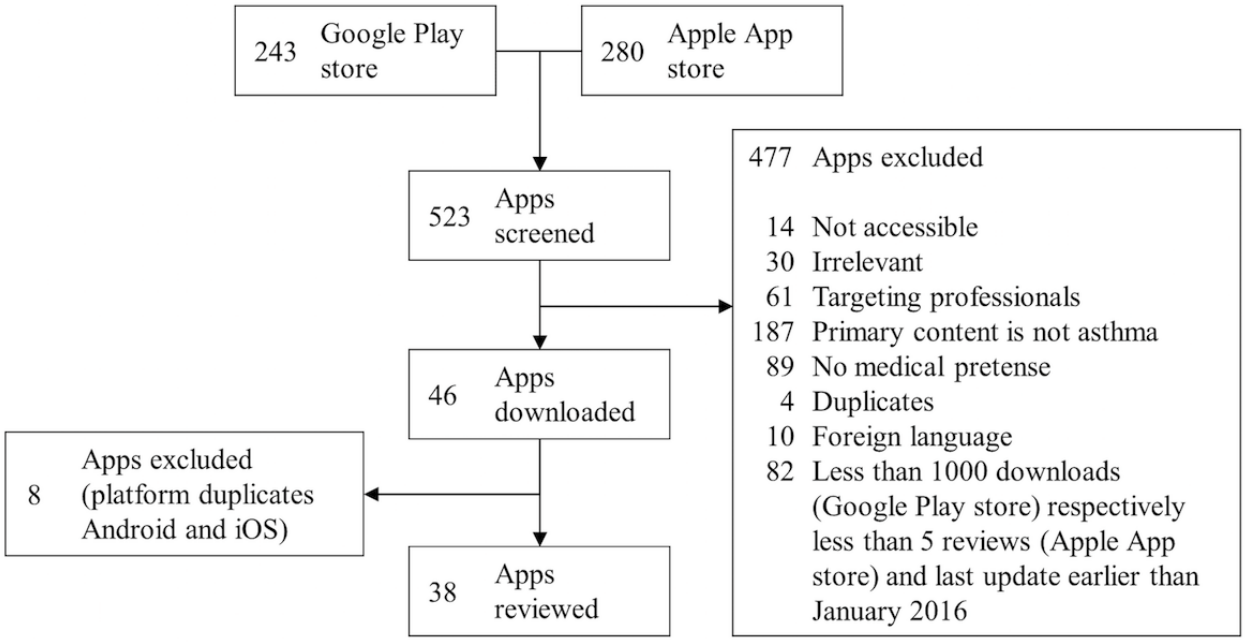
More specifically, this review addresses (R1) by reviewing available functions (similar to [14,15]), (R2) by applying a taxonomy of behavior change techniques [21], (R3) by rating a taxonomy of gamification components [22], and (R4) by using the Mobile Application Rating Scale (MARS; [24]).

Systematic Search

The second author (RJ) searched the term "asthma" in the US versions of the two most popular app stores: Google Play store and Apple App store [25]. The latest app search was conducted on April 10, 2016. In total, 523 apps were identified (Figure 1).

XSL•FO
RenderX

Figure 1. Systematic search and exclusion criteria.



Inclusion and Exclusion Criteria

The strategy for app selection was to include publicly available and well-adopted apps that are designed specifically for asthma. To identify these apps, the following exclusion criteria were applied: no access (eg, only accessible through a special health program), irrelevance (ie, apps completely unrelated to asthma), target group are professionals (eg, apps supporting physicians), primary content is not asthma (eg, general weather or nutrition apps with very limited ancillary functions dedicated to asthmatics), no medical pretense (eg, yoga and acupressure apps), duplicates (ie, different versions of the same app with marginally different functions), and foreign language (ie, apps not in English). To ensure that relatively well-adopted apps are reviewed, apps with less than 1000 downloads (Google Play store) or less than five reviews (Apple App store) were excluded.

Because the number of downloads or reviews of an app depends on its release date, apps with less than 1000 downloads (Google Play store) or five reviews (Apple App store) were not excluded if they were updated in 2016. Otherwise, a selection bias against recently released apps would have occurred. In fact, ten otherwise excluded apps remained in this review due to this criterion (ie, four Android and six iPhone operating system [iOS] apps).

Based on the specified criteria, 477 apps were excluded. The remaining 46 apps were checked for duplicates between the two platforms (ie, Android and iOS). After excluding eight duplicates, the systematic search yielded 38 apps to review.

Review Process

The review process consisted of two main steps. First, basic information about the app was gathered from the corresponding app store (eg, app description, date of last update, developer, number of downloads if available, number of user reviews, and average rating). Only publicly available data, which was directly extracted from the app stores, was used in this process. Additionally, an app's website was consulted if available.

Second, the four requirements for app potential were evaluated. Except for the MARS framework, for which multiple mobile phones and tablets were used (ie, Huawei P9 Lite, HTC M9, LG Leon 4G LTE, iPad Air 2, and iPhone 6), apps were reviewed on a LG Leon 4G LTE (Android) and iPhone 6 (iOS). In the next section, the measures of the four app potential requirements are introduced in more detail.

Measures of App Potential Requirements

R1: App Functions

An exhaustive list of all existing functions was developed through a bottom-up review approach: whenever a function was encountered that was not yet specified in the list, it was expanded accordingly. In this way, a total of 42 functions was identified. Afterwards, the available functions were assessed in all apps based on the complete list.

Additionally, functions were grouped post hoc into four categories: tracking (eg, of peak flow values), information (eg, about asthma symptoms), assessment (eg, indicating asthma control by color-coding peak flow values), and notification functions (eg, medication intake reminder). These function categories are conceptually identical to prior app reviews [14,15], except for one difference: the category of therapeutic tools was replaced with notification functions. Reasons for this are the low prevalence of therapeutic tools in asthma apps [15] and the potential utility of notification functions for asthma self-management (eg, as reminders for physician appointments [26]).

R2: Potential to Change Behavior

Many reviews have applied behavior change technique taxonomies as measures of behavior change efforts in health interventions and apps (eg, [27-29]). However, to the best of

XSL•FO
RenderX

the authors' knowledge, they have not yet been applied to the domain of asthma apps.

In this review, the application of the 26 behavior change techniques proposed by Abraham and Michie [21] was rated for each app. Ratings were dummy coded (1="technique fully applied", 0="technique not applied"). The first and second author (PT and RJ) agreed on explicit rating criteria (see Multimedia Appendix 1 for an exemplary criteria list) and related functions for all behavior change techniques (eg, social network functions enable the application of the behavior change techniques "provide information about other's approval," "provide opportunities for social comparison," and "plan social support or social change"). This approach addressed the issue of mediocre interrater reliability for some of the behavior change techniques (ie, whereas the mean kappa values ranged from .74 to .82 in the original publication, 20 out of 71 kappa scores (28%) were below .70 [21]). By applying explicit and objective rating criteria, the process of rating behavior change techniques was facilitated, and rating ambiguity was reduced. Differences in ratings between the first and second author were resolved in discussions.

In some instances, a behavior change technique was rated as partially applied (.50). This was the case if an app fulfilled the application criteria only partially (eg, nonadjustable task reminder for "time management") or if an agreement between the first and second author could not be reached.

R3: Potential to Promote App Use

Gamification is a promising way to promote the use of information technology (IT) systems such as mobile apps [22,30]. A taxonomy of 31 gamification components [22] was applied to rate an app's potential for motivating app use. Similar to the rating process of behavior change techniques, the first and second author agreed on an explicit list of relevant rating criteria and related functions for gamification components (see Multimedia Appendix 2 for an exemplary criteria list). Again, dummy coding was used for rating gamification components and partially implemented components were considered (1="gamification component fully applied," .50="gamification component partially applied," and 0="gamification component not applied"). Differences in ratings were resolved in discussions between the first and second author and may have resulted in rating a gamification component as "partially applied."

In addition to listing individual gamification components, the applied taxonomy groups individual gamification components into five categories: system design (seven components), challenges (three components), rewards (six components), social influences (eleven components), and user specifics (four components). The original publication provides comprehensive definitions of each category [22].

R4: App Quality

Four of the authors (PT, RJ, JK, and FB) participated in rating app quality by means of the MARS framework [24]. However, the ratings of the first and second author were averaged due to their close cooperation in the rating process, thereby preserving independency of ratings. The MARS framework contains 19 items on a 5-point Likert scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent) to assess app quality. Items are grouped into four subscales: engagement (5 items), functionality (4 items), aesthetics (3 items), and information quality (7 items). The average of the four scales determines the app quality score.

One item from the information quality scale requires a literature search regarding clinical efficacy. PubMed and Google Scholar were searched using the following term: [%app name%] AND ["randomized controlled trial"/"RCT"] OR ["study"]. Literature inclusion criteria were published in a peer-reviewed journal, clearly defined outcome construct that is a valid measure for asthma (eg, peak flow values) and description of the applied study design.

Additionally, MARS contains a subjective quality scale that measures the subjectively perceived app quality. This scale consists of four 5-point Likert scale questions regarding personal recommendation, potential future usage, willingness to pay for an app, and an overall star rating.

MARS is an objective (intraclass correlation coefficient [ICC] between raters=.79) and highly reliable scale (alpha=.90). As recommended by the authors of MARS [24], all raters participated in a workshop regarding the application of the framework. This workshop was based on the MARS training video [31].

Statistical Methods

To assess the status quo of app potential in asthma apps, ratings were analyzed descriptively. ICC between raters were computed as a measure for interrater reliability of the MARS scales. Finally, correlations between the aggregated framework scores were calculated.

Results

General App Characteristics

Out of the 38 reviewed apps, 13 apps were available for both iOS and Android (34%). The remaining apps were exclusively developed for either iOS (16/38, 42%) or Android (9/38, 24%). The vast majority of apps was free of charge (36/38, 95%). In-app purchases were offered in seven apps (18%, 7/38). Additionally, eight apps contained some form of advertising (21%, 8/38).

Most apps were developed in the United States (14/38, 37%), followed by Switzerland (4/38, 11%), Australia, India, Spain, the United Kingdom (each 2/38, 5%), Germany, Portugal, Lithuania, Latvia, and the Czech Republic (each 1/38, 3%). Private companies accounted for the development of half of the apps (n=19), independent developers for nine apps (24%, 9/38), and university hospitals and health-related foundations for four apps each (11%, 4/38). Both nonprofit organizations and governmental nonprofit bodies were responsible for one app each (3%, 1/38). Almost half of the apps claimed that they involved health care entities in the development process (16/38, 42%).

XSL•FO
RenderX

Table 1. Overview general app characteristics.

App characteristics	n (%)	Mean (SD[a])	Median (IQR[b])	Range
Paid apps and price (USD)	2 (5)	3.49 (2.12)	3.49 (1.50)	1.99-4.99
Average user rating (iOS)	15 (39)	3.53 (0.90)	3.50 (1.00)	1.50-5.00
Average user rating (Android)	22 (58)	3.94 (0.51)	3.90 (0.50)	3.00-5.00
Number of ratings (iOS)	29 (76)	22.52 (58.26)	5 (16)	0-304
Number of ratings (Android)	23 (61)	47.35 (87.72)	17 (48.50)	0-419
Number of downloads (Android)[c]	23 (61)	5275.43 (8189.73)	3000 (4725)	30-30,000
Days since last update	38 (100)	498.63 (601.65)	177 (922)	2-2231

[a]SD: standard deviation.

[b]IQR: interquartile range.

[c]The Google Play store reports download numbers in ranges (eg, “10-50”). These values were standardized by calculating the mean of the minimum and maximum (eg, 10-50 downloads equals 30 downloads).

In general, a wide range in the distinct app characteristics became apparent (Table 1). The range was particularly noticeable in the category “days since last update”: one app was updated just two days before the review deadline (“e-symptoms”), whereas “Asthma Journal Free” was last modified over 6 years ago. On average, apps were not updated for approximately more than 1 year and 4 months.

Evaluating App Potential

R1: App Functions

Tracking functions were the most common function category. Out of all 38 apps, 30 apps (79%) offered at least one of the following functions: peak flow tracking (23/38, 61%), medication tracking (22/38, 58%), symptom tracking (19/38, 50%), trigger tracking (18/38, 47%), notes (15/38, 39%), clinical asthma questionnaires (14/38, 37%), sleep tracking (11/38, 29%), health parameter tracking (9/38, 24%), medical appointment tracking (8/38, 21%), location tracking (6/38, 16%), asthma attack tracking (4/38, 11%), accomplishment tracking (3/38, 8%), snapshots (3/38, 8%), score tracking of games and quizzes (2/38, 5%), tracking through Apple or Google Health (2/38, 5%), and medication package size tracking (1/38, 3%). Except for apps using Apple or Google Health for physical activity tracking and location tracking over the global positioning system (GPS), all parameters had to be entered manually. Additionally, 26 apps (68%, 26/38) provided functions to record the tracked values over time through diaries (eg, calendars) and logs (eg, lists and charts).

A total of 26 apps (68%, 26/38) incorporated at least one of the following information functions: therapeutic information or instruction (22/38, 58%), general asthma information (20/38, 53%), medication information or instruction (16/38, 42%), first aid information or instruction (10/38, 26%), inhaler technique guidance (10/38, 26%), app tutorial (9/38, 24%), peak expiratory flow information or instruction (9/38, 24%), news (8/38, 21%), air quality (7/38, 18%), information through doctor dashboard (4/38, 11%), information through games or quizzes (3/38, 8%), and clinic or pharmacy locator (2/38, 5%).

Assessment functions can be considered an extension of tracking functions by providing an interpretation or evaluation of recorded values. Twenty apps (53%, 20/38) offered one or more of these functions: decision support (14/38, 37%), peak flow classification (9/38, 24%), adjustable asthma action plan (5/38, 13%), standardized asthma action plan (4/38, 11%), and symptom forecast (1/38, 3%).

The least common category was notification functions. A total of 18 apps (47%, 18/38) notified the user in at least one of the following contexts or events: medication intake (12/38, 32%), asthma questionnaire reminder (7/38, 18%), peak flow reminder (7/38, 18%), alert notification to health care professionals (7/38, 18%), medical appointment reminder (6/38, 16%), news notification (4/38, 11%), adjustable task reminder (3/38, 8%), weather and pollen notification (1/38, 3%), and medication package depletion (1/38, 3%).

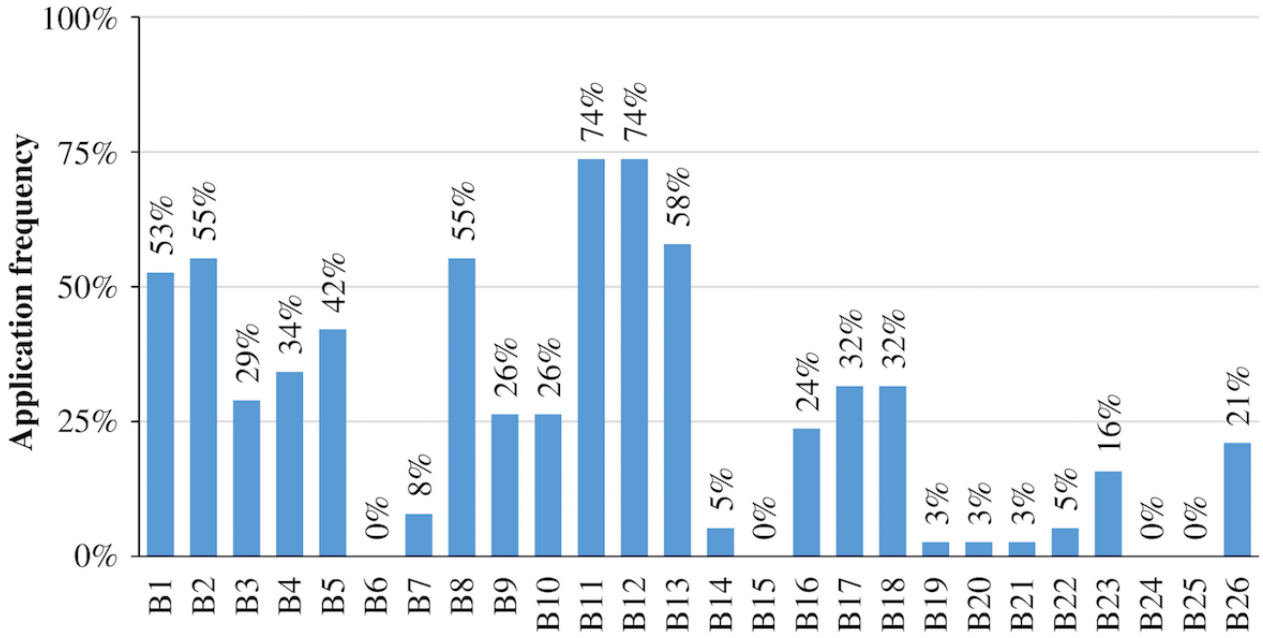
In summary, the most widespread functions in asthma apps were tracking and information functions. More specifically, providing therapeutic information or instruction, general asthma information, as well as tracking of peak flow values, medications and symptoms were implemented by at least half of all apps.

R2: Potential to Change Behavior

On average, apps applied 7.12 (standard deviation [SD]=4.46) of the 26 behavior change techniques (B1-B26). The amount of applied techniques ranged from 1 (“Asthma Treatment”) to 19 (“Wizdy Pets”).

XSL•FO
RenderX

Figure 2. Percentage of asthma apps fully applying the corresponding behavior change technique (B1-B26; N=38).



The fraction of behavior change techniques rated as "partially applied" was negligible (27/988 ratings, 2.7%). For the sake of simplicity and interpretability, only behavior change techniques rated as "fully applied" are reported in this section and in Figure 2. The prevalence of behavior change techniques in asthma apps varied considerably between techniques. To aggregate findings, behavior change techniques were divided into four categories based on their application frequency: no application at all (ie,. technique not applied by any app resulting in an application frequency of 0%), seldom applied (ie, technique applied by one to nine apps resulting in an application frequency between 0-25%), considerable application rates (ie, technique applied by ten to 19 apps resulting in an application frequency between 25-50%), and frequent application (ie, technique applied by 20 to 28 apps resulting in an application frequency between 50-75%).

Four behavior change techniques were not applied at all: provide general encouragement (B6), teach to use prompts or cues (B15), stress management (B24), and motivational interviewing (B25).

The nine seldom applied techniques were set graded tasks (B7; 3/38, 8%), provide contingent rewards (B14; 2/38, 5%), agree on behavioral contract (B16; 9/38, 24%), provide opportunities for social comparison (B19; 1/38, 3%), plan social support or social change (B20; 1/38, 3%), prompt identification as a role model (B21; 1/38, 3%), prompt self-talk (B22; 2/38, 5%), relapse prevention (B23; 6/38, 16%), and time management (B26; 8/38, 21%). Seven techniques were applied with considerable frequency: provide information about other's approval (B3; 11/38, 29%), prompt intention formation (B4; 13/38, 34%), prompt barrier identification (B5; 16/38, 42%), model or demonstrate the behavior (B9; 10/38, 26%), prompt specific goal setting (B10; 10/38, 26%), prompt practice (B17; 12/38, 32%), and use follow-up prompts (B18; 12/38, 32%).

The list of the most commonly applied behavior change techniques in mHealth asthma apps contained six techniques: provide information about behavior-health link (B1; 20/38, 53%), provide information on consequences (B2; 21/38, 55%), provide instructions (B8; 21/38, 55%), prompt review of behavioral goals (B11; 28/38, 74%), prompt self-monitoring of behavior (B12; 28/38, 74%), and provide feedback on performance (B13; 22/38, 58%).

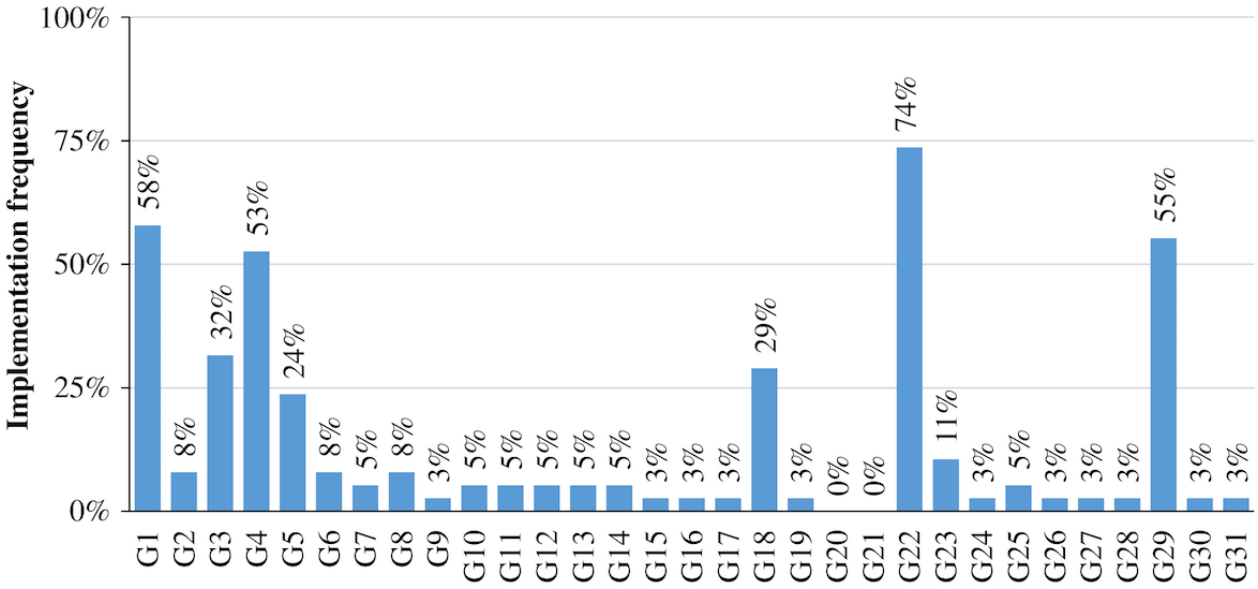
R3: Potential to Promote App Use

Asthma apps contained on average 4.89 gamification components (SD=4.21) of a total of 31 components (G1-G31). The app with the most implemented gamification components was "WizdyPets" (n=24.50), whereas "choO" integrated the fewest components (n=0.50). Similar to the ratings of behavior change techniques, the amount of gamification components rated as "partially applied" can be considered insignificant (31/1178 ratings, 2.6%). Following the reporting logic from the previous section, only gamification components rated as "fully applied" are illustrated in Figure 3 and covered in this section.

Again, gamification components were grouped based on their prevalence into four categories: not implemented at all (ie, component was not implemented by any app resulting in an implementation frequency of 0%), seldom implemented (ie, component was implemented by one to nine apps resulting in an implementation frequency between 0-25%), considerably often implemented (ie, component was implemented by ten to 19 apps resulting in an implementation frequency between 25-50%), and frequently implemented (ie, component was implemented by 20 to 28 apps resulting in an implementation frequency between 50-75%).

XSL•FO
RenderX

Figure 3. Percentage of asthma apps fully applying the corresponding gamification component (G1-G31; N=38).



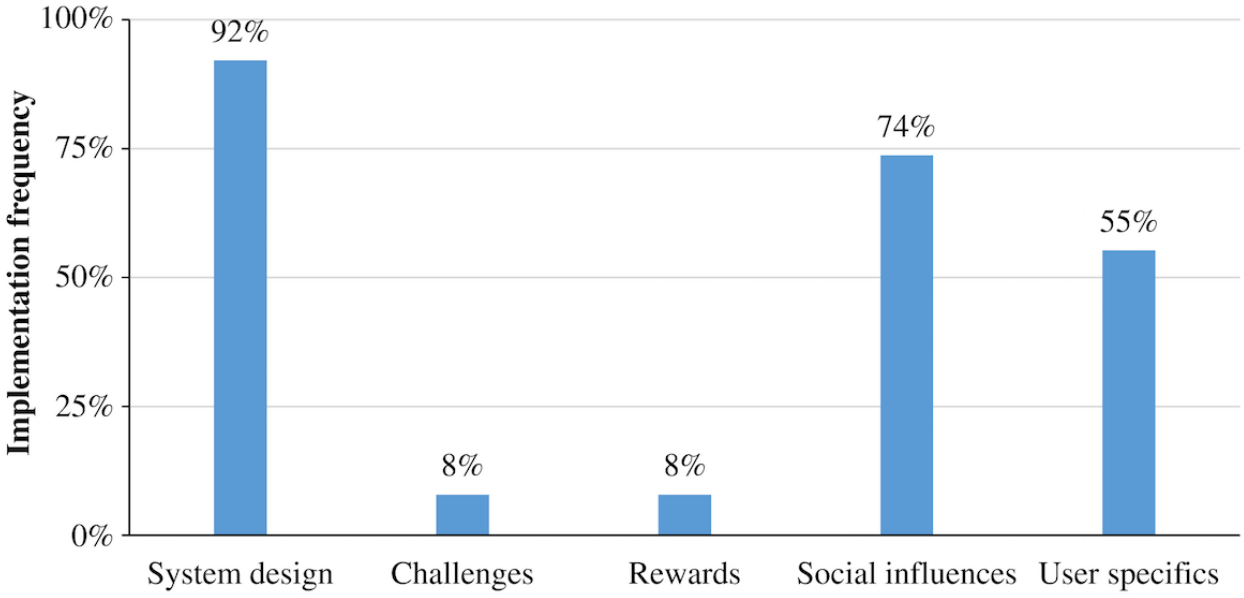
In total, four gamification components were frequently implemented in the reviewed apps: feedback (G1; 22/38, 58%), meaning (G4; 20/38, 53%), shadowing (G22; 28/38, 74%), and ideological incentives (G29; 21/38, 55%).

Two gamification components were considerably often implemented: reminder (G3; 12/38, 32%) and collaboration (G18; 11/38, 29%).

However, the vast majority of gamification components was seldom implemented: audible feedback (G2; 3/38, 8%), interaction concepts (G5; 9/38, 24%), visually resembling existing games (G6; 3/38, 8%), fantasy (G7; 2/38, 5%), goals (G8; 3/38, 8%), time pressure (G9; 1/38, 3%), progressive disclosure (G10; 2/38, 5%), ownership (G11; 2/38, 5%), achievement (G12; 2/38, 5%), point system (G13; 2/38, 5%), badges (G14; 2/38, 5%), bonus (G15; 1/38, 3%), loss aversion (G16; 1/38, 3%), status (G17; 1/38, 3%), reputation (G19; 1/38, 3%), social facilitation (G23; 4/38, 11%), conforming behavior (G24; 1/38, 3%), leaderboards (G25; 2/38, 5%), altruism (G26; 1/38, 3%), virtual goods (G27; 1/38,3%), user levels (G28; 1/38, 3%), virtual characters (G30; 1/38, 3%), and self-expression (G31; 1/38, 3%). Competition (G20) and envy (G21) were not implemented in the reviewed apps.

In addition to listing individual gamification components, the taxonomy groups individual gamification components into five categories: system design, challenges, rewards, social influences, and user specifics [22]. Figure 4 illustrates how many apps implemented at least one component of the corresponding category.

Figure 4. Percentage of apps which applied at least one component fully from the corresponding gamification component category.



XSL•FO
RenderX

Table 2. Descriptive results of Mobile Application Rating Scale (MARS) scores (N=38).

MARS scales	ICC[a,b,c]	Mean (SD[d])	Range
Engagement	.92	2.77 (0.78)	1.00-4.73
Functionality	.72	3.55 (0.57)	1.75-4.42
Aesthetics	.72	3.12 (0.68)	2.00-4.78
Information quality	.88	3.24 (0.66)	1.40-4.33
Subjective quality	.85	2.65 (0.87)	1.00-4.50
App quality	.88	3.17 (0.58)	1.54-4.55

[a]ICC: intraclass correlation coefficient.

[b]Intraclass correlations were calculated as a measure for interrater reliability in a two-way random model evaluating consistency among the three sets of ratings.

[c]Interrater reliability was assessed based on the ratings of 37 apps. One app ("Asthma") was removed from the app store before it was rated by all raters.

[d]SD: standard deviation.

In terms of gamification component category, system design components (eg, feedback, reminder, and meaning) were implemented by almost every app (35/38, 92%). The majority of apps also made use of components from the categories social influences (28/38, 74%; eg, leaderboards, competition, and reputation) and user specifics (21/38, 55%; eg, user levels, virtual character, and self-expression). In contrast, challenges (ie, goals, time pressure, and progressive disclosure) and rewards (eg, achievements, point system, and badges) were applied very infrequently (each 3/38, 8%).

R4: App Quality

The average quality of asthma apps can be considered marginally above acceptable (mean=3.17), whereas average subjective app quality laid between poor and acceptable (mean=2.65). In terms of MARS subscales, asthma apps performed worst in user engagement (mean=2.77) and best in app functionality (mean=3.55). Ratings for aesthetics (mean=3.12) and information quality (mean=3.24) were rather mediocre (Table 2).

In the literature review, no study was identified that evaluated the efficacy of any of the included asthma apps. The only peer-reviewed papers to mention apps were app reviews or clinical communications (eg, [32]). Hence, the information quality mean was calculated from a maximum of 6 items. Additionally, eleven out of the 38 apps (29%) could not be assessed in some of the information quality items due to a lack of corresponding functions (eg, no visualized information was provided that made it impossible to rate the visual information quality).

Considerable quality differences between apps emerged in the analysis. The range of app quality ratings and subjective app quality ratings reached from inadequate or poor (eg, "Asthma" scored a 1.54 and a 1.00 in terms of app quality and subjective app quality) to almost excellent (eg, "Wizdy Pets" scored a 4.55 and 4.50 in terms of app quality and subjective app quality).

Summary of Requirements (R1-R4)

Overall, the results demonstrate vast differences between apps in all of the investigated requirements of app potential (Table 3; see Multimedia Appendix 3 for the complete aggregated ranking table).

To synthesize the findings, associations between the different review frameworks were considered by calculating the corresponding correlation coefficients between the aggregated scores (Table 4). All correlations between review frameworks were significant. This pattern of correlations implies consistency of ratings between different review frameworks. In other words, apps that scored well in one review framework tended to also score well in all the other frameworks.

XSL•FO
RenderX

Table 3. Summary of all rated app potential requirements. Behavior change techniques or gamification components rated as "fully applied" and "partially applied" were included in the analysis.

App potential requirements	Mean (SD[a])	Range	Scale
R1: functions			
Tracking functions	4.21 (3.39)	0-11.00	0-16.00
Information functions	3.21 (2.63)	0-8.00	0-12.00
Assessment functions	0.87 (0.98)	0-3.00	0-5.00
Notification functions	1.26 (1.65)	0-5.00	0-5.00
Functions (overall)	9.55 (5.73)	1.00-24.00	0-42.00
R2: behavior change techniques	7.12 (4.46)	1.00-19.00	0-26.00
R3: gamification components			
System design	2.20 (1.35)	0.50-4.00	0-7.00
Challenges	0.22 (0.60)	0-3.00	0-3.00
Rewards	0.26 (1.07)	0-6.00	0-6.00
Social influences	1.37 (1.24)	0-5.00	0-11.00
User specifics	0.84 (0.74)	0-4.00	0-4.00
Gamification components (overall)	4.89 (4.21)	0.50-24.50	0-31.00
R4: MARS[b] scales			
Engagement	2.77 (0.78)	1.00-4.73	1.00-5.00
Functionality	3.55 (0.57)	1.75-4.42	1.00-5.00
Aesthetics	3.12 (0.68)	2.00-4.78	1.00-5.00
Information quality	3.24 (0.66)	1.40-4.33	1.00-5.00
Subjective quality	2.65 (0.87)	1.00-4.50	1.00-5.00
App quality	3.17 (0.58)	1.54-4.55	1.00-5.00

[a]SD: standard deviation.

[b]MARS: Mobile Application Rating Scale.

Table 4. Pearson correlations between aggregated review frameworks (N=38). All ratings of behavior change techniques and gamification components were included in the analysis (including techniques and components rated as "partially applied or implemented").

Aggregated review frameworks	Number of functions (R1)	Number of behavior change techniques (R2)	Number of gamification components (R3)
Number of behavior change techniques (R2)	*r*=.73 (*P*<.001)		
Number of gamification components (R3)	*r*=.33 (*P*=.04)	*r*=.80 (*P*<.001)	
MARS[a] app quality (R4)	*r*=.38 (*P*=.02)	*r*=.56 (*P*<.001)	*r*=.62 (*P*<.001)

[a]MARS: Mobile Application Rating Scale.

Discussion

Principal Findings

This review analyzed whether asthma apps have the potential to improve a patient's asthma self-management. For this purpose, different requirements of app potential were considered, namely available functions (R1), applied behavior change techniques (R2), implemented gamification components (R3), and general app quality (R4).

In terms of available functions (R1), asthma apps offered functions associated with active ingredients of effective self-management to a considerable extent (ie, asthma education, self-monitoring of symptoms or peak flow values, regular review of treatment, and an action plan [8]). However, with regard to self-monitoring of symptoms and peak flow values, not a single app offered options to track asthma-related parameters through mobile phone sensors. Without exploiting the full potential of today's technology (eg, through automated tracking functions

XSL•FO
RenderX

[33]), therapeutic advantages of mobile phones cannot be fully realized (eg, delivery of just in time adaptive interventions [34]).

The investigated apps applied a number of behavior change techniques (R2). This finding implies that the apps could in principle enable behavior change relevant for asthma self-management. However, techniques related to stress management and motivational interviewing were not implemented at all. This is particularly striking because stress management is recommended by asthma guidelines [23], and asthmatics tend to trivialize their disease symptoms [35], which might result in a lack of long-term motivation for disease management.

In general, asthma apps also seemed to be able to motivate users through gamification components (R3) but with considerable differences between apps: the relatively high standard deviation of gamification components relative to its mean and its wide range imply that only a few apps have implemented gamification features to a significant extent. For example, the category "rewards" was rarely implemented: only three out of all 38 apps (8%) implemented at least one gamification component of this category. This is somewhat surprising as rewards usually belong to the most commonly implemented gamification components [22]. Point systems, badges, and achievements can be implemented comparatively easy, and their effectiveness is backed by operant conditioning theory, one of psychology's most recognized theories [36].

Furthermore, app quality varied widely (R4). In terms of MARS subscales, engagement scored worst. This may call into question the long-term engagement with asthma apps, a crucial factor for determining their value [7]. In other domains like diabetes self-management, researchers have observed that long-term engagement of app users is generally limited [37]. However, chronic diseases like asthma require long-term self-management [23]. One potential way to improve long-term engagement, which has been successfully applied to physical activity, are interactions with virtual coaches [38]. Thus, developers of upcoming asthma apps might consider the implementation of virtual coaches (eg, chat bots) to enhance long-term engagement.

In summary, great variation in all of the investigated requirements of app potential was found. Therefore, an unequivocal conclusion whether currently available asthma apps have the potential to improve asthma self-management is neither reasonable nor possible. However, due to the high correlations between review frameworks, this review has shown that ratings are consistent across the four requirements of app potential. This implies that high quality apps tend to score well in all requirements for app potential. For example, "Asthma Health by Mount Sinai," which was ranked second best out of all 38 apps in terms of MARS app quality, performed also very well in terms of behavior change techniques (rank 2/38), gamification components (rank 4/38), and available functions (rank 3/38). On the other hand, one of the apps with the lowest MARS app quality ratings ("Asthma Treatment"; rank 36/38) also scored extremely poorly in terms of behavior change techniques (rank 38/38), gamification components (rank 37/38), and available functions (rank 38/38). The consistency between frameworks facilitates the decision which asthma apps to recommend to patients: a promising decision strategy might be to pick one of the best rated apps (see Multimedia Appendix 3 for aggregated app ratings); no trade-off between app potential requirements is necessary. In this way, the asthma apps with the potential to improve a patient's self-management can be selected for clinical practice.

Limitations

In the definition of app potential used in this paper, no measure for long-term engagement was included. As mentioned before, long-term engagement is highly relevant for asthma apps in specific and mHealth apps in general. To the best of the authors' knowledge, there is no valid proxy that predicts long-term use reliably. App reviews and developers would highly benefit from the identification of such proxies, making it a promising field for future research.

Furthermore, apps that are not publicly available were excluded. This limits the generalizability of the results. For example, two apps, which were restricted to study participants and thus were not publicly available, have indicated promising results regarding efficacy [9,12]. Excluding not publicly available apps might have led to a systematic selection bias against apps with a greater potential for improving asthma self-management.

Finally, this review did not contain a content analysis. It only assessed whether a function was implemented, but conclusions regarding its medical value are beyond the scope of this review. However, such an analysis was already conducted in prior reviews [14,15].

Comparison With Prior Work

In general, this review extends the findings of previous reviews by focusing on the potential of apps for asthma self-management. Some of the findings are in line with prior research [13-15,39]: the biggest concern regarding currently available apps is the lack of clinical validation. Not a single study demonstrating efficacy for any of the investigated apps could be identified. However, as argued above, this constraint might be limited to apps that are publicly available. Nevertheless, this review suggests that even without scientific proof for efficacy, at least some of the investigated apps still have the potential to improve asthma self-management.

Therefore, a somewhat more positive conclusion regarding the value of asthma apps can be cautiously drawn in comparison with prior reviews (eg, [14,15]). Among other factors, this may be due to the bottom-up research perspective focusing on app potential instead of an app's evidence base and the development that asthma apps have undergone since the last reviews have been conducted.

Conclusions

In conclusion, this review has found that the potential of asthma apps for improving asthma self-management varies considerably between apps. Physicians and asthmatics should therefore carefully read app reviews before deciding which app to recommend or to use. Additionally, currently available asthma apps do not take full advantage of today's technology. Developers should address the research-practice gaps outlined in the discussion.

XSL•FO
RenderX

Acknowledgments

The first, third, fourth, and fifth author are part of the CSS Health Lab, which is funded by the CSS Insurance (Switzerland). The CSS Insurance was not involved in any stage of this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Exemplary rating criteria for behavior change techniques in mHealth asthma apps.

[PDF File (Adobe PDF File), 25KB - mhealth_v5i8e113_app1.pdf]

Multimedia Appendix 2

Exemplary rating criteria for gamification components in mHealth asthma apps.

[PDF File (Adobe PDF File), 25KB - mhealth_v5i8e113_app2.pdf]

Multimedia Appendix 3

Complete rating table of all 38 asthma apps containing app name, developer, tested app version, app characteristics, app functions, sum scores of behavior change techniques, gamification components, and aggregated MARS ratings.

[XLSX File (Microsoft Excel File), 19KB - mhealth_v5i8e113_app3.xlsx]

References

1. World Health Organization. WHO. Geneva, Switzerland; 2013 Nov. Asthma URL: http://www.who.int/mediacentre/factsheets/fs307/en/ [accessed 2016-11-11] [WebCite Cache ID 6lwfJjosb]
2. Global Asthma Network. Global Asthma Network. Auckland, New Zealand; 2014. The global asthma report 2014 URL: http://www.globalasthmanetwork.org/publications/Global_Asthma_Report_2014.pdf [accessed 2016-12-15] [WebCite Cache ID 6mm0R99VR]
3. Centers for Disease Control Prevention (CDC). CDC. Atlanta, USA; 2014. National health interview survey: most recent asthma data URL: http://www.cdc.gov/asthma/most_recent_data.htm [accessed 2016-11-11] [WebCite Cache ID 6lwfV0iIz]
4. Moorman JE, Akinbami LJ, Bailey CM, Zahran HS, King ME, Johnson CA, et al. National surveillance of asthma: United States, 2001-2010. Vital Health Stat 3 2012 Nov(35):1-58 [FREE Full text] [Medline: 24252609]
5. Agarwal R, Gao GD, DesRoches C, Jha AK. The digital transformation of healthcare: current status and the road ahead. Inf Syst Res 2010 Dec;21(4):796-809. [doi: 10.1287/isre.1100.0327]
6. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol 2016 Mar 10;34(3):239-246. [doi: 10.1038/nbt.3495] [Medline: 26963544]
7. Wu AC. The promise of improving asthma control using mobile health. J Allergy Clin Immunol Pract 2016;4(4):738-739. [doi: 10.1016/j.jaip.2016.04.003] [Medline: 27393782]
8. Gibson PG, Powell H, Coughlan J, Wilson AJ, Abramson M, Haywood P, et al. Self-management education and regular practitioner review for adults with asthma. Cochrane Database Syst Rev 2003(1):CD001117. [doi: 10.1002/14651858.CD001117] [Medline: 12535399]
9. Merchant RK, Inamdar R, Quade RC. Effectiveness of population health management using the propeller health asthma platform: a randomized clinical trial. J Allergy Clin Immunol Pract 2016;4(3):455-463. [doi: 10.1016/j.jaip.2015.11.022] [Medline: 26778246]
10. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34 Suppl:1209-1219. [doi: 10.1037/hea0000306] [Medline: 26651462]
11. Ahmed S, Ernst P, Bartlett SJ, Valois M, Zaihra T, Paré G, et al. The effectiveness of web-based asthma self-management system, my asthma portal (MAP): a pilot randomized controlled trial. J Med Internet Res 2016 Dec 01;18(12):e313 [FREE Full text] [doi: 10.2196/jmir.5866] [Medline: 27908846]
12. Cook KA, Modena BD, Simon RA. Improvement in asthma control using a minimally burdensome and proactive smartphone application. J Allergy Clin Immunol Pract 2016 Jul;4(4):730-7.e1. [doi: 10.1016/j.jaip.2016.03.005] [Medline: 27107690]
13. Marcano BJ, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev 2013 Nov 27(11):CD010013. [doi: 10.1002/14651858.CD010013.pub2] [Medline: 24282112]

XSL•FO
RenderX

14. Huckvale K, Car M, Morrison C, Car J. Apps for asthma self-management: a systematic assessment of content and tools. BMC Med 2012;10:144 [FREE Full text] [doi: 10.1186/1741-7015-10-144] [Medline: 23171675]
15. Huckvale K, Morrison C, Ouyang J, Ghaghda A, Car J. The evolution of mobile apps for asthma: an updated systematic assessment of content and tools. BMC Med 2015;13:58 [FREE Full text] [doi: 10.1186/s12916-015-0303-x] [Medline: 25857569]
16. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing clinical research. Philadelphia, PA: Lippincott Williams & Wilkins; 2013.
17. Apple Inc. Apple. 2016. Apple store review guidelines URL: https://developer.apple.com/app-store/review/guidelines/ [accessed 2016-12-12] [WebCite Cache ID 6mhlxS034]
18. The Food and Drug Administration. FDA. 2015 Feb 09. Mobile medical apps: guidance for industry and food and drug administration staff URL: http://www.fda.gov/downloads/MedicalDevices/.../UCM263366.pdf [accessed 2016-12-12] [WebCite Cache ID 6mhllpUVG]
19. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
20. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
21. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
22. Thiebes S, Lins S, Basten D. Gamifying information systems - a synthesis of gamification mechanics and dynamics. 2014 Presented at: The 22nd European Conference on Information Systems (ECIS); June 9-11, 2014; Tel Aviv, Israel URL: http://aisel.aisnet.org/ecis2014/proceedings/track01/4/
23. Global Initiative for Asthma (GINA). Ginasthma. 2016. Global strategy for asthma management and prevention (2016 update) URL: http://ginasthma.org/wp-content/uploads/2016/04/wms-GINA-2016-main-report-final.pdf [accessed 2017-07-13] [WebCite Cache ID 6rvXcqBxx]
24. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
25. IDC Research Inc. IDC. 2016. Smartphone OS market share Q2/2016 URL: http://www.idc.com/prodserv/smartphone-os-market-share.jsp [accessed 2016-11-18] [WebCite Cache ID 6m0rxO0ls]
26. Ornstein SM, Garr DR, Jenkins RG, Rust PF, Arnon A. Computer-generated physician and patient reminders. Tools to improve population adherence to selected preventive services. J Fam Pract 1991 Jan;32(1):82-90. [Medline: 1985140]
27. Martin J, Chater A, Lorencatto F. Effective behaviour change techniques in the prevention and management of childhood obesity. Int J Obes (Lond) 2013 Oct;37(10):1287-1294. [doi: 10.1038/ijo.2013.107] [Medline: 23756676]
28. Gardner B, Smith L, Lorencatto F, Hamer M, Biddle SJ. How to reduce sitting time? a review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. Health Psychol Rev 2015 Sep 16;70(10):1618-1625. [doi: 10.1080/17437199.2015.1082146] [Medline: 26315814]
29. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR Mhealth Uhealth 2016;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
30. Hamari J, Koivisto J, Sarsa H. Does gamification work? -- a literature review of empirical studies on gamification. 2014 Presented at: 2014 47th Hawaii International Conference on System Sciences; January 6-9, 2014; Waikoloa, HI URL: http://ieeexplore.ieee.org/document/6758978/ [doi: 10.1109/HICSS.2014.377]
31. Stoyanov SR. Youtube. Youtube; 2016. MARS training video URL: https://www.youtube.com/watch?v=25vBwJQIOcE [accessed 2016-06-16] [WebCite Cache ID 6iIc6rURc]
32. Wu AC, Carpenter JF, Himes BE. Mobile health applications for asthma. J Allergy Clin Immunol Pract 2015;3(3):446-8.e1. [doi: 10.1016/j.jaip.2014.12.011] [Medline: 25725939]
33. Barata F, Kowatsch T, Tinschert P, Filler A. Personal MobileCoach: tailoring behavioral interventions to the needs of individual participants. 2016 Presented at: UbiComp '16 Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; September 12-16, 2016; Heidelberg, Germany URL: http://dl.acm.org/citation.cfm?doid=2968219.2972713 [doi: 10.1145/2968219.2972713]
34. Tinschert P, Barata F, Kowatsch T. Enhancing asthma control through IT: design, implementation and planned evaluation of the mobile asthma companion. 2017 Presented at: 13th International Conference on Wirtschaftsinformatik; February 12-15, 2017; St Gallen, Switzerland URL: http://aisel.aisnet.org/wi2017/track13/paper/6/
35. Rabe KF, Vermeire PA, Soriano JB, Maier WC. Clinical management of asthma in 1999: the asthma insights and reality in Europe (AIRE) study. Eur Respir J 2000 Nov;16(5):802-807 [FREE Full text] [Medline: 11153575]
36. Staddon JE, Cerutti DT. Operant conditioning. Annu Rev Psychol 2003;54:115-144 [FREE Full text] [doi: 10.1146/annurev.psych.54.101601.145124] [Medline: 12415075]
37. Tatara N, Arsand E, Skrøvseth SO, Hartvigsen G. Long-term engagement with a mobile self-management system for people with type 2 diabetes. JMIR Mhealth Uhealth 2013;1(1):e1 [FREE Full text] [doi: 10.2196/mhealth.2432] [Medline: 25100649]

XSL•FO
RenderX

38. Watson A, Bickmore T, Cange A, Kulshreshtha A, Kvedar J. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. J Med Internet Res 2012;14(1):e1 [FREE Full text] [doi: 10.2196/jmir.1629] [Medline: 22281837]
39. Al-Durra M, Torio M, Cafazzo JA. The use of behavior change theory in internet-based asthma self-management interventions: a systematic review. J Med Internet Res 2015;17(4):e89 [FREE Full text] [doi: 10.2196/jmir.4110] [Medline: 25835564]

Abbreviations

GPS: global positioning system
ICC: intraclass correlation coefficient
IOS: iPhone operating system
IQR: interquartile range
IT: information technology
MARS: Mobile Application Rating Scale
SD: standard deviation

Edited by G Eysenbach; submitted 16.12.16; peer-reviewed by E Dogan, K Ng; comments to author 19.02.17; revised version received 02.03.17; accepted 13.06.17; published 02.08.17.

Please cite as:
Tinschert P, Jakob R, Barata F, Kramer JN, Kowatsch T
The Potential of Mobile Apps for Improving Asthma Self-Management: A Review of Publicly Available and Well-Adopted Asthma Apps
JMIR Mhealth Uhealth 2017;5(8):e113
URL: http://mhealth.jmir.org/2017/8/e113/
doi:10.2196/mhealth.7177
PMID:28768606

©Peter Tinschert, Robert Jakob, Filipe Barata, Jan-Niklas Kramer, Tobias Kowatsch. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 02.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Impact of Text Messaging on Medication Adherence and Exercise Among Postmyocardial Infarction Patients: Randomized Controlled Pilot Trial

Avinash Pandey[1*]; Alexis A Krumme[2,3*], MS; Tejal Patel[4*], PharmD; Niteesh K Choudhry[2,3*], MD, PhD

[1]University of Western Ontario, London, ON, Canada

[2]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

[3]Center for Healthcare Delivery Sciences, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

[4]School of Pharmacy, University of Waterloo, Waterloo, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Niteesh K Choudhry, MD, PhD
Brigham and Women's Hospital and Harvard Medical School
1620 Tremont St Suite 3030
Boston, MA, 02120
United States
Phone: 1 617 278 0930
Fax: 1 617 232 8602
Email: nkchoudhry@bwh.harvard.edu

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2019/4/e9254/

Abstract

Background: Adherence to evidence-based therapies such as medications and exercise remains poor among patients after a myocardial infarction (MI). Text message reminders have been shown to improve rates of adherence to medication and exercise, but the existing studies have been of short duration.

Objective: Two single-center randomized controlled pilot trials were conducted to evaluate the impact of text message reminders over 12 months on adherence to cardiac medications and exercise among patients receiving cardiac rehabilitation after hospitalization for MI.

Methods: In the medication adherence trial, 34 patients were randomized to receive usual care alone or usual care plus daily text message reminders delivered at the time of day at which medications were to be taken. In the exercise adherence trial, 50 patients were randomized to receive usual care alone or usual care plus 4 daily text messages reminding them to exercise as directed.

Results: The text message reminders led to a mean 14.2 percentage point improvement in self-reported medication adherence over usual care (P<.001, 95% CI 7-21). In the exercise trial, text message reminders resulted in an additional 4.2 days (P=.001, 95% CI 1.9-6.4) and 4.0 hours (P<.001, 95% CI 2.4-5.6) of exercise per month over usual care and a nonsignificant increase of 1.2 metabolic equivalents (METS; P=.06) in exercise capacity as assessed by a BRUCE protocol at 12 months.

Conclusions: Text message reminders significantly increased adherence to medication and exercise among post-MI patients receiving care in a structured cardiac rehabilitation program. This technology represents a simple and scalable method to ensure consistent use of evidence-based cardiovascular therapies.

Trial Registration: Clinicaltrials.gov NCT02783287; https://clinicaltrials.gov/ct2/show/NCT02783287 (Archived by WebCite at http://www.webcitation.org/6sBnvNb05)

(JMIR Mhealth Uhealth 2017;5(8):e110) doi:10.2196/mhealth.7144

XSL•FO
RenderX

KEYWORDS

medication adherence; cardiac rehabilitation; exercise; randomized controlled trial

Introduction

Acute myocardial infarction (MI) remains a leading cause of death and disability worldwide [1]. The use of evidence-based therapies, including exercise and medications, has contributed to substantial reductions in cardiovascular morbidity and mortality [2]. Whereas the prescription of these highly effective therapies is now nearly universal, many patients do not adhere to their medications or exercise regimens over the long term, which in turn is associated with negative consequences for cardiovascular outcomes and spending [3-5].

Although many factors are associated with such gaps in care, simple forgetfulness has been identified as a key contributor [6-8]. Strategies to address this barrier include improvements to pillboxes and bottles, including manual and electronic reminder systems; however, only a few have been rigorously evaluated and, in some cases, their complexity, cost, and the need for patients to purchase and set them up to use them may limit their ultimate value [9-13].

With the widespread use of mobile phones, including the recent increase in uptake among elderly patients, text message–based reminders may represent a potentially effective avenue for intervention [14-16]. Although some clinical benefit from text message reminders has been established in improving both medication and exercise adherence, there is little evidence about the effectiveness of this strategy for patients recently discharged after MI [17-20], which represents a particularly high-risk patient subgroup [21]. The few studies that do exist have been of limited duration, with the longest study among post-MI patients being of an 8-week duration only. Accordingly, we sought to assess the impact of structured daily text message reminders on adherence to post-MI medications and exercise in two randomized controlled pilot trials.

Methods

To assess the impact of text message reminders on adherence to medications and exercise, we conducted two single-center, open-label, randomized controlled pilot trials in patients recently discharged from hospital after MI. This research was approved by the University of Waterloo Research Ethics Committee (20980) and is registered with clinicaltrials.gov (NCT02783287).

The trials were conducted at Cambridge Cardiac Rehab in Ontario, Canada, which enrolled consecutive patients aged 18 years and older, who within the preceding 2 weeks had been discharged from hospital after MI and enrolled in a structured cardiac rehabilitation program. In the medication adherence trial, we included patients who were receiving treatment with medications from all four of the following classes: antiplatelets, beta-blockers, angiotensin converting enzyme inhibitors-I or angiotensin receptor blockers, and 3-hydroxy-3-methylglutaryl-coenzyme reductase enzyme inhibitors (statins). As the trial involved daily text messaging, patients taking medications in dosing regimens of more than once daily were excluded. In the exercise adherence trial, we included patients who had been prescribed an exercise regimen comprising 45 min of exercise 5 days a week. In both trials, individuals without a mobile phone, those who were unable to read and write in English or provide informed consent, or those who were incarcerated were excluded.

In both trials, following recruitment and written informed consent, we collected demographic information, including age, gender, medical diagnosis, medication list, and educational status. Participants were randomized using a Web-based random number generator in a 1:1 ratio to intervention or control. Patients were not permitted to participate in both trials.

In the medication adherence trial, patients randomized to intervention received daily text messages at the time they preferred to take their medications. The text messages simply indicated that patients should remember to take their medications and contained no identifiable information such as medications names or classes. For example, patients who took their medication in the morning received a text message that read, “Please remember to take your morning medications now.” Patients randomized to the control arm of the study did not receive text reminders. Automated text message reminder software was developed with Microsoft Small Basic, a simplified version of Microsoft Visual Basic, and did not utilize bidirectional contact.

For the exercise adherence trial, patients randomized to the intervention group were sent text message reminders 4 times daily, at 7:30 am, 12:00 pm, 6:00 pm, and 9:00 pm. The text message reminders simply read, “Please remember to exercise for 30 minutes today.” These timings were identified as common timings for exercise by the staff of the cardiac rehabilitation center where the study was conducted. Patients randomized to the control arm of the study did not receive any text message reminders.

All patients in both trials participated in the structured outpatient cardiac rehabilitation program in which they were enrolled. The program comprised 1 week of education about diet, proper exercise technique, heart disease and MI, smoking cessation, and stress management and was followed by 2 sessions per week of on-site exercise under the supervision of kinesiologists and nurses. The study protocol was conducted with the help of the staff at the study site. Cardiac rehabilitation nurses and kinesiologists distributed and collected logbooks and enrolled patients into the text message reminder system. After completing the initial 3 months of the program, patients continued their rehabilitation at home but were instructed to return for a follow-up assessment at 1 year. In both trials, patients were followed up for up to 12 months, including the 3 initial months of the program, and patients and their health care providers were aware of the arm to which they had been randomized.

In the adherence trial, all study participants were asked to use a logbook to record the name and timing of the medications they had taken on a daily basis. The logbooks were based on

XSL•FO
RenderX

the standard logbooks used by the cardiac rehabilitation center as suggested by The Canadian Association of Cardiovascular Prevention and Rehabilitation. The logbooks had 1 page per day, with separate rows for different times of the day. Patients were instructed not to log anything if they missed a medication or did not exercise on a specific day. Missing entries were interpreted as a missed dose. Logbooks were collected monthly. From these logs, absolute medication adherence was calculated as the percentage of total prescribed doses that were actually taken each month. For example, if a patient took 27 out of the 30 prescribed doses for a month, they were deemed to be 90% adherent. The primary outcome for this trial was average adherence over 12 months, calculated as the mean of each of the 12 monthly measurements of the percentage of days covered (PDC). The secondary outcome was "full adherence," which was a measure of whether average adherence over the 12 months of follow-up was greater than 80%, which in turn represented a level of use above which patients with coronary artery disease benefit from statins and the threshold used by most quality measures [21-23].

In the exercise trial, all participants were asked to log the time when they started and finished exercising every day. Logbooks were collected on a monthly basis. The primary outcome for this trial was the average number of days per month that a patient exercised over 12 months, calculated as the mean of each of the 12 monthly measurements of the days per months exercised. Secondary outcomes included the average number of hours of exercise per month and cardiopulmonary fitness as assessed by a BRUCE protocol 12 months after randomization [24,25]. Cardiopulmonary fitness was evaluated with an exercise stress test conducted according to the BRUCE protocol, which is the most widely used method employed in routine clinical practice. In this approach, patients walk on a treadmill and the speed and inclination are adjusted based on a preset schedule. Patients walk until they are unable to continue or are instructed to stop by the technician conducting the test. The duration of time that elapses before stopping is then converted to metabolic equivalents (METS) based on standard calculations.

Sample sizes for the trials were estimated based on the results of two preparatory studies conducted at the Cambridge Cardiac Care Center. The medication adherence study was a 2-month crossover study of 30 patients who were recently discharged from hospital after admission for coronary artery disease; these patients were randomized to receive text message reminders for 1 month and to receive no text message reminders the other month. The exercise adherence study was a 2-month parallel group study of 16 post-MI patients who were randomized to either receive daily text message reminders or not to receive daily text message reminders for 2 months. The medication adherence pilot found a 10.8 percentage point (standard deviation [SD] 8.9) increase in adherence from text message reminders. On the basis of this increase in adherence, and assuming an alpha of .05, we estimated that we could achieve 80% power with a sample size of at least 32 patients. The exercise adherence pilot trial found an increase of 10.3 (SD 18.7) days per month of exercise; on the basis of this increase and assuming an alpha of .05, we estimated that we could achieve 80% power with a sample size of 50 patients.

All analyses were performed based on the intention-to-treat principles. We calculated means and frequencies of prerandomization variables separately by study arm. For the medication adherence trial, we evaluated the impact of text messaging on average adherence and optimal adherence using linear and logistic regression, respectively. For the exercise adherence trial, all outcomes were evaluated using linear regression. After performing these analyses, we ran multivariable models adjusting for characteristics that were imbalanced by chance at baseline.

We additionally conducted hypothesis-generating subgroup analyses on the following factors: sex, age >65 years versus ≤65 years, postsecondary education versus up to postsecondary education, and self-reported depression versus no depression at baseline. Specifically, we evaluated whether model-based interaction terms were statistically significant for the primary outcome. We used Statistical Analysis System version 9.4 for all statistical analyses.

Results

Medication Adherence

We screened a total of 90 patients discharged from hospital after acute MI, of whom 56 (63%) did not meet study inclusion criteria or declined to participate (Figure 1). The remaining 34 were randomly assigned to receive usual care alone (control) or usual care plus text message reminders (intervention). One control patient withdrew from the study during the first month of the study. Baseline characteristics of the study participants are shown in Table 1. Intervention subjects were slightly older, more likely to be smokers, taking more medications at baseline, and female. Baseline adherence before the beginning of the study was not measured. At month 1, patients in the control group achieved 90% medication adherence.

The impact of text messaging on medication adherence is shown in Figure 1 and Table 2. The average medication adherence during follow-up for patients randomized to usual care was 80% (95% CI 73-86). The mean difference in PDC between the text message and control groups was 14.2 percentage points (95% CI 7-21, $P<.001$; Table 2). All intervention patients were optimally adherent to their prescribed medications during follow-up compared with 50% (8/16) of control patients ($P<.001$). The results remained unchanged in a multivariable model adjusting for differences in sex, the only characteristic that was statistically significantly imbalanced at baseline.

XSL•FO
RenderX

Figure 1. Monthly percentage of days covered (PDC) by study arm.

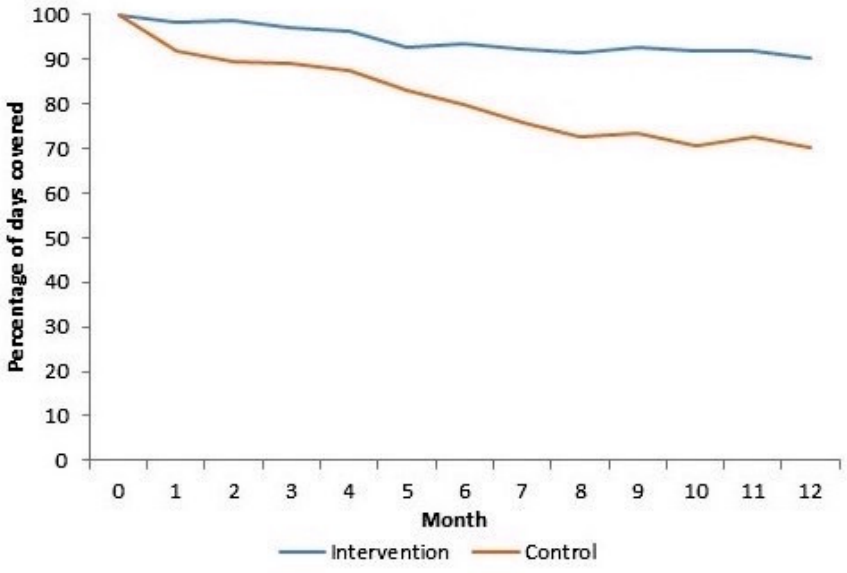


Table 1. Baseline characteristics.

Characteristic	Adherence trial		Exercise trial	
	Text message reminders (N=17)	Usual care (N=16)	Text message reminders (N=25)	Usual care (N=25)
Demographic				
Age in years, mean (SD)	64.6 (11.5)	62.1 (11.0)	64.3 (10.7)	63.7 (9.6)
Male gender, n (%)	6 (35)	14 (88)	15 (60)	11 (44)
Education[a]				
Secondary school, n (%)	10 (67)	8 (50)	17 (71)	16 (64)
Postsecondary school, n (%)	5 (33)	8 (50)	7 (29)	9 (36)
Clinical characteristics				
Cancer, n (%)	1 (6)	1 (6)	2 (8)	1 (4)
COPD[b], n (%)	3 (18)	3 (19)	4 (16)	5 (20)
Dementia, n (%)	3 (18)	2 (13)	5 (20)	4 (16)
Depression (self-report), n (%)	6 (35)	6 (38)	12 (48)	10 (40)
Diabetes, n (%)	6 (35)	7 (44)	10 (40)	8 (32)
Heart failure[c], n (%)	-	-	10 (40)	9 (36)
Chronic kidney disease, n (%)	2 (12)	3 (19)	6 (24)	5 (20)
Smoker, n (%)	7 (41)	4 (25)	10 (40)	9 (36)
Stroke, n (%)	4 (24)	1 (6)	4 (16)	3 (12)
Baseline BRUCE[c], mean (SD)	-	-	5.0 (1.7)	5.1 (1.9)
Number of medications[d], mean (SD)	10.1 (4.5)	8.0 (5.2)	10.5 (4.6)	9.0 (6.1)

[a]2 patients missing from Adherence trial, intervention arm; 1 patient missing from Exercise trial, intervention arm.

[b]COPD: chronic obstructive pulmonary disease.

[c]Not collected for Adherence trial.

[d]1 patient missing from Adherence trial, control arm.

XSL•FO
RenderX

Table 2. Impact of text messaging on study outcomes.

Trial and outcome	Parameter	Text message reminders (95% CI)	Usual care (95% CI)	Difference (95% CI)	*P* value[a]
Medication adherence					
Primary	Average percentage of days covered during follow-up	94 (92-96)	80 (73-86)	14 (7-21)	<.001
Secondary	Percentage (N) fully adherent during follow-up	100 (17)	50 (8)	50	<.001[b]
Exercise adherence					
Primary	Mean days of exercise per month during follow-up	17.2 (16.0-18.5)	13.1 (11.1-15.0)	4.2 (1.9-6.4)	.001
Secondary	Mean hours of exercise per month during follow-up	12.5 (11.5-13.6)	8.5 (7.3-9.8)	4.0 (2.4-5.6)	<.001
	Cardiopulmonary fitness at month 12 (METS)[c]	7.4 (6.5-8.3)	6.2 (5.2-7.2)	1.2 (−0.1 to 2.5)	.06

[a]Satterthwaite method, except where noted.

[b]Fisher exact test.

[c]METS: metabolic equivalents, where 1 MET is equal to a whole-body oxygen consumption of 3.5 mL O_2/kg/min.

Table 3. Subgroup analysis.

Trial and subgroup	Text message reminders (N)	Usual care (N)	Absolute difference	*P* value[a]
Adherence outcome: percentage of days covered				
Male	93 (6)	80 (14)	13	.95
Female	94 (11)	81 (12)	13	.95
Age >65 years	91 (7)	73 (8)	18	.01
Age ≤65 years	96 (10)	85 (8)	11	.01
>12 years of education	94 (5)	82 (8)	12	.60
<12 years of education	93 (10)	77 (8)	16	.60
Depression (self-report)	95 (6)	79 (6)	16	.95
No depression	94 (11)	80 (10)	14	.95
Exercise outcome: days of exercise per month				
Male	16.8 (15)	11 (11)	5.8	.38
Female	19.2 (10)	10.5 (14)	8.7	.38
Age >65 years	18.2 (12)	12 (11)	6.2	.15
Age ≤65 years	16 (13)	14.2 (14)	1.8	.15
>12 years of education	17.5 (7)	13.9 (9)	3.6	.74
<12 years of education	17.1 (17)	12.6 (16)	4.5	.74
Depression (self-report)	18.4 (12)	12.6 (10)	5.8	.31
No depression	15.8 (13)	12.7 (15)	3.1	.31

[a]Regression model-based interaction term.

Adherence was consistently higher for patients randomized to receive text messages across all subgroups but was statistically significantly only for age (Table 3). Text messages increased adherence by an average of 18% among patients ≥65 years of age compared with 11% among patients younger than 65 years (*P* value for interaction .009).

Exercise Adherence

We screened a total of 92 patients discharged from hospital after an acute MI (Figure 2). Of these, 42 were ineligible to participate in the study. The remaining 50 patients were randomized to either receive usual care or usual care in addition to text message reminders.

Baseline characteristics of the study participants are shown in Table 1. As in the medication adherence trial, intervention subjects tended to be older, more likely to smoke, and, on average, taking more medications at baseline. Additionally, intervention subjects were more likely to be female but were similar to controls with respect to cardiopulmonary fitness.

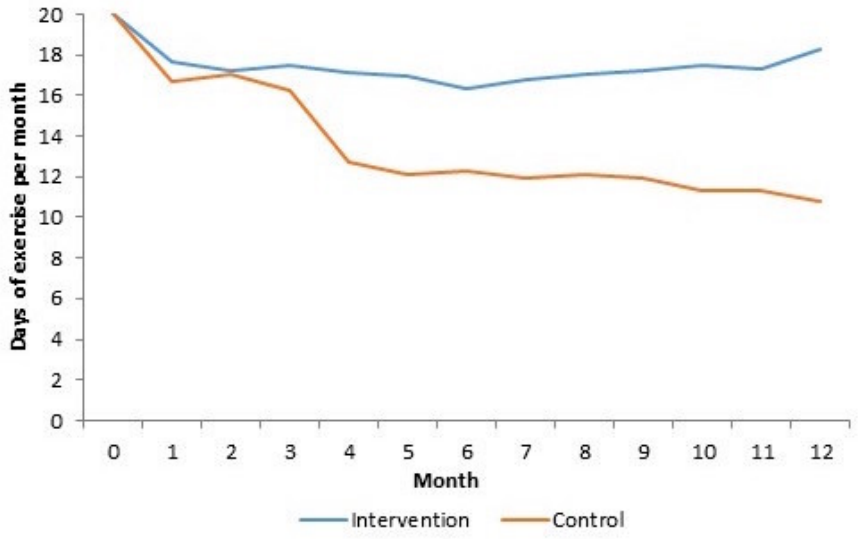
XSL•FO
RenderX

Baseline exercise behavior before the beginning of the study was not observed. At month 1, patients in the control group exercised on an average of 16.7 days per month.

The impact of text messaging on exercise adherence is shown in Figure 2 and Table 2. The patients assigned to usual care exercised on an average of 13.1 days per month. In the text message group, the average number of days of exercise per month increased by 4.2 days (95% CI 1.9-6.4) to 17.2 days (P<.001). In multivariable models adjusting for imbalances in baseline characteristics, the results remained unchanged.

Patients receiving text messages also exercised an additional 4.0 hours per month (P<.001) and had a nonstatistically significant increase of 1.2 METS (P=.06) of exercise capacity. In subgroup analyses, days of exercise per month were consistently higher across all strata of patients randomized to the intervention arm (Table 3).

Figure 2. Monthly days of exercise by study arm.



Discussion

Principal Findings

In this pair of randomized controlled pilot trials of text message reminders delivered to post-MI patients participating in cardiac rehabilitation, we found significant improvement in both adherence to medications and the frequency and duration of exercise over 12 months of follow-up as measured by patient self-report. In addition, the secondary outcome of the trial focusing on exercise adherence showed a nonsignificant trend to improved cardiopulmonary fitness as measured by exercise stress testing.

Several studies have demonstrated the potential benefits of external reminders, including text messaging, for improving medication adherence [17,26]. A recent meta-analysis of 16 randomized studies of adult patients with chronic disease, including coronary artery disease, HIV, and epilepsy, found that the odds of full adherence was more than doubled for those who received text messages [17]. Of the studies included, only 3 explicitly studied patients taking medications for secondary prevention after acute coronary syndrome, of which only 2 were ultimately included in the meta-analysis [19,20]. Both studies followed patients for only 30 days, sending daily personalized reminders to patients in the intervention arms to take their cardiovascular medications. Moreover, neither of these studies was conducted within the context of a comprehensive post-MI cardiac rehabilitation program, in which approximately 40% of patients post-MI in the United States enroll [27-29], and which could conceivably have augmented or extinguished the impact of text message reminders. In contrast, our study, despite being small in size and conducted at a single center, found sustained improvements in medication-taking from text message reminders delivered in addition to cardiac rehabilitation, as compared with cardiac rehabilitation alone that persisted over the course of 12 months of follow-up.

Several studies have demonstrated the beneficial effects of text message reminders on exercise adherence among patients with heart disease [26,30,31]. For example, the TEXT-MI trial enrolled patients with documented cardiovascular disease and randomized them to usual care or to receive text messages that provided advice, motivational reminders, and support to change a wide range of lifestyle behaviors, one of which was physical activity [26,30,31]. This trial found clinically meaningful and statistically significant improvements in exercise, low-density lipoprotein cholesterol, blood pressure, body mass index, and smoking even though follow-up ended after 6 months. A few existing studies have specifically enrolled patients in the immediate post-MI setting or have been conducted concurrently with cardiac rehabilitation. Frederix et al [32] studied the impact of augmenting cardiac rehabilitation with a “tele-rehabilitation” program in which emails and texts were sent by a coach based upon patient’s individual accelerometer data. The intervention resulted in significant improvements in physical fitness and quality of life over 24 weeks of follow-up. In contrast, our intervention relied on fully automated messages, without the need to receive and process accelerometer data, and followed patients for 12 months. The improvements in self-reported

XSL•FO
RenderX

exercise adherence translated to higher exercise tolerance, as measured by the BRUCE protocol exercise stress test administered at 12 months. Although these results did not reach statistical significance, they demonstrate a strong long-term trend that bears confirmation in a larger study.

There are many similarities between medication adherence and exercise adherence behaviors that suggest why a highly similar intervention may have been effective for both behaviors. In specific, medication and exercise adherence require daily engagement and have multiple barriers to their performance, including motivation, literacy or knowledge, or simple forgetfulness. Moreover, studies of text messaging to improve medication adherence or exercise adherence have reported high patient satisfaction, with the vast majority of patients reporting that the short message service messages were useful and easy to understand [19,26,31].

A greater drop-off in the days of exercise per month was noted in the control group at month 3. This corresponds to the end of the on-site cardiac rehabilitation program. For the first 3 months of cardiac rehabilitation, patients exercised on-site for 2 sessions per week. Following this, patients continued to exercise at home but did not exercise on-site. No similar drop-off was noted in the text message arm of the study.

In our subgroup analyses, we observed consistent effects of texting on medication and exercise adherence. Although we had hypothesized a priori that access and knowledge of the technology may have been a barrier to effective use of this intervention among the elderly, our observations revealed the contrary. These results are consistent with studies that found other electronic interventions to be as or more effective among older individuals [33,34]. This being said, more research is required to confirm these results in larger and more diverse patient populations.

Patients were surveyed following the end of the study to discuss their experience with daily text message reminders. The majority reported that these reminders were helpful in maintaining adherence to their therapies and did not find them to be overly intrusive. They reported their preference for the text messages over alarms, which could be turned off more easily. Patients also reported liking that the reminders came from an outside source and likened them to calls from a nurse or other health care provider. More formalized polling of patient participants in future studies may provide further insight into patient experience with reminder systems such as the one examined in this study.

Limitations

There are several limitations to our trials. First, both trials had small sample sizes and were conducted at a single cardiac rehabilitation facility in Ontario, Canada. Future studies with larger sample sizes could seek to evaluate other end points such as hospital readmissions. Data collected in this study were largely self-reported, although this has been the primary mode of outcome evaluation in the majority of existing studies evaluating text reminders on medication adherence [17]. The fact that the improvement in exercise adherence we observed was correlated with objective improvements in aerobic fitness measured by stress testing suggests that our outcome measurement was reliable and any mismeasurement was nondifferential with respect to the exposure. Our medication adherence trial included only patients on once-a-day regimens of the four classes of medications studied. Whereas all four medication classes we studied are widely available with this dosing strategy, our results may not be generalizable to all post-MI patients. Additionally, most of the patients in the study had one or more chronic diseases and were taking medications for these diseases. The message sent to patients reminding them to take their medications read, “Please remember to take your morning medications now.” Although we expect this nonspecific messaging would have encouraged patients to take all of their therapies as prescribed, we only evaluated adherence to the four post-MI medications. Also, the long follow-up of the study during which no one dropped out and which required patients to complete logbooks on a daily basis may limit generalizability to other patient groups. In both trials, logbooks were used to measure medication and exercise adherence because this method, previously validated, was already in use at the cardiac rehabilitation center where we conducted the study, and we hoped to integrate study procedures into the cardiac rehabilitation program. It is possible that patients may have taken a medication or exercised but forgotten to log this in the logbook; however, we do not believe that this has significantly impacted our results [35]. Finally, there were baseline imbalances in the treatment arms in potentially important characteristics, notably sex, although our overall results remained unchanged in multivariable and subgroup analyses. Further research should be conducted in larger and more diverse populations.

Conclusions

In conclusion, structured text message reminders were found to significantly improve adherence to both medications and exercise. Although further research is required to validate our results in larger and more diverse settings, text messaging appears to represent a simple and scalable strategy for improving adherence to medications and exercise among post-MI patients.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V1.6.1).

XSL•FO
RenderX

[PDF File (Adobe PDF File), 445KB - mhealth_v5i8e110_app1.pdf]

References

1. Oliveira GB, Avezum A, Roever A. Cardiovascular disease burden: evolving knowledge of risk factors in myocardial infarction and stroke through population-based research and perspectives in global prevention. Front Cardiovasc Med 2015;2:32 [FREE Full text] [doi: 10.3389/fcvm.2015.00032] [Medline: 26664903]
2. Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, Kottke TE, et al. Explaining the decrease in U.S. deaths from coronary disease, 1980-2000. N Engl J Med 2007 Jun 07;356(23):2388-2398. [doi: 10.1056/NEJMsa053935] [Medline: 17554120]
3. Langer Research Associates. National Community Pharmacists Association. 2013. Medication adherence in America: a national report card URL: http://www.ncpa.co/adherence/AdherenceReportCard_Full.pdf [WebCite Cache ID 6rrkbl2zA]
4. Choudhry NK, Setoguchi S, Levin R, Winkelmayer W, Shrank W. Trends in adherence to secondary prevention medications in elderly post-myocardial infarction patients. Pharmacoepidemiol Drug Saf 2008 Dec;17(12):1189-1196 [FREE Full text] [doi: 10.1002/pds.1671] [Medline: 18956426]
5. Fonarow GC, French W, Frederick P. Trends in the use of lipid-lowering medications at discharge in patients with acute myocardial infarction: 1998 to 2006. Am Heart J 2009 Jan;157(1):185-94.e2. [doi: 10.1016/j.ahj.2008.09.001] [Medline: 19081417]
6. Martin LR, Williams S, Haskard K, Dimatteo M. The challenge of patient adherence. Ther Clin Risk Manag 2005 Sep;1(3):189-199 [FREE Full text] [Medline: 18360559]
7. Gadkari AS, McHorney C. Unintentional non-adherence to chronic prescription medications: how unintentional is it really? BMC Health Serv Res 2012 Jun 14;12:98 [FREE Full text] [doi: 10.1186/1472-6963-12-98] [Medline: 22510235]
8. Wroe AL. Intentional and unintentional nonadherence: a study of decision making. J Behav Med 2002 Aug;25(4):355-372. [Medline: 12136497]
9. Choudhry NK, Krumme A, Ercole P, Girdish C, Isaman D, Matlin O, et al. Rationale and design of the randomized evaluation to measure improvements in non-adherence from low-cost devices (REMIND) trial. Contemp Clin Trials 2015 Jul;43:53-59. [doi: 10.1016/j.cct.2015.05.006] [Medline: 25979316]
10. Checchi KD, Huybrechts KF, Avorn J, Kesselheim AS. Electronic medication packaging devices and medication adherence: a systematic review. JAMA 2014 Sep 24;312(12):1237-1247 [FREE Full text] [doi: 10.1001/jama.2014.10059] [Medline: 25247520]
11. Rosen MI, Rigsby M, Salahi J, Ryan C, Cramer J. Electronic monitoring and counseling to improve medication adherence. Behav Res Ther 2004 Apr;42(4):409-422. [doi: 10.1016/S0005-7967(03)00149-9] [Medline: 14998735]
12. Vervloet M, Linn AJ, van Weert JC, de Bakker DH, Bouvy ML, van DL. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature. J Am Med Inform Assoc 2012;19(5):696-704 [FREE Full text] [doi: 10.1136/amiajnl-2011-000748] [Medline: 22534082]
13. Zedler BK, Kakad P, Colilla S, Murrelle L, Shah N. Does packaging with a calendar feature improve adherence to self-administered medication for long-term use? a systematic review. Clin Ther 2011 Jan;33(1):62-73. [doi: 10.1016/j.clinthera.2011.02.003] [Medline: 21397775]
14. Friedman RH, Kazis LE, Jette A, Smith MB, Stollerman J, Torgerson J, et al. A telecommunications system for monitoring and counseling patients with hypertension. Impact on medication adherence and blood pressure control. Am J Hypertens 1996 Apr;9(4):285-292. [Medline: 8722429]
15. Orrell C, Cohen K, Mauff K, Bangsberg DR, Maartens G, Wood R. A randomized controlled trial of real-time electronic adherence monitoring with text message dosing reminders in people starting first-line antiretroviral therapy. J Acquir Immune Defic Syndr 2015 Dec 15;70(5):495-502. [doi: 10.1097/QAI.0000000000000770] [Medline: 26218411]
16. Lauffenburger JC, Choudhry N. Text messaging and patient engagement in an increasingly mobile world. Circulation 2016 Feb 09;133(6):555-556 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.116.021182] [Medline: 26769741]
17. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
18. Khonsari S, Subramanian P, Chinna K, Latif LA, Ling LW, Gholami O. Effect of a reminder system using an automated short message service on medication adherence following acute coronary syndrome. Eur J Cardiovasc Nurs 2015 Apr;14(2):170-179. [doi: 10.1177/1474515114521910] [Medline: 24491349]
19. Park LG, Howie-Esquivel J, Chung ML, Dracup K. A text messaging intervention to promote medication adherence for patients with coronary heart disease: a randomized controlled trial. Patient Educ Couns 2014 Feb;94(2):261-268. [doi: 10.1016/j.pec.2013.10.027] [Medline: 24321403]
20. Quilici J, Fugon L, Beguin S, Morange PE, Bonnet J, Alessi M, et al. Effect of motivational mobile phone short message service on aspirin adherence after coronary stenting for acute coronary syndrome. Int J Cardiol 2013 Sep 20;168(1):568-569. [doi: 10.1016/j.ijcard.2013.01.252] [Medline: 23462636]

XSL•FO
RenderX

21. Choudhry NK, Glynn R, Avorn J, Lee J, Brennan T, Reisman L, et al. Untangling the relationship between medication adherence and post-myocardial infarction outcomes: medication adherence and clinical outcomes. Am Heart J 2014 Jan;167(1):51-58. [doi: 10.1016/j.ahj.2013.09.014] [Medline: 24332142]
22. Centers for Medicare & Medicaid Services. 2014 Nov. Trends in part C & D star rating measure cut points URL: https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/Downloads/2015-Trends-in-Part-C-and-D-Star-Rating-Measure-Cut-Points-.pdf [accessed 2015-11-09] [WebCite Cache ID 6rtSHT5lc]
23. NCQA.: National Committee for Quality Assurance Summary table of measures, product lines and changes URL: https://www.ncqa.org/Portals/0/HEDISQM/Hedis2015/List_of_HEDIS_2015_Measures.pdf [accessed 2015-10-12] [WebCite Cache ID 6rtSp6LaJ]
24. Bruce RA. Exercise testing of patients with coronary heart disease. Principles and normal standards for evaluation. Ann Clin Res 1971 Dec;3(6):323-332. [Medline: 5156892]
25. Bruce RA, DeRouen T, Peterson D, Irving J, Chinn N, Blake B, et al. Noninvasive predictors of sudden cardiac death in men with coronary heart disease. Predictive value of maximal stress testing. Am J Cardiol 1977 May 26;39(6):833-840. [Medline: 871109]
26. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
27. Menezes AR, Lavie CJ, Milani RV, Forman DE, King M, Williams MA. Cardiac rehabilitation in the United States. Prog Cardiovasc Dis 2014 Apr;56(5):522-529. [doi: 10.1016/j.pcad.2013.09.018] [Medline: 24607017]
28. Dalal HM, Doherty P, Taylor R. Cardiac rehabilitation. BMJ 2015 Sep 29;351:h5000 [FREE Full text] [Medline: 26419744]
29. Jelinek MV, Thompson D, Ski C, Bunker S, Vale M. 40 years of cardiac rehabilitation and secondary prevention in post-cardiac ischaemic patients. Are we still in the wilderness? Int J Cardiol 2015 Jan 20;179:153-159. [doi: 10.1016/j.ijcard.2014.10.154] [Medline: 25464436]
30. Maddison R, Whittaker R, Stewart R, Kerr A, Jiang Y, Kira G, et al. HEART: heart exercise and remote technologies: a randomized controlled trial study protocol. BMC Cardiovasc Disord 2011 May 31;11:26 [FREE Full text] [doi: 10.1186/1471-2261-11-26] [Medline: 21624142]
31. Pfaeffli DL, Whittaker R, Dixon R, Stewart R, Jiang Y, Carter K, et al. Acceptability of a mobile health exercise-based cardiac rehabilitation intervention: a randomized trial. J Cardiopulm Rehabil Prev 2015;35(5):312-319. [doi: 10.1097/HCR.0000000000000125] [Medline: 26181037]
32. Pfaeffli DL, Whittaker R, Dixon R, Stewart R, Jiang Y, Carter K, et al. Acceptability of a mobile health exercise-based cardiac rehabilitation intervention: a randomized trial. J Cardiopulm Rehabil Prev 2015;35(5):312-319. [doi: 10.1097/HCR.0000000000000125] [Medline: 26181037]
33. Buckwalter KC, Wakefield B, Hanna B, Lehmann J. New technology for medication adherence: electronically managed medication dispensing system. J Gerontol Nurs 2004 Jul;30(7):5-8. [Medline: 15287321]
34. Mira JJ, Navarro I, Botella F, Borrás F, Nuño-Solinís R, Orozco D, et al. A Spanish pillbox app for elderly patients taking multiple medications: randomized controlled trial. J Med Internet Res 2014;16(4):e99 [FREE Full text] [doi: 10.2196/jmir.3269] [Medline: 24705022]
35. Jeffrey BA, Hannan M, Quinn E, Zimmerman S, Barton B, Rubin C, et al. Self-reported adherence with the use of a device in a clinical trial as validated by electronic monitors: the VIBES study. BMC Med Res Methodol 2012 Nov 14;12:171 [FREE Full text] [doi: 10.1186/1471-2288-12-171] [Medline: 23150931]

Abbreviations

METS: metabolic equivalents
MI: myocardial infarction
PDC: percentage of days covered

Edited by L Buis; submitted 13.12.16; peer-reviewed by M Vervloet, J Gatwood, B Holtz; comments to author 03.02.17; revised version received 11.05.17; accepted 05.06.17; published 03.08.17.

Please cite as:
Pandey A, Krumme AA, Patel T, Choudhry NK
The Impact of Text Messaging on Medication Adherence and Exercise Among Postmyocardial Infarction Patients: Randomized Controlled Pilot Trial
JMIR Mhealth Uhealth 2017;5(8):e110
URL: http://mhealth.jmir.org/2017/8/e110/
doi:10.2196/mhealth.7144
PMID:28778843

XSL•FO
RenderX

©Avinash Pandey, Alexis A Krumme, Tejal Patel, Niteesh K Choudhry. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 03.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Feature-Free Activity Classification of Inertial Sensor Data With Machine Vision Techniques: Method, Development, and Evaluation

Jose Juan Dominguez Veiga[1], BSc (Hons), MSc; Martin O'Reilly[2], BEng (Hons), HDip; Darragh Whelan[2], BSc (Hons), MSc; Brian Caulfield[2], BSc (Hons), MSc, PhD; Tomas E Ward[1], BE, MEngSc, PhD

[1]Insight Centre for Data Analytics, Department of Electronic Engineering, Maynooth University, Maynooth, Ireland

[2]Insight Centre for Data Analytics, University College Dublin, Dublin, Ireland

Corresponding Author:
Jose Juan Dominguez Veiga, BSc (Hons), MSc
Insight Centre for Data Analytics
Department of Electronic Engineering
Maynooth University
Bioscience and Engineering Building, North Campus
Co. Kildare
Maynooth,
Ireland
Phone: 353 17086000
Fax: 353 16289063
Email: josejuan.dominguezveiga@nuim.ie

Abstract

Background: Inertial sensors are one of the most commonly used sources of data for human activity recognition (HAR) and exercise detection (ED) tasks. The time series produced by these sensors are generally analyzed through numerical methods. Machine learning techniques such as random forests or support vector machines are popular in this field for classification efforts, but they need to be supported through the isolation of a potentially large number of additionally crafted features derived from the raw data. This feature preprocessing step can involve nontrivial digital signal processing (DSP) techniques. However, in many cases, the researchers interested in this type of activity recognition problems do not possess the necessary technical background for this feature-set development.

Objective: The study aimed to present a novel application of established machine vision methods to provide interested researchers with an easier entry path into the HAR and ED fields. This can be achieved by removing the need for deep DSP skills through the use of transfer learning. This can be done by using a pretrained convolutional neural network (CNN) developed for machine vision purposes for exercise classification effort. The new method should simply require researchers to generate plots of the signals that they would like to build classifiers with, store them as images, and then place them in folders according to their training label before retraining the network.

Methods: We applied a CNN, an established machine vision technique, to the task of ED. Tensorflow, a high-level framework for machine learning, was used to facilitate infrastructure needs. Simple time series plots generated directly from accelerometer and gyroscope signals are used to retrain an openly available neural network (Inception), originally developed for machine vision tasks. Data from 82 healthy volunteers, performing 5 different exercises while wearing a lumbar-worn inertial measurement unit (IMU), was collected. The ability of the proposed method to automatically classify the exercise being completed was assessed using this dataset. For comparative purposes, classification using the same dataset was also performed using the more conventional approach of feature-extraction and classification using random forest classifiers.

Results: With the collected dataset and the proposed method, the different exercises could be recognized with a 95.89% (3827/3991) accuracy, which is competitive with current state-of-the-art techniques in ED.

Conclusions: The high level of accuracy attained with the proposed approach indicates that the waveform morphologies in the time-series plots for each of the exercises is sufficiently distinct among the participants to allow the use of machine vision approaches. The use of high-level machine learning frameworks, coupled with the novel use of machine vision techniques instead of complex manually crafted features, may facilitate access to research in the HAR field for individuals without extensive digital signal processing or machine learning backgrounds.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2017;5(8):e115) doi:10.2196/mhealth.7521

KEYWORDS

machine learning; exercise; biofeedback

Introduction

Background

Inertial sensors are ubiquitous in everyday objects such as mobile phones and wristbands and can provide large amounts of data regarding movement activity. Analysis of such data can be diverse, but in general terms can be characterized as complex operations using a broad range of machine learning techniques and highly sophisticated signal processing methods. The latter is required to extract salient features that can improve recognition performance. These features are not only complex to calculate, but also making a priori reasoned arguments toward their effectiveness in improving overall results is difficult. The temptation to include additional features in an attempt to improve classification accuracy may result in pipelines (infrastructure) with excessive complexity, yielding slower processing and increased resource usage. To counter this proliferation of features, it is common to use dimensionality reduction techniques including linear approaches such as principal component analysis and increasingly common nonlinear methods principally based on manifold learning algorithms.

In contrast to this complex tool, we propose a method to classify human activity from inertial sensor data based on images and using deep learning-based machine vision techniques. This approach reduces the amount of deep domain knowledge needed in terms of digital signaling processing (DSP), down to some basic steps of preprocessing and segmentation, substituting instead a neural network that can learn the appropriate features independent of a user-driven feature candidature step. Convolutional networks are not trivial to work with, but the recent availability of higher level deep learning frameworks such as TensorFlow [1] and the use of transfer learning, a technique to reuse already trained convolutional neural networks (CNNs), considerably reduces the skills needed to set up and operate such a network.

In this study, we sought to demonstrate a novel application of machine vision techniques as a classification method for inertial measurement unit (IMU) data. The main goal of this work was to develop a novel data analysis pathway for researchers who are most interested in this type of work, such as medical and exercise professionals. These individuals may not have the technical background to implement existing state-of-the-art data analysis pathways. We also aimed to evaluate the efficacy of our new classification technique by attempting to detect five commonly completed lower-limb exercises (squats, deadlifts, lunges, single-leg squats, and tuck jumps) using the new data analysis pathway. The accuracy, sensitivity, and specificity of the pathway were compared with recently published work on the same dataset.

Related Work

The three main topics in this section are as follows: (1) a brief overview of the current human activity recognition (HAR) and exercise detection (ED) literature, (2) an account of some of the newer advances in the field that are using neural networks for certain parts of the feature discovery and reduction process, and (3) an introduction to transfer learning, highlighting its benefits in terms of time and resource savings, and working with smaller datasets.

Activity Classification for Inertial Sensor Data

Over the past 15 years, inertial sensors have become increasingly ubiquitous due to their presence in mobile phones and wearable activity trackers [2]. This has enabled countless applications in the monitoring of human activity and performance spanning applications in general HAR, gait analysis, the military field, the medical field, and exercise recognition and analysis [3-6]. Across all these application spaces, there are common challenges and steps which must be overcome and implemented to successfully create functional motion classification systems.

Human activity recognition with wearable sensors usually pertains to the detection of gross motor movements such as walking, jogging, cycling, swimming, and sleeping [5,7]. In this field of motion tracking with inertial sensors, the key challenges are often considered to be (1) the selection of the attributes to be measured; (2) the construction of a portable, unobtrusive, and inexpensive data acquisition system; (3) the design of feature extraction and inference methods; (4) the collection of data under realistic conditions; (5) the flexibility to support new users without the need for retraining the system; and (6) the implementation in mobile devices meeting energy and processing requirements [3,7]. With the ever-increasing computational power and battery life of mobile devices, many of these challenges are becoming easier to overcome.

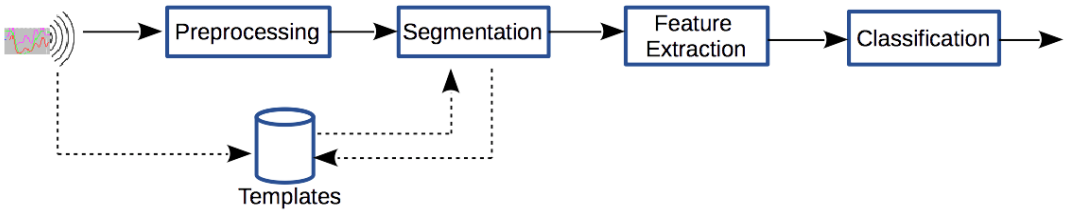
Whereas system functionality is dependent on hardware constraints, the accuracy, sensitivity, and specificity of HAR systems are most reliant on building large, balanced, labeled datasets; the identification of strong features for classification; and the selection of the best machine learning method for each application [3,8-10]. Investigating the best features and machine learning methods for each HAR application requires an individual or team appropriately skilled in signal processing and machine learning and a large amount of time. They must understand how to compute time-domain, frequency-domain, and time-frequency domain features from inertial sensor data and train and evaluate multiple machine learning methods (eg, random forests [11], support vector machines [12], k-nearest neighbors [13], and logistical regression [14]) with such features [3-5]. This means that those who may be most interested in the output of inertial sensor based activity recognition systems (eg, medical professionals, exercise professionals, and biomechanists) are unable to design and create the systems

XSL•FO
RenderX

without significant engagement with machine learning experts [4].

The above challenges in system design and implementation are replicated in activity recognition pertaining to more specific or acute movements. In the past decade, there has been a vast amount of work in the detection and quantification of specific rehabilitation and strength and conditioning exercises [15-17]. Such work has also endeavored to detect aberrant exercise technique and specific mistakes that system users make while exercising, which can increase their chance of injury or decrease their body's beneficial adaptation due to the stimulus of exercise [17,18]. The key steps in the development of such systems have been recently outlined as (1) inertial sensor data collection, (2) data preprocessing, (3) feature extraction, and (4) classification (Figure 1) [4]. Whereas the first step can generally be completed by exercise professionals (eg, physiotherapists and strength and conditioning coaches), the remaining steps require skills outside that included in the training of such experts. Similarly, when analyzing gait with wearable sensors, feature extraction and classification have been highlighted as essential in the development of each application [19,20]. This again limits the type of professional who can create such systems and the rate at which hypotheses for new systems can be tested.

Figure 1. Steps involved in the development of an inertial measurement unit (IMU)-based exercise classification system.



Neural Networks and Activity Recognition

In the past few years, CNNs have been applied in a variety of manners to HAR, in both the fields of ambient and wearable sensing. Mo et al applied a novel approach utilizing machine vision methods to recognize twelve daily living tasks with the Microsoft Kinect. Rather than extract features from the Kinect data streams, they developed 144×48 images using 48 successive frames from skeleton data and 15×3 joint position coordinates and 11×3×3 joint rotation matrices. These images were then used as input to a multilayer CNN which automatically extracted features from the images that were fed in to a multilayer perceptron for classification [21]. Stefic and Patras utilized CNNs to extract areas of gaze fixation in raw image training data as participants watched videos of multiple activities [22]. This produced strong results in identifying salient regions of images that were then used for action recognition. Ma et al also combined a variety of CNNs to complete tasks, such as segmenting hands and objects from first-person camera images and then using these segmented images and motion images to train an action-based and motion-based CNN [23]. This novel use of CNNs allowed an increase in activity recognition rates of 6.6%, on average. These research efforts demonstrated the power of utilizing CNNs in multiple ways for HAR.

Research utilizing CNNs for HAR with wearable inertial sensors has also been published recently. Zeng et al implemented a method based on CNNs which captures the local dependency and scale invariance of an inertial sensor signal [24]. This allows features for activity recognition to be identified automatically. The motivation for developing this method was the difficulties in identifying strong features for HAR. Yang et al also highlighted the challenge and importance of identifying strong features for HAR [25]. They also employed CNNs for feature learning from raw inertial sensor signals. The strength of CNNs in HAR was again demonstrated here as its use in this circumstance outperformed other HAR algorithms, on multiple datasets, which utilized heuristic hand-crafting of features or shallow learning architectures for feature learning. Radu et al also recently demonstrated that the use of CNNs to identify discriminative features for HAR when using multiple sensor inputs from various mobile phones and smartwatches, which have different sampling rates, data generation models, and sensitivities, outperforms classic methods of identifying such features [26]. The implementation of such feature learning techniques with CNNs is clearly beneficial but is complex and may not be suitable for HAR system developers without strong experience in machine learning and DSP. From a CNN perspective, these results are interesting and suggest significant scope for further exploration for machine learning researchers. However, for the purposes of this paper, their inclusion is to both succinctly acknowledge that CNN has been applied to HAR previously and to distinguish the present approach which seeks to use well developed CNN platforms tailored for machine vision tasks in a transfer learning context for HAR recognition using basic time series as the only user created features.

Transfer Learning in Machine Vision

Deep learning-based machine vision techniques are used in many disciplines, from speech, video, and audio processing [27], through to HAR [21] and cancer research [28].

Training deep neural networks is a time consuming and resource intensive task, not only needing specialized hardware (graphics processing unit [GPU]) but also large datasets of labeled data. Unlike other machine learning techniques, once the training work is completed, querying the resulting models to predict results on new data is fast. In addition, trained networks can be repurposed for other specific uses which are not required to be known in advance of the initial training [29]. This arises from the generalized vision capabilities that can emerge with suitable training. More precisely, each layer of the network learns a number of features from the input data and that knowledge is refined through iterations. In fact, the learning that happens at different layers seems to be nonspecific to the dataset, including the identification of simple edges in the first few layers, the

XSL•FO
RenderX

subsequent identification of boundaries and shapes, and growing toward object identification in the last few layers. These learned visual operators are applicable to other sets of data [30]. Transfer learning then is the generic name given to a classification effort when a pretrained network is reused for a task for which it was not specifically trained for. Deep learning frameworks such as Caffe [31] and TensorFlow can make use of pretrained networks, many of which have been made available by researchers in repositories such as the Caffe Model Zoo, available in their github repository.

Retraining requires not only a fraction of the time that a full training session would need (min/h instead of weeks), but more importantly in many cases, allows for the use of much smaller datasets. An example of this is the inception model provided by Google, whose engineers reportedly spent several weeks training on ImageNet [32] (a dataset of over 14 million images in over 2 thousand categories), using multiple GPUs and the TensorFlow framework. In their example [33], they use in the order of 3500 pictures of flowers in 5 different categories to retrain the generic model, producing a model with a fair accuracy rating on new data. In fact, during the retraining stage, the network is left almost intact. The final classifier is the only part that is fully replaced, and "bottlenecks" (the layer before the final one) are calculated to integrate the new training data into the already "cognizant" network. After that, the last layer is trained to work with the new classification categories. This happens in image batches of a size that can be adapted to the needs of the new dataset (alongside other hyperparameters such as learning rate and training steps).

Each step of the training process outputs values for training accuracy, validation accuracy, and cross entropy. A large difference between training and validation accuracy can indicate potential "overfitting" of the data, which can be a problem especially with small datasets, whereas the cross entropy is a loss function that provides an indication of how the training is progressing (decreasing values are expected).

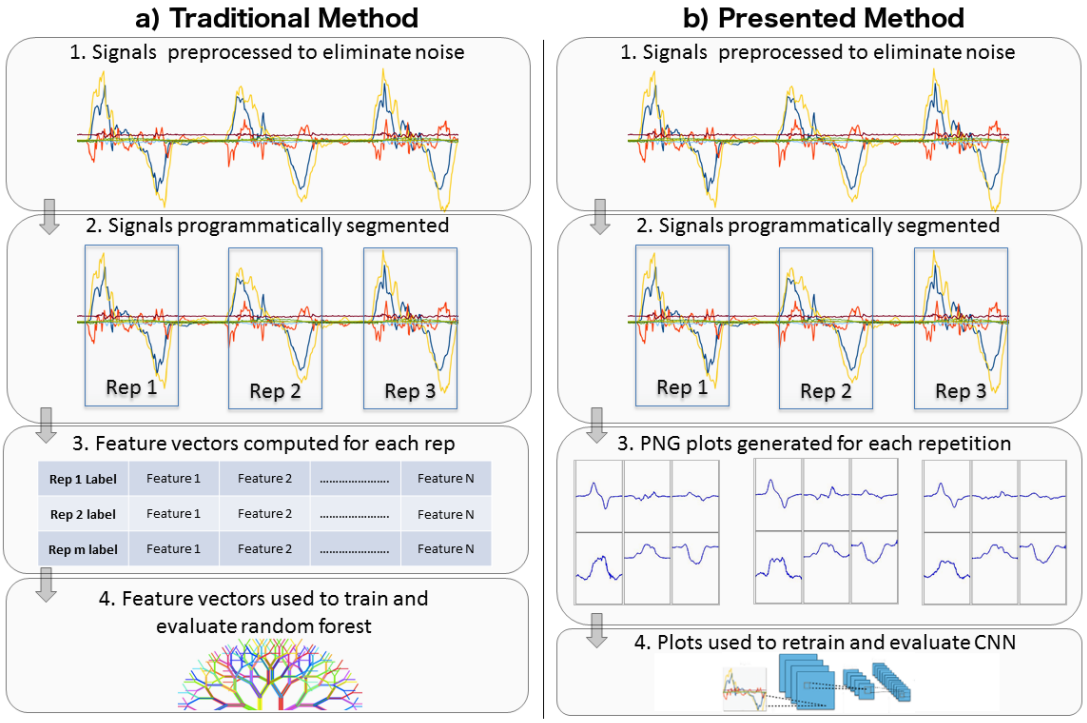
Methods

Study Design

Given the potential advantages of transfer learning in machine vision for the purposes of HAR, we next describe an exemplar study where we apply these ideas for the purposes of classifying exercise data from inertial sensors. This very specific example is sufficiently comprehensive in scale, and scope to represent a typical use case for the approach which to reiterate will use pretrained CNNs with one lightweight additional training step, to classify inertial sensor data based on images generated from the raw data (Figure 2). The level of DSP skills to perform this analysis will be shown to be much lower compared with other methods of classifying this type of data with other machine learning techniques that rely on engineered features (Figure 1).

This section contains all the details required to replicate this approach, focusing on how the data was collected, and how our system was set up and used.

Figure 2. Depiction of the changes between traditional methods and the one presented in this paper, in particular steps 3 and 4.



XSL•FO
RenderX

Data Collection

Participants

A total of 82 healthy volunteers aged 16-38 years (59 males, 23 females, age: 24.68 years [SD (standard deviation) 4.91], height: 1.75m [SD 0.09], body mass: 76.01kg [SD 13.29]) were recruited for the study. Participants did not have a current or recent musculoskeletal injury that would impair performance of multi-joint, lower-limb exercises. All participants had been completing each of the five exercises as part of their training regime for at least one year. The human research ethics committee at University College Dublin approved the study protocol and written informed consent was obtained from all participants before testing. In cases where participants were under the age of 18 years, written informed consent was also obtained from a parent or guardian.

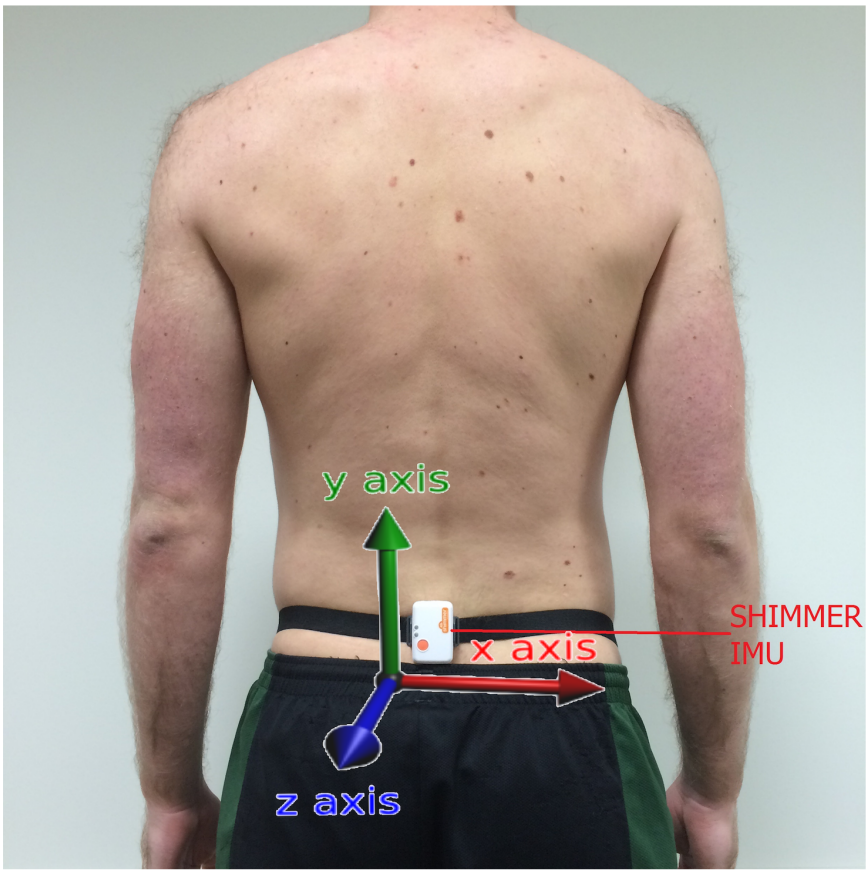
Procedures

The testing protocol was explained to participants upon their arrival at the laboratory. Following this, they completed a 10-min warm-up on an exercise bike (Lode BV, Groningen, The Netherlands), maintaining a power output of 100W at 75-85 revolutions per min. Next, an IMU (SHIMMER, Dublin, Ireland) was secured on the participant by a chartered physiotherapist at the spinous process of the 5th lumbar vertebra (Figure 3). The orientation and location of all the IMUs was consistent for all the study participants across all exercises.

A pilot study was used to determine an appropriate sampling rate and the ranges for the accelerometer and gyroscope on board the IMU. In the pilot study, squat, lunge, deadlift, single-leg squat, and tuck jump data were collected at 512 samples/s. A Fourier transform was then used to determine signal and noise characteristics of the signal that were all found to be less than 20 Hz. Therefore, a sampling rate of 51.2 samples/s was deemed appropriate for this study based upon the Shannon sampling theorem and the Nyquist criterion [34]. The Shimmer IMU was configured to stream tri-axial accelerometer (±16 g) and gyroscope (±500 ˚/s) data with the sensor ranges chosen based upon data from the pilot study. Each IMU was calibrated for these specific sensor ranges using the Shimmer 9DoF Calibration application.

After completion of their warm up, participants proceeded to do one set of 10 repetitions of bodyweight squats, barbell deadlifts at a load of 25kg, bodyweight lunges, and bodyweight single-leg squats (Figure 4). A chartered physiotherapist demonstrated the correct technique for each of the exercises. Participants familiarized themselves with each exercise, and their technique was assessed to be correct by the physiotherapist. Correct technique for squats, lunges, and deadlifts was defined using guidelines from the National Strength and Conditioning Association [35]. Single leg squats were completed according to the scoring criteria outlined by Whatman et al [36]. Finally, each participant completed the 10-second tuck jump test while attempting to maintain good form throughout [37].

Figure 3. Inertial measurement unit (IMU) position: the spinous process of the 5th lumbar vertebra.



XSL•FO
RenderX

Figure 4. The five exercises completed for this study: bodyweight squat (upper left), bodyweight lunge (upper middle), barbell deadlift (lower left), single leg squat (lower middle), and tuck jump (right).

Preparation for Transfer Learning

Based on the previous design for an IMU-based exercise classification system (Figure 1), with this new method the feature extraction step is not needed (Figure 2) as the CNN will take care of automatically both training the model and discovering the features by itself. The segmentation process is directly followed by the classification task (training and inference).

Convolutional Neutral Network (CNN) Infrastructure

Working with convolutional networks is not a trivial task. Fortunately, since the advent of deep learning in the last few years, a number of frameworks such as TensorFlow and Caffe have appeared in the market and are readily available for researchers. Most of these frameworks are open source, supported by large companies or universities, and provide not only helper libraries for numerical computation and machine learning but also a flexible architecture and the possibility to almost trivially use multiple central processing units (CPUs) and GPUs if available.

The authors used TensorFlow for the particular results provided in this paper, but any other framework or higher level library would suffice. Installing TensorFlow can be cumbersome, but Google provides a Docker container [38] with all the components to run TensorFlow out of the box. Documentation and scripts are also provided to retrain [39] networks and query [40] the new classifier. The aforementioned Docker container and scripts were used in this paper with minimal modifications.

The preprocessing and segmentation of inertial data to create the images that are fed into the CNN were prepared with MATLAB (2012, The MathWorks), as explained in the following section.

Data Preparation

Six signals were collected from the IMU; accelerometer *x, y, and z;* and gyroscope *x, y, and z*. Data were analyzed using MATLAB. To ensure the data analyzed applied to each participant's movement and to eliminate unwanted high-frequency noise, the six signals were low pass filtered at f_c=20 Hz using a Butterworth filter of order n=8.

The filtered signals were then programmatically segmented into epochs that relate to single, full repetitions of the completed exercises. Many algorithms are available to segment human motion during exercise. These include the sliding window algorithm, top-down, bottom-up algorithms, zero-velocity crossing algorithms, template-base matching methods, and combination algorithms of the above [4]. These algorithms all have advantages and disadvantages. For the purpose of the creation of a functioning exercise detection classifier, a simple peak-detection algorithm was used on the gyroscope signal with the largest amplitude for each exercise. The start and end points of each repetition were found by looking for the corresponding zero-crossing points of the gyroscope signal leading up to and following the location of a peak in the signal. Example results of the segmentation algorithm used on the gyroscope *x* signal, from an IMU positioned on the spine during 3 repetitions of the deadlift exercise, are provided (Figure 5).

XSL•FO
RenderX

Each extracted repetition of exercise data was resampled to a length of 250 samples. The six signals were then plotted using the MATLAB subplot function. The first subplot, gyroscope *x* (sagittal plane) was plotted between the y-axis range of ±250 °/s. Subplots 2 and 3, gyroscope *y* and *z* (frontal and transverse plane) were plotted between the y-axis range of ±100 °/s. Accelerometer *x* (subplot 4) was plotted in the y-axis range of ±3 m/s^2 and accelerometer *y* and *z* (subplots 5 and 6) were plotted in the range ±15 m/s^2. Axes labels and markers were programmatically hidden, and the blank space between each subplot was minimized. Following this, the graphs were saved as 470x470 JPEG files. Examples of the generated JPEG files are provided (Figure 6).

Retraining and Using the New Model

Transfer learning is the main technique used in this paper. This reuses an already trained CNN for classification purposes. In this case, the framework TensorFlow was used, which provides access to a model called "inception" trained on over 14 million images and also provides example scripts to retrain the network, that is, discarding the provided classifier and adjusting the values of the last layer of the network according to the new data provided. The retraining scripts expect to find the images in a particular folder (passed as a parameter) and layout (Figure 7), that is, a folder for each category that the new classifier will learn to identify, containing training pictures in jpg format. During training, the network will automatically identify the features to use to create the classifier.

There are a number of hyperparameters that can be changed depending on the new data used to retrain, such as the validation and training split of data to be used, the size of the batches to train on, or the learning rate applied (probably the most important of all for fine tuning and avoiding extra computation). The only parameter changed in this work was the number of steps, from a default 4000 iterations to 96,000 steps. This number provides high accuracy without showing signs of overfitting (see Results section).

The output of the training phase is simply two files, one with new weights (the retrained network) and a second file with labels for the data trained (the default names are retrained_graph.pb and retrained_labels.txt). These two files are all that is needed to predict results coming from new data. The classifier can be queried with the classify_image script mentioned previously.

Retraining and querying are actions that can be performed in a multitude of ways, with different frameworks and in different configurations. This work is about making things accessible and available. The Docker container for Tensorflow, with the documentation and helper scripts, was the simplest route the authors could find.

Figure 5. Detection of peak, start, and end points of exercise repetitions (neighboring zero crossing values to the peak locations).

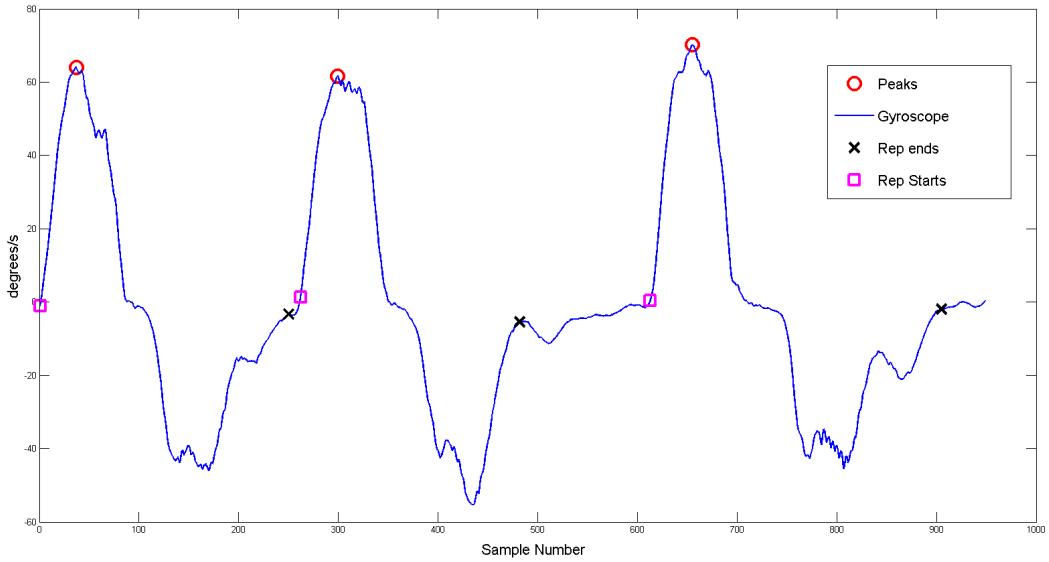
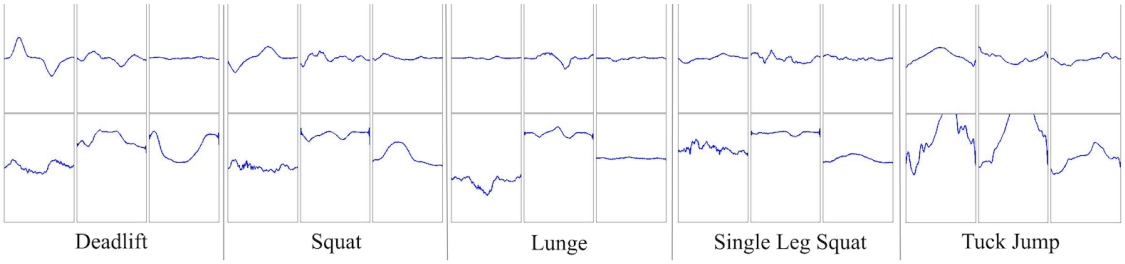
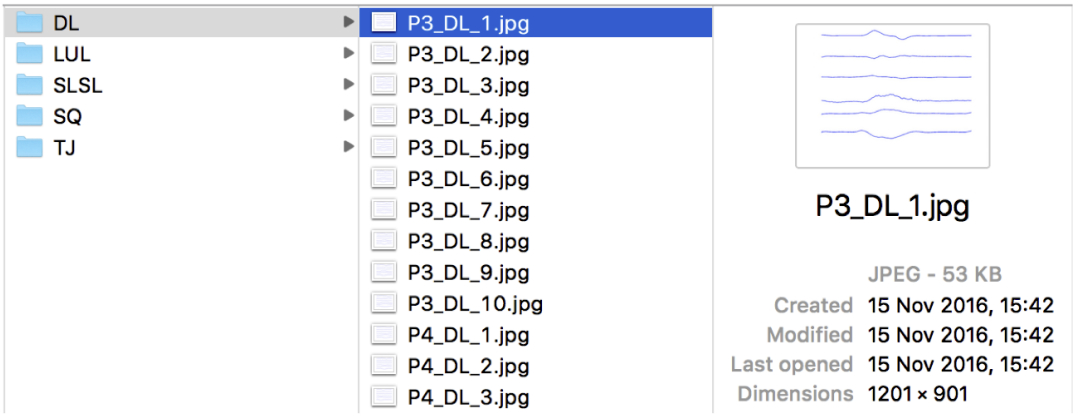


Figure 6. Samples of the generated plots (JPEG files) which were used as training and test data in this study.



XSL•FO
RenderX

Figure 7. Folders containing images for the five exercises (Bodyweight squat: SQ, bodyweight lunge: LUL, barbell deadlift: DL, single leg squat: SLSL, and tuck jump: TJ).



Results

As mentioned in the previous section, each training batch outputs training and validation accuracy and a cross entropy (loss function) amount, alongside with final validation accuracy. Rolled averages for those four values for training sessions of 96,000 steps are shown (Figure 8).

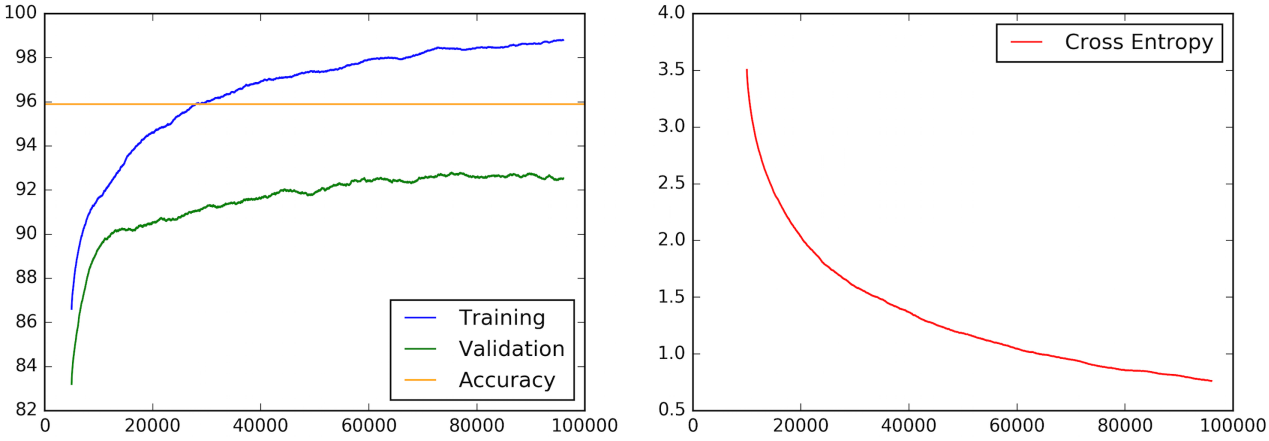
As observed, the cross entropy keeps falling steadily, and the average difference between training and testing is not very large, so overfitting is not an issue. Averaged over 5 runs of training, the final accuracy result was a 95.89% (3928/3991) for 96,000 steps. Figure 9 shows a confusion matrix for this method.

Figure 10 is an illustration of a misclassified plot. Part (a) of the image shows a typical lunge signal, whereas part (c) shows a typical single leg squat signal. Part (b) in the middle shows an example of a lunge repetition misclassified as a single leg squat. The issue seems to be concentrated in the top part of the image. The most likely reason for the odd lunge signal shape is that the subject may have looked over their shoulder or twisted for some reason during the repetition, and the final result is confusing the classifier, as it would confuse an expert looking directly at the plot.

These results are equivalent with a recently published method on the same dataset whereby the accuracy was found to be 94.1% [17]. Figure 11 shows the confusion matrix for this feature-based classification effort, and as it can be seen, the results are similar. However, leave-one-subject-out-cross-validation was used in this instance, so the results are not directly comparable.

The emphasis on this work, though, is in the ease of setup by using transfer learning and the need of only basic digital processing skills to prepare the data, when compared with other methods in this area.

Figure 8. Training (blue) and validation (green) accuracy during training phase, with final accuracy (orange) and cross entropy (red) for 96,000 steps.



XSL•FO
RenderX

Figure 9. Confusion matrix for the machine vision-based classification method.

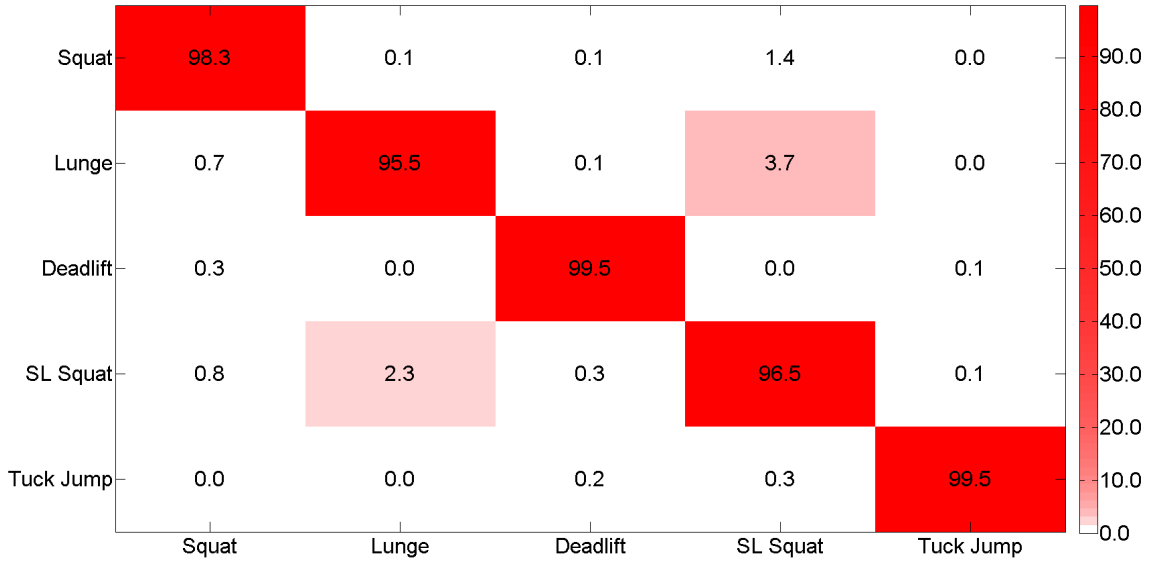


Figure 10. A lunge signal (a), a lunge signal misclassified as a single leg squat (b), and a single leg squat signal (c) for comparison.

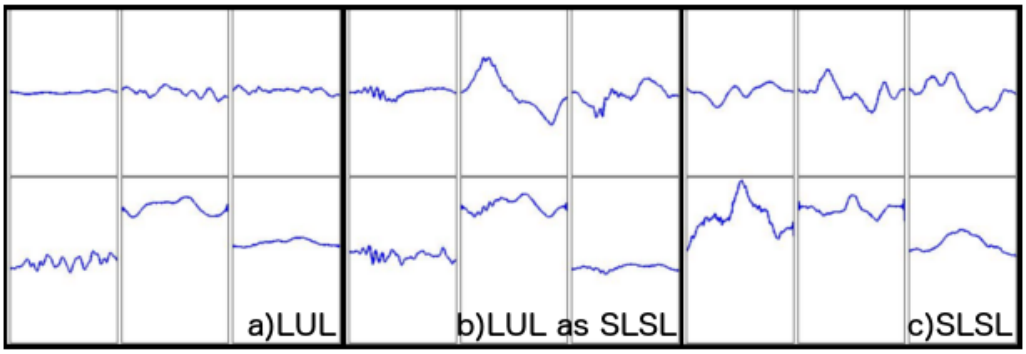
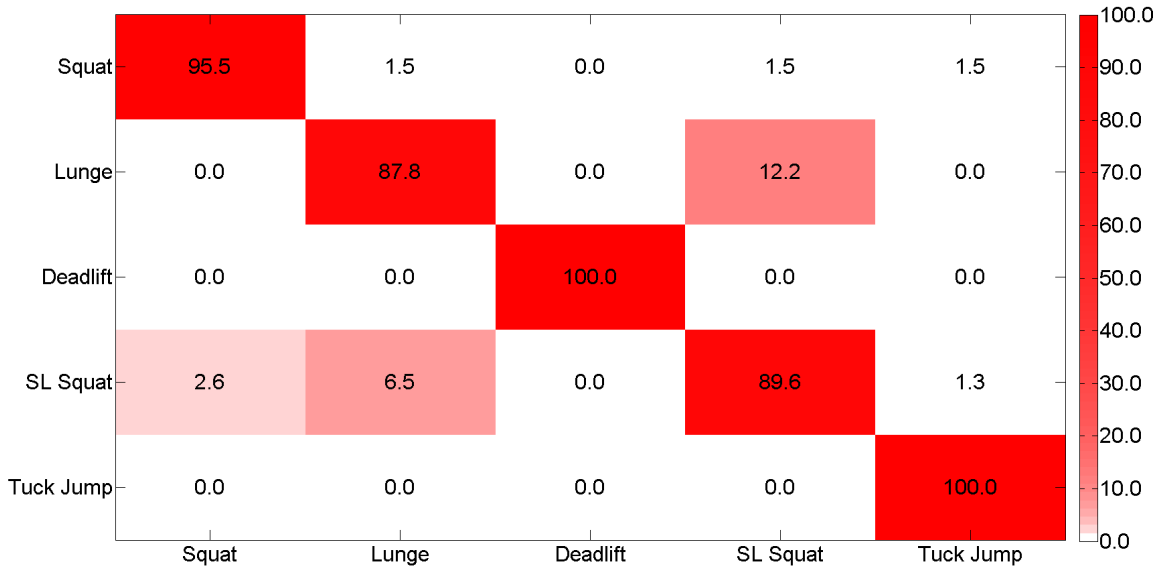


Figure 11. Confusion matrix for the feature based classification method.



XSL•FO
RenderX

Discussion

Principal Findings

An analysis of the data collected with the method proposed obtained an average 95.89% (3827/3991) classification accuracy, which is competitive with current state-of-the-art techniques. This high level of accuracy indicates that the distinctive waveforms in the plots for each of the exercises can be generalized among different participants, and the patterns created are appropriate for classification efforts. These results are coupled with the underlying recurrent theme for this work—to enable a more approachable entry path into the HAR and ED fields. To do so, high-level machine learning frameworks, coupled with a novel use of machine vision techniques, are used in two main ways: first, to avoid the complexity of manually crafted features only available through advanced DSP techniques, and second, to facilitate dimensionality reduction by allowing the CNN to take care of both feature extraction and classification tasks.

Comparison With Prior Work

The methodology employed and the results achieved in this paper can be directly compared with a recently published ED paper on the exact same dataset [17]. In this recently published work, identical filtering and segmentation methodologies were employed. However, a vast amount of additional signals and data processing were required to achieve classification with the lumbar worn IMU. As well as the 6 signals from the accelerometer and gyroscope used in this paper, 12 additional signals were used for classification. These were magnetometer *x*, *y*, and *z*, magnitude of acceleration, magnitude of rotational velocity, and the IMU's three-dimensional (3-D) orientation as represented by a rotation quaternion (W, X, Y, and Z) and Euler angles (pitch, roll, and yaw). Furthermore, 19 features were then computed from the segmented epochs of the 18 signals. These features were namely "mean," "RMS," "standard deviation," "kurtosis," "median," "skewness," "range," "variance," "max," "index of max," "min," "index of min," "energy," "25th percentile," "75th percentile," "level crossing rate," "fractal dimension," and the "variance of both the approximate and detailed wavelet coefficients using the Daubechies 4 mother wavelet to level 7." This resulted in a total of 342 features per exercise repetition. These features and their associated exercise label were used to evaluate and train a random forests classifier with 400 trees. Following leave-one-subject-out-cross-validation an accuracy result of 94.64% was achieved with this method. This recent work also demonstrated the laborious process of identifying the most important features for classification that can improve the efficiency of the reported technique used.

Although the accuracy result achieved in this recent work (94.64%) is slightly less than that presented in this paper (95.89%), the results should not be directly compared. This is because the additional signals used by O'Reilly et al [17] and the different method of cross-validation utilized in both studies to compute accuracy mean it is not a perfectly like-for-like comparison. However, it can be stated that similar levels of accuracy have been achieved with both methods. Most importantly, the ease of implementation of the classification method presented here greatly exceeds that presented by O'Reilly et al [17]. Most notably, the need to use additional signals and derive many features from them has been eliminated. This minimizes the signal processing and machine learning experience needed by the person investigating the possibility of creating a classifier. This is in line with the core objective of this paper.

Limitations

Simplicity was of utmost importance when designing this novel classification method for accelerometer and gyroscope data. Subsequently, maximal possible accuracy may not have been achieved. Utilizing a better understanding on how to parameterize the retraining effort and other techniques such as fine tuning (a method to reuse certain parts of a pretrained network instead of simply changing the last layer and classifier), could produce better results. A better understanding on how to deal with the type of data we are using could be beneficial. In general, machine vision work is plagued with issues such as partial occlusion, deformation, or viewpoint variation, which the data in this work does not suffer from. Due to that, and also to make the baseline of this work as simple as possible, no data augmentation or any kind of image processing techniques has been used. The results reported have been obtained only with resources from readily available frameworks, mostly on default settings.

It should also be noted that the presented method of classifying inertial sensor data with machine vision techniques has only been evaluated on exemplar samples of exercises that were conducted in a laboratory setting. Results are of high accuracy and competitive, with recent work on the same dataset [17] and therefore, act as a proof of concept for the method. However, the method has not yet been evaluated in classifying inertial sensor data arising from free-living activities and other HAR classification tasks. Future work should investigate the method's efficacy in such areas. Of key importance will be to simplify each application's preprocessing and segmentation of the inertial sensor data.

Conclusions

This paper has described a novel application approach for the classification of inertial sensor data in the context of HAR. There are two stand-out benefits of the machine vision approach described. The first is the ease of setting up the infrastructure for the CNNs involved through the use of transfer learning. The second is the reduction in the depth of digital signal processing expertise required on the part of the investigator. Due to the many difficulties in creating inertial sensor based activity recognition systems, the authors believe there is a need for a system development path which is easier to use for people who lack significant background in signal processing and machine learning. In particular, the new development pathway should eliminate the most difficult tasks conventionally identified with this area, that is, feature development or extraction and dimensionality reduction for the best machine learning method for each new application (Figure 1). The new development pathway, although eliminating these steps, does not compromise the attainment of high quality classification accuracy, sensitivity,

XSL•FO
RenderX

and specificity which is currently achieved through their successful implementation by appropriate experts (Figure 4). The exemplar study described here illustrates that the method is very competitive in comparison with customized solutions. Either way, the new pathway, at the very least, will allow for the easier testing of hypotheses relating to new inertial sensor-based activity classification systems, that is, is the classification possible at all based on the collected dataset? Ideally, it should also achieve equivalent.

Whereas the presented method does successfully eliminate the need for feature crafting and identification of optimal classification algorithms, it does not eliminate the process of signal preprocessing and signal segmentation before performing classification. Therefore, there remains some complexity in the process of achieving exercise classification when using the machine vision technique. However, the authors consider the process of filtering, segmenting, and plotting inertial sensor signals considerably less complex than identifying and computing strong features and an optimal classification method for the classification of inertial sensor data.

Future Work

Even though the current infrastructure used is readily available, certain skills such as familiarity with Docker or with Python data science stacks and basic DSP skills are still needed. The creation of a full package that could be installed on the researcher's machine could be an avenue to explore. Also the preprocessing and segmentation steps to prepare the data could be simplified by providing a set of scripts.

A number of professional machine vision companies exist in the market, and some provide online services that allow retraining of their custom models and could also be used for this type of work, avoiding the need for setting up the CNN infrastructure locally.

The availability of this technology on Android mobile devices is something that the authors are also pursuing. TensorFlow may provide some initial support in this area. Finally, although this paper emphasizes the lack of a necessity to present features other than the basic time series, it is clear that augmentation with derived features presents further opportunities for performance tweaking. For researchers more comfortable with such feature development, this application avenue is worth exploring.

Acknowledgments

The Insight Centre for Data Analytics is supported by Science Foundation Ireland under grant number SFI/12/RC/2289. This project was also partly funded by the Irish Research Council as part of a Postgraduate Enterprise Partnership Scheme with Shimmer (EPSPG/2013/574).

Conflicts of Interest

None declared.

References

1. Arxiv. 2016 Mar 16. TensorFlow: large-scale machine learning on heterogeneous distributed systems URL: https://arxiv.org/abs/1603.04467 [accessed 2017-07-25] [WebCite Cache ID 6sDHItHpX]
2. Perez A, Labrador M, Barbeau S. G-Sense: a scalable architecture for global sensing and monitoring. IEEE Network 2010 Jul;24(4):57-64. [doi: 10.1109/MNET.2010.5510920]
3. Lara OD, Labrador MA. A survey on human activity recognition using wearable sensors. IEEE Commun Surv Tutorials 2013;15(3):1192-1209. [doi: 10.1109/SURV.2012.110112.00192]
4. Whelan DF, O'Reilly MA, Ward TE, Delahunt E, Caulfield B. Technology in rehabilitation: evaluating the single leg squat exercise with wearable inertial measurement units. Methods Inf Med 2017 Mar 23;56(2):88-94. [doi: 10.3414/ME16-02-0002] [Medline: 27782290]
5. Preece SJ, Goulermas JY, Kenney LP, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors--a review of classification techniques. Physiol Meas 2009 Apr;30(4):R1-33. [doi: 10.1088/0967-3334/30/4/R01] [Medline: 19342767]
6. Stoppa M, Chiolerio A. Wearable electronics and smart textiles: a critical review. Sensors 2014;14(7):11957-11992. [doi: 10.3390/s140711957]
7. Kim E, Helal S, Cook D. Human activity recognition and pattern discovery. IEEE Pervasive Comput 2010;9(1):48 [FREE Full text] [doi: 10.1109/MPRV.2010.7] [Medline: 21258659]
8. Sugawara E, Nikaido H. Properties of AdeABC and AdeIJK efflux systems of Acinetobacter baumannii compared with those of the AcrAB-TolC system of Escherichia coli. Antimicrob Agents Chemother 2014 Dec;58(12):7250-7257 [FREE Full text] [doi: 10.1128/AAC.03728-14] [Medline: 25246403]
9. Moreno-Torres JG, Saez JA, Herrera F. Study on the impact of partition-induced dataset shift on k-fold cross-validation. IEEE Trans Neural Netw Learn Syst 2012 Aug;23(8):1304-1312. [doi: 10.1109/TNNLS.2012.2199516] [Medline: 24807526]
10. Razavi HA, Kurfess TR. Detection of wheel and workpiece contact/release in reciprocating surface grinding. J Manuf Sci Eng 2003;125(2):394. [doi: 10.1115/1.1559160]

XSL•FO
RenderX

11. Flaxman AD, Vahdatpour A, Green S, James SL, Murray CJ, Population Health Metrics Research Consortium (PHMRC). Random forests for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards. Popul Health Metr 2011 Aug 04;9:29 [FREE Full text] [doi: 10.1186/1478-7954-9-29] [Medline: 21816105]
12. Singh KP, Basant N, Gupta S. Support vector machines in water quality management. Anal Chim Acta 2011 Oct 10;703(2):152-162. [doi: 10.1016/j.aca.2011.07.027] [Medline: 21889629]
13. Dudani SA. The distance-weighted k-nearest-neighbor rule. IEEE Trans Syst, Man, Cybern 1976 Apr;SMC-6(4):325-327. [doi: 10.1109/TSMC.1976.5408784]
14. Bishop CM. Pattern Recognition and Machine Learning. New York, NY: Springer; 2006.
15. Giggins O, Kelly D, Caulfield B. Evaluating rehabilitation exercise performance using a single inertial measurement unit. 2013 Presented at: Pervasive Computing Technologies for Healthcare; May 5-8, 2013; Venice, Italy p. 49-56. [doi: 10.4108/icst.pervasivehealth.2013.252061]
16. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 2012;9:21 [FREE Full text] [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
17. O'Reilly M, Whelan D, Ward T, Delahunt E, Caulfield B. Technology in S&C: tracking lower limb exercises with wearable sensors. J Strength Cond Res 2017 Feb 15:- Epub ahead of print. [doi: 10.1519/JSC.0000000000001852] [Medline: 28234711]
18. Bassett SF. The assessment of patient adherence to physiotherapy. NZ J Physiother 2003;31(2):60-66.
19. Nyan MN, Tay FE, Seah KH, Sitoh YY. Classification of gait patterns in the time-frequency domain. J Biomech 2006;39(14):2647-2656. [doi: 10.1016/j.jbiomech.2005.08.014] [Medline: 16212968]
20. Tao W, Liu T, Zheng R, Feng H. Gait analysis using wearable sensors. Sensors 2012;12(2):2255-2283. [doi: 10.3390/s120202255]
21. Mo L, Li F, Zhu Y, Huang A. Human physical activity recognition based on computer vision with deep learning model. 2016 Presented at: Instrumentation and Measurement Technology Conference; May 23-26, 2016; Taipei, Taiwan p. 1-6. [doi: 10.1109/I2MTC.2016.7520541]
22. Stefic D, Patras I. Action recognition using saliency learned from recorded human gaze. Image Vis Comput 2016 Aug;52:195-205. [doi: 10.1016/j.imavis.2016.06.006]
23. Ma M, Fan H, Kitani KM. Going deeper into first-person activity recognition. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 27-30, 2016; Seattle, WA p. 1894-1903. [doi: 10.1109/CVPR.2016.209]
24. Zeng M, Nguyen LT, Yu B, Mengshoel OJ, Zhu J, Wu P, et al. Convolutional neural networks for human activity recognition using mobile sensors. 2014 Presented at: 6th International Conference on Mobile Computing, Applications and Services; November 6-7, 2014; Austin, Texas p. 197-205. [doi: 10.4108/icst.mobicase.2014.257786]
25. Yang, J., Nguyen, M.N., San, P.P., Li, X. and Krishnaswamy, S., Yang JB, Nguyen MN, San PP, Li XL, Krishnaswamy S. Deep convolutional neural networks on multichannel time series for human activity recognition. 2015 Presented at: Proceedings of the 24th International Conference on Artificial Intelligence; July 25-31, 2015; Buenos Aires, Argentina p. 3995-4001.
26. Radu V, Lane ND, Bhattacharya S, Mascolo C, Marina M, Kawsar F. Towards multimodal deep learning for activity recognition on mobile devices. 2016 Presented at: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; September 12-16, 2016; Heidelberg, Germany p. 185-188. [doi: 10.1145/2968219.2971461]
27. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
28. Cruz-Roa AA, Arevalo OJ, Madabhushi A, González OF. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. Med Image Comput Comput Assist Interv 2013;16(Pt 2):403-410. [Medline: 24579166]
29. Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S. Arxiv. 2014. CNN features off-the-shelf: an astounding baseline for recognition URL: https://arxiv.org/abs/1403.6382 [accessed 2017-07-25] [WebCite Cache ID 6sDKO67Go]
30. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? 2014 Presented at: Proceedings of the 27th International Conference on Neural Information Processing Systems; December 08-13, 2014; Montreal, Canada p. 3320-3328.
31. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: convolutional architecture for fast feature embedding. 2014 Presented at: Proceedings of the 22nd ACM international conference on Multimedia; November 03-07, 2014; Orlando, FL p. 675-678. [doi: 10.1109/72.279181]
32. Deng J, Dong W, Socher R, Li LJ, Li K, Li FF. ImageNet: a large-scale hierarchical image database. 2009 Presented at: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009); June 20-25, 2009; Miami, FL. [doi: 10.1109/CVPR.2009.5206848]
33. Tensorflow. TensofFlow image retraining example URL: https://www.tensorflow.org/how_tos/image_retraining/ [accessed 2017-02-07] [WebCite Cache ID 6o6I1SMkw]
34. Jerri A. The Shannon sampling theorem - its various extensions and applications: a tutorial review. Proc IEEE 1977 Nov;65(11):1565-1596. [doi: 10.1109/PROC.1977.10771]

XSL•FO
RenderX

35. Earle RW, Baechle TR. NSCA's Essentials of Personal Training. Champaign, IL: Human Kinetics; 2004.
36. Whatman C, Hing W, Hume P. Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests. Phys Ther Sport 2012 May;13(2):87-96. [doi: 10.1016/j.ptsp.2011.07.001] [Medline: 22498149]
37. Myer GD, Ford KR, Hewett TE. Tuck jump assessment for reducing anterior cruciate ligament injury risk. Athl Ther Today 2008 Sep 01;13(5):39-44 [FREE Full text] [Medline: 19936042]
38. Docker. Tensorflow docker conatiner URL: https://hub.docker.com/r/tensorflow/tensorflow/ [accessed 2017-02-07] [WebCite Cache ID 6o6I63rVr]
39. Github. Image retraining script URL: https://github.com/tensorflow/tensorflow/blob/master/tensorflow/examples/image_retraining/retrain.py [accessed 2017-02-07] [WebCite Cache ID 6o6IAPV4K]
40. Github. Image classification script URL: https://github.com/tensorflow/models/blob/master/tutorials/image/imagenet/classify_image.py [accessed 2017-02-07] [WebCite Cache ID 6o6IE43KZ]

Abbreviations

3-D: three-dimensional
CNN: convolutional neural network
CPU: central processing unit
DSP: digital signal processing
ED: exercise detection
GPU: graphics processing unit
HAR: human activity recognition
IMU: inertial measurement unit
SD: standard deviation

Edited by G Eysenbach; submitted 16.02.17; peer-reviewed by L Mo, G Norman; comments to author 30.03.17; revised version received 10.05.17; accepted 31.05.17; published 04.08.17.

Please cite as:
Dominguez Veiga JJ, O'Reilly M, Whelan D, Caulfield B, Ward TE
Feature-Free Activity Classification of Inertial Sensor Data With Machine Vision Techniques: Method, Development, and Evaluation
JMIR Mhealth Uhealth 2017;5(8):e115
URL: http://mhealth.jmir.org/2017/8/e115/
doi:10.2196/mhealth.7521
PMID:28778851

©Jose Juan Dominguez Veiga, Martin O'Reilly, Darragh Whelan, Brian Caulfield, Tomas E Ward. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 04.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mobile Phone Detection of Semantic Location and Its Relationship to Depression and Anxiety

Sohrab Saeb[1,2], PhD; Emily G Lattie[1], PhD; Konrad P Kording[2], PhD; David C Mohr[1], MD, PhD

[1]Center for Behavioral Intervention Technologies (CBITs), Department of Preventive Medicine, Northwestern University, Chicago, IL, United States
[2]Rehabilitation Institute of Chicago, Department of Physical Medicine and Rehabilitation, Northwestern University, Chicago, IL, United States

Corresponding Author:
Sohrab Saeb, PhD
Center for Behavioral Intervention Technologies (CBITs)
Department of Preventive Medicine
Northwestern University
10th Fl.
750 N Lake Shore Dr.
Chicago, IL, 60611
United States
Phone: 1 3125034626
Fax: 1 3129089588
Email: s-saeb@northwestern.edu

Abstract

Background: Is someone at home, at their friend's place, at a restaurant, or enjoying the outdoors? Knowing the semantic location of an individual matters for delivering medical interventions, recommendations, and other context-aware services. This knowledge is particularly useful in mental health care for monitoring relevant behavioral indicators to improve treatment delivery. Local search-and-discovery services such as Foursquare can be used to detect semantic locations based on the global positioning system (GPS) coordinates, but GPS alone is often inaccurate. Mobile phones can also sense other signals (such as movement, light, and sound), and the use of these signals promises to lead to a better estimation of an individual's semantic location.

Objective: We aimed to examine the ability of mobile phone sensors to estimate semantic locations, and to evaluate the relationship between semantic location visit patterns and depression and anxiety.

Methods: A total of 208 participants across the United States were asked to log the type of locations they visited daily, using their mobile phones for a period of 6 weeks, while their phone sensor data was recorded. Using the sensor data and Foursquare queries based on GPS coordinates, we trained models to predict these logged locations, and evaluated their prediction accuracy on participants that models had not seen during training. We also evaluated the relationship between the amount of time spent in each semantic location and depression and anxiety assessed at baseline, in the middle, and at the end of the study.

Results: While Foursquare queries detected true semantic locations with an average area under the curve (AUC) of 0.62, using phone sensor data alone increased the AUC to 0.84. When we used Foursquare and sensor data together, the AUC further increased to 0.88. We found some significant relationships between the time spent in certain locations and depression and anxiety, although these relationships were not consistent.

Conclusions: The accuracy of location services such as Foursquare can significantly benefit from using phone sensor data. However, our results suggest that the nature of the places people visit explains only a small part of the variation in their anxiety and depression symptoms.

(JMIR Mhealth Uhealth 2017;5(8):e112) doi:10.2196/mhealth.7297

KEYWORDS

semantic location; geographic positioning systems; mobile phone; classification; decision tree ensembles; extreme gradient boosting; depression; anxiety

XSL•FO
RenderX

Introduction

Passive and unobtrusive detection of the physical location of individuals has been made possible over the years by embedding global positioning system (GPS) systems into commonly used devices, such as mobile phones. Physical location alone is usually not very useful for understanding human activity, or the motivations that underlie that activity. In contrast to physical location, semantic location carries additional information about the meaning of the location [1]. For example, semantic location might tell us if a location is a home, place of work, dining establishment, or place of worship, thereby infusing the geographic location with human relevance.

A growing number of papers have shown that a variety of location features, measured by GPS, can detect mental health problems such as depression [2-6], bipolar disorder [7], and social anxiety [8]. It is unclear at this point why these GPS location features may be related to depression or anxiety. It may be that the nature of the places and the meaning inherent in different locations affect how we feel. Previous research has shown that there is a relationship between mood and certain activities, such as religious practice [9], participating in social activity [10], and spending excess sedentary time at home [11]. Improving the ability to detect locations affiliated with these activities could offer not just a greater understanding of the behavioral and environmental contributors to depression and anxiety, but also unique methods for prompting just-in-time adaptive interventions (JITAIs) using mobile technologies. This approach could add value beyond that gathered using other JITAI triggers (eg, self-reported difficulties, GPS location, and electro-cardiogram signals), and may enable us to determine if a person with a history of depression is relapsing, or if a person is about to have a panic attack [12].

Local search-and-discovery services, such as Foursquare, can estimate semantic locations based on GPS coordinates and the data they have globally collected from populated areas in the world. When these services are embedded in a mobile app using an application programming interface (API), they can passively provide location-specific information for the locations that users visit. Since its launch in 2009, Foursquare has been used in research applications to accomplish diverse tasks, ranging from the analysis of individuals' food and drink habits across cultures [13] to the examination of the popularity of venues and identifying factors contributing to venue popularity [14]. Foursquare has tapped into a new model of location-based advertising such that users can be notified of businesses in their immediate vicinity, and can receive benefits such as discounts and coupons for "checking in" to these businesses.

However, asking search-and-discovery services such as Foursquare about semantic locations, based on a given GPS coordinates, has limitations. First, GPS can be inaccurate, and particularly in denser urban environments, variability in GPS may lead to the detection of false locations. For example, one might be at a restaurant within a shopping mall, and the search-and-discovery service may classify the person as at a shop rather than a restaurant. Second, although these services can detect "residential" locations, they cannot distinguish a person's home from another home they are visiting. These limitations prevent such services from being a reliable source of information, especially for behavioral sensing and intervention, where it is crucial to know exactly when a person is at home, work, a friend's home, or other locations.

In addition to GPS, mobile phones can sense many more variables in the environment, such as light, sound, and Wi-Fi signals. Using a mobile phone, we can also determine what type of physical activity an individual is performing, how much time they spend in a location, and how they interact with their phones. Semantic locations may have distinct signatures, such as the length of time a person spends at a location, time of the day and day of the week that they visit, type of activities that they perform, and the sound and light conditions in the environment. These features may help us to determine if the place is home, a grocery store, place of worship, or a library. As an obvious example, a place that a person spends time over night is most likely home, and a bright place visited during the day, with intermittent walks and stops, is likely a store. Therefore, detection of semantic locations using mobile phone sensors seems feasible.

The aim of this paper was first to develop methods for improving mobile phone-based detection of semantic locations by incorporating sensors beyond the simple GPS. We developed methods for detecting semantic locations, and compared their accuracy to that of Foursquare. While improving semantic location detection is worthwhile and could further serve clinical and consumer-driven purposes, our second aim was to explore the relationship between semantic location detection and depression and anxiety. We specifically investigated the relationship between semantic location visits and the severity of depression and anxiety symptoms, as well as the differences between individuals with and without those symptoms.

Methods

Participant Recruitment

We recruited participants between October 28, 2015 and February 12, 2016. The recruitment was done in collaboration with Focus Pointe Global (FPG), a company that specializes in market and scientific research strategies and participant recruitment and retention [15]. FPG maintains a panel of 1.5 million potential participants from the general population. For our study, FPG sent out emails to potential participants with links to the screener questionnaire. Additionally, FPG used phone calls to contact potential participants from their in-house registries.

Interested individuals from the general population of the United States contacted FPG and were screened for eligibility using a brief questionnaire. Individuals were eligible for our study if they were at least 18 years old, able to read and understand English, owned a mobile phone with Android 4.4 through 5.1, and had access to Wi-Fi for at least one 3-hour period per day. We excluded individuals who indicated on self-report that they were diagnosed with any psychotic disorders, were unable to walk more than half a mile (4 city blocks), or had positive screens for alcohol abuse (Alcohol Use Disorders Identification

XSL•FO
RenderX

Test [16] score >16), drug abuse (Drug Abuse Screening Test-10 [17] score >6), suicidal ideation (Beck Depression Inventory-II [18] item 9 rating >2), or bipolar disorder (Mood Disorder Questionnaire [19] question 1 score 7, an endorsement of question 2, and a response of 2 or 3 for question 3). We also excluded individuals who shared their phone with others.

Depressive symptoms were measured using the Patient Health Questionnaire, 9-item (PHQ-9) [20]. On the PHQ-9, participants are prompted to indicate how frequently they have experienced specific symptoms over the past two weeks, such as "feeling down, depressed, or hopeless" and "feeling tired or having little energy". Participants respond on a four-point Likert-type scale, ranging from 0 indicating "not at all" to 4 indicating "nearly every day." PHQ-9 scores range between 0-27. We also used the cut-off point of 10 to divide participants into those who screened positive for depression (PHQ-9 >10; termed *depressed* in this paper) and those screened negative (PHQ-9 <10; termed *nondepressed*). This cut-off point has been shown to maximize the sum of sensitivity and specificity for depression diagnosis [20].

For anxiety assessment, we used the Generalized Anxiety Disorder, 7-item (GAD-7) [21]. The GAD-7 is structured similarly to the PHQ-9, and participants are prompted to indicate how frequently they have experienced symptoms such as, "feeling nervous, anxious, or on edge" and, "being so restless that it's hard to sit still" over the past two weeks on the same four-point Likert-type scale. GAD-7 scores range between 0-21. We used the cut-off point of 10 to separate those participants who screened positive for GAD (GAD-7 >10; termed *anxious* in this paper) from those who screened negative (GAD-7 <10; termed *nonanxious*). At this cut-off point, the sum of sensitivity and specificity is maximized [21].

We wanted to have a wide range of depression and anxiety symptoms in our sample, and therefore we selected roughly equal numbers of participants in four groups, based on their screening assessments: depressed and anxious, depressed and nonanxious, nondepressed and anxious, and nondepressed and nonanxious. In addition to assessment at baseline, we also assessed each participant's depression and anxiety at week 3 and week 6.

Participant Enrollment

Eligible participants were consented using procedures approved by the Northwestern University Institutional Review Board. Consenting was done using a website: participants were directed to a webpage that contained information about the study procedures, benefits, and potential risks. Specifically, participants were informed about the sensor data that were going to be collected from their mobile phones, the types of questions that would be asked throughout the study, and the procedures undertaken to protect their private information. After digitally signing the consent form, participants were enrolled in our study.

Each participant was enrolled for a period of 6 weeks. First, a study identification (ID) number was assigned to the participant by FPG. Participants were then asked to complete an online questionnaire regarding their demographic information, which consisted of their age, gender, race, and ethnicity, along with their US state of residence, and information about various aspect of their lives that could impact movement patterns (eg, health difficulties, number of jobs, and job locations). Finally, participants downloaded two apps: *Purple Robot* [22], which collected sensor data from their phones; and *EMA app*, which asked them questions about the places they visited. Participants were compensated between US $25 and $270.40 depending on how long they stayed in the study and how many of the daily questionnaires they answered.

Mobile Phone Data Collection

After participants were enrolled, we started collecting two categories of data from their mobile phones: (1) sensor data, which contained data from the physical sensors as well as software services such as phone and short message service (SMS) communications; and (2) ecological momentary assessment (EMA) data, which consisted of daily questions that showed up on participants' phones asking them about the locations they visited throughout the day.

The phone sensor data were captured using the *Purple Robot* [22] app. Purple Robot is a multi-purpose, open-source Android app that we have developed for passive collection of mobile phone sensor data [3]. This app gathers data from the sensors and services available on the phone, including light, sound, GPS, accelerometer, phone and SMS communications, screen, and Wi-Fi. The app initially stores sensor data on the device, and then transmits them as network connectivity becomes available. This strategy allows us to collect data in a variety of wireless connectivity scenarios with the confidence that intermittent network access does not affect the nature, quality, or quantity of the collected data.

For the collection of EMA data, we used a second Android app, *EMA app*, which asked participants questions about the locations they visited throughout the day. The app was specifically developed for this study. Each evening, the app analyzed the GPS data collected over the previous 24 hours. The EMA app first clustered the GPS data using an adaptive *k*-means clustering method [3], considering a maximum radius of 100 meters for each cluster, and then removed the clusters that the user visited for a duration of less than 10 minutes. This second step removed clusters that were not actual locations, but were generated because the user was moving slowly (eg, they were stuck in the traffic). After detecting the visited locations, the EMA app provided the participant with a map identifying each location, the time they were at the location, and asked the following questions: "What is the name of this place?" and, "What kind of place is this?"

What is the Name of This Place?

A list of likely location names was provided to the user to choose from. This list was obtained from the Foursquare location API. The participant could also enter their own location name if it was not provided.

What Kind of Place is This?

This list was adapted from Foursquare venue categories, and included Arts & Entertainment, Food, Nightlife Spot, Outdoors & Recreation, Professional or Medical Office, Spiritual, Shop

XSL•FO
RenderX

or Store, Travel or Transport, and Home. In addition, we added Work, Another's Home, and Other. If the participant answered Other, they were asked to enter the location type. The EMA app saved the cluster center corresponding to each detected location, the visit times, and the participant's answers to the questions regarding that location. Table 1 lists the location categories we used in the EMA app, and how they matched Foursquare's high-level location categories.

Table 1. Location category labels reported by our study participants (left) and their corresponding high-level Foursquare location categories.

EMA app Location Category	Foursquare Location Category
Nightlife Spot (Bar, Club)	Nightlife Spot
Outdoors & Recreation	Outdoors & Recreation
Arts & Entertainment (Theater, Music Venue, Etc.)	Arts & Entertainment
Professional or Medical Office	Professional & Other Places
Food (Restaurant, Cafe)	Food
Home	Residence
Shop or Store	Shop & Service
Travel or Transport (Airport, Bus Stop, Train Station, Etc.)	Travel & Transport
Work	-
Another's Home	-

Purple Robot and EMA app anonymized any sensitive information before storage and transmission. Specifically, the apps used an MD5 hashing algorithm [23] to anonymize the study participant identifiers. Once the data was anonymized, it was transmitted to the data collection server, and the local copy was deleted from the device. The data residing on the server could be linked with other information gathered during the study only if the unique identifiers used by the participants and the study-specific keys used to encrypt the data were known.

Foursquare Evaluation

We wanted to assess how well Foursquare could predict the type of locations that users reported daily. To do so, we used the Foursquare wrapper library [24] in Python, and queried the type of location for each location that participants visited. These queries used 4 parameters: latitude, longitude, database version date, and limit. For latitude and longitude, we used the GPS coordinates of the visited location that was saved by the EMA app. For the database version date, we used the current date at the time of the query, which was 2016/8/10, so that we had the latest version of the data. The limit parameter indicated the number of guesses, for which we used 1, so that it returned the best match. We performed these queries for each of the visited locations recorded by EMA app.

Foursquare's response to our queries was in JavaScript object notation (JSON) format [25], and contained the place ID, name, contact information, address, distance from the queried coordinates, and the location category. From this information, we only saved location category and distance.

The location category returned by the Foursquare website was too specific, being as detailed as "Cambodian Restaurant" or "College Math Building". Since we did not need this level of detail in our study, we used Foursquare's Category Hierarchy [26] to translate these low-level categories into high-level ones. This category hierarchy can be obtained in JSON format using the HTTPS query detailed in Textbox 1. The response contains the whole category hierarchy.

Textbox 1. HTTPS query for category hierarchy.

https://api.foursquare.com/v2/venues/categories?oauth_token=<TOKEN>&v=<VERSION>

Where TOKEN can be obtained from Foursquare's developers' website, and VERSION is the database version date in YYYYMMDD format

After querying the Foursquare category for each location cluster, we compared it to the category reported by the participant, and calculated the accuracy (see section: Classifier Evaluation). We skipped locations reported as Work, Another's Home, or Spiritual for this comparison, since these did not exist in Foursquare categories. The calculated accuracy gave us the performance of Foursquare in predicting semantic locations.

Detecting Semantic Location from Phone Sensor Data

Sensor Features

To classify semantic locations from phone sensor data, we first calculated their features. These features were extracted from all sensor data that were gathered during a visit to a location. In this way, for every location visit, we obtained one feature vector. This vector consisted of 45 features, which will be described in the following sections.

XSL•FO
RenderX

Light Features

Light features were calculated from light intensity, in *lux*, sampled by the light sensor at 10 Hz. This sampling frequency could vary from device to device, so light features were designed such that they did not depend on the sampling frequency. These features consisted of basic statistics including mean, variance, skewness, and kurtosis. In addition, we calculated the percentage of time the light sensor output was zero, and the number of times that it crossed its mean value in 1 second.

Sound Features

Sound features captured different aspects of the sound in the environment. Specifically, we sampled the audio using the phone's microphone every 5 minutes, each time for 15 seconds. From each 15 second audio recording, we extracted the power and the dominant frequency. Power was calculated as described in Figure 1.

To calculate the dominant frequency, we obtained the amplitude of the fast Fourier transform of the audio signal, and found the frequency that maximized the amplitude.

Screen Features

We used screen activity to measure the amount of participants' interaction with their phones. We calculated the number of times the screen state transitioned from *OFF* to *ON*, as well as the average and the standard deviation (SD) of the duration that the screen was *ON* each time.

Activity Features

We used the physical activity states provided by the Android Activity Recognition API. We sampled this API every 10 seconds. The Physical Activity API uses the accelerometer sensor to detect the following physical activities: *Still*, *Walking*, *Running*, *Tilting*, *On Bike*, *In Vehicle*, *Unknown*. We calculated the percentage of time that the participant was in *Still*, *Tilting*, *Walking*, and *Unknown* states. In addition, we calculated the percentage of transitions for a number of state transitions that we expected to be informative about the type of location the participant was visiting. These transitions included *Still* to *Walking*, *Still* to *Tilting*, *Still* to *Unknown*, and *Walking* to *Unknown*.

Communication Features

Communication features consisted of the total number of incoming, outgoing, and missed phone calls. In addition, we derived the number of incoming and outgoing SMS text messages.

GPS Features

These features were calculated from the latitude and longitude values provided by the GPS sensor, sampled every 5 minutes. GPS features included average latitude, average longitude, and *location variance* defined as the equation in Figure 2.

In addition to these features, by filtering out the data points that were outside the 50-meter radius of a location's average latitude and longitude during a visit, we approximated the *visit frequency* to that location, and the *mean time interval* between the visits.

Wi-Fi Features

We sampled the current access point's media access control address and the number of available Wi-Fi networks every 5 minutes. We only used the number of Wi-Fi networks as a feature.

Time Features

We calculated the visit duration, the *timespan* of the visit, the visit mid-time in hour, and the day of the week at the start and the end time of the visit. Visit duration was defined as the total time a participant spent at a location on a given day, while visit timespan was the time from when they entered that location first on a given day to the time they left it on the same day.

Weather Features

We obtained the weather conditions at the location and time of visits. For this data, we used the Weather Underground service [27]. For each detected location, we queried Weather Underground for the history of weather data in that location, which returned those data for the past year from the date of query. The responses were in JSON format, with each entry corresponding to one weather report. We searched for the report that was closest to the time the user visited that location, and used the temperature, dew point, and weather condition as features.

Classifier Architecture

We wanted to see how successfully we could detect semantic locations, reported by the participants, using the sensor features that were passively collected from their mobile phones. For this classification problem, we used ensembles of decision trees with the gradient boosting optimization method [28], also known as extreme gradient boost (XGBoost). These classifiers have been shown to outperform other classification methods in high-dimensional machine learning problems [29]. In this study, we particularly chose XGBoost because these classifiers perform well when the dimensionality of the data relative to the number of samples is large [30], and that they can deal with missing values.

A decision tree, shown in Figure 3, determines the class of a feature vector by making sequential, individual decisions on the elements of that vector. Each decision is made at a *node*, where the value of one feature is compared to a threshold value. The node has two outgoing branches that reach next-level nodes. Depending on whether the feature value is larger or smaller than the threshold, one of the branches is chosen. One branch is also designated to the condition where the feature value is missing.

Each decision tree in the ensemble is assigned to one class, and provides a *prediction score* at its leaf node (Figure 3, boxes) for the class it belongs to. The ensemble's prediction score for each class is calculated by summing over the prediction scores of all trees in that class, as detailed in Figure 4.

The final class probabilities are calculated as a softmax function of the predictions scores using the equation shown in Figure 5.

Therefore, for each given feature vector, the ensemble provides a probability distribution over the classes.

XSL•FO
RenderX

Figure 1. Sound power calculation; where S(n) is the sound amplitude (dB) at sample n, and N is the total number of samples.

$$P = \frac{1}{N}\sum_{n=1}^{N} S(n)^2$$

Figure 2. Location variance feature; calculated as the logarithm of the sum of variances in latitude and longitude values.

$$LV = \log(\sigma_{lat}^2 + \sigma_{long}^2)$$

Figure 3. An example of a single decision tree in the ensemble of decision trees. Each circle is a tree node, where a decision is made by comparing a feature value fxx to a threshold. For a given feature vector, depending on which path is taken, a single prediction score is generated, shown in the boxes. Note that one of the outgoing branches of each node is also dedicated to the situation where the data is missing.

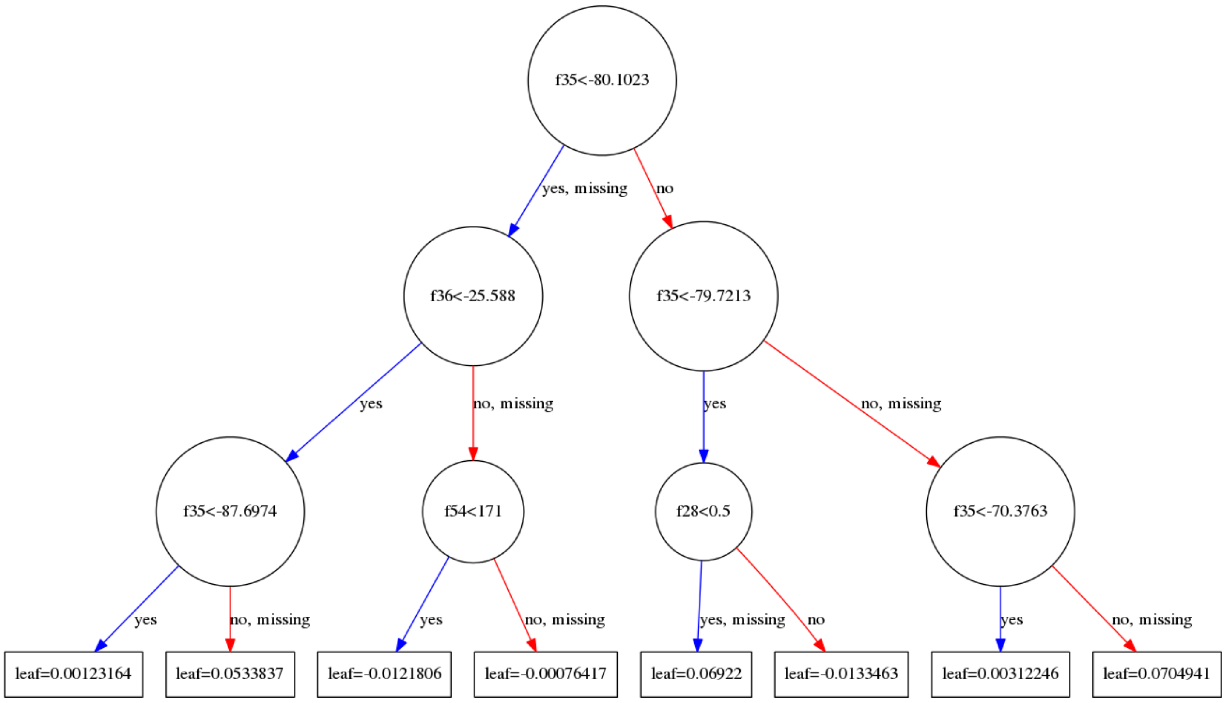


Figure 4. Aggregation of prediction scores made by individual trees; where gk,m represents the decision tree k in class m, that maps a feature vector x to a prediction score gk,m(x), and ym is the ensemble prediction score for class m. K is the total number of trees for each class.

$$y_m = \sum_{k=1}^{K} g_{k,m}(x)$$

Classifier Training

The goal of training is to push the class probabilities p_m (Figure 5) as close as possible to the true classes in the training data. However, we also wanted to avoid overfitting to the training data. Therefore, our training objective should also prevent the model from becoming too complex. Accounting for these two objectives, the XGBoost optimization algorithm uses the cost function explained in Figure 6.

While the logistic loss term in Figure 6 (leftmost term) penalizes the discrepancy between the ensemble's prediction and the ground truth, the rest of the terms prevent trees from overfitting by penalizing the number of nodes (T) as well as the magnitude of their prediction scores (y).

Figure 5. Class probability calculation; where pm represents the probability for class m, ym is the ensemble prediction score for class m, and M is the number of classes.

$$p_m = \frac{e^{y_m}}{\sum_{n=1}^{M} e^{y_n}}$$

XSL•FO
RenderX

Figure 6. General form of the cost function; where l(yi,yi*) is the logarithmic loss [31] between the prediction scores yi and the true prediction scores yi*, T is the number of nodes in a tree, N is the number of training samples, γ, λ, and α are constants, and ‖ . ‖ 1 and ‖ . ‖ 2 are L1 and L2 norms, respectively.

$$L = \sum_{i=1}^{N} l(y_i, y_i^*) + \gamma T + \frac{1}{2}\lambda\|\boldsymbol{y}\|_2 + \alpha\|\boldsymbol{y}\|_1$$

Figure 7. Cost function for training a new tree added at iteration t; where gt(.) is the prediction score provided by the new tree. See Figure 6 caption for more details on the parameters.

$$L^{(t)} = \sum_{i=1}^{N} l\left(y_i^{(t-1)} + g_t(x_i), y_i^*\right) + \gamma T + \frac{1}{2}\lambda\|\boldsymbol{y}\|_2 + \alpha\|\boldsymbol{y}\|_1$$

In the gradient boosting method, trees are added to the ensemble one by one. The ensemble starts with one tree, which is fit to the training data using the cost function in Figure 6. At each iteration, a new tree is added to the ensemble such that it fits to the residual error of the existing trees on the training data. Concisely, the new tree complements the existing trees such that, at iteration *t*, the cost function in Figure7 is minimized.

The parameters of the new tree are chosen such that $L^{(t)}$ is minimized. In this way, the ensemble gradually fits to the training data. To find out when to stop adding new trees to the ensemble, we calculated the cross-validation error within the training dataset at each iteration. As the number of trees increase, this error decreases. However, after a certain point, the error starts to increase due to overfitting. We stopped adding new trees at that point, and evaluated the resulting classifier on the test set (see Classifier Evaluation).

Hyperparameter Tuning

We tuned the hyperparameters of the XGBoost classifier by grid search, and used data from 10% of participants. Within this subset of data, we performed a 10-fold cross-validation to estimate the area under the curve (AUC; see Classifier Evaluation). We chose the set of parameters on the grid that maximized this AUC.

The parameters included in hyperparameter tuning were γ, L1 regularization weight (α), L2 regularization weight (λ), learning rate, maximum tree depth, subsampling fraction (*r*), and feature subsampling fraction (*s*). Subsampling fraction, $r \in (0,1)$, determines the fraction of training data samples that are seen by each tree during training, while features subsampling fraction, $s \in (0,1)$, is the fraction of features that are seen by each tree node. After finding the optimal value of these hyperparameters, we trained and evaluated classifiers on the whole dataset.

Classifier Evaluation

Our goal was to create algorithms that could determine the semantic locations for unseen individuals, so we trained and evaluated the classifiers using a subject-wise cross-validation scheme. Specifically, we randomly selected 70% of the subjects to train the classifier, and used the remaining 30% to evaluate its prediction accuracy. We repeated this procedure 100 times. The distribution of prediction errors on held-out participants used as *test* provides an unbiased estimate of the prediction error of the algorithm for the population from which our dataset is sampled [32]. Therefore, we could tell how well our classifier would generalize to new, unseen individuals.

To calculate the prediction error in each round of cross-validation, we estimated the receiver operating characteristic curve, and calculated the AUC. The AUC ranges between 0 and 1, with 0.5 indicating chance level performance. The advantage of using AUC is that it is robust to the imbalance in the number of samples in the classes. Therefore, by iterating over all participants as *test*, we obtained a good estimate of the classifier's accuracy.

Relationship Between Semantic Location and Depression and Anxiety

We evaluated the relationship between the amount of time participants spent at each semantic location and their level of depressive and anxious symptoms, measured by PHQ-9 and GAD-7, respectively. We performed two analyses. First, we calculated Pearson's correlation between the scores and the time spent in each location, across all participants. For the second analysis, we divided participants into depressed and nondepressed, as well as anxious and nonanxious, based on their scores. For depression, we defined the two groups by considering participants who consistently had PHQ-9 <10 (termed nondepressed) or PHQ-9 >10 (termed depressed) across all three assessment time points. Likewise, for anxiety, we defined the two groups by considering participants who consistently had GAD-7 <10 (termed nonanxious) or GAD-7 >10 (termed anxious). Therefore, in both analyses, we excluded the participants who crossed the PHQ-9=10 or GAD-7=10 thresholds. The main reason was that these participants could not be reliably classified. Furthermore, if we had included them, it would have added two additional categories (those who improved and those who got worse), which would have reduced power. It is also unclear how we would interpret any relationships with participants transitioning from one clinical state to another. After dividing subjects into these groups, we compared the duration of time that participants spent at each semantic location between the groups, using two-sample *t*-tests.

Results

Participant Statistics

A total of 208 individuals passed the eligibility criteria for participating in our study, and were recruited. One participant did not install the software on their phone, and another had invalid GPS data. These two participants were removed from all analyses. Of the remaining 206 participants, 22 (10.7%) stopped providing data before the end of the 6-week period. However, many continued to send data after the end of 6 weeks, with 27 (13.1%) providing more than 60 days of data.

XSL•FO
RenderX

The 206 participants included in the analyses were 170 females (82.5%) and 36 males (17.5%). Participants' ages ranged between 18 and 66 years, with a mean of 39.3 (SD 10.3). The participants' locations were diverse, covering most of the populated states and major cities in the United States. Most of these locations (86.8%, 178/206) were in "mostly urban" areas, as defined by the United States Census Bureau [33], while 12.1% (25/206) were in "mostly rural" areas. The rural or urban condition for the location of the remaining 3 participants could not be determined. The average depression score (PHQ-9) was 9.72 (SD 5.10), and the anxiety score (GAD-7) was 9.01 (SD 5.41). These values show that our participants had a wide distribution of depression and anxiety symptoms.

In response to a question on employment status, 61.2% (126/206) indicated that they were employed, 20.9% (43/206) were unemployed, 8.3% (17/206) had a disability which prevented them from working, and 1.9% (4/206) were retired. Sixteen participants (7.8%, 16/206) did not specify their employment status. Of the 126 employed participants, 98 (77.8%) had one, 23 (18.3%) had two, 4 (3.2%) had three, and one (0.8%) had four jobs. In addition, of these 126 participants, 36 (28.6%) worked in more than one location.

Semantic Location Self-Reports

The semantic locations reported by the participants were diverse. While most participants reported the predefined locations in Purple Robot, as the example in Figure 8 A shows for one participant, many participants defined their own semantic locations by selecting "Other" and typing in their desired semantic location name. The total number of distinct location types reported by all participants was 370; however, only a small fraction of these locations was consistently reported by most participants (Figure 8 B). Therefore, apart from a few categories which need to be considered for future studies (eg, School and Library), most of the visited locations were among the locations that we had considered in the initial design of our mobile app.

Classifier Hyperparameters

The optimized hyperparameters for the XGBoost classifier were the following: for sensor-only classification, we set the number of trees to 200, the fraction of samples seen by each tree to 0.2, and the fraction of features to 0.5. For classification based on both sensor and Foursquare features, these three parameters were set to 300, 0.25, and 0.2, respectively. In both scenarios, we set γ=0.4, λ=1, α=0, the maximum depth of decision trees to 4, and learning rate to 0.025. Given these parameter values, our training procedure was substantially regularized.

Predicting Semantic Location

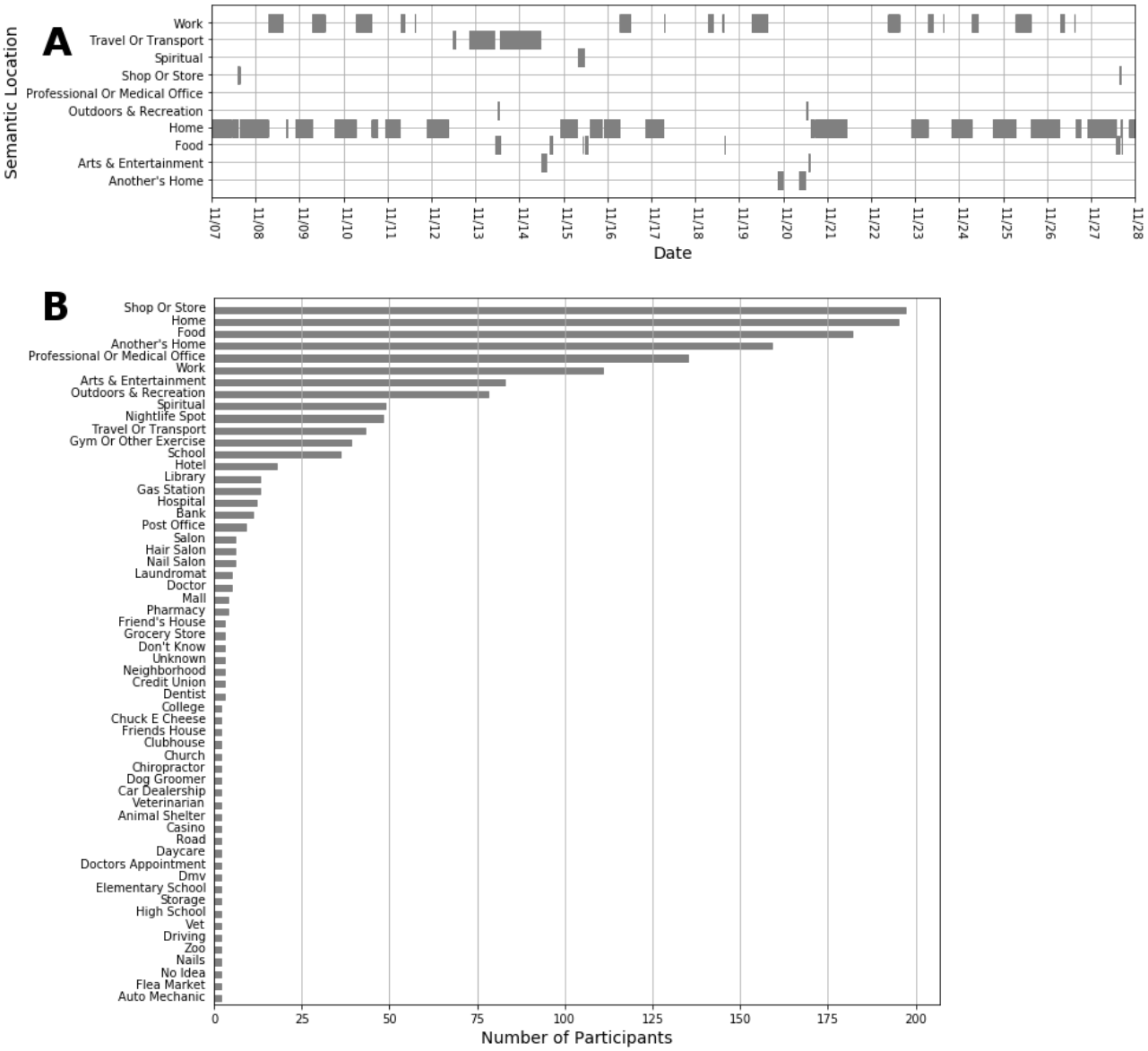
We first measured how accurately Foursquare could detect the semantic locations reported by participants. To obtain the locations detected by Foursquare, we used the GPS coordinates of that location, and queried Foursquare about its closest match to that location. We then compared the results to the locations reported by participants, and calculated the AUC for each category. The results are shown in the left column of Table 2. While Foursquare could detect Shop or Store with an average AUC 0.76, its AUC for Home was close to the chance level. Foursquare did not have location categories equivalent to Work, Another's Home, or Spiritual, and therefore the AUCs for these categories could not be calculated. On average, the accuracy of Foursquare in detecting 8 semantic locations was approximately 0.62.

Table 2. Mean area under the curve (AUC) in detecting each location category, using Foursquare only, mobile phone sensor features only, and both. Note that we could not use Foursquare to detect Work, Another's Home, or Spiritual locations; hence there are no results.

Semantic Location	Foursquare	Sensor	Sensor+Foursquare
Travel or Transport, mean (CI)	0.54 (0.49-0.60)	0.79 (0.72-0.86)	0.84 (0.78-0.91)
Nightlife Spot, mean (CI)	0.61 (0.53-0.72)	0.87 (0.78-0.94)	0.89 (0.79-0.95)
Spiritual, mean (CI)	N/A	0.82 (0.75-0.88)	0.87 (0.80-0.92)
Outdoors & Recreation, mean (CI)	0.59 (0.53-0.64)	0.81 (0.71-0.88)	0.86 (0.75-0.92)
Arts & Entertainment, mean (CI)	0.67 (0.61-0.73)	0.88 (0.85-0.91)	0.92 (0.88-0.95)
Work, mean (CI)	N/A	0.86 (0.82-0.90)	0.87 (0.83-0.91)
Professional or Medical Office, mean (CI)	0.65 (0.58-0.73)	0.85 (0.80-0.91)	0.88 (0.83-0.93)
Another's Home, mean (CI)	N/A	0.77 (0.69-0.82)	0.83 (0.75-0.89)
Food, mean (CI)	0.64 (0.59-0.68)	0.79 (0.74-0.83)	0.83 (0.78-0.87)
Home, mean (CI)	0.53 (0.51-0.56)	0.96 (0.95-0.97)	0.96 (0.95-0.97)
Shop or Store, mean (CI)	0.76 (0.73-0.79)	0.86 (0.82-0.90)	0.89 (0.85-0.92)
Mean AUC	0.62	0.84	0.88

XSL•FO
RenderX

Figure 8. (A) Location report data from one example participant, collected between 11/07/2015 and 11/28/2015. Each rectangle shows the period of time the participant has been in a specific location. The sensor data during that time period is used to create a feature vector, which is then used to detect that semantic location. (B) Top locations visited by all participants, sorted by how many participants visited them. As the total number of unique reported locations was 370, we only included the ones that had been visited by at least two participants.



We wanted to determine whether mobile phone sensors alone could detect the semantic location of participants. We used 45 features that were extracted from a variety of sensors during the time that the participant was visiting a location (see section: Sensor Features). We trained the XGBoost classifiers to map these features to semantic locations, and tested these classifiers on participants that they had not seen during training. Compared to Foursquare, the AUC of detecting certain locations was considerably higher (Table 2, middle column). Specifically, using the sensors instead of Foursquare yielded AUCs that were on average more than 20% greater (Table 2, middle column). This increase was mostly evident for Home, Nightlife Spot, and Travel or Transport categories. Overall, the average AUC for all semantic locations increased to 0.84. Therefore, not only could we use phone sensors alone to detect semantic locations, but their performance was considerably better than Foursquare.

Next, we used both Foursquare and phone sensor data to see if this approach could further increase the accuracy of our classifiers. To this end, we added two extra features to the 45 features that we previously used for training the classifiers: the Foursquare location type, which was represented by a binary vector with 9 elements (each corresponding to one category); and the distance to the nearest Foursquare location. Therefore, the total number of features increased to 55. Using this new feature set further increased the average AUC to 0.88 (Table 2, right column). This increase was mostly evident in detecting Food, Shop or Store, Art & Entertainment, and Spiritual categories. Therefore, augmenting mobile phone sensor features with Foursquare data made our classifiers better at detecting semantic locations.

XSL•FO
RenderX

Finally, we asked which features contributed the most to detecting semantic locations by estimating their *importance*. To obtain feature importance for each feature, we removed that feature from the training data and calculated the resulting change in the cross-validated AUC. The results are shown in Figure 9, with features sorted by their importance. While features such as Visit Timespan, Location Variance, Latitude, Number of Wi-Fi Networks, Visit Duration, and Visit Frequency have the highest importance, several features have close to zero or negative importance, meaning that their removal does not affect (or even slightly improves) the performance of the classifiers. These features include some of the sensor features as well as Foursquare features. However, one should note that each of these effects are generated by removing only one feature from the feature set, and the collective effect of removing multiple features might be different. Nevertheless, it seems that most sensor and Foursquare features are useful in distinguishing semantic locations.

Relationship Between Semantic Location and Depression and Anxiety

We evaluated the relationship between the time spent at different semantic locations and the level of depression and anxiety symptoms, measured by PHQ-9 and GAD-7, respectively. First, we evaluated the linear correlation between these two groups of variables (Table 3). When considering individual correlations, some were statistically significant ($P<.05$). Notably, the duration of time spent at Spiritual locations is negatively correlated with depression and anxiety scores, for 3 of 6 assessments. When we consider the total number of 66 comparisons between all semantic locations and depression and anxiety scores, we cannot rule out the possibility that these significant correlations are generated by chance. However, because these calculations are not independent, conservative corrections (such as a Bonferroni correction) may not be appropriate [34].

Table 3. Linear correlation coefficients (*r*) between time spent at semantic locations and depression (PHQ-9) and anxiety (GAD-7) scores. Values show the median of 1000 bootstrap estimates of *r*. Italicized values indicate coefficients that are significantly (*P*<.05) different from zero. However, by correcting for multiple comparisons (66 comparisons here) we cannot rule out the possibility that these correlations are a result of chance.

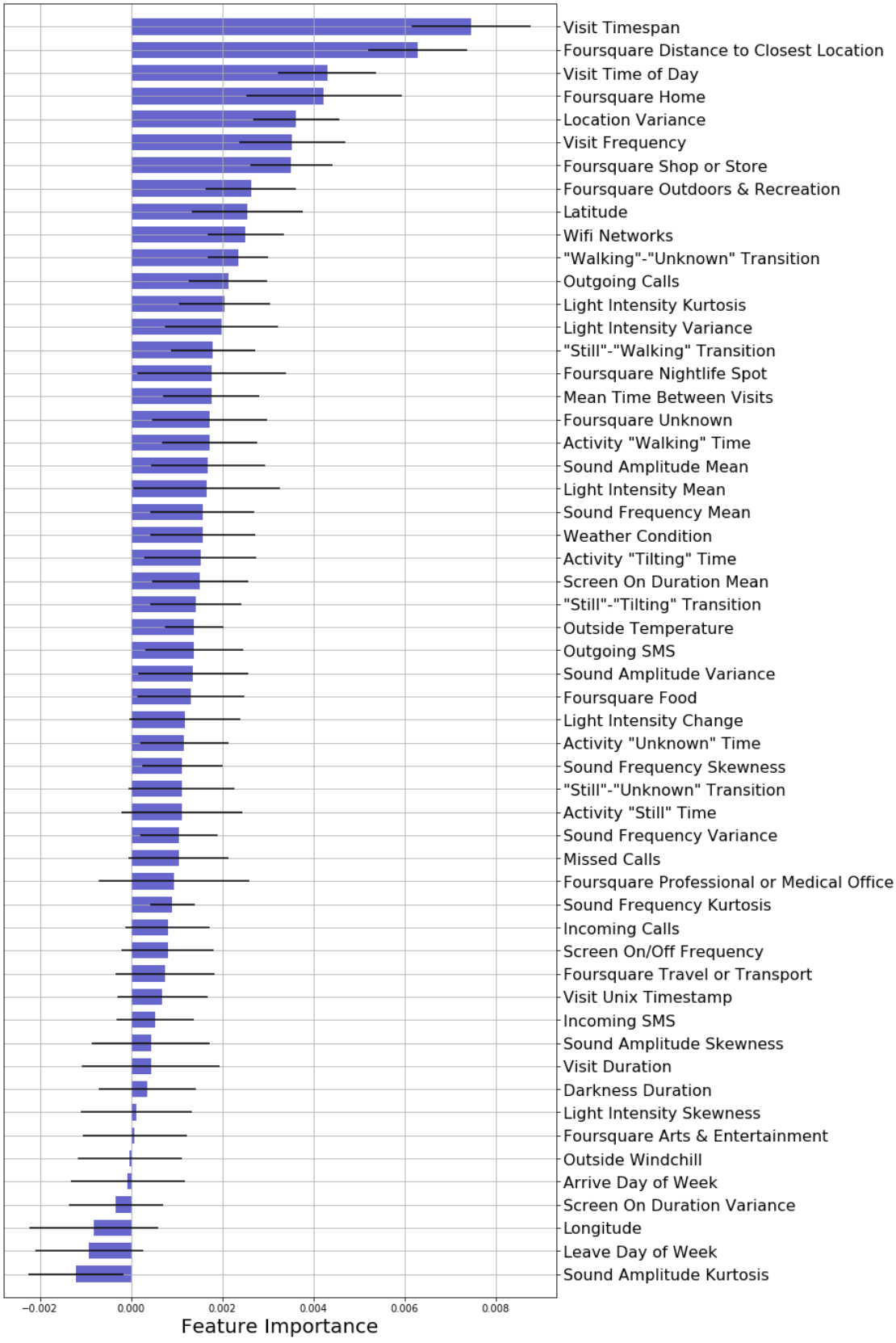
	PHQ-9 Week 0	PHQ-9 Week 3	PHQ-9 Week 6	GAD-7 Week 0	GAD-7 Week 3	GAD-7 Week 6
Home	0.057	0.073	0.089	0.083	0.101	0.097
Shop or Store	-0.010	0.0183	-0.020	0.001	-0.030	-0.038
Work	-0.084	-0.139	-0.140	-0.083	*-0.176*	-0.085
Food	-0.088	-0.093	*-0.152*	-0.089	-0.086	-0.115
Another's Home	0.046	-0.065	-0.064	-0.016	-0.003	0.000
Professional or Medical Office	0.029	*0.096*	0.049	-0.069	0.019	0.051
Outdoors & Recreation	0.016	-0.123	-0.101	-0.065	-0.131	-0.109
Arts & Entertainment	*-0.172*	-0.092	-0.090	-0.044	-0.055	-0.057
Travel or Transport	-0.070	-0.037	*-0.113*	0.082	0.012	*-0.088*
Spiritual	-0.041	-0.078	*-0.147*	-0.094	*-0.143*	*-0.168*
Nightlife Spot	*-0.126*	*-0.173*	-0.045	0.041	-0.063	-0.045

We also performed a group difference analysis, by dividing the participants into two groups (once based on their depression scores, and another time based on their anxiety scores). We compared the duration of time participants spent at each semantic location between these groups. For depression, the nondepressed group consisted of 51 participants and the depressed group consisted of 68 participants. The remaining 88 participants crossed the PHQ-9=10 threshold between the assessments, and were excluded from this analysis because they could not be clearly classified. For anxiety, the nonanxious group consisted of 51 participants while the anxious group consisted of 61 individuals. The remaining 96 participants crossed the GAD-7=10 threshold and were excluded.

The results for depression are shown in Figure 10A. While the depressed and nondepressed groups seemed to have different distributions of time spent across locations, these differences were significant ($P<.05$) only for two locations: the nondepressed group spent significantly more time at Work, while the depressed group had more time spent at a Professional or Medical Office. For the anxious versus nonanxious comparison (Figure 10B), the difference was only significant for the Spiritual category, with the nonanxious group spending more time in this location category, on average. Therefore, it seems that time spent at semantic locations contains some information about depression and anxiety, but these findings are not consistent.

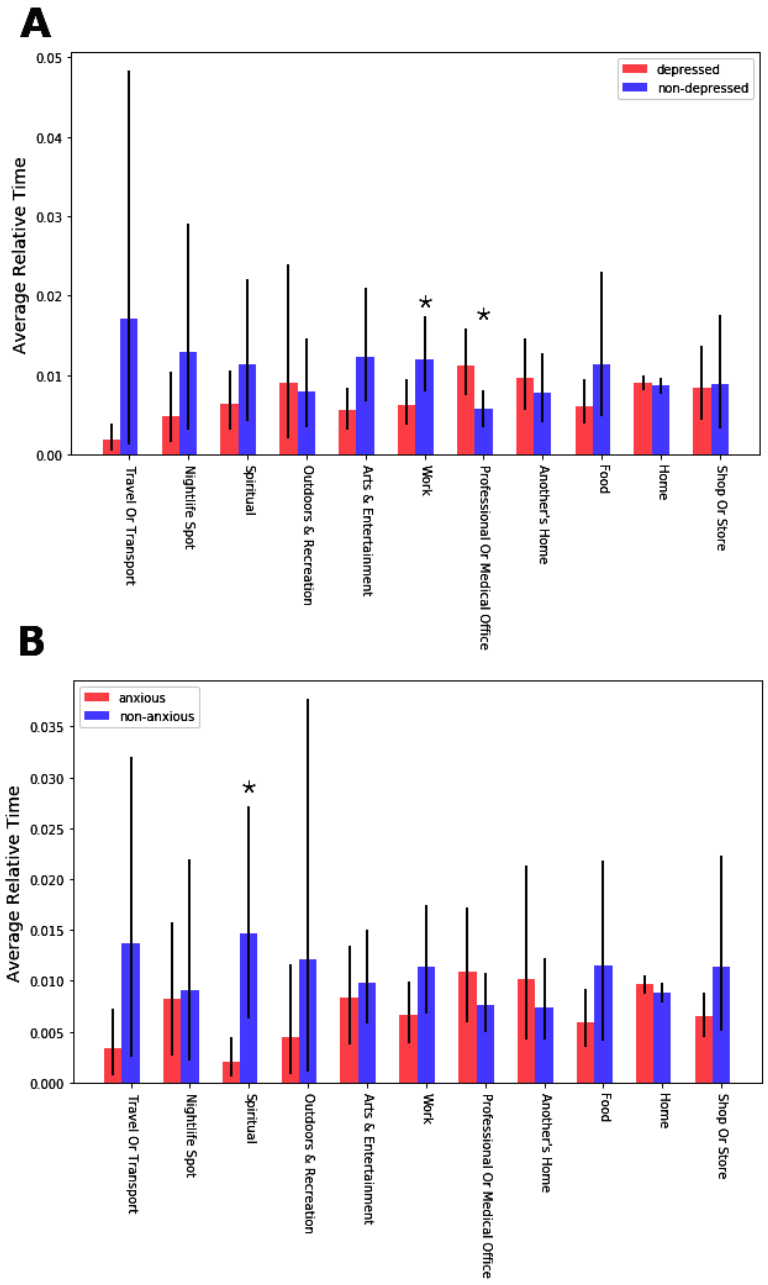
XSL•FO
RenderX

Figure 9. Mobile phone sensor feature importance in detecting semantic locations. Features are sorted based on their importance, from top to bottom. The importance of each feature is calculated by computing the decrease in the cross-validated area under the curve when that feature is removed from the feature set. Negative values indicate an increase in performance. Each value is the mean feature importance across cross-validation folds, and error bars show the standard error of the mean.



XSL•FO
RenderX

Figure 10. The relationship between semantic location visit duration, depression, and anxiety. Each bar shows the average time spent at each location by (A) depressed versus nondepressed and (B) anxious versus non-anxious groups, relative to the total time spent by all participants in that location. Error bars show 95% CIs. Both mean and CIs are obtained by bootstrapping over 1000 iterations. Stars indicate significant difference between the means, obtained using a 2-sample t-test at the P<.05 level. However, adjusting for multiple comparisons, these differences are all nonsignificant.



Discussion

Principal Results

In this paper, we were able to detect the type of locations that individuals visited, using data passively collected from their mobile phones. The phone sensor data were especially crucial in detecting these semantic locations. Sensor features alone produced accuracies that were more than 20% greater than those reported by Foursquare, and combining the sensor features with Foursquare produced even greater accuracy. This result is not surprising since detecting semantic location based on GPS alone is not necessarily accurate, especially in urban areas [35], and can lead to detecting nearby locations instead of the actual location. Sensors, which are available on most mobile phones, can provide valuable information about the type of locations phone users visit, and can significantly improve the accuracy of these services.

The performance of the classifiers considerably varied across the location types. While Home could be detected with an AUC of above 0.95, the classification AUC for Another's Home and Food was 0.83. This variability may have multiple causes. First, visits to certain locations, such as home or work, are more regular in time, which makes them easier to detect based on the time of visit. Another cause might be that some semantic locations such as Travel or Transport were less represented in the data, since participants visited those locations less often.

XSL•FO
RenderX

This factor has likely made it difficult for classifiers to find the feature patterns that are distinct indicators of those locations. Finally, while some locations (eh, Home) have a clear definition, participants may have been confused about which location type to report for some other locations. For example, a participant might have had food in a store, and have reported that location as either "Food" or "Shop or Store". Overall, although classification performance varied across different semantic locations, it significantly benefited from incorporating mobile phone sensor data.

While we could detect the types of locations, we found only few significant relationships between the amount of time spent in those locations and self-reported symptoms of depression and anxiety. Furthermore, these few relationships were weak and inconsistent. This failure may have multiple explanations. First, our categorization of semantic locations was based largely on Foursquare categories, which was not developed with mental health or wellness in mind and may not be accurate, useful, or relevant to mental health. These categories were also often imprecise (eg, "Professional or Medical Office"). For mental health research, we may need to create location categories that are mostly relevant to the factors that influence mental health.

Second, the lack of a consistent relationship between semantic location and depression or anxiety may reflect larger problems in the literature. Past research has examined smaller, discrete samples of participants, such as university students [2,4,8,36] or residents of the same city [3,37]. This study sample was geographically diverse, with a broader sample of the American population. This diversity in location enriched our dataset by including people from rural and urban areas, and different climates, cultures, and lifestyles. While this diversity helped us to obtain a better estimate of the accuracy of location detection in real-world applications, it may also reflect problems with increasing dimensionality, as this area of research moves towards more generalizable samples.

It is possible that this finding is accurate: that the kinds of places we go is *not* related to our level of depression or anxiety. This theory would suggest that the relationship between movement through geographic space and depression or anxiety [2-4,37] may be related to some other aspect of mobility patterns. For example, it may be that depression or anxiety is more related to the processes of getting to various locations, such as physical activity [10,38,39], than the actual locations themselves. Furthermore, low motivation in depressed individuals may decrease the likelihood of moving from a commonly visited location (such as home or work) and a less frequently visited place (such as a store or movie theater), but may have very little to do with moving from a less frequently visited place to home or work [40].

Limitations

There are a number of limitations that need to be mentioned. First, when detecting semantic locations we did not consider the transitions between locations. Knowing the transition probabilities can be useful; for example, it may be more likely to visit Home after Shop or Store. One reason for not considering transitions was that we only considered the top 11 most-visited locations for the classification problem, and therefore the sequence of semantic locations in the training data were not necessarily consecutive in time. Another reason was the existence of gaps in the data, which caused further separation between consecutive visits. Incorporating transition probabilities in detecting semantic locations, when possible, will likely increase the classification accuracy of the resulting algorithms.

Second, semantic locations may have signatures that we failed to capture through our phone sensors. For example, the type of phone apps people use, or individuals who they contact, can be a good distinguishing feature between locations. Using such sources of information as features in future studies may improve the performance of semantic location detection.

Third, our study participants differed from the general population in a few aspects. Approximately 83% of the participants were women, significantly different from 50.8% in the general population of the United States [41]. Furthermore, nearly 21% of the participants were unemployed, compared to the nationwide estimate of 5% unemployment [42]. Finally, we only included individuals who owned smartphones, while approximately 28% of Americans do not own such phones [43]. In addition, our inclusion of people with only Android phones excluded 41% of smartphone users who use phones with other operating systems [44]. Census data shows that owing a smartphone is associated with certain demographic variables such as age, education, and income [43]. Therefore, our inclusion criteria might have affected the study sample.

Fourth, the assessment of depression and anxiety in this study was based on self-report, and therefore may not generalize to assessments based on diagnostic interview. A clinical diagnosis usually involves an in-depth interview and consideration of confounding factors, based on the criteria in the Diagnostic and Statistical Manual of Mental Disorders [45,46]. In our study, the assessment was solely based on Web-based self-reported PHQ-9 and GAD-7 scores, and therefore our study sample may be different from a clinical sample. It is likely that we would find a stronger relationship between mental health state and the type of visited locations in a clinical sample, compared to what we found in this study. Nevertheless, electronic assessment of depression has been used and validated by many previous studies [47,48].

Fifth, data collection took place from late October to early February, and thus most participants were providing data during the winter holiday season. While the geographic diversity of the sample allows us to account for variations in weather (eg, participants from Florida experienced a much different climate than those in Minnesota), we recognize that holiday-related travel, such as spending time at other family members' homes, and holiday-related time away from work presents a departure from an individual's typical behavior. The holiday season may have served as a confounder, as participants may have been engaged in activities not representative of how they would behave during other times of the year. Furthermore, the 6-week study period may not have been long enough to detect changes or meaningful relationships between behavioral patterns and mood. Ultimately, we aim to develop models to ascertain the relative components of these factors. However, as this is a relatively new field of inquiry, the timing and length of this

XSL•FO
RenderX

study protocol may have interfered with our ability to detect true signals.

Conclusions

In conclusion, mobile phone sensors promise considerably more accurate estimations of individuals' daily life behaviors. In this study, we have shown that semantic location (the type of locations that people visit) can be detected using a combination of phone sensors and a mapping service such as Foursquare. We performed this study in a sample that was diverse in terms of geographic location, climate, education, employment, and lifestyle. However, there were no consistent relationships between the time spent at different locations and depression or anxiety. Future research should focus on those semantic locations that are more likely to be relevant to depression or anxiety. In addition, longer studies that extend across seasons, and larger studies that are more adequately powered to manage the level of dimensionality in human subject data, will be better positioned to investigate the relationships between semantic locations and mental health. The advancement of mobile phone technology will facilitate the design of these future studies.

Acknowledgments

This study was supported by the following National Institute of Health grants: 5R01NS063399, P20MH090318, and R01MH100482. The authors would like to thank Weather Underground for providing access to weather history data.

Conflicts of Interest

None declared.

References

1. Hu H, Lee D. IEEE International Conference on Mobile Data Management.: IEEE; 2004. Semantic location modeling for location navigation in mobile environment URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.10.9272&rep=rep1&type=pdf [accessed 2017-07-23] [WebCite Cache ID 6sB8gG7pI]
2. Farhan A, Yue C, Morillo R, Ware S, Lu J, Bi J. Behavior vs introspection: refining prediction of clinical depression via smartphone sensing data. 2016 Oct 25 Presented at: 7th Conference on Wireless Health; October 25-27, 2016; Bethesda, MD, USA.
3. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Dec;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26701262]
4. Saeb S, Zhang M, Kwasny M, Karr C, Kording K, Mohr D. The relationship between clinical, momentary, and sensor-based assessment of depression. 2015 May 20 Presented at: 9th International Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health); May 20–23, 2015; Istanbul, Turkey.
5. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
6. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 2016 Mar 29;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]
7. Gruenerbl A, Osmani V, Bahle G, Carrasco J, Oehler S, Mayora O, et al. Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients. 2014 Mar 07 Presented at: 5th Augmented Human International Conference ACM; March 7-9, 2014; Kobe, Japan p. 38.
8. Huang Y, Xiong H, Leach K, Zhang Y, Chow P, Fua K. Assessing social anxiety using GPS trajectories and point-of-interest data. 2016 Sep 12 Presented at: International Joint Conference on Pervasive and Ubiquitous Computing ACM; September 12-16, 2016; Heidelberg, Germany p. 898-903.
9. Smith TB, McCullough ME, Poll J. Religiousness and depression: evidence for a main effect and the moderating influence of stressful life events. Psychol Bull 2003 Jul;129(4):614-636. [Medline: 12848223]
10. Clark LA, Watson D. Mood and the mundane: relations between daily life events and self-reported mood. J Pers Soc Psychol 1988 Feb;54(2):296-308. [Medline: 3346815]
11. Hamer M, Coombs N, Stamatakis E. Associations between objectively assessed and self-reported sedentary time with mental health in adults: an analysis of data from the Health Survey for England. BMJ Open 2014 Mar 20;4(3):e004580 [FREE Full text] [doi: 10.1136/bmjopen-2013-004580] [Medline: 24650807]
12. Nahum-Shani I, Smith S, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2016 Sep 23. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
13. Silva TH, de Melo PO, Almeida J, Musolesi M, Loureiro A. arXiv. 2014. You are what you eat (and drink): identifying cultural boundaries by analyzing food & drink habits in Foursquare URL: https://arxiv.org/pdf/1404.1009.pdf [accessed 2017-07-23] [WebCite Cache ID 6sB9VKiAT]

XSL•FO
RenderX

14. Li Y, Steiner M, Wang L, Zhang Z, Bao J. Exploring venue popularity in Foursquare. : IEEE; 2013 Apr 19 Presented at: INFOCOM; April 14-19, 2013; Turin, Italy p. 3357-3362.
15. Focus Pointe Global. URL: http://www.focuspointeglobal.com/ [accessed 2017-07-13] [WebCite Cache ID 6rv0f7Pg2]
16. Bohn MJ, Babor TF, Kranzler HR. The Alcohol Use Disorders Identification Test (AUDIT): validation of a screening instrument for use in medical settings. J Stud Alcohol 1995 Jul;56(4):423-432. [Medline: 7674678]
17. Skinner HA. The drug abuse screening test. Addict Behav 1982;7(4):363-371. [Medline: 7183189]
18. Beck A, Steer R, Brown G. BDI-II, Beck depression inventory: manual. San Antonio, TX: Harcourt Brace; 1996.
19. Hirschfeld RM, Williams JB, Spitzer RL, Calabrese JR, Flynn L, Keck PE, et al. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. Am J Psychiatry 2000 Nov;157(11):1873-1875. [doi: 10.1176/appi.ajp.157.11.1873] [Medline: 11058490]
20. 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]
21. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
22. Center for Behavioral Intervention Technologies. PurpleRobot. 2015. URL: https://tech.cbits.northwestern.edu/purple-robot/2015/ [accessed 2017-07-13] [WebCite Cache ID 6rv386AlG]
23. NIST-FIPS Standard. Federal Information Processing Standards Publication. 2001. Announcing the advanced encryption standard (AES) URL: http://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.197.pdf [accessed 2017-07-23] [WebCite Cache ID 6sBA4TjOU]
24. Lewis M. Python Software Foundation. 2016 Sep 12. Foursquare 1! URL: https://pypi.python.org/pypi/foursquare [accessed 2017-07-14] [WebCite Cache ID 6rwSEN2x3]
25. Bray T. Internet Engineering Task Force (IETF).: Internet Engineering Task Force (IETF); 2014. The JavaScript object notation (JSON) data interchange format URL: https://tools.ietf.org/html/rfc7159 [accessed 2017-07-23] [WebCite Cache ID 6sBAEltRN]
26. Foursquare. Foursquare Category Hierarchy. 2017. URL: https://developer.foursquare.com/categorytree [accessed 2017-07-14] [WebCite Cache ID 6rwTDpTVd]
27. Weather Underground. 2017. URL: https://www.wunderground.com/ [accessed 2017-07-13] [WebCite Cache ID 6rwTGJv3p]
28. Friedman JH. The Annals of Statistics. 2001. Greedy function approximation: a gradient boosting machine URL: https://projecteuclid.org/download/pdf_1/euclid.aos/1013203451 [accessed 2017-07-23] [WebCite Cache ID 6sBAUJz8Q]
29. Chen T, He T. Higgs boson discovery with boosted trees. 2014 Dec 08 Presented at: 2014 International Conference on High-Energy Physics and Machine Learning; December 8-13, 2014; Montreal, Canada p. 69-80.
30. Skurichina M, Duin R. Bagging, boosting and the random subspace method for linear classifiers. Pattern Anal Appl 2002 Jun 7;5(2):121-135. [doi: 10.1007/s100440200011]
31. Murphy K. Machine Learning: A Probabilistic Perspective. Cambridge, MA, USA: MIT Press; 2012.
32. Xu G, Huang JZ. Asymptotic optimality and efficient computation of the leave-subject-out cross-validation. Ann Statist 2012 Dec;40(6):3003-3030. [doi: 10.1214/12-Aos1063]
33. Ratcliffe M, Burd C, Holder K, Fields A. Defining Rural at the U.S. Census Bureau.: United States Census Bureau; 2016. URL: http://www2.census.gov/geo/pdfs/reference/ua/Defining_Rural.pdf [accessed 2017-07-14] [WebCite Cache ID 6rwU7yrKQ]
34. Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998 Apr 18;316(7139):1236-1238 [FREE Full text] [Medline: 9553006]
35. Modsching M, Kramer R, Hagen K. Field trial on GPS accuracy in a medium size city: the influence of built-up. 2006 Mar 16 Presented at: 3rd Workshop on Positioning, Navigation and Communication; March 16, 2006; Hannover, Germany.
36. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. 2014 Sep 13 Presented at: 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 13-17, 2014; Seattle, Washington, USA p. 3-14. [doi: 10.1145/2632048.2632054]
37. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. 2015 Sep 07 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp); September 7-11, 2015; Osaka, Japan p. 1293-1304.
38. Harris AH, Cronkite R, Moos R. Physical activity, exercise coping, and depression in a 10-year cohort study of depressed patients. J Affect Disord 2006 Jul;93(1-3):79-85. [doi: 10.1016/j.jad.2006.02.013] [Medline: 16545873]
39. Goodwin RD. Association between physical activity and mental disorders among adults in the United States. Prev Med 2003 Jun;36(6):698-703. [Medline: 12744913]
40. Papandrea M, Jahromi KK, Zignani M, Gaito S, Giordano S, Rossi GP. On the properties of human mobility. Comput Commun 2016 Aug;87:19-36. [doi: 10.1016/j.comcom.2016.03.022]
41. Howden L, Meyer J. United States Census Bureau. 2011. Age and sex composition: 2010 URL: https://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf [accessed 2017-07-23] [WebCite Cache ID 6sBBG8v7r]

XSL•FO
RenderX

42. Labor Force Statistics from the Current Population Survey.: Bureau of Census for the Bureau of Labor Statistics URL: https://www.bls.gov/cps/ [accessed 2017-07-13] [WebCite Cache ID 6rwVDI4sf]
43. Poushter J. Pew Research Center. 2016. Smartphone ownership and Internet usage continues to climb in emerging economies URL: http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2017-07-14] [WebCite Cache ID 6rwVL5e67]
44. Elmer-DeWitt P. Fortune. 2016. About Apple's 40% share of the U.S. smartphone market URL: http://fortune.com/2016/02/11/apple-iphone-ios-share/ [accessed 2017-07-14] [WebCite Cache ID 6rwVWhtY4]
45. Goldman LS, Nielsen NH, Champion HC. Awareness, diagnosis, and treatment of depression. J Gen Intern Med 1999 Sep 21;14(9):569-580. [doi: 10.1046/j.1525-1497.1999.03478.x]
46. Rapee R. Generalized anxiety disorder: a review of clinical features and theoretical concepts. Clin Psychol Rev 1991 Jan 21;11(4):419-440. [doi: 10.1016/0272-7358(91)90116-C]
47. Fann JR, Berry DL, Wolpin S, Austin-Seymour M, Bush N, Halpenny B, et al. Depression screening using the Patient Health Questionnaire-9 administered on a touch screen computer. Psychooncology 2009 Jan;18(1):14-22 [FREE Full text] [doi: 10.1002/pon.1368] [Medline: 18457335]
48. van Duinen M, Rickelt J, Griez E. Validation of the electronic Visual Analogue Scale of Anxiety. Prog Neuropsychopharmacol Biol Psychiatry 2008 May 15;32(4):1045-1047. [doi: 10.1016/j.pnpbp.2008.02.002] [Medline: 18343008]

Abbreviations

API: application programming interface
AUC: area under the curve
EMA: ecological momentary assessment
FPG: Focus Pointe Global
GAD-7: Generalized Anxiety Disorder, 7-item
GPS: global positioning system
ID: identification
JITAI: just-in-time adaptive intervention
JSON: JavaScript object notation
PHQ-9: Patient Health Questionnaire, 9-item
SD: standard deviation
SMS: short message service
XGBoost: extreme gradient boost

Edited by G Eysenbach; submitted 10.01.17; peer-reviewed by A Mosa, J Torous, F Wahle, L Barnes, M Faurholt-Jepsen, A Paglialonga; comments to author 18.03.17; revised version received 09.05.17; accepted 17.06.17; published 10.08.17.

Please cite as:
Saeb S, Lattie EG, Kording KP, Mohr DC
Mobile Phone Detection of Semantic Location and Its Relationship to Depression and Anxiety
JMIR Mhealth Uhealth 2017;5(8):e112
URL: http://mhealth.jmir.org/2017/8/e112/
doi:10.2196/mhealth.7297
PMID:28798010

©Sohrab Saeb, Emily G Lattie, Konrad P Kording, David C Mohr. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 10.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Quantifying Human Movement Using the Movn Smartphone App: Validation and Field Study

Ralph Maddison[1,2*], PhD; Luke Gemming[2*], PhD; Javier Monedero[3*], PhD; Linda Bolger[3*], BSc; Sarahjane Belton[3*], PhD; Johann Issartel[3*], PhD; Samantha Marsh[2*], PhD; Artur Direito[2*], PhD; Madeleine Solenhill[4*], PhD; Jinfeng Zhao[5*], PhD; Daniel John Exeter[5*], PhD; Harshvardhan Vathsangam[6*], PhD; Jonathan Charles Rawstorn[1,2*], PhD

[1]Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Burwood, Australia

[2]National Institute for Health Innovation, School of Population Health, University of Auckland, Auckland, New Zealand

[3]School of Health & Human Performance, Dublin City University, Dublin, Ireland

[4]Center of Research on Welfare Health and Sport, School of Health and Welfare, Halmstad University, Halmstad, Sweden

[5]Department of Epidemiology and Biostatistics, School of Population Health, University of Auckland, Auckland, New Zealand

[6]Robotic Embedded Systems Laboratory, Robotics and Autonomous Systems Center, University of Southern California, Los Angeles, CA, United States

[*]all authors contributed equally

Corresponding Author:
Jonathan Charles Rawstorn, PhD
Institute for Physical Activity and Nutrition
School of Exercise and Nutrition Sciences
Deakin University
221 Burwood Highway
Burwood, 3125
Australia
Phone: 61 3 924 68461
Email: jonathan.rawstorn@deakin.edu.au

Abstract

Background: The use of embedded smartphone sensors offers opportunities to measure physical activity (PA) and human movement. Big data—which includes billions of digital traces—offers scientists a new lens to examine PA in fine-grained detail and allows us to track people's geocoded movement patterns to determine their interaction with the environment.

Objective: The objective of this study was to examine the validity of the Movn smartphone app (Moving Analytics) for collecting PA and human movement data.

Methods: The criterion and convergent validity of the Movn smartphone app for estimating energy expenditure (EE) were assessed in both laboratory and free-living settings, compared with indirect calorimetry (criterion reference) and a stand-alone accelerometer that is commonly used in PA research (GT1m, ActiGraph Corp, convergent reference). A supporting cross-validation study assessed the consistency of activity data when collected across different smartphone devices. Global positioning system (GPS) and accelerometer data were integrated with geographical information software to demonstrate the feasibility of geospatial analysis of human movement.

Results: A total of 21 participants contributed to linear regression analysis to estimate EE from Movn activity counts (standard error of estimation [SEE]=1.94 kcal/min). The equation was cross-validated in an independent sample (N=42, SEE=1.10 kcal/min). During laboratory-based treadmill exercise, EE from Movn was comparable to calorimetry (bias=0.36 [−0.07 to 0.78] kcal/min, t_{82}=1.66, P=.10) but overestimated as compared with the ActiGraph accelerometer (bias=0.93 [0.58-1.29] kcal/min, t_{89}=5.27, P<.001). The absolute magnitude of criterion biases increased as a function of locomotive speed ($F_{1,4}$=7.54, P<.001) but was relatively consistent for the convergent comparison ($F_{1,4}$=1.26, P<.29). Furthermore, 95% limits of agreement were consistent for criterion and convergent biases, and EE from Movn was strongly correlated with both reference measures (criterion r=.91, convergent r=.92, both P<.001). Movn overestimated EE during free-living activities (bias=1.00 [0.98-1.02] kcal/min, t_{6123}=101.49, P<.001), and biases were larger during high-intensity activities ($F_{3,6120}$=1550.51, P<.001). In addition, 95% limits of agreement for convergent biases were heterogeneous across free-living activity intensity levels, but Movn and ActiGraph measures were

XSL•FO
RenderX

strongly correlated (r=.87, P<.001). Integration of GPS and accelerometer data within a geographic information system (GIS) enabled creation of individual temporospatial maps.

Conclusions: The Movn smartphone app can provide valid passive measurement of EE and can enrich these data with contextualizing temporospatial information. Although enhanced understanding of geographic and temporal variation in human movement patterns could inform intervention development, it also presents challenges for data processing and analytics.

(JMIR Mhealth Uhealth 2017;5(8):e122) doi:10.2196/mhealth.7167

KEYWORDS

telemedicine; smartphone; validation studies; geographic information systems; locomotion; physical activity; humans

Introduction

The World Health Organization (WHO) recognizes physical inactivity as one of the leading global risk factors for morbidity and premature mortality [1]. Despite the considerable benefits of regular physical activity (PA; [2,3]), it has been estimated that 21.4% of the global population is inactive (perform little or no activity), with greater prevalence of physical inactivity among most developed countries (27.8%; [4]).

To move forward in PA research, it has been suggested that "more of the same is not enough" [5]. Different approaches are needed to reduce the burden of disease associated with physical inactivity. Technological innovations such as smartphones and wearable sensors offer potential to improve the reach, enhance delivery (greater frequency of contact and duration of intervention), and increase effectiveness of interventions to improve PA levels. Despite their potential, it is unclear whether these new devices provide research-grade precision measurement. To address this concern, a number of validation studies have been conducted [6-8].

Compared with the frequently used ActiGraph accelerometer, studies have demonstrated acceptable levels of agreement over periods of 7 days against the Fitbit Zip wearable sensor [6] and CalFit smartphone app [8]. Despite the acceptable measurement properties of new wearable sensors, the benefit of smartphones is that they are carried by most people, most of the time. At the population level, their ubiquitous use, the big data (millions of data points) and geospatial information generated by smartphones may offset any potential measurement inaccuracies by providing valuable insight into behavioral patterns (eg, temporal stability) and their contexts. Moreover, the potential of these technologies are vast. For example, coupled with the Internet of Things [9], smartphones can be used to track people's movement within cities or environments, thereby providing a rich source of contextual information and the potential to deliver "just-in-time" interventions. Compared with stand-alone accelerometers, smartphones offer advantages in terms of usability and integration of supplementary data. These features support the delivery of more responsive, engaging, and context-specific interventions that could improve uptake, adherence, and effectiveness.

Few published studies have explored the utility of smartphones to measure both PA and human movement. A recent convergent validity study comparing an Android smartphone activity tracker against the ActiGraph accelerometer found acceptable associations and agreement in both laboratory and free-living environments [10]. To extend the evidence, further validation work against criterion measures is required [11]. Furthermore, health is geospatial, and if we can see trends in behavior spatially, we can monitor (and improve) population and individuals' needs [12]. To illustrate, González et al [13] used anonymized cellular phone data from 100,000 users in New York, United States, to capture people's position over a period of 6 months. They showed that human trajectories had a high degree of temporal and spatial regularity. In other words, humans follow simple reproducible patterns; this in turn has important implications for interventions to enhance human mobility [13]. In PA research, the feasibility of linking global positioning system (GPS) and accelerometer data has been well established [14-16]. However, these studies have typically involved the use of 2 separate devices (accelerometer and GPS) for limited periods (7-28 days). Smartphone apps offer advantages over these approaches; they are relatively cheap, readily available, incorporate native sensors (GPS, gyroscopes, and accelerometers), and permit passive data collection—thus requiring minimal input from participants. This has clear advantages in terms of reducing participant burden for research.

We aimed to examine the validity of the Movn smartphone app for estimating PA energy expenditure (EE) and quantifying human movement patterns. Two validation studies were conducted against criterion and convergent methods; a supporting cross-validated study assessed the consistency of activity data when collected across different devices. GPS and geographic information system (GIS) data were integrated to demonstrate the feasibility of geospatial analysis.

Methods

A dual-phase cross-sectional study was conducted to determine the validity of the Movn smartphone app (Moving Analytics) for assessing EE and human movement patterns during laboratory-based and free-living daily activities among a convenience sample of healthy adults. Phase 1 comprised laboratory-based treadmill exercise at light to vigorous levels of intensity and free-living daily activities. Phase 2 comprised a cross-validation during laboratory-based activities among a separate sample of participants. EE was the main measurement of interest; it is the most appropriate outcome for validating accelerometers as it can be directly related to the accepted methods of PA categorization, compared with robust criterion data collected via indirect calorimetry, and allows standardized comparison with other accelerometer devices [17].

XSL•FO
RenderX

Study Participants and Recruitment

In phase 1, a total of 21 adults (13 female), aged 20 to 55 years, were recruited in Dublin, Ireland (see Table 1). Participants were recruited via direct contact through the university and by word of mouth. Adults were eligible for inclusion provided they met the following criteria: aged 18 to 65 years, able to give written informed consent, and able to communicate in English. In Phase 2, a total of 42 adults (27 male) aged between 18 and 33 years were recruited from the Greater Los Angeles Area. Participants were recruited via direct contact through the University of Southern California and by word of mouth. Phase 1 and 2 study protocols were approved by the Dublin City University Research Ethics Committee and the institutional review board of the University of Southern California, respectively.

Phase 1 Procedures

Upon arrival to the laboratory, participants completed demographic information, including age and sex, and the Physical Activity Readiness Questionnaire [18]; all were deemed safe to exercise. Anthropometric measurements were taken; height was measured to the nearest 0.1 cm with a portable stadiometer, and weight was measured to the nearest 0.1 kg on an electronic scale (Seca).

Reference accelerometry was quantified using the GT1M (ActiGraph Corp), a dual-axis accelerometer with established reliability and validity [19]. Epoch duration was set to 1 s during laboratory-based activities and 10 s during free-living activities. ActiGraph devices were fitted to an elastic belt on participants' right hip at the midaxillary line for the duration of the test. Comparison accelerometry was quantified using the Moto G first-generation smartphone (Motorola Mobility LLC) running Android version 4.3 (Google Inc) and a research version of the Movn app. Movn is a commercially available app—for Android and Apple iPhone operating systems—that uses inbuilt smartphone accelerometers to passively quantify time spent in moderate-to vigorous-intensity physical activities such as walking and running. Raw accelerometer data were captured at the maximum frequency permitted by the phone hardware (at/above 200 Hz) and downsampled to 50 Hz. The Movn app samples GPS data every 30 min unless movement is detected, and every minute during periods of movement; this sampling approach was adopted to balance sampling frequency and power consumption. Movn also allows users to set daily PA goals and can issue prompts throughout the day to facilitate goal achievement. The smartphone was secured in a phone holder and positioned adjacent to the ActiGraph accelerometer. The same phone was used for all the participants.

Following familiarization with the smartphone and accelerometer, the K4b2 portable indirect calorimeter (Cosmed) was used to assess resting and exercise EE [20]. A 2-point calibration procedure was conducted before each testing session according to the manufacturer's guidelines. Calibration of the oxygen (O_2) and carbon dioxide (CO_2) sensors was performed with standard gases of known concentrations (gas 1: O_2=20.93%, CO_2=0.04%; gas 2: O_2=15.00%, CO_2=5.00%). Respiratory volume was calibrated using a 3-L syringe. The rate of EE was estimated using the following formula as calculated by the K4b2 system: EE (kcal/min)=(3.781 × $V\dot{}O_2$) + (1.237 × $V\dot{}CO_2$) if UN (urea nitrogen)=0, where $V\dot{}O_2$=oxygen uptake (L/min) and $V\dot{}CO_2$=carbon dioxide production (L/min) [20].

A face mask (Hans Rudolf) held in place by a nylon harness covered the participants' nose and mouth. The mask was attached to a bidirectional digital turbine flow meter to measure the volumes of inspired and expired air. Heart rate data were captured using the FT1, a chest-worn sensor (Polar Electro Oy).

Once instrumented, participants remained seated for 15 min while physiological data were recorded; the last 5 min of data were averaged to calculate resting EE. Participants then completed four discrete bouts of walking and running on a motorized treadmill (Quasar Med, H/P Cosmos Sports & Medical GmbH). Participants completed 5-min bouts of exercise at light (4 km/h and 6 km/h), moderate (10 km/h), and vigorous intensity levels (≥12 km/h), separated by 3-min bouts of passive recovery.

Following laboratory-based activities, participants were instructed to wear the ActiGraph accelerometer (as described above) and carry the smartphone (as they would normally carry their personal phone) for 24 hours during free-living daily activities. Participants were instructed to remove the devices during water-based activities such as swimming, showering, and bathing. An optional sports armband carry case was provided to participants who preferred this carry method during free-living exercise. After 24 hours, participants returned the phone and accelerometer.

Phase 2 Procedures

Procedures for the cross-validation sample have been described elsewhere [21]. In brief, each participant wore a Samsung Galaxy Nexus S phone, Android version 2.3.3 with Movn app, installed on the right iliac crest with a belt holder to record movement. EE was measured by the Oxycon portable indirect calorimeter (CareFusion) worn in a backpack fitted to the comfort of the participant. Participants completed three 6-min bouts of treadmill walking (4, 5, and 6 km/h), separated by 2 min of passive recovery.

Data Handling

Heart rate, calorimeter, and ActiGraph data were downloaded using the manufacturers' software; smartphone data were downloaded using a text reader, exported for manual analysis, and synchronized in postprocessing.

Mean values were calculated during the last 2 min of each laboratory-based activity bout, following similar procedures in phase 1 and 2. Free-living accelerometer data were cleaned by removing nonwear time, equivalent to >60 min of continuous zero counts. A minimum of 10 hours of available free-living data were required for the analysis. Data were processed to generate comparable units of EE for analysis [17].

Statistical Analysis

Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS) version 21 for Windows (IBM Corp).

XSL•FO
RenderX

To derive estimates of EE from Movn activity counts, multivariate regression was conducted following established methods [19] to identify the strongest relationship between Movn activity counts, participant characteristics, and EE measured via indirect calorimetry. Estimates of EE were then used to identify Movn activity count thresholds associated with accepted classifications for light (<3 metabolic equivalent of task [MET]), moderate (3-6 MET), hard (6-9 MET), and very hard (>9 MET) activity intensity levels [19]. The regression equation was also applied to phase 2 laboratory-based data to cross-validate the accuracy of EE estimation among an independent sample using different smartphone hardware.

EE measurement validity was evaluated following guidelines proposed by Welk et al [17], who suggest that agreement between two measurement methods requires demonstration of three unique characteristics: equivalent group estimates, association between measurements, and absence of systematic and/or heterogeneous bias. Furthermore, supplementary analyses at an individual level are also recommended to determine whether group-level agreement is consistent across individuals.

The criterion and convergent validity of Movn EE measurement were assessed compared with EE measured via indirect calorimetry during phase 1 laboratory-based activities and with EE estimated from ActiGraph movement counts during phase 1 laboratory and free-living activities, respectively; *t*-tests were conducted to detect systematic criterion and convergent group-level measurement biases. Simple analyses of variance were also conducted to determine whether group-level criterion or convergent measurement biases were affected by activity intensity level during laboratory-based activities; significant main effects were explored with Bonferroni-corrected paired comparisons (ie, least significant difference $P \times N(N-1)/2$ paired comparisons). Furthermore, 95% limits of agreement for biases were calculated to assess absolute measurement agreement and homogeneity of biases across the measurement range [22,23]. Relationships between Movn and reference measurement methods were assessed by calculating Pearson correlation coefficients and two-way random effects intraclass correlation coefficients for absolute agreement (ICC).

Supplementary analyses compared time-synchronized group-level measurements throughout the laboratory-based activity protocol to determine agreement between measurement patterns [17]. Furthermore, individual-level biases were calculated to determine whether group-level agreement was consistent across the sample.

To determine the feasibility of using smartphone sensor data to ascertain the geographic location of activity and patterns of human movement, accelerometer and GPS data were combined to provide an indication of the location and intensity of PA. Data were imported into ArcGIS version 10.2.2 (Esri) transformed to location points and interpolated into two-dimensional (2D) spatial paths and three-dimensional (3D) spatiotemporal trajectories.

Descriptive data are reported as mean (standard deviation); bias data are reported as mean and 95% CI; alpha=.05 for all hypothesis tests.

Results

Phase 1 participants included 21 adults (13 female) aged 20 to 55 years (see Table 1); phase 2 participants included 42 adults (27 male) aged 18 to 33 years. All participants were at normal weight.

Estimating Energy Expenditure

A multivariate regression model including Movn activity counts and participant body mass was the strongest predictor of measured EE (Equation 1, r^2=.83; SEE=1.94 kcal/min). The relationship between measured and predicted EE is shown in Figure 1.

Equation 1: EE (kcal)=0.00063 × activity level (counts/min) + 0.121 × body mass (kg) − 5.66

Equation 1 was used to identify Movn activity count thresholds that correspond with accepted classifications of light (<3 MET), moderate (3-6 MET), hard (6-9 MET), and very hard (≥9 MET) activity intensity levels [19]; thresholds derived from phase 1 laboratory-based data are presented in Table 2. A correction factor (2121) was applied to the smartphone accelerometer data to facilitate scale congruence and allow comparison between Movn and ActiGraph activity counts.

When applied to the cross-validation sample (phase 2), Equation 1 estimation accuracy was lower during laboratory-based walking (r^2=.24; SEE=1.10 kcal/min).

Table 1. Participant characteristics.

Demographics	Phase 1 Mean (SD)	Phase 2 Mean (SD)
N (men/women)	21 (8/13)	42 (27/15)
Age, in years	27 (7.9)	26 (3.8)
Height (cm)	171.2 (7.3)	172 (8)
Weight (kg)	70.5 (11.6)	68 (12.0)
Body mass index (kg/m^2)	23.1 (2.6)	22.0 (3.0)

XSL•FO
RenderX

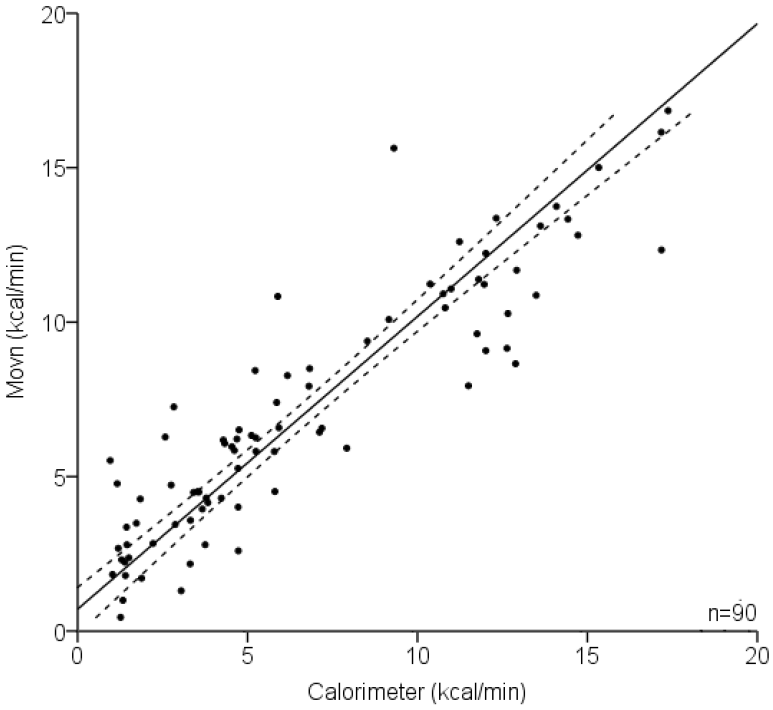
Table 2. Movn activity count thresholds for classifying activity intensity level.

Activity intensity level[a]	MET[b] range	Activity (counts/min)
Light	<3	<1253
Moderate	3-6	1253-1272
Hard	6-9	1273-6987
Very hard	≥9	>6987

[a]Activity intensity level classification adapted from Freedson et al [19].

[b]MET: metabolic equivalent of task.

Figure 1. Relationship between measured energy expenditure (EE; indirect calorimetry) and Movn app EE derived from multivariate regression of Movn activity counts on measured EE during phase 1 laboratory-based activities.



Laboratory-Based Activities

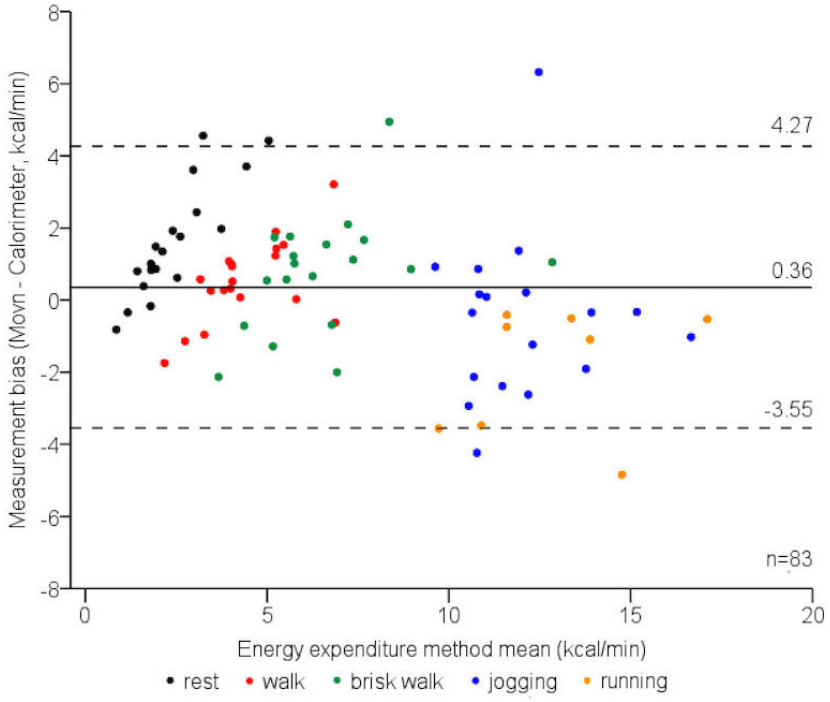
Criterion Validity

Table 3 summarizes EE during phase 1 laboratory-based treadmill exercise, assessed with indirect calorimetry, Movn, and ActiGraph. Movn overestimated EE compared with the criterion indirect calorimetry method (Table 3), but the small magnitude did not represent a systematic measurement bias (t_{82}=1.66, P=.10). A statistically significant main effect of activity intensity level was detected on measurement biases ($F_{1,4}$=7.54, P<.001), indicating systematic variance across laboratory-based activity levels. The Movn app overestimated EE at rest and slower locomotive speeds and underestimated EE at faster locomotive speeds (Table 3). Bonferroni-corrected paired comparisons revealed that the absolute EE measurement bias at 12 km/h was statistically significantly larger than all other speeds with the exception of 10 km/h. Furthermore, the EE measurement bias at 10 km/h was statistically significantly smaller than during rest.

The 95% limits of agreement for criterion EE measurement biases were moderate at most activity intensity levels (Figure 2), indicating acceptable absolute measurement agreement. Biases were relatively consistent across the measurement range; however, variance was slightly wider at faster locomotive speeds (10-12 km/h, Table 3). Finally, Movn and criterion EE measures were strongly correlated (r=.91, ICC=.95, both P<.001), indicating excellent relative measurement agreement.

XSL•FO
RenderX

Figure 2. The 95% limits of agreement for phase 1 laboratory-based criterion energy expenditure measurement biases, categorized by activity intensity level.



Convergent Validity

Compared with EE derived from the convergent reference ActiGraph device during phase 1 laboratory-based activities, Movn systematically overestimated EE (t_{89}=5.27, P<.001, Table 3). There was no statistically significant main effect of activity intensity level ($F_{1,4}$=1.26, P<.29), indicating convergent measurement biases were relatively consistent across laboratory-based activity levels.

The 95% limits of agreement for convergent EE measurement biases were also moderate at all activity intensity levels (Figure 3), indicating acceptable absolute measurement agreement. Once again, measurement error was relatively consistent across the measurement range; however, similar to the criterion analysis, variance was slightly wider at faster locomotive speeds (10-12 km/h, Table 3). Finally, Movn and criterion EE measures were strongly correlated (r=.92, ICC=.93, both P<.001), indicating excellent relative measurement agreement.

Table 3.

Activity	Energy expenditure (kcal/min)			Biases (kcal/min)	
	Calorimeter	Movn	ActiGraph	Criterion	Convergent
	Mean (SD)			Mean (95% CI)	
Rest	1.75 (0.64)	3.25 (1.80)	1.83 (1.68)	1.60 (0.85-2.35)[d,e]	1.30 (0.65-1.96)
4 km/h	4.15 (1.20)	4.66 (1.69)	4.10 (1.83)	0.52 (−0.04 to 1.08)[e]	0.59 (0.22-0.95)
6 km/h	6.42 (2.13)	6.95 (2.45)	6.67 (2.30)	0.74 (−0.05 to 1.52)[e]	0.40 (−0.02 to 0.82)
10 km/h	11.94 (2.22)	11.78 (2.18)	10.26 (2.52)	−0.53 (−1.67 to 0.60)[a]	1.08 (0.01-2.15)
12 km/h	14.15 (2.68)	11.76 (2.97)	10.87 (1.86)	−1.90 (−3.37 to −0.42)[a-c]	1.05 (−0.81 to 2.92)
Total	7.08 (4.79)	7.45 (4.15)	6.54 (3.90)	0.36 (−0.07 to 0.78)	0.93 (0.58-1.29)[f]

[a-e]Systematic difference in bias between locomotive speeds (P<.001-.01, Bonferroni-corrected).

[a]Rest.

[b]4km/h.

[c]6km/h.

[d]10km/h.

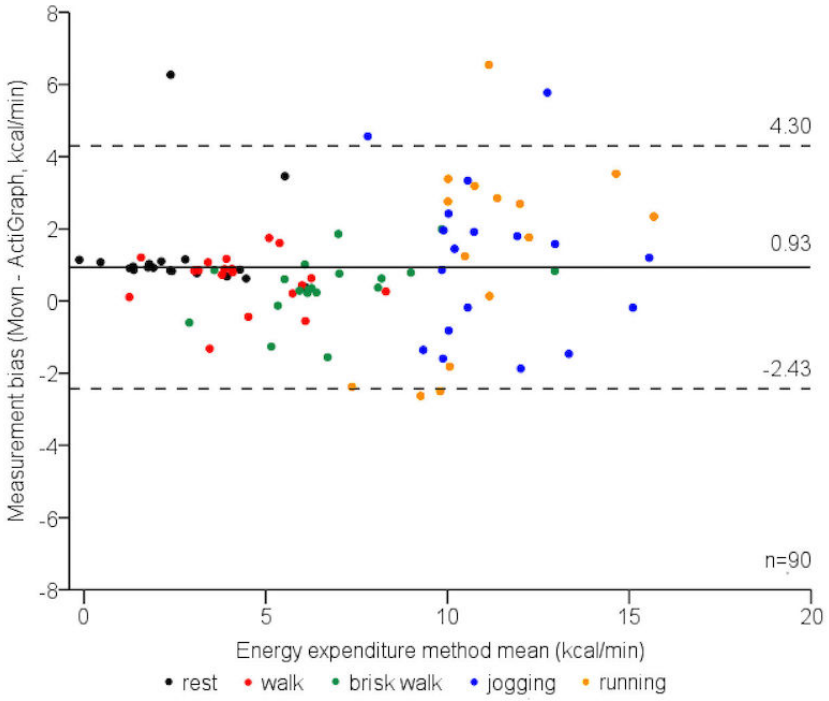
[e]12km/h.

[f]Overall systematic bias compared with the ActiGraph device (P<.001).

Energy expenditure=average during third and fourth min of each intensity bout.

XSL•FO
RenderX

Figure 3. The 95% limits of agreement for phase 1 laboratory-based convergent energy expenditure measurement biases, categorized by activity intensity level.



Supplementary Analyses

Supplementary analyses comparing time-synchronized group-level measurements throughout the phase 1 laboratory-based activity protocol indicate that Movn and calorimetry EE measurement patterns were very similar at low and moderate activity intensity levels (Figure 4; [17]); however, two important trends were identified. First, the Movn EE measurement pattern diverts substantially below the criterion calorimetry measurement late in the exercise protocol during the fastest locomotive speeds; differing measurement patterns at faster speeds indicate that the Movn app may be less valid for quantifying high-intensity locomotive activities. Second, notable asynchronicity between Movn and calorimeter EE measurement patterns reflects the expected latency between changes in energy demands (ie, instant change in locomotive speed) and physiological EE (ie, gradual increase in oxygen consumption).

Examination of individual-level measurement biases revealed relatively small overestimation of total EE among the majority of participants; however, EE was substantially overestimated for 2 participants and underestimated for 6 participants (Figure 5; [17]). This inconsistency suggests Movn may have more utility for group-level surveillance tool than individual-level measurement.

XSL•FO
RenderX

Figure 4. Supplementary analysis of the relative accuracy of measured (indirect calorimetry) and estimated (Movn app) energy expenditure (EE) during phase 1 laboratory-based activities. Minute-by-minute EE measured by the Movn app and criterion reference calorimeter.

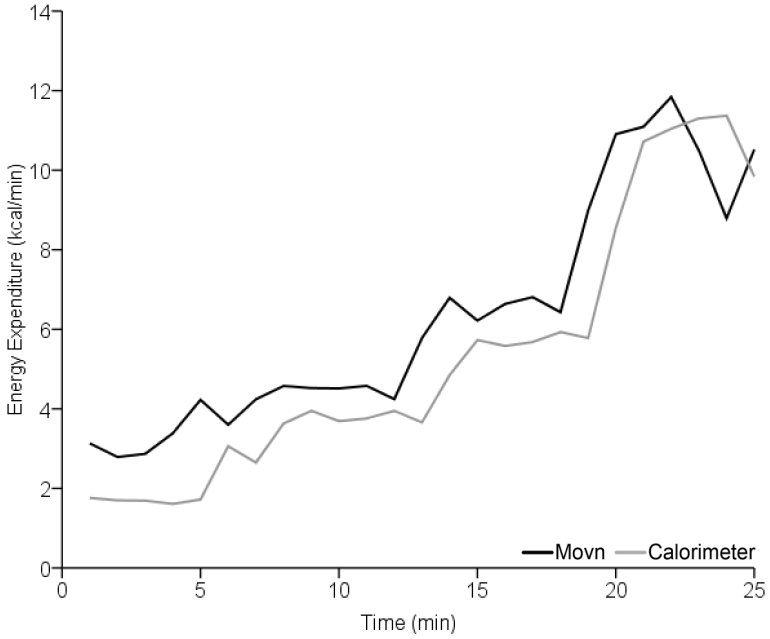
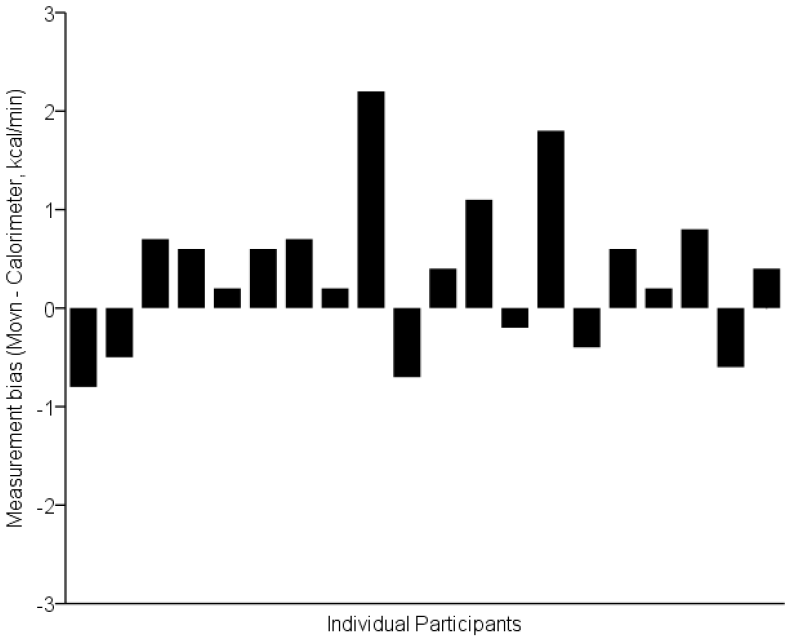


Figure 5. Supplementary analysis of the relative accuracy of measured (indirect calorimetry) and estimated (Movn app) energy expenditure (EE) during phase 1 laboratory-based activities. Mean EE measurement biases across individual participants.



Free-Living Activities

Convergent Validity

Sensor wear time compliance was lower than anticipated during phase 1 free-living activities. After removing nonwear time, only 8 participants met minimum data requirements. Participants recorded an average of 766 (SD 189) min of activity on valid days, yielding a total free-living convergent reference sample of 6124 min. Table 4 summarizes EE during phase 1 free-living activities, assessed with Movn and ActiGraph devices. Compared with the convergent reference ActiGraph, Movn overestimated EE during free-living activities (Table 4); the mean bias was larger than during laboratory-based activity and represented a systematic bias between devices (t_{6123}=101.49, P<.001). Consistent with laboratory-based activities, a statistically significant effect of activity intensity level was also detected on free-living measurement biases ($F_{3,6120}$=1550.51, P<.001), indicating systematic variance in measurement biases across commonly used levels for classifying free-living activity intensity (Table 2). Movn overestimated EE during all free-living activity levels and, with the exception of light- and moderate-intensity activity levels, bias magnitudes grew with activity intensity level (Table 4). Heterogeneous measurement biases during free-living activity are presented in Figure 6,

XSL•FO
RenderX

which has been graphically categorized by individual participant (n=8) to highlight both the variance in resting EE, and relatively consistent pattern of bias heterogeneity between individuals. Positive and negative biases appear relatively symmetrical within participants (Figure 6). Further investigation indicates negative measurement biases may reflect periods of smartphone noncarry time; however, as it was not possible to validate this assumption, we conservatively treated the data as measurement error.

Finally, despite the systematic bias, EE measures from Movn and ActiGraph were strongly correlated (r=.87, ICC=.83, both P<.001), indicating excellent relative measurement agreement during free-living activities.

Table 4. Energy expenditure during phase 1 free-living activities.

Level of intensity	Energy expenditure (kcal/min)		Bias (kcal/min)
	Movn Mean (SD)	ActiGraph Mean (SD)	Mean (95% CI)
Light (<3 metabolic equivalent of task [MET])	2.43 (1.15)	1.61 (1.34)	0.83 (0.81-0.84)[c,d]
Moderate (3-6 MET)	2.32 (0.86)	1.01 (0.65)	1.31 (0.69-1.93)[c,d]
Hard (6-9 MET)	4.19 (1.49)	2.24 (1.6)	1.96 (1.87-2.04)[a,b,d]
Very hard (≥9 MET)	7.66 (1.19)	4.25 (2.01)	3.41 (3.16-3.66)[a-c]
Total	2.73 (1.51)	1.73 (1.45)	1.00 (0.98-1.02)[e]

[a-d]Systematic difference in biases between activity intensity levels (P<.001-.02, Bonferroni-corrected).

[a]Light.

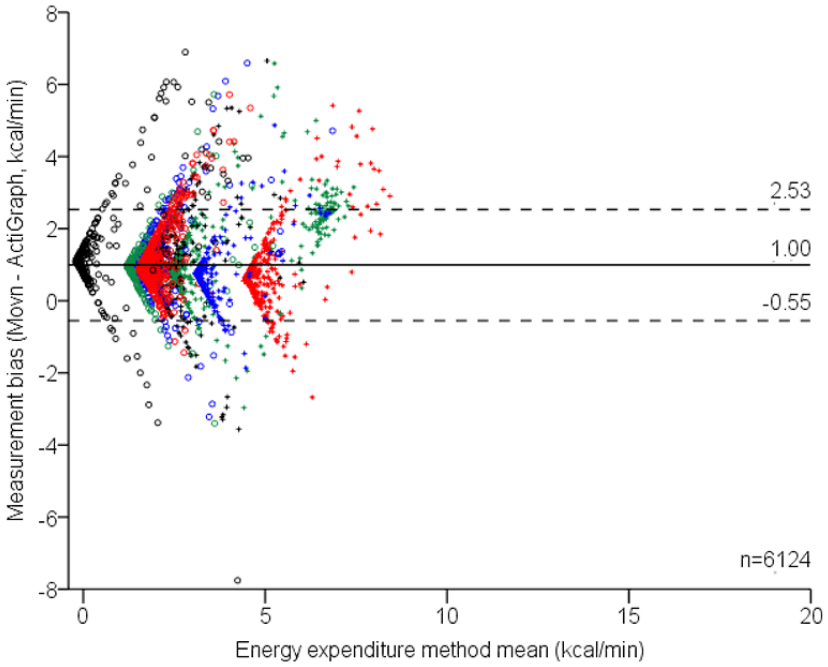
[b]Moderate.

[c]Hard.

[d]Very hard.

[e]Overall systematic bias compared with the ActiGraph device (P<.001).

Figure 6. The 95% limits of agreement for phase 1 free-living convergent energy expenditure measurement biases; categorized by individual participants who recorded ≥10 hours activity data per day.



XSL•FO
RenderX

Table 5. Accuracy of global positioning system (GPS) location data.

Accuracy radius (m)[a]	Location samples n (%)	Median accuracy radius (m)
≤25	246 (32.3)	19
26-50	307 (40.3)	30
51-75	29 (3.8)	57
76-100	45 (5.9)	96
>100	134 (17.6)	2370
Total	761 (100.0)	30

[a]68% probability of the true position lying within specified radii of the recorded location coordinates.

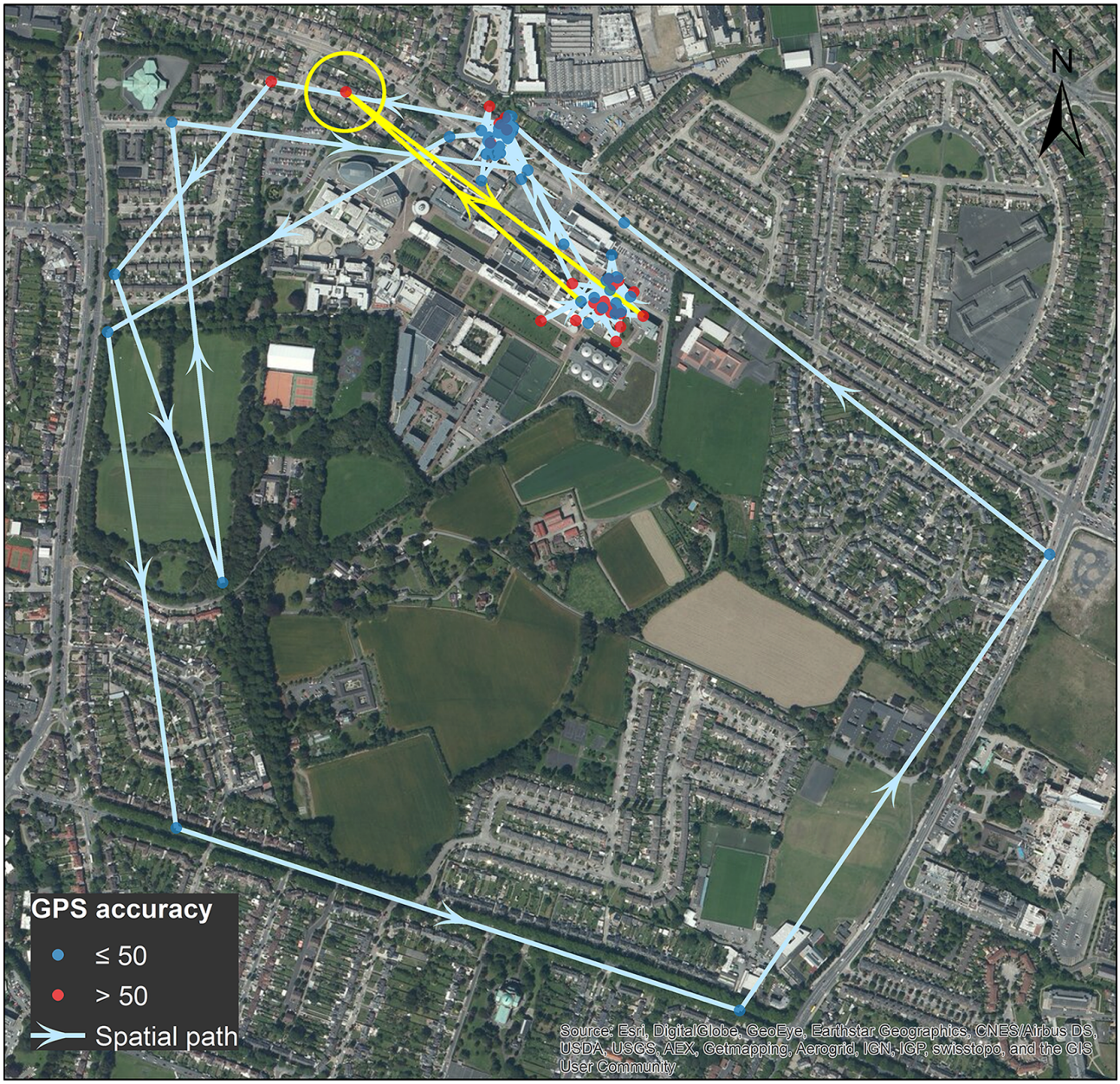
Understanding Geospatial Human Movement

Free-living accelerometry yielded 30,666 records; during this time, Movn logged 761 GPS coordinate pairs. The majority of GPS data (553/761, 72.7%) had acceptable location measurement accuracy (68% probability of true position lying within a 50-m radius of recorded location; [24]) and the median location accuracy radius was 30 m (Table 5); however, almost 20% of GPS location data were characterized by low measurement accuracy (accuracy radius>100 m, median=2370 m).

Figure 7 presents an example of mapped daily location data for 1 participant; blue and red markers indicate accuracy radii of ≤50 m and >50 m, respectively. Straight paths were interpolated between sequential locations; arrows show movement direction. The circled location is likely erroneous given the significant deviation from previous and subsequent locations over a relatively short time and sharp turning angle (see the yellow highlighted path).

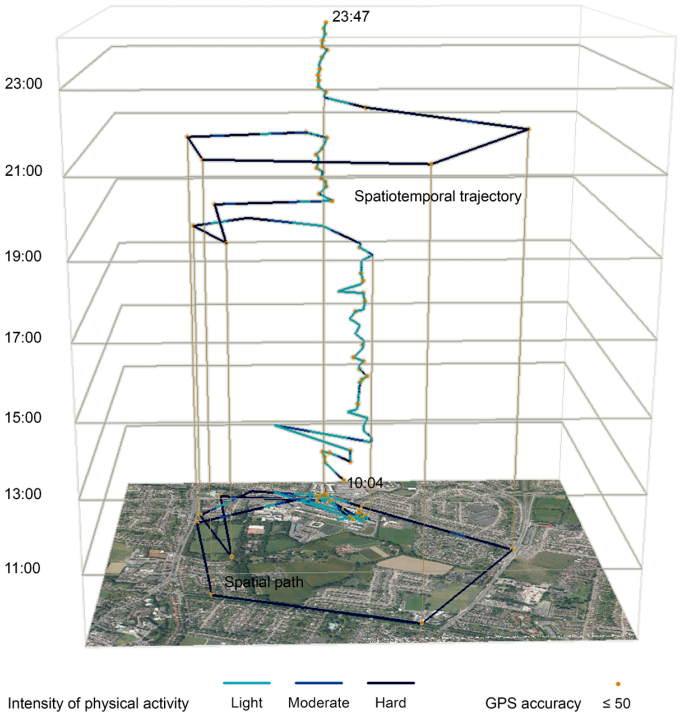
XSL•FO
RenderX

Figure 7. Example of participant location markers and interpolated spatial path.



XSL•FO
RenderX

Figure 8. Detailed example of a participant's daily movement. Position is depicted on the two-dimensional plane, and time is depicted on the vertical plane. Graduated shading of the movement path indicates activity level (dark shading=higher activity level). Orange point markers represent locations with an accuracy of >50 m.



The temporal variability of activity intensity level is not readily displayed by the two dimensions, and dense concentrations of location markers make it difficult to resolve movement direction sequences. To address this, a time-geography approach [25] was used to present an enhanced 3D illustration of activity patterns; position is depicted on a typical 2D plane, time on the vertical plane, and intensity level using graduated shading (Figure 8). Duration, sequence, and movement information can be observed or inferred from such spatiotemporal trajectories; for example, with the exception of the segments that link to less accurate locations (ie, vertices with orange marker points), the main concentration of points and paths indicate that this nominal participant spent the majority of the day (approximately 11:00 AM-17:40 PM) conducting light-intensity activities around a central location. In the evening (approximately 17:40 PM-18:40 PM), the participant traveled a relatively long distance with higher-intensity PA.

Discussion

Principal Findings

This study sought to examine the validity of the Movn smartphone app for assessing PA and human movement patterns. In agreement with previous research, Movn activity counts were strongly related to measured EE, and the SEE for the predictive function compared favorably with that reported by Freedson et al [19]. Activity counts associated with thresholds for categorizing intensity levels were lower than those reported by Freedson et al [19], possibly because of differences between algorithms that convert raw accelerometer output signals (ie, voltage) into activity counts. Compared with indirect calorimetry in a controlled laboratory setting, an absence of systematic overall measurement bias, moderate limits of agreement, and very strong correlations demonstrated good agreement for Movn-derived estimates of EE. Agreement was not as strong when using a different model smartphone in the cross-validation model, and this may be explained by differences between integrated solid state accelerometer chips.

Together, these data suggest that Movn could provide acceptable PA measurement precision in controlled laboratory settings; however, greater error variability at higher activity intensity levels suggests that it may be better suited to quantifying general daily PAs than higher intensity exercise training. These findings are comparable with the recent validations of Android smartphone-based activity measurement tools, which demonstrated strong correlations but larger measurement error at higher activity intensity levels when compared with the ActiGraph GTX3 accelerometer [7,8].

During 24 hours of free-living activity, Movn overestimated EE compared with the commonly used and validated ActiGraph accelerometer; biases were, once again, larger and more variable during higher intensity activities. Nonetheless, Movn and reference estimates of EE were strongly related, and these data compare favorably to other data comparing smartphones and the Fitbit Zip wearable sensor with the ActiGraph in free-living settings [6,7]. Differences in measurement units preclude more detailed comparisons of measurement validity.

Interindividual variability revealed in exploratory supplementary analyses indicated the Movn app may be most suitable for group-level monitoring. As smartphone apps can be more feasibly and affordably distributed across large populations than wearable sensors such as the ActiGraph, this approach holds considerable potential for large-scale population-based research.

XSL•FO
RenderX

It was feasible to integrate smartphone-collected geospatial and accelerometry data within a GIS to provide insight into temporal movement patterns (location and intensity). In combination with the potential for large-scale population surveillance, the capability to augment EE data with mapped movement patterns highlights the potential for smartphone-based measurement tools to support novel research evaluating the drivers and effects of interventions and policies that impact PA. However, such large-scale analyses also present challenges for data processing and analytics.

Collectively, this study and previous works highlight the value of smartphones as an acceptable measure of PA in both laboratory-based and free-living contexts. Given their ubiquity, integrated sensor features (accelerometry, gyroscope, inclinometer, and GPS), and passive data collection, smartphones offer additional value compared with existing measurement approaches. Further value is added by the capacity to generate large datasets that could be used to understand temporal, location, and contextual factors that affect PA. Such information could be used to provide point of decision prompts or behavior change strategies to increase PA and decrease sedentary behavior.

The integration of geospatial and accelerometry data within a GIS in this study demonstrates the potential of using smartphones for describing the context (ie, time and location of activity) of human movement patterns. This approach has been demonstrated but has typically relied on separate devices to collect activity and geospatial data, which adds participant burden and limits the duration of observation (7-14 days). Passive smartphone data collection reduces participant burden and permits sustained data collection, for as long as the app is installed and individuals continue to charge their smartphones. This is exemplified by many commercially available smartphone apps that currently harness these features (eg, Moves, Human, and Movn). Smartphones offer pragmatic advantages over stand-alone GPS devices, although the maximum data capture frequency is typically lower (approximately 1 Hz vs 5+ Hz) [26]. Increased sampling frequencies have the potential to improve GPS measurement accuracy but would have a detrimental effect on smartphone battery life and may only offer additional benefits during activities with rapidly fluctuating movement patterns [26]. In this study, the Movn app was configured to record GPS location every 30 min, or when activity was detected via accelerometry to optimize smartphone battery life. More frequent sampling would increase the resolution of geospatial movement patterns but may limit the maximum recording duration because of faster power consumption.

This study has important implications for future research. Smartphones can passively measure PA in large population groups without the use of dedicated measurement tools and provide opportunities to enrich traditional PA measurement with contextualizing temporal and geopositional data. Future work could leverage existing integration between smartphones, smartwatches, and consumer-grade wearable activity trackers to capture PA during periods of smartphone noncarry time. Temporal and positional contextual data provide opportunities to understand peoples' movement patterns, and this may help to identify and capitalize on optimal opportunities for PA intervention. Utilizing the big data generated by smartphone apps offers opportunities to provide detailed information on how and where people move and whether these patterns are stable over time. If patterns can be predicted, then interventions could be delivered via smartphones to promote PA and reduce sedentary behavior. Geospatial data could be linked with other sensors or data sources via the Internet of Things to provide "just-in-time" interventions such as promoting active transport options (nearby cycling or walking routes) rather than driving a motor vehicle. Such an approach would be consistent with the notion of smart city research, which harnesses the Internet of Things for public and environmental health surveillance [9].

Smart cities focus investment on digital infrastructure, including information and communication technology rather than traditional physical infrastructure. Whereas smart city technologies typically focus on system efficiencies (traffic and waste management, etc), they offer potential to promote PA and health [9,27]. To fully maximize the benefits of smartphone data, it will be necessary to develop big data analytical methods to extract, process, and interpret large quantities of data. To achieve this, PA researchers need to develop expertise in these techniques or collaborate with people who have appropriate expertise in big data processing and analytics.

Strengths and Limitations

Strengths of this study include the use of indirect calorimetry as a criterion reference for the validation of EE estimation, the use of a cross-validation sample to assess the validity of EE estimation on different smartphones, and the integration of geospatial and accelerometry data within a GIS. A limitation of this research relates to the data reduction methods of smartphone data. Unlike research-grade accelerometers, there are no established methods for managing smartphone accelerometer data. Periods of negative EE measurement bias may have indicated smartphone noncarry time; however, as methods for classifying phone and individual inactivity have yet to be determined, it was not possible to validate this assumption. We applied criteria typically used for accelerometers (removing >60 min of consecutive zero values), which may have assisted with agreement between methods but do not differentiate smartphone noncarry time from true inactivity. Further research is needed to characterize data patterns that can distinguish these use patterns, and this will have a significant impact on how smartphone data are interpreted. The increasingly common inclusion of embedded gyroscopic sensors could help to overcome this problem. It should be noted that stand-alone wearable accelerometers are also subject to this limitation, and prolonged nonwear time may result in greater data loss than intermittent smartphone noncarry time. Researchers should consider these limitations and select the type of measurement tool that best suits their experimental objectives. Additional limitations include the number and characteristics of participants and the limited period of free-living activity monitoring. As the primary sampling unit in this study is sensor observations not individuals (ie, 50 Hz accelerometer data), the study is appropriately powered to achieve validation objectives. Because participants were healthy adults, these results may not generalize to populations with

XSL•FO
RenderX

musculoskeletal limitations or medical conditions that modify the energetic demands of comparable PAs. Finally, the short 24-hour free-living validation period may limit the range of activities and activity intensity levels included in the free-living analysis. Although high frequency sensor sampling enables ample statistical power for validation objectives, future research may consider longer data collection periods that may include a wider range of activities.

Conclusions

The Movn smartphone app provided valid measurement of physical activity EE at low and moderate activity intensity levels; however, measurement validity was reduced during higher intensity activities. Given their ubiquity, integrated sensors, passive data collection, and potential to connect with external data streams, smartphones provide an ideal opportunity to enhance understanding of the nature and context of human movement, particularly at a population level. This presents future challenges for data processing and analytics, as well as opportunities to inform novel, responsive, and individualized interventional strategies.

Acknowledgments

We would like to acknowledge all the participants from Dublin City University for their involvement. This research was conducted as part of a European Union FP7 Research Exchange. Dr Maddison was supported by a Health Research Council, Sir Charles Hercus Fellowship.

Conflicts of Interest

Dr Harshvardhan Vathsangam owns a majority stake in Moving Analytics. Moving Analytics is the creator of the Movn smartphone app used in this study. This was an investigator-led study, and Moving Analytics did not have any influence over the design and conduct of the study or the interpretation of the data. The remaining authors do not have any conflicts of interest.

References

1. WHO. 2011. Physical inactivity: a global public health problem URL: http://www.who.int/dietphysicalactivity/factsheet_inactivity/en/ [accessed 2017-08-08] [WebCite Cache ID 6sYhmgDeg]
2. Nocon M, Hiemann T, Muller-Riemenschneider F, Thalau F, Roll S, Willich S. Association of physical activity with all-cause and cardiovascular mortality: a systematic review and meta-analysis. Eur J Cardiovasc Prev Rehabil 2008;15(3):239-246. [doi: 10.1097/HJR.0b013e3282f55e09] [Medline: 18525377]
3. Wen C, Wai J, Tsai M, Yang YC, Cheng TY, Lee M, et al. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. The Lancet 2011 Oct;378(9798):1244-1253. [doi: 10.1016/S0140-6736(11)60749-6] [Medline: 21846575]
4. Dumith S, Hallal P, Reis R, Kohl H. Worldwide prevalence of physical inactivity and its association with human development index in 76 countries. Prev Med 2011;53(1-2):24-28. [doi: 10.1016/j.ypmed.2011.02.017] [Medline: 21371494]
5. Hallal P, Bauman A, Heath G, Kohl IH, Lee I, Pratt M. Physical activity: more of the same is not enough. The Lancet 2012 Jul;380(9838):190-191. [doi: 10.1016/S0140-6736(12)61027-7]
6. Tully M, McBride C, Heron L, Hunter R. The validation of Fibit ZipTM physical activity monitor as a measure of free-living physical activity. BMC Res Notes 2014;7:952. [doi: 10.1186/1756-0500-7-952] [Medline: 25539733]
7. Hekler EB, Buman M, Grieco L, Rosenberger M, Winter S, 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]
8. Donaire-Gonzalez D, de Nazelle A, Seto E, Mendez M, Nieuwenhuijsen MJ, Jerrett M. Comparison of physical activity measures using mobile phone-based CalFit and Actigraph. J Med Internet Res 2013 Jun 13;15(6):e111 [FREE Full text] [doi: 10.2196/jmir.2470] [Medline: 23896156]
9. Kamel Boulos MN, Al-Shorbaji NM. On the internet of things, smart cities and the WHO healthy cities. Int J Health Geogr 2014 Mar 27;13:10 [FREE Full text] [doi: 10.1186/1476-072X-13-10] [Medline: 24669838]
10. Pfaeffli DL, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and internet support for coronary heart disease self-management: results from the text4heart randomized controlled trial. J Med Internet Res 2015 Oct 21;17(10):e237 [FREE Full text] [doi: 10.2196/jmir.4944] [Medline: 26490012]
11. Welk G, editor. Physical activity assessment for health-related research. Champaign, IL: Human Kinetics; 2002.
12. Percival G. Opengeospatial. 2013. Smart cities depend on smart location URL: http://www.opengeospatial.org/blog/1886 [accessed 2016-12-05] [WebCite Cache ID 6mW9CiOq1]
13. González MC, Hidalgo CA, Barabási AL. Understanding individual human mobility patterns. Nature 2008 Jun 05;453(7196):779-782. [doi: 10.1038/nature06958] [Medline: 18528393]
14. Maddison R, Jiang Y, Vander HS, Exeter D, Mhurchu CN, Dorey E. Describing patterns of physical activity in adolescents using global positioning systems and accelerometry. Pediatr Exerc Sci 2010 Aug;22(3):392-407. [Medline: 20814035]

XSL•FO
RenderX

15. Carlson J, Schipperijn J, Kerr J, Saelens B, Natarajan L, Frank L, et al. Locations of physical activity as assessed by GPS in young adolescents. Pediatrics 2016 Jan;137(1):- [FREE Full text] [doi: 10.1542/peds.2015-2430] [Medline: 26647375]
16. Takemoto M, Carlson J, Moran K, Godbole S, Crist K, Kerr J. Relationship between objectively measured transportation behaviors and health characteristics in older adults. Int J Environ Res Public Health 2015 Oct 30;12(11):13923-13937 [FREE Full text] [doi: 10.3390/ijerph121113923] [Medline: 26528999]
17. Welk GJ, McClain J, Ainsworth BE. Protocols for evaluating equivalency of accelerometry-based activity monitors. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S39-S49. [doi: 10.1249/MSS.0b013e3182399d8f] [Medline: 22157773]
18. Canadian Society for Exercise Physiology. CSEP. 2002. PAR-Q & you: physical activity readiness questionnaire (revised 2002) URL: http://www.csep.ca/CMFiles/publications/parq/par-q.pdf [WebCite Cache ID 6mWL06wY7]
19. 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]
20. Elia M, Livesey G. Energy expenditure and fuel selection in biological systems: the theory and practice of calculations based on indirect calorimetry and tracer methods. World Rev Nutr Diet 1992;70:68-131. [Medline: 1292242]
21. Vathsangam H, Emken A, Schroeder E, Spruijt-Metz D, Sukhatme G. Determining energy expenditure from treadmill walking using hip-worn inertial sensors: an experimental study. IEEE Trans Biomed Eng 2011 Oct;58(10):2804-2815 [FREE Full text] [doi: 10.1109/TBME.2011.2159840] [Medline: 21690001]
22. Altman D, Bland J. Measurement in medicine: the analysis of method comparison studies. The Statistician 1983 Sep;32(3):307-317. [doi: 10.2307/2987937]
23. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307-310. [Medline: 2868172]
24. Android Developers. Android. 2016. Android APIs: android.location getaccuracy URL: https://developer.android.com/reference/android/location/Location.html#getAccuracy() [WebCite Cache ID 6mWJZhvKu]
25. Hägerstrand T. What about people in regional science? Pap Reg Sci 1970;24(1):7-24. [doi: 10.1111/j.1435-5597.1970.tb01464.x]
26. Rawstorn JC, Maddison R, Ali A, Foskett A, Gant N. Rapid directional change degrades GPS distance measurement validity during intermittent intensity running. PLoS One 2014;9(4):e93693 [FREE Full text] [doi: 10.1371/journal.pone.0093693] [Medline: 24733158]
27. Graham D, Hipp J. Emerging technologies to promote and evaluate physical activity: cutting-edge research and future directions. Front Public Health 2014;2:66 [FREE Full text] [doi: 10.3389/fpubh.2014.00066] [Medline: 25019066]

Abbreviations

CO_2: carbon dioxide
EE: energy expenditure
GIS: geographic information software
GPS: global positioning system
ICC: intraclass correlation coefficient
MET: metabolic equivalent of task
O_2: oxygen
PA: physical activity
SPSS: Statistical Package for the Social Sciences
3D: three-dimensional
2D: two-dimensional
$\dot{V}O_2$: oxygen uptake
WHO: World Health Organization

Edited by G Eysenbach; submitted 14.12.16; peer-reviewed by D Donaire-Gonzalez, R Hunter; comments to author 26.01.17; revised version received 13.03.17; accepted 02.04.17; published 17.08.17.

Please cite as:
Maddison R, Gemming L, Monedero J, Bolger L, Belton S, Issartel J, Marsh S, Direito A, Solenhill M, Zhao J, Exeter DJ, Vathsangam H, Rawstorn JC
Quantifying Human Movement Using the Movn Smartphone App: Validation and Field Study
JMIR Mhealth Uhealth 2017;5(8):e122
URL: http://mhealth.jmir.org/2017/8/e122/
doi:10.2196/mhealth.7167
PMID:28818819

XSL•FO
RenderX

©Ralph Maddison, Luke Gemming, Javier Monedero, Linda Bolger, Sarahjane Belton, Johann Issartel, Samantha Marsh, Artur Direito, Madeleine Solenhill, Jinfeng Zhao, Daniel John Exeter, Harshvardhan Vathsangam, Jonathan Charles Rawstorn. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 17.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

How Accurate Is Your Activity Tracker? A Comparative Study of Step Counts in Low-Intensity Physical Activities

Parastoo Alinia[1*]; Chris Cain[1*]; Ramin Fallahzadeh[1*]; Armin Shahrokni[2*]; Diane Cook[1*]; Hassan Ghasemzadeh[1*]

[1]School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, United States

[2]Geriatrics / Gastrointestinal Oncology Service, Memorial Sloan-Kettering Cancer Center, New York, NY, United States

[*]all authors contributed equally

Corresponding Author:
Parastoo Alinia
School of Electrical Engineering and Computer Science
Washington State University
Pullman, Washington
Pullman, WA, 99164
United States
Phone: 1 509 335 3564
Fax: 1 509 335 3818
Email: parastoo.alinia@wsu.edu

Abstract

Background: As commercially available activity trackers are being utilized in clinical trials, the research community remains uncertain about reliability of the trackers, particularly in studies that involve walking aids and low-intensity activities. While these trackers have been tested for reliability during walking and running activities, there has been limited research on validating them during low-intensity activities and walking with assistive tools.

Objective: The aim of this study was to (1) determine the accuracy of 3 Fitbit devices (ie, Zip, One, and Flex) at different wearing positions (ie, pants pocket, chest, and wrist) during walking at 3 different speeds, 2.5, 5, and 8 km/h, performed by healthy adults on a treadmill; (2) determine the accuracy of the mentioned trackers worn at different sites during activities of daily living; and (3) examine whether intensity of physical activity (PA) impacts the choice of optimal wearing site of the tracker.

Methods: We recruited 15 healthy young adults to perform 6 PAs while wearing 3 Fitbit devices (ie, Zip, One, and Flex) on their chest, pants pocket, and wrist. The activities include walking at 2.5, 5, and 8 km/h, pushing a shopping cart, walking with aid of a walker, and eating while sitting. We compared the number of steps counted by each tracker with gold standard numbers. We performed multiple statistical analyses to compute descriptive statistics (ie, ANOVA test), intraclass correlation coefficient (ICC), mean absolute error rate, and correlation by comparing the tracker-recorded data with that of the gold standard.

Results: All the 3 trackers demonstrated good-to-excellent (ICC>0.75) correlation with the gold standard step counts during treadmill experiments. The correlation was poor (ICC<0.60), and the error rate was significantly higher in walker experiment compared to other activities. There was no significant difference between the trackers and the gold standard in the shopping cart experiment. The wrist worn tracker, Flex, counted several steps when eating (P<.01). The chest tracker was identified as the most promising site to capture steps in more intense activities, while the wrist was the optimal wearing site in less intense activities.

Conclusions: This feasibility study focused on 6 PAs and demonstrated that Fitbit trackers were most accurate when walking on a treadmill and least accurate during walking with a walking aid and for low-intensity activities. This may suggest excluding participants with assistive devices from studies that focus on PA interventions using commercially available trackers. This study also indicates that the wearing site of the tracker is an important factor impacting the accuracy performance. A larger scale study with a more diverse population, various activity tracker vendors, and a larger activity set are warranted to generalize our results.

(JMIR Mhealth Uhealth 2017;5(8):e106) doi:10.2196/mhealth.6321

KEYWORDS

activities of daily living; activity tracker; mobility limitations; mobile health

XSL•FO
RenderX

Introduction

Increasing physical activity (PA) level and maintaining a healthy diet are among the important factors to sustain and improve cardiovascular health [1-3]. Measuring daily PA is important to objectively monitor an individual's health. A simple, yet effective, approach to measure PA is to count the number of steps an individual takes in a given day [4]. Based on a common PA guideline, healthy adults are recommended to take 10,000 steps per day to maintain physical fitness and health [5]. However, common recommended levels of PA are challenging targets for less active individuals [6]. Various accelerometer-based devices such as Jawbone, Fitbit, Misfit, and Garmin have been developed for PA monitoring. These devices are small, non-invasive, easy-to-use, and provide an objective indicator of PA behavior [7]. Furthermore, they intend to avoid common sources of error in subjective measurement (eg, self-reporting) [4,8-10]. As researchers continue to utilize the commercially available PA trackers in clinical trials, there is a need to assess the accuracy of these trackers, particularly for low-intensity PAs and those that involve assistive devices [11-13]. Currently, the research community remains uncertain about how reliable off-the-shelf trackers are while performing the aforementioned PAs [14].

There have been several studies analyzing these trackers of which some are aimed at older adults and patients with chronic diseases [4,5,7-10,12,15-25]. These studies evaluated validity and reliability of activity trackers such as Fitbit (ie, Zip, One, Ultra, and Flex) during slow, moderate, and brisk walking in laboratory settings or free living environments. These research studies mostly utilized Yamax SW200 pedometer (Yamax Corporation, Tokyo, Japan), and ActiGraph GT3X (ActiGraph LLC,

Pensacola, FL, USA) as the gold standard step counters. A study presented by Lauritzen [22], investigated the accuracy of the commercially available trackers (eg, Fitbit) during low-intensity activities in older adults with reduced mobility. However, the study focused on a small number of participants (ie, 5 to 7 in each activity group) and wearing sites. Furthermore, although activity trackers are designed to be worn on certain positions on the body, the relationship between the wearing site and types of PAs that are best tracked by activity trackers is unknown till date. Prior research studied potential impacts of wearing site on the performance of the PA trackers [26,27]. Furthermore, the effect of sensor localization on activity monitoring performance has been studied in the past [28,29]. However, such studies did not investigate the influence of the wearing site of activity trackers on step count accuracy. By analyzing step count data gathered from trackers worn on different body positions, we can demonstrate if the wearing site can improve the step count accuracy for one or a subset of activities.

Our primary focus in this paper is to evaluate the reliability of commercially available trackers during low-intensity PAs and those activities requiring assistive tools. Specifically, our goals are to (1) determine the accuracy of 3 Fitbit devices, Zip, One, and Flex, worn at different wearing sites including pants pocket, chest, and wrist during walking at 2.5 km/h, 5 km/h, and 8 km/h performed by healthy adults on a treadmill; (2) determine the accuracy of the Fitbit trackers worn at different sites (ie, pants pocket, chest, and wrist) during real-life activities including walking with a shopping cart, walking with a walker, and eating; and (3) examine whether the intensity of PAs impacts the optimal wearing site of the tracker.

Methods

Pre-Study Experiment

Fitbit trackers use micro-electro-mechanical systems (MEMS) three-axis accelerometers to capture motion signals from users. It is important to mention that any disagreement among step numbers reported by the trackers was due to the wearing site of the tracker rather than the embedded signal-processing algorithm that infers step counts from the accelerometer signals. Therefore, prior to the main experiment, we performed a series of experiments to identify if the step counting algorithms of different trackers were similar in their performance while worn on the same wearing site.

The wearing site of each tracker was determined based on their specifications and convenience. Fitbit Flex is typically worn on the wrist. The extra movements of the upper body can contribute to an inability to detect steps correctly using Flex. To solve this problem, one can change the hand settings from "non-dominant" to "dominant." The dominant option in Fitbit decreases the sensitivity of step counting [30]. We activated the dominant hand option on Fitbit Flex for left-handed participants. Fitbit Zip is often worn on locations such as a shirt pocket, bra, pants pocket, belt, and waist. Fitbit One can be worn comfortably in a pocket or on a bra.

In this experiment, one of the participants wore Fitbit Zip, One and Flex on the same wearing site for 4 hours performing daily living routines in a free-living setting. The experiment was repeated 3 times with the following scenarios: wearing all 3 Fitbit trackers (1) on the left wrist, (2) on the left pants pocket, and (3) on the chest. After each experiment, we compared the number of the steps each tracker counted in each 5-minute interval with those of the other trackers. The 5-minute interval is the shortest time interval in which the Fitbit tracker numbers can be obtained through the online dashboard [31].

Main Experiment

We conducted a study with healthy adults who performed 6 PAs while wearing 3 different Fitbit trackers including Zip, One, and Flex on 3 different sites including chest, pants pocket, and wrist. Two sets of PAs were included in our experiment, including (1) walking on a treadmill at 3 different speeds, 2.5 km/h, 5 km/h, and 8 km/h and (2) real-life activities including walking with a shopping cart, walking with a walker, and eating an apple while sitting. In the treadmill experiment, participants were asked to walk on a treadmill at 3 different speeds for 5 minutes each. Our goal was to investigate the performance of the trackers in normal walking activities. In the shopping cart and walker experiments, participants walked with a shopping cart and a walker for 5 minutes each. In the eating activity, participants were asked to eat an apple while sitting on a chair.

XSL•FO
RenderX

To capture data from participants, 2 data collection methods were used: a motion sensor based activity tracker for recording the number of steps and a camera to record videos to keep track of the actual number of steps. In our analysis, the steps measured by the trackers were compared against the gold standard number extracted from the videos. We set the dominancy option of the Fitbit Flex of the left-handed and right-handed participant to "dominant hand" and "non-dominant hand," respectively. During the eating experiment, Fitbit Flex was worn on the dominant wrist. Therefore, we set its dominancy option to "dominant hand" for all the participants during the eating experiment. Figure 1 shows a participant wearing the trackers on left wrist, chest, and left pants pocket during a treadmill experiment.

Recruitment and Participants

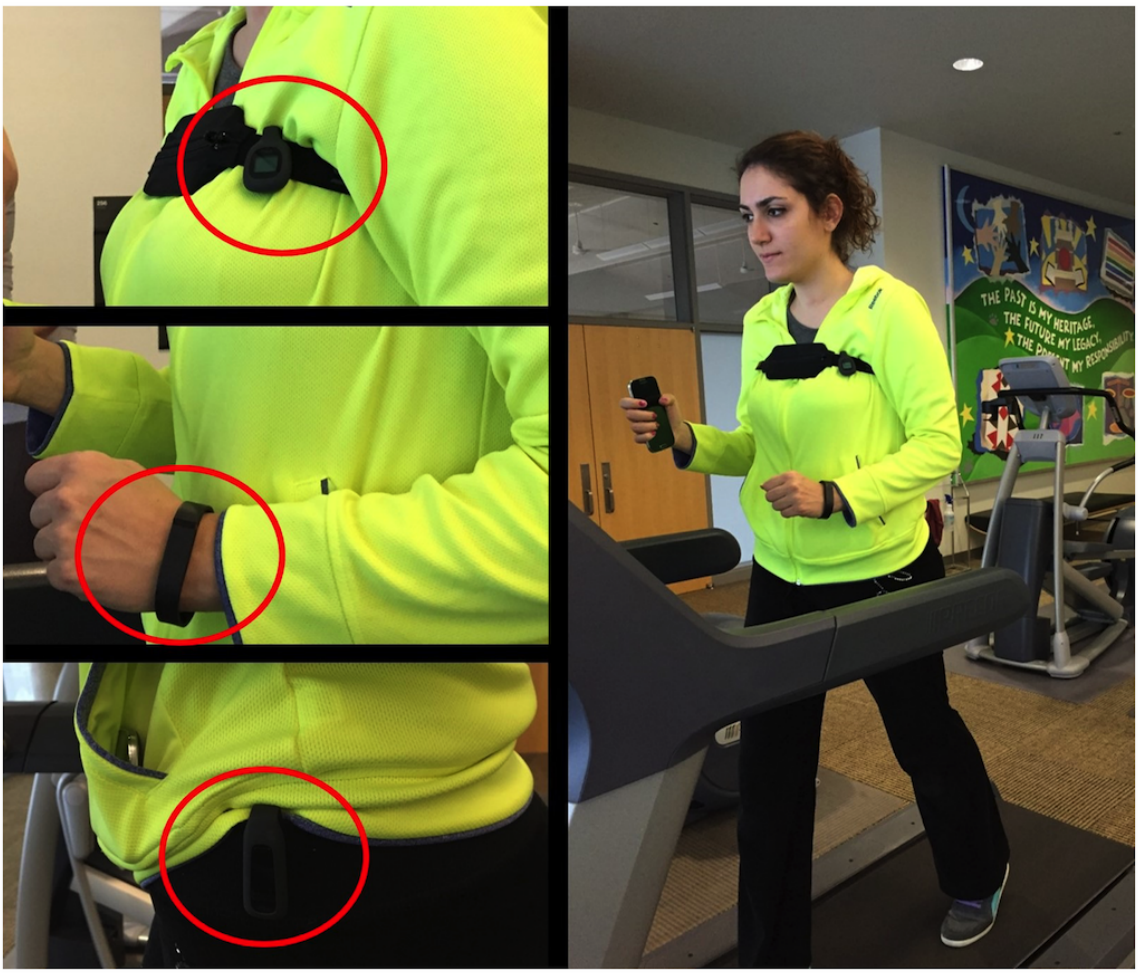
The study protocol was reviewed and approved by the Washington State University (WSU) Institutional Review Board (IRB). Inclusion criteria included the absence of gait affecting conditions such as fractures and broken bones as well as neurological impairments. Exclusion criteria included inability to walk with an assistive device, inability to walk on treadmill, and inability to conduct 30 minutes of light to moderate PA (MET<6) with multiple rests in between. Prior to data collection, all participants were informed by the study coordinator regarding the study aims, testing procedure, and methods. Participants completed an eligibility questionnaire regarding physical condition, age, and gender. The participants were recruited through direct contact as well as through an advertisement at the WSU School of EECS (Electrical Engineering & Computer Science).

Statistical Analysis

We first assessed systematic differences between trackers on 3 wearing sites and the gold standards by a one-way analysis of variance (ANOVA) test. We also defined 2 error parameters: *errors per minute* and *error rate*. Error per minute is defined as the difference between the number of steps recorded by the trackers and the actual number of steps in 1 minute. Error rate refers to the percentage of the misestimated. We note that for the eating experiment, given that the actual number of steps is 0, the absolute error is not reported.

Second, we evaluated the test-retest reliability of the tracker on each wearing site by computing the intraclass correlation coefficient (ICC) (2-way random, absolute agreement, single measures with 95% CI). We used common cut-off points for reliability assessment. The cut-off points were >0.90 (excellent), 0.75-0.90 (good), 0.60-0.75 (moderate), and <0.60 (low) [5].

Figure 1. A participant wearing 3 trackers on the wrist, chest, and pants pocket while walking on a treadmill.



XSL•FO
RenderX

Results

Pre-Study Results

A total of 15 healthy adults, including 7 females and 8 males, aged between years 21 and 31, were recruited to participate in this experimental study. Table 1 shows demographic information as well as PA statistics for the participant groups.

Table 2 shows the results of the *t* test analysis on pairs of the Fitbit trackers and gold standard during the pre-study test. The results show no significant difference in the number of steps recorded during the test experiment. Therefore, we conclude that any possible differences in the number of steps reported by various trackers used in the main study must be due to the tracker's wearing site for that specific task.

Main Results

Descriptive Statistics Analysis

Figure 2 shows the results of the ANOVA test (95% CI 0.55-1.00) on the mean value of the step counts captured by the trackers on each wearing site compared to gold standard values. There are no significant differences in recorded steps in each of the 3 wearing sites (chest, pants pocket, and wrist) for walking on treadmill and shopping cart experiments. In the walker experiment, all trackers exhibit a significant difference from the actual step count ($P<.001$). In the eating experiment, the wrist tracker occasionally counted steps due to the hand movements while eating the apple.

Systematic Difference and Mean Absolute Error Analysis

The performance of the trackers in step counting was compared using the average error rate and error per minute values of all the participants during the experiment. Figure 3 illustrates the performance of the trackers in different wearing sites during activities of daily living. This figure compares error per minute and error rate in counting steps. The error rate and error per minute metrics for the various trackers are detailed in Table 3.

Furthermore, the chest tracker recorded 0 steps during the eating experiment, the wrist tracker counted 4.8 steps per minute in the eating experiment, and the pants pocket tracker counted 0 steps during the eating experiment.

Table 1. Physical and demographical characteristics of the participants.

Variables	All (N=15)	Female (n=7)	Male (n=8)
Age in years	21-31	24-26	23-31
Height (cm)	155-189	155-178	175-189
Body mass (kg)	47-86	47-75	65-86
Body Mass Index (kgm^{-2})	19.0-24.8	19.0-24.8	20.1-24.8
Average number of steps taken in treadmill experiments	610	583	643
Average number of steps taken in walker experiment	236	231	240
Average number of steps taken in shopping cart experiment	417	400	418

Table 2. *t* test results on the number of the steps different Fitbit trackers count when used on the same wearing sites.

Wearing site	Trackers	Hypothesis	Probability
Wrist	Zip and One	0	.65
	Zip and Flex	0	.79
	One and Flex	0	.83
Chest	Zip and One	0	.69
	Zip and Flex	0	.83
	One and Flex	0	.85
Pants pocket	Zip and One	0	.84
	Zip and Flex	0	.78
	One and Flex	0	.93

XSL•FO
RenderX

Figure 2. The result of the ANOVA tests on the steps recorded by trackers worn on different wearing sites (chest, pocket, and wrist) with the gold standard number of steps.

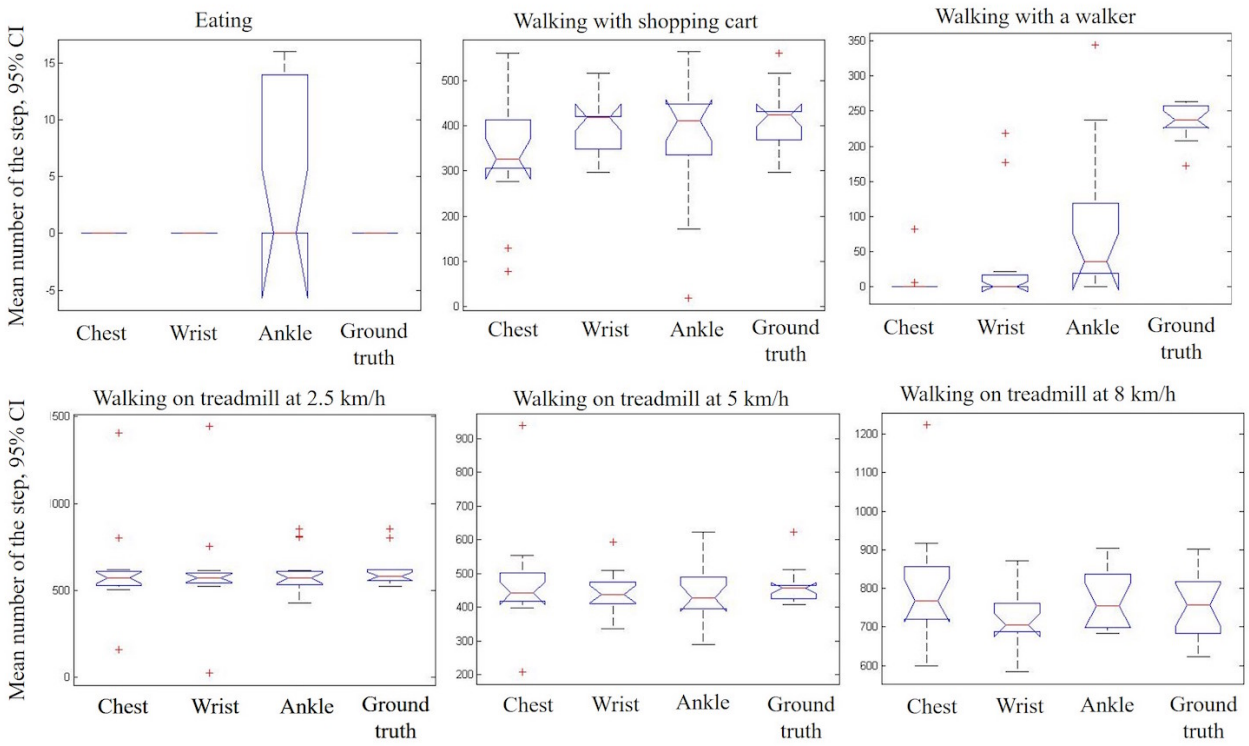
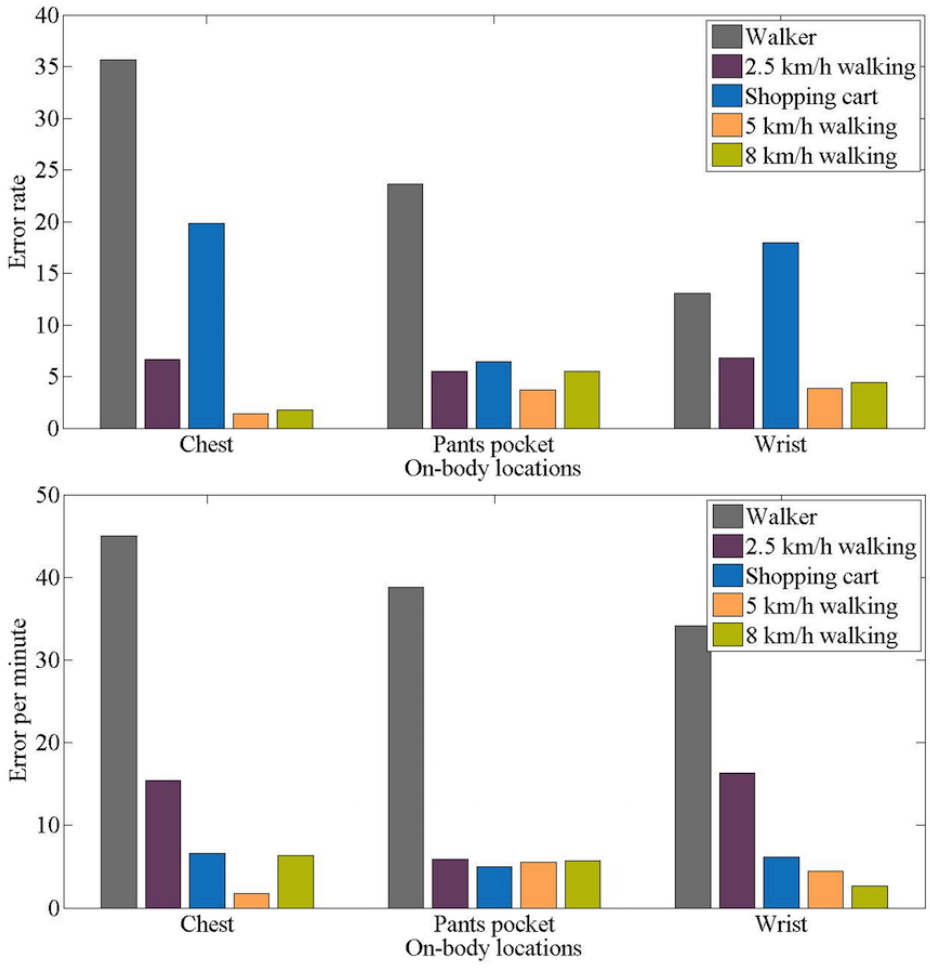


Figure 3. Error rate (left) and error per minute (right) of activity trackers in different wearing sites during the activities of daily living except eating.



XSL•FO
RenderX

Table 3. Error rate and error per minute values for trackers.

Wearing site	Activity	Error rate	Error per minute
Wrist	Walking with walker	73.1%	34.2
	Walking at 2.5 km/h	6.8%	6.2
	Walking with shopping cart	19.8%	16.4
	Walking at 5 km/h	3.9%	4.4
	Brisk walking at 8 km/h	4.4%	2.6
Chest	Walking with walker	95.6%	45.0
	Walking at 2.5 km/h	6.7%	6.6
	Walking with shopping cart	19.8%	15.4
	Walking at 5 km/h	1.4%	1.8
	Brisk walking at 8 km/h	1.8%	6.3
Pants pocket	Walking with walker	83.6%	38.8
	Walking at 2.5 km/h	5.5%	5.0
	Walking with shopping cart	6.4%	5.9
	Walking at 5 km/h	3.7%	5.5
	Brisk walking at 8 km/h	5.5%	5.7

Table 4. Intraclass correlation coefficients between the wearing sites and gold standard values.

Wearing site	Activity	ICC	95% CI
Wrist	Treadmill walking at 2.5 km/h	0.76	0.61-0.94
	Treadmill walking at 5 km/h	1.07	0.91-0.99
	Treadmill brisk walking at 8 km/h	1.07	0.90-0.98
	Walking with the walker	<0.01 (0.003)	0.68-0.99
	Walking with the shopping cart	0.10	0.73-0.87
	Eating	0.00	0.67-1.00
Chest	Treadmill walking at 2.5 km/h	0.70	0.56-0.93
	Treadmill walking at 5 km/h	1.37	0.97-0.99
	Treadmill brisk walking at 8 km/h	0.68	0.55-0.92
	Walking with the walker	<0.01 (0.006)	0.87-0.97
	Walking with the shopping cart	0.61	0.74-0.79
	Eating	N/A[a]	N/A
Pants pocket	Treadmill walking at 2.5 km/h	0.91	0.73-0.97
	Treadmill walking at 5 km/h	0.83	0.67-0.95
	Treadmill brisk walking at 8 km/h	0.71	0.58-0.93
	Walking with the walker	0.02	0.82-0.95
	Walking with the shopping cart	0.74	0.83-0.87
	Eating	N/A	N/A

[a]N/A: not available.

Correlation Analysis

Table 4 shows the ICC between the trackers on each wearing site and the gold standard for each PA type. The ICC values ranged from 0.56 (chest) to 0.97 (pants pocket) for walking on the treadmill at 2.5 km/h, 0.55 (chest) to 0.99 (wrist) for walking on the treadmill at 5 km/h, 0.58 (pants pocket) to 0.98 (wrist) walking on the treadmill at 8 km/h, 0.68 (wrist) to 0.99 (wrist) for walking with the shopping cart, 0.74 (chest) to 0.87 (wrist) for walking with the walker, and 0.67 (wrist) to 1.00 (wrist) for eating.

XSL•FO
RenderX

Discussion

We evaluated the performance of 3 Fitbit trackers during 2 sets of PAs: (1) Easy to monitor activities such as walking on a treadmill, and (2) real-life activities that are potentially harder to monitor such as walking with a walker, walking with a shopping cart, and eating. Among these activities, the ones performed with lower intensity such as walking with a walker and walking at 2.5 km/h may represent some of the activities that older adults perform regularly. While we acknowledge that a limitation of our study is the lack of older adult participants, this study may have implications for utilizing activity trackers in populations that routinely perform PAs with a lower intensity or those that involve carrying walking aids.

The fact that the utilized trackers in this study were most accurate during treadmill walking can be explained by the controlled nature of human gait during treadmill walking as well as the fact that in normal gait the body will not exert extra movements to control the balance. As a result, trackers can easily detect each strike during treadmill walking. We also observed that decreasing or increasing the walking pace from normal speed reduces the accuracy of step counting. In non-treadmill activities, step detection becomes less accurate because movement patterns deviate from typical human gait patterns.

Our systematic differences analysis revealed that the intensity of PA impacts the choice of optimal wearing site of the tracker. Our results showed that the chest was the best site for more intense activities such as moderate walking at 5 km/h and brisk walking at 8 km/h, while the same site was least accurate in low intensity activities such as walking at 2.5 km/h and walking with the walker. This is consistent with the results by other researchers who reported the waist as the least accurate site for tracking low-intensity activities [4]. Furthermore, our results show that wrist was the most promising site during less intense activities such as walking at 2.5 km/h and walking with a walker. These results confirm prior findings by Diaz et al [15], who discovered that a wrist tracker had the biggest difference during slow, moderate and brisk walking on the treadmill. Moreover, our results showed that the pants pocket was a better wearing site in terms of step counting accuracy while pushing a shopping cart. Yet, this wearing site was least accurate during intense activities. This finding is also consistent with the results obtained in several prior studies such as a study by Mammen et al [4].

Looking at the result of the error-per-minute for all of the activities, one could conclude that there are 2 triggers to accurately detect the steps in PA. First, more intense activities are easier to detect. Second, steps in activities that better imitate normal walking such as walking at 5 km/h and 8 km/h are better identifiable compared to activities with abnormal walking patterns such as walking with a walker. Crouter et al [16], examined the accuracy of 5 electronic pedometers and found that pedometers are less accurate at slower walking speeds. In another work performed by Thorup et al [23], Fitbit Zip demonstrated high accuracy (absolute error <3%) on the walking speeds of 3.6 km/h and higher.

In our ICC analysis, the wrist site (ie, Flex) showed good-to-excellent correlation (ICC>0.75) during treadmill walking. Several prior studies indicated good correlation (ICC 0.75-0.90) between the wrist tracker (Flex) and the gold standard as well as the experiments conducted by Kooiman et al [5] on the accuracy of 10 consumer level activity trackers.

A study by Beets et al [32] on the accuracy of the pedometers in youth with developmental disabilities indicated a low correlation (ICC<0.60) during a shopping cart experiment [7,8]. In our study, the site pants pocket (ie, One) demonstrated good-to-excellent correlation (ICC>0.75) as well. This amount of correlation is smaller than the prior results by Beets et al. They identified that a hip worn tracker demonstrated excellent correlation (ICC 0.97-0.99) with the gold standard values [8]. This can be explained by the fact that a hip tracker is potentially more stationary in body coordinates compared to a pants pocket tracker than can be potentially loose during human gait. We observed a moderate correlation (0.60<ICC<0.75) with the gold standard while walking with the shopping cart.

In our analysis, the chest site had a moderate correlation in treadmill walking at 2.5 km/h and 5 km/h, while it was excellent during walking on the treadmill at 8 km/h. In a research study, the upper body Fitbit tracker had a low correlation (ICC 0.1-0.4) for low-intensity treadmill walking (1.7-2.7 km/h) [4]. It showed moderate correlation (0.60<ICC<0.75) with gold standard during the shopping cart experiment.

Our data show that accelerometer-based trackers provide reliable measures for moderate and brisk walking on treadmill. The performance of these trackers decline as the walking speed increases or decreases from the normal walking pace. Moreover, the accelerometer-based trackers demonstrate a major performance drop in step detection during activities that deviate from a normal walking pattern. An example of such abnormal walking patterns is when walking with a walker. This may suggest that utilizing commercially available trackers in studies that involve low-speed activities and those involving walking aids requires either careful selection of an activity tracker robust to such activities or exclusion of participants whose routine behavior involves these activities.

Furthermore, our result, in particular those related to walker and shopping cart experiments, may suggest a need for developing more advanced algorithmic solutions that consider tracker wearing sites as well as activity intensity into account while computing step counts. Accelerometer-based step detection algorithms usually try to find one stride by detecting the minimum and maximum peaks from the accelerometer sensor signal. Finding a stride pattern in the acceleration signal of irregular walking patterns such as walking with a walker or a shopping cart can be more challenging that requires real-time adaptation of the underlying algorithms to personalize the step counting methods for each individual and with respect to the changing context of the user [23]. Developing such an algorithm first requires detecting the user's activity type in real-time and second identifying the wearing site of the sensor to obtain the highest accuracy. Using advances discussed in recent studies, we are able to locate the sensor on the user's body coordinate by identifying the wearing site [28,29]. Furthermore, the study

XSL•FO
RenderX

presented by Krishnan et al [33]and other researchers on activity recognition suggest that we can use machine learning algorithms to identify PA types and wearing sites of the sensors from accelerometer sensor data.

Conflicts of Interest

None declared.

References

1. Dontje M, van der Wal MH, Stolk R, Brügemann J, Jaarsma T, Wijtvliet P, et al. Daily physical activity in stable heart failure patients. J Cardiovasc Nurs 2014;29(3):218-226. [doi: 10.1097/JCN.0b013e318283ba14] [Medline: 23416939]
2. DeFina LF, Haskell W, Willis B, Barlow C, Finley C, Levine B, et al. Physical activity versus cardiorespiratory fitness: two (partly) distinct components of cardiovascular health? Prog Cardiovasc Dis 2015;57(4):324-329. [doi: 10.1016/j.pcad.2014.09.008] [Medline: 25269066]
3. Johns DJ, Hartmann-Boyce J, Jebb SA, Aveyard P, Behavioural Weight Management Review Group. Diet or exercise interventions vs combined behavioral weight management programs: a systematic review and meta-analysis of direct comparisons. J Acad Nutr Diet 2014 Oct;114(10):1557-1568 [FREE Full text] [doi: 10.1016/j.jand.2014.07.005] [Medline: 25257365]
4. Mammen G, Gardiner S, Senthinathan A, McClemont L, Stone M, Faulkner G. Is this Bit Fit? Measuring the quality of the Fitbit step-counter. HFJC 2012;5(4):30-39 [FREE Full text]
5. Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
6. Varma VR, Tan EJ, Wang T, Xue Q, Fried LP, Seplaki CL, et al. Low-intensity walking activity is associated with better health. J Appl Gerontol 2014 Oct;33(7):870-887 [FREE Full text] [doi: 10.1177/0733464813512896] [Medline: 24652915]
7. Gusmer R, Bosch TA, Watkins A, Ostrem JD, Dengel DR. Comparison of FitBit® Ultra to ActiGraph™ GT1M for assessment of physical activity in young adults during treadmill walking. TOSMJ 2014 Apr 04;8(1):11-15 [FREE Full text] [doi: 10.2174/1874387001408010011]
8. Lee J, Welk G. Validity of consumer-based physical activity monitors and calibration of smartphone for prediction of physical activity energy expenditure. Med Sci Sports Exerc 2013;46(9):1840-1848 [FREE Full text] [doi: 10.1249/MSS.0000000000000287] [Medline: 1500564508]
9. Gusmer J, Bosch T, Ostrem J, Dengel D, Ph D. conservancy.umn. 2012. Validation of the FitBit ® for Physical Activity Quantification URL: https://conservancy.umn.edu/handle/11299/139972 [accessed 2017-06-26] [WebCite Cache ID 6rV8Fg2R2]
10. 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 Dec 23;7:952 [FREE Full text] [doi: 10.1186/1756-0500-7-952] [Medline: 25539733]
11. West DS, Monroe CM, Turner-McGrievy G, Sundstrom B, Larsen C, Magradey K, et al. A technology-mediated behavioral weight gain prevention intervention for college students: controlled, quasi-experimental study. J Med Internet Res 2016 Jun 13;18(6):e133 [FREE Full text] [doi: 10.2196/jmir.5474] [Medline: 27296086]
12. Wang J, Cadmus-Bertram L, Natarajan L, White M, Madanat H, Nichols J, et al. Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed J E Health 2015 Oct;21(10):782-792 [FREE Full text] [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]
13. Cadmus-Bertram L, Marcus B, Patterson R, Parker B, Morey B. Randomized trial of a Fitbit-based physical activity intervention for women. Am J Prev Med 2015 Sep;49(3):414-418 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]
14. Cadmus-Bertram L. Using fitness trackers in clinical research: what nurse practitioners need to know. J Nurse Pract 2017;13(1):34-40. [doi: 10.1016/j.nurpra.2016.10.012]
15. Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, et al. Fitbit®: an accurate and reliable device for wireless physical activity tracking. Int J Cardiol 2015 Apr 15;185:138-140 [FREE Full text] [doi: 10.1016/j.ijcard.2015.03.038] [Medline: 25795203]
16. Crouter SE, Schneider PL, Bassett DR. Spring-levered versus piezo-electric pedometer accuracy in overweight and obese adults. Med Sci Sports Exerc 2005 Oct 1;37(10):1673-1679 [FREE Full text] [doi: 10.1249/01.mss.0000181677.36658.a8]
17. Adam NJ, Spierer DK, Gu J, Bronner S. Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. J Med Eng Technol 2013 Oct;37(7):456-462. [doi: 10.3109/03091902.2013.831135] [Medline: 24007317]
18. Fulk G, Combs S, Danks K, Nirider C, Raja B, Reisman D. Accuracy of 2 activity monitors in detecting steps in people with stroke and traumatic brain injury. Phys Ther 2014;94(2):222-229 [FREE Full text] [doi: 10.2522/ptj.20120525]

XSL•FO
RenderX

19. Guo F, Li Y, Kankanhalli MS, Brown MS. An evaluation of wearable activity monitoring devices. : ACM; 2013 Presented at: Proceedings of the 1st ACM International Workshop on Personal Data Meets Distributed Multimedia; 2013; Barcelona, Spain URL: http://dl.acm.org/citation.cfm?doid=2509352.2512882 [doi: 10.1145/2509352.2512882]
20. Storm F, Heller B, Mazzà C. Step detection and activity recognition accuracy of seven physical activity monitors. PLoS One 2015 Jan;10(3) [FREE Full text] [doi: 10.1371/journal.pone.0118723]
21. Fortune E, Lugade V, Morrow M, Kaufman K. Validity of using tri-axial accelerometers to measure human movement--Part II: Step counts at a wide range of gait velocities. Med Eng Phys 2014;36(6):659-669 [FREE Full text] [doi: 10.1016/j.medengphy.2014.02.006]
22. Lauritzen J, Muñoz A, Luis SJ, Civit A. The usefulness of activity trackers in elderly with reduced mobility: a case study. Stud Health Technol Inform 2013;192:759-762. [Medline: 23920659]
23. Thorup C, Andreasen J, Sørensen EE, Grønkjær M, Dinesen BH, Hansen J. Accuracy of a step counter during treadmill and daily life walking by healthy adults and patients with cardiac disease. BMJ Open 2017 Mar 31;7(3):e011742 [FREE Full text] [doi: 10.1136/bmjopen-2016-011742] [Medline: 28363918]
24. Block V, Pitsch E, Tahir P, Cree B, Allen D, Gelfand J. Remote physical activity monitoring in neurological disease: a systematic review. PloS one 2016;11(4) [FREE Full text] [doi: 10.1371/journal.pone.0154335]
25. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015;12:42. [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
26. Cleland I, Kikhia B, Nugent C, Boytsov A, Hallberg J, Synnes K, et al. Optimal placement of accelerometers for the detection of everyday activities. Sensors (Basel) 2013;13(7):9183-9200 [FREE Full text] [doi: 10.3390/s130709183] [Medline: 23867744]
27. Alinia P, Saeedi R, Mortazavi B, Rokni A, Ghasemzadeh H. Impact of sensor misplacement on estimating metabolic equivalent of task with wearables. 2015 Presented at: IEEE 12th Int Conf Wearable Implant Body Sens Networks Internet IEEE; 9-12 June, 2015; Cambridge, MA, USA URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7299385
28. Saeedi R, Purath J, Venkatasubramanian K, Ghasemzadeh H. Toward seamless wearable sensing: Automatic on-body sensor localization for physical activity monitoring. In: Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE.: IEEE; 2014 Presented at: Engineering in Medicine and Biology Society (EMBC); 26-30 August, 2014; Chicago, IL, USA p. 5385-5388 URL: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6944843 [doi: 10.1109/EMBC.2014.6944843]
29. Saeedi R, Amini N, Ghasemzadeh H. Patient-centric on-body sensor localization in smart health systems. 2014 Presented at: 48th Asilomar Conf Signals, Syst Comput Internet IEEE; 2-5 November, 2014; Pacific Grove, CA, USA URL: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=7094840
30. help.fitbit. How do I measure and adjust my stride length? URL: http://help.fitbit.com/articles/en_US/Help_article/1135 [accessed 2017-07-11] [WebCite Cache ID 6rs0wXgUP]
31. Fitbit. URL: https://www.fitbit.com/home [accessed 2017-06-26] [WebCite Cache ID 6rWAVyXIY]
32. Beets MW, Combs C, Pitetti KH, Morgan M, Bryan RR, Foley JT. Accuracy of pedometer steps and time for youth with disabilities. Adapt Phys Activ Q 2007 Jul;24(3):228-244. [Medline: 17916919]
33. Krishnan N, Cook D. Activity recognition on streaming sensor data. Pervasive Mob Comput Internet 2014 Mar 1;10:138-154 [FREE Full text] [doi: 10.1016/j.pmcj.2012.07.003]

Abbreviations

ICC: intraclass correlation coefficient
MEMS: micro-electro-mechanical systems
PA: physical activity

Edited by G Eysenbach; submitted 06.07.16; peer-reviewed by D Glance, S Gomersall, C Thorup; comments to author 19.11.16; revised version received 07.02.17; accepted 11.05.17; published 11.08.17.

Please cite as:
Alinia P, Cain C, Fallahzadeh R, Shahrokni A, Cook D, Ghasemzadeh H
How Accurate Is Your Activity Tracker? A Comparative Study of Step Counts in Low-Intensity Physical Activities
JMIR Mhealth Uhealth 2017;5(8):e106
URL: http://mhealth.jmir.org/2017/8/e106/
doi:10.2196/mhealth.6321
PMID:28801304

XSL•FO
RenderX

©Parastoo Alinia, Chris Cain, Ramin Fallahzadeh, Armin Shahrokni, Diane Cook, Hassan Ghasemzadeh. Originally published in JMIR Mhealth and Uhealth (http://mhealth.jmir.org), 11.08.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX