JMIR mHealth and uHealth

Impact Factor (2022): 5.0
Volume 3 (2015), Issue 2 ISSN 2291-5222 Editor in Chief: Laurie Buis, PhD, MSI

Contents

Viewpoints

Original Papers

Short Paper

Review

Tutorial

Viewpoint

Integrating mHealth and Systems Science: A Combination Approach to Prevent and Treat Chronic Health Conditions

Nicolas Michel Oreskovic[1,2], MD, MPH; Terry T Huang[3], CPH, MPH, PhD; Jon Moon[4], PhD

[1]Division of General Academic Pediatrics, Department of Pediatrics, Massachusetts General Hospital, Boston, MA, United States

[2]Department of Pediatrics, Harvard Medical School, Boston, MA, United States

[3]School of Public Health, City University of New York, New York, NY, United States

[4]MEI Research, Ltd., Edina, MN, United States

Corresponding Author:
Nicolas Michel Oreskovic, MD, MPH
Division of General Academic Pediatrics
Department of Pediatrics
Massachusetts General Hospital
100 Cambridge Street, 5th Fl.
Boston, MA, 02116
United States
Phone: 1 617 726 0593
Fax: 1 617 726 1886
Email: noreskovic@mgh.harvard.edu

Abstract

Chronic health conditions are a growing global health concern and account for over half of all deaths worldwide. Finding ways to decrease the burden of and resources allotted to chronic health conditions is of primary importance. Recent advances in technology and insights into modeling techniques offer promising approaches, which if combined, represent a novel direction that would further advance the prevention and treatment of chronic health conditions.

(JMIR mHealth uHealth 2015;3(2):e62) doi:10.2196/mhealth.4150

KEYWORDS

mHealth; systems science; chronic health conditions; obesity; physical activity

Introduction

Chronic health conditions (CHC) account for over half of all deaths worldwide. In the United States, 1 in 4 adults and 1 in 15 children suffer from two or more CHC, with 86% of all health care dollars spent on the treatment of CHC [1]. Obesity alone is a major population burden, with 2 out of 3 adults and 1 out of 3 youth overweight or obese in the United States [2]. Many CHC, such as obesity, are difficult to treat due to their complex etiologies; managing such complicated multifactorial health conditions continues to tie up valuable health care resources and accounts for substantial health care costs. Finding ways to decrease the costs and resources allotted to CHC is of primary importance. Recent advances in technology and insights into the use of systems science are promising approaches. Systems science is used to understand complex connections between structure and behavior in a system over time. Mobile health technology (mHealth) and systems science are two fields that have separately been applied to address complex problems with some success. A combined mHealth and systems science approach would represent a novel direction that would further advance the prevention and treatment of CHC.

mHealth

mHealth is increasingly recognized as a tool to manage CHC, reduce health disparities, and address complex health problems [3-5]. In part, mHealth accomplishes this by automating processes that are currently resource-heavy. mHealth describes an array of technologies that encompass wireless sensors, software, and mobile phones worn and accessed by caregivers, patients, and individuals interested in being engaged in their own health to facilitate collecting and communicating health-related data. The nearly universal presence of mHealth technology throughout most areas of health care and ubiquity of mobile phones in the population, including underserved populations, hold great promise for increasing the integration of empirical real-time data into clinical practice and expanding the reaches of health care delivery, all to the benefit of patients'

health [3,6]. mHealth technology also offers unique advantages by being able to collect objective, patient-generated sensor data, such as activity and location, that are particularly important for CHC, given the underlying importance of health habits in most CHC [7]. Beyond using location and activity data, mHealth has the added ability to adapt to changes in a user's location and activity, to changes in the user's environment (eg, seasons), and to adapt to users (based on continuous feedback). As mHealth offers the power to dramatically expand access to care through objective data collection, increased efficiency, and enhanced data-driven practice, the next step is to embed systems science insights from diverse fields into mHealth technology, so that we can enhance the capability of mHealth as an effective and scalable tool for the prevention and treatment of CHC. One practical example of this combined approach is in obesity care; while an abundance of obesity- and weight management-related mHealth applications have surfaced, these applications have yet to incorporate systems science approaches.

Systems Science

While mHealth increases efficiency through engineering solutions, systems science provides a theory-based approach to improve both efficiency and effectiveness. Systems science is a particularly useful frame to understand complex relationships that are inherent in many CHC affected by biologic, social, and environmental factors [8]. A systems-oriented approach provides several key techniques to change behavior-an important factor in the outcome of most CHC-including information feedback, ownership, collaboration, competition, accountability, and rewards. A systems approach works by helping to identify which "levers" to pull, and in which order, in a series of often nonlinear complex relationships, thus allowing complicated and dynamic relationships to be understandable, actionable, and efficient while providing the optimal impact. There is growing interest in the scientific community in applying systems techniques to understand and disentangle the complex relationships and etiologies underlying CHC, exemplified in two 2010 Institute of Medicine reports, one on obesity prevention [9] and one on improving population health [10].

mHealth-Systems Science

Both mHealth and systems science offer to advance solutions for complex CHC by streamlining efforts and focusing on the most actionable levers. The two fields are aligned naturally and offer complementary approaches to improving the delivery of CHC treatments. While using mHealth alone can impact efficiency by automating steps and scaling up reach, an approach already being employed to treat various chronic diseases including obesity, diabetes, and tobacco addiction [11-14], it is not sufficient to marshal mHealth resources to automatize complex health problems. To truly impact the morbidity and mortality associated with complex CHC, a multi-step staged approach is required, an approach capable of identifying the steps worth automating and those which can be omitted, an approach that will increase data processing and data analysis speeds, increase efficiency, and decrease human error without impacting the desired primary outcome. The starting point in this engineering approach might be to identify all the known and anticipated steps in an intervention, including data collection, data processing, and decision points [15]. A weight analysis may follow, wherein the steps that are deemed to be most salient are prioritized for automation. The exact determination of weights may vary depending on the CHC, intended treatment, and resources available. For example, importance may be placed on cost, human labor, processing time requirements, patient preference, or anticipated prevalence of uptake. Once steps have been identified and prioritized, sequential automation can ensue until all the steps which have been deemed worth automating have been automated or until available resources (eg, time, money) have been depleted. The specific sequence and methods used for each staged approach will vary depending on the CHC and the treatment being delivered; however, the common goal should be to create an automated solution that is efficient, parsimonious (using the fewest possible resources necessary to achieve a desired outcome), and tailored.

The extent of automation may be integrated or complete. In integrated automation, the goal is to enhance and facilitate human-based care; creating a complementary approach where complex and resource-heavy programs are made more efficient through automation to facilitate CHC treatments. In complete automation, the goal of automation is to remove the health care provider from the equation altogether, such that all aspects of CHC management, including data collection, integration, and analysis with feedback loops, are managed electronically. Although both scenarios are achievable, this article focuses on integrated approaches and solutions.

We recently designed, and are pilot testing, an intervention to increase physical activity in obese adolescents with behavior change counseling guided by providing the adolescents with objective information on their location and activity [16]. The intervention was proven feasible and is effective among adolescents. The heavy manual burden and various decisions surrounding processing steps and behavioral counseling options, however, illustrate the complexity of promoting healthy behavior change and reveal the study's limited scalability in the absence of enhanced mHealth capabilities. In this case, mHealth can be expanded to automate key processing steps (eg, combining objective location and activity data) and decision points, in addition to improving communication with participants.

The capabilities of mHealth technology can be augmented by incorporating systems insights directly within it. For example, intervening directly on information and social feedbacks is a central strategy in a systems approach. Obesity has been identified as one CHC where using systems insights to tailor feedback can be particularly impactful [17]. These feedbacks can maximize successful behavior change in several ways: (1) participant ownership in a health intervention can be increased by directly allowing users to see and use their own data. In addition, ownership can be further enhanced by allowing participants to customize their mHealth application and design alternative solutions for health behavior goals; (2) collaboration can be improved by allowing study participants to be linked to their health care provider or health coach directly via their mHealth application and select whom they wish to have involved

in counseling sessions (eg, family, parents, friends, health care providers, or health coaches). In addition, the technology platform can also connect participants who have similar activity goals, thus creating emergent social networks that are goal-oriented and provide additional support to participants; (3) competition can be further leveraged by allowing individuals or groups to accumulate points over time, with top performers viewable to everyone on a common mHealth application; (4) goal-setting is often used to arrive at a personalized health behavior goal, mutually agreed upon by the participant, parent, and pediatrician. Accountability can be further enhanced using ecologic momentary assessment techniques [18], by sending reminders and connecting with participants to re-enforce commitment to the program. Further, the technology platform can also allow parents, health providers or other members of participants' online social network to directly check in with participants when noncompliance with a goal is detected. Rewards can be built into each system lever as incentives to maximize the likelihood of changing behavior. Several of these techniques, including data organization, competition, and feedback, have previously been proposed as part of a gaming approach to obesity self-management [19]. Gaming traditionally limits its focus on the use of "controlling" devices (eg, keyboard, mouse, joystick), and involves tight but limited and inflexible loops for the user. A combined systems-science mHealth approach offers the ability to broaden the approach, by also employing "monitoring" devices (eg, sensors), to collect objective patient-generated data, and feeding this data into feedback and counseling. A combined approach has the further ability to tailor feedback, providing an optimized combination of human and machine intelligence, with content and feedback means adapted to the user.

A synergistic "mHealth-systems science" approach should be part of the next-generation public health interventions for CHC. Leveraging systems insights within mHealth technology can further optimize behavior change strategies. As technology becomes ever more affordable, this combined approach offers a promising vehicle for scaling up systems-oriented behavior change interventions to the whole population.

Conflicts of Interest

Jon Moon is an employee of MEI Research, Ltd. The authors have no financial or other conflicts or interests to report.

References

1. Centers for Disease ControlPrevention (CDC). Chronic disease prevention and health promotion (12/01). 2014. Chronic disease prevention and health promotion (12/01) URL: http://www.cdc.gov/chronicdisease/overview/index.htm [accessed 2015-05-27] [WebCite Cache ID 6YqHSdibM]
2. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011-2012. JAMA 2014 Feb 26;311(8):806-814. [doi: 10.1001/jama.2014.732] [Medline: 24570244]
3. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]
4. Rivers BM, Bernhardt JM, Fleisher L, Green BL. Opportunities and challenges of using technology to address health disparities. Future Oncol 2014 Mar;10(4):519-522. [doi: 10.2217/fon.14.17] [Medline: 24754582]
5. Arya M, Kumar D, Patel S, Street RL, Giordano TP, Viswanath K. Mitigating HIV health disparities: the promise of mobile health for a patient-initiated solution. Am J Public Health 2014 Dec;104(12):2251-2255. [doi: 10.2105/AJPH.2014.302120] [Medline: 25322292]
6. Smith A. Pew Research Center. Smartphone Ownership 2013 Update URL: http://www.pewinternet.org/2013/06/05/smartphone-ownership-2013/ [accessed 2015-05-27] [WebCite Cache ID 6Yq1o50xC]
7. Stanley KG, Osgood ND. The potential of sensor-based monitoring as a tool for health care, health promotion, and research. Ann Fam Med 2011 Aug;9(4):296-298 [FREE Full text] [doi: 10.1370/afm.1292] [Medline: 21747099]
8. Huang TT, Drewnosksi A, Kumanyika S, Glass TA. A systems-oriented multilevel framework for addressing obesity in the 21st century. Prev Chronic Dis 2009 Jul;6(3):A82 [FREE Full text] [Medline: 19527584]
9. Institute of Medicine (IOM). Bridging the gap in obesity medicine: A framework to inform decision making. Washington, DC: The National Academies Press; 2010. URL: https://www.iom.edu/~/media/Files/Report%20Files/2010/Bridging-the-Evidence-Gap-in-Obesity-Prevention/Bridging%20the%20Evidence%20Gap%202010%20%20Report%20Brief.pdf [accessed 2015-05-27] [WebCite Cache ID 6Yq203vUA]
10. Institute of Medicine (IOM). For the public's health: The role of measurement in action and accountability. Washington, DC: The National Academies Press; 2011. URL: https://www.iom.edu/~/media/Files/Report%20Files/2010/For-the-Publics-Health-The-Role-of-Measurement-in-Action-and-Accountability/For%20the%20Publics%20Health%202010%20Report%20Brief.pdf [accessed 2015-05-27] [WebCite Cache ID 6Yq254Nr4]
11. Woolford SJ, Clark SJ, Strecher VJ, Resnicow K. Tailored mobile phone text messages as an adjunct to obesity treatment for adolescents. J Telemed Telecare 2010;16(8):458-461 [FREE Full text] [doi: 10.1258/jtt.2010.100207] [Medline: 20959393]

12. Brendryen H, Drozd F, Kraft P. A digital smoking cessation program delivered through internet and cell phone without nicotine replacement (happy ending): randomized controlled trial. J Med Internet Res 2008;10(5):e51 [FREE Full text] [doi: 10.2196/jmir.1005] [Medline: 19087949]
13. Knight E, Stuckey MI, Petrella RJ. Health promotion through primary care: enhancing self-management with activity prescription and mHealth. Phys Sportsmed 2014 Sep;42(3):90-99. [doi: 10.3810/psm.2014.09.2080] [Medline: 25295771]
14. Khairat S, Garcia C. Developing an mHealth framework to improve diabetes self-management. Stud Health Technol Inform 2014;205:533-537. [Medline: 25160242]
15. Aurum A, editor. Engineering and Managing Software Requirements. Springer-Verlag Berlin Heidelberg: Springer; 2005:1-15.
16. Oreskovic NM, Goodman E, Park ER, Robinson AI, Winickoff JP. Design and implementation of a physical activity intervention to enhance children's use of the built environment (the CUBE study). Contemp Clin Trials 2015 Jan;40:172-179. [doi: 10.1016/j.cct.2014.12.009] [Medline: 25533728]
17. Finegood DT, Merth Thomas D N, Rutter H. Implications of the foresight obesity system map for solutions to childhood obesity. Obesity (Silver Spring) 2010 Feb;18 Suppl 1:S13-S16. [doi: 10.1038/oby.2009.426] [Medline: 20107455]
18. Dunton GF, Whalen CK, Jamner LD, Henker B, Floro JN. Using ecologic momentary assessment to measure physical activity during adolescence. Am J Prev Med 2005 Nov;29(4):281-287. [doi: 10.1016/j.amepre.2005.07.020] [Medline: 16242590]
19. Giabbanelli PJ, Crutzen R. Supporting self-management of obesity using a novel game architecture. Health Informatics J 2014 Feb 19. [doi: 10.1177/1460458214521051] [Medline: 24557604]

Abbreviations

CHC: chronic health conditions
mHealth: mobile health technology

Edited by G Eysenbach; submitted 16.12.14; peer-reviewed by P Giabbanelli, S Khairat; comments to author 19.02.15; revised version received 10.03.15; accepted 23.03.15; published 02.06.15.

Please cite as:
Oreskovic NM, Huang TT, Moon J
Integrating mHealth and Systems Science: A Combination Approach to Prevent and Treat Chronic Health Conditions
JMIR mHealth uHealth 2015;3(2):e62
URL: http://mhealth.jmir.org/2015/2/e62/
doi:10.2196/mhealth.4150
PMID:26036753

Viewpoint

Can Standards and Regulations Keep Up With Health Technology?

Christopher James Vincent[1*], BSc, PhD; Gerrit Niezen[2*], BEng, MEng, PhD; Aisling Ann O'Kane[1*], BASc, MSc, PEng; Katarzyna Stawarz[1*], BSc, MSc

[1]UCL Interaction Centre, University College London, London, United Kingdom

[2]Future Interaction Technologies Lab, Department of Computer Science, Swansea University, Swansea, Wales, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Christopher James Vincent, BSc, PhD
UCL Interaction Centre
University College London
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 44 (0)203 108 7057 ext 57057
Fax: 44 (0)20 7387 1397
Email: c.vincent@ucl.ac.uk

Abstract

Technology is changing at a rapid rate, opening up new possibilities within the health care domain. Advances such as open source hardware, personal medical devices, and mobile phone apps are creating opportunities for custom-made medical devices and personalized care. However, they also introduce new challenges in balancing the need for regulation (ensuring safety and performance) with the need to innovate flexibly and efficiently. Compared with the emergence of new technologies, health technology design standards and regulations evolve slowly, and therefore, it can be difficult to apply these standards to the latest developments. For example, current regulations may not be suitable for approaches involving open source hardware, an increasingly popular way to create medical devices in the maker community. Medical device standards may not be flexible enough when evaluating the usability of mobile medical devices that can be used in a multitude of different ways, outside of clinical settings. Similarly, while regulatory guidance has been updated to address the proliferation of health-related mobile phone apps, it can be hard to know if and when these regulations apply. In this viewpoint, we present three examples of novel medical technologies to illustrate the types of regulatory issues that arise in the current environment. We also suggest opportunities for support, such as advances in the way we review and monitor medical technologies.

(JMIR mHealth uHealth 2015;3(2):e64) doi:10.2196/mhealth.3918

KEYWORDS

governmental regulations; health services; medical devices; mHealth; mobile phones; open source initiative; software; standards; technology

Introduction

In recent years, there has been a rapid, major, and continued advance in scientific discovery and technology proliferation that provides the means to support health care in new ways. For example, the percentage of UK adults who own a mobile phone has risen from 39% to 51% in just 1 year [1]. The proliferation of mobile phones and pervasiveness of health apps [2] allow patients to manage and track their health conditions on the go, which turns mobile phones into a tool for health-related behavior change [3]. This means that care can be provided outside of clinical settings [4,5] and technology can be used to address growing health care demands, such as an increasing prevalence of chronic conditions and aging populations [6].

A growing number of medical and health-related technologies becoming available can be adapted to support personal care, both in terms of customized hardware and software. For example, electronic devices are not only becoming ubiquitous, but are also easier to make; three-dimensional printers are becoming significantly cheaper (the market is predicted to grow by 500% in 5 years [7]). Three-dimensional printers are devices that create three-dimensional objects based on an electronic

data source containing a three-dimensional model. As a result, these printers have opened up the possibility to produce custom-made medical devices as needed, where needed [8], which is sometimes referred to as "hyperlocal micro manufacturing" [9]. These types of advances will continue to provide solutions to health care problems that seemed near impossible to solve a decade ago, and they generate their own unique considerations about how these technologies fit into existing regulatory frameworks.

The need for regulation has long been established in the health care domain and has led to manufacturers considering safety during the design and evaluation of medical devices [10] (for a US perspective on ethical standards, see [11]). Medical device manufacturers often use medical device standards (eg, [12]) to guide their design and production processes and to comply with regulatory requirements. For example, the IEC 60601 [13] series puts in place requirements relating to safety and effectiveness, focusing on various aspects of the product (eg, electrical integrity, alarms). The IEC 62366 [14] standard describes a usability engineering process, to satisfy similar requirements, which is linked to a risk analysis standard (ISO 14971 [15]).

Unfortunately, novel and personal medical technologies do not always fit into the process specified in standards because they move away from what is currently and generally accepted as good practice to a situation in which there may be little or no precedent for comparison. Health care technology innovation may be hurt by the current regulatory system [16]. Sometimes standards do not provide sufficient design and evaluation criteria for novel technologies that differ significantly from equivalent predecessors; sometimes regulation may stifle innovation to the point where new technology cannot benefit the health care system (eg, through the time or cost constraints); however, sometimes existing systems may not be applied at all, or they are not applicable when it comes to modern technology.

In this paper, we open a discussion about the challenges to existing regulatory systems posed by novel and personal health care technologies. By presenting three examples that we have encountered as part of our research, we highlight some of the issues. First, we describe an open source infusion pump that raises questions about how to control the quality of custom-made medical devices. Next, we present our research on mobile medical devices that challenges the methods of evaluation set out in current medical usability standards. Finally, we discuss the design of a mobile phone app for medication adherence that may or may not be governed by the existing regulations. Although papers focusing on regulatory challenges have already been published (eg, [17]), we contribute to the discussion by introducing three case studies, outlining the issues with standards and regulations, and proposing ways to address these issues.

Novel Health Care Technologies

Overview

The following examples describe the tension between health care innovation and regulation. They come from research conducted as a part of the Computer-Human Interaction for Medical Devices (CHI + MED) project, focusing on developing tools to support safe and usable health technology (medical devices). The following section presents three technologies: open source hardware, mobile health care technologies, and health-related mobile phone apps. It describes the regulatory challenges that may be encountered in the development of these kinds of devices, and identifies opportunities for addressing these issues.

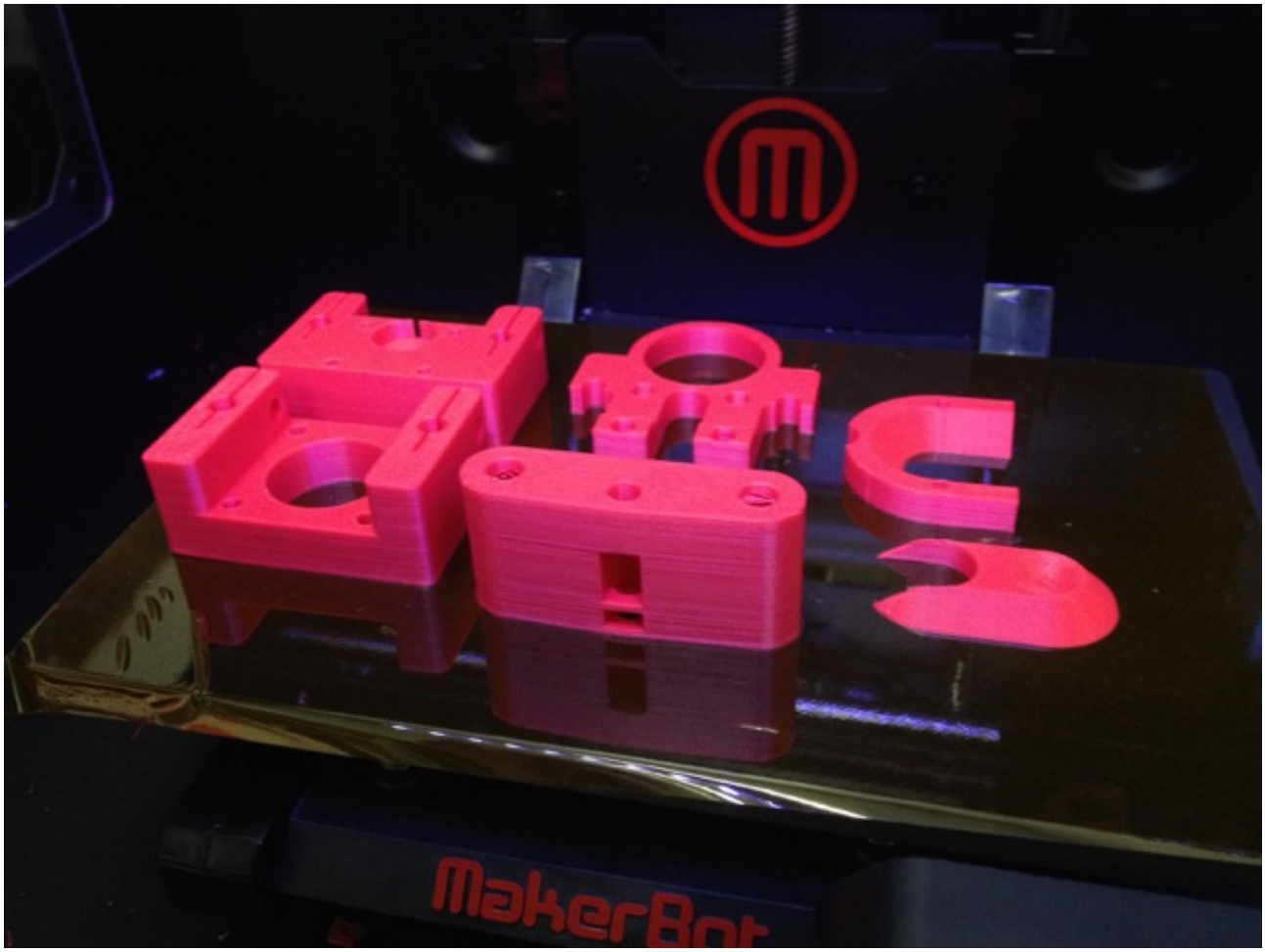
Example 1: Open Source Medical Devices

Background

Open source hardware is an emerging business model where the design files of a product, including the circuit schematics, source code, and physical design, are made publicly available under a license so that anyone can study, modify, distribute, make, and sell the design or hardware based on that design. In recent years, three-dimensional printers have made possible a rapid production of customized medical devices [8], from fitted mechanical limbs [18] and mobile phone-connected microscopes [19] to parts for syringe pumps (as shown in Figure 1). Coupled with an open source approach, more can be achieved with less cost, because production can occur in-house, based on a freely available design.

Building on work of the Michigan Tech Open Sustainability Technology (MOST) group [20], we are demonstrating the process of building an open source syringe pump that implements design principles and interface guidelines published as part of the CHI + MED project. We are creating a complete open platform for further research and development [21]. Design files and software made public by MOST, under an open source license, are at the core of the project. The approach not only leads to economic savings through a reduction in the life-cycle cost [22], but also it allows others to improve on the design, share the improvements with others, and get rapid feedback from the end user. This openness can benefit multiple stakeholders and lead to effective technology and improved patient outcomes (for equivalent arguments relating to open source software, see [23]). It can also allow staff from hospital departments such as medical physics and clinical engineering to repair and customize their own devices, reducing a reliance on external providers.

Figure 1. Three-dimensional printing technology (image credit Gerrit Niezen; image license CC-BY).



Regulatory Challenges

The ability to modify someone else's designs and the ease of rapid and unique production could interfere with formal quality-control processes, implicit in the existing medical device regulation. Although three-dimensional printing is a tool for prototyping and not for long-term use and reuse, it is possible to see how this technology could be used for the latter purpose.

Existing standards may not be practical as the documentation required for review and approval may be disproportionate when the design is limited to a very small number of production units. While the steps that are followed during the development and testing of medical devices are specified and controlled by standards (in the European Union, those listed in [12]), the process followed during the aforementioned activities may be ad hoc. For example, processes specified in medical device standards were created with traditional manufacturing process in mind. At a certain point, a design would be frozen and considered complete. This is not the case when devices can be continuously improved upon by the creator and others. The need for documentation and testing that closely adheres to standards may remove the flexibility that novel approaches bring. Repeated design changes and the requirement for oversight from a review body may be cumbersome on both sides. Although there are many advantages to realizing the benefits of existing process, systems need to be made agile and proportionate.

Opportunities

With an open source approach, there is an opportunity to share the rationale behind the design, the process used to derive the design, as well as the design itself. For example, online documentation tools such as wikis and version control software can be adapted to provide a better overview of the development workflow and process that has been followed. It is also possible to share evaluation results; if a component or design has evolved over time, knowing how and why this has happened could help those at distance understand the constraints of a solution. Through sharing and periodically updating these documents, duplication of effort can be avoided. For example, it does not always make sense for the same component to undergo the same testing by multiple parties. Moreover, documentation can be scrutinized by multiple specialists, without being confounded by the proprietary nature of "closed" solutions. Therefore, rather than requiring the same documentation as that for traditionally manufactured devices, regulation could involve transparent records of all components and changes made to those components, including the rationale and assumptions. This would help to support the quality of such devices without stifling innovation, with the onus being on those implementing a solution to check and review these documents.

Example 2: Mobile Health Care Technologies and Home Use Devices

Background

Given the need to support care outside of clinical settings, technologies are also being developed to provide increasing autonomy and self-care. Our research on CHI + MED investigates how people use technology to manage their health care needs during their day-to-day lives. One such set of technologies includes devices used in the self-management of type 1 diabetes, a complex chronic condition that requires significant personal responsibility over a lifetime [24]. A common form of diabetes technology is the glucose meter that is used to measure blood glucose levels for everyday medication dose calculations, as well as for identifying high and low blood sugar levels, which are dangerous in the long term and potentially fatal in the short term, respectively [25]. Thanks to advances in technology and human factors engineering, glucose meters can be used by people with minimal or no training. They are easier to carry, easier to use, and can store results. This empowers people with diabetes and grants independence [26].

However, our work shows that complexities of everyday life such as people's work life, romantic life, friendships, hobbies, travel, holidays [27], or whom they are with [28,29] impact on the use of these devices (Figure 2). Understanding these factors is incredibly complex [30], but necessary to ensure glucose meters are reliable and meet users needs. The problem is that the evaluation methods suggested by standards are not adequate in addressing everyday use, as they have been developed with a focus on technology used in clinical environments, where there is more certainty about the characteristics of the work place and levels of training.

Figure 2. Everyday life and a glucose meter (image credit Aisling Ann O'Kane; image license CC-BY).

Regulatory Challenges

Mobile health care technologies used for self-management, such as glucose meters used by people with diabetes, are medical devices and are regulated as such. Standard usability engineering process applies, such as IEC 62366 [14], for these cases.

These personal health care devices are used in the context of people's everyday lives, yet the design and evaluation involves the same usability standards as medical devices found in hospitals. The definition of usability is the same (see [14]). The focus is on the device's effectiveness, efficiency, ease of learning, and user satisfaction, in what is assumed to be a controlled context. This standardized usability evaluation practice makes it difficult to support the range of individual needs of users outside clinical environments: in their homes, on the go, as a part of their everyday life. This can be seen by accounts of the application of standard usability engineering process [31], such as IEC 62366 [14]. Although such standards are voluntary, and the techniques are illustrative, many

companies feel that they have little option but to adhere to their content [32].

Although standards such as IEC 62366 [14] suggest taking context into account using techniques such as task analysis, contextual inquiry, functional analysis, testing in simulated clinical environments, and field testing, how they might be adapted or tailored to represent the unstable context of everyday life is not elaborated on. These methods may not capture the influences that a person's life might have on the safe use and adoption of these devices, outside the confines of a hospital. Influences could include the emotional aspects of self-care or the social impact of bystanders, when using a device in front of others.

Likewise, it can be unmanageable to scope everyday health care technologies, for the purposes of making assumptions about their use. Pervasive technologies are used by all sorts of people, in all sorts of situations, and in all sorts of contexts. Individual differences are inevitable, and they have been shown to impact users' experience and challenge the design and development of products [33]. This raises concerns for health care technologies where differences might result in safety risks.

These concerns can be addressed by limiting the scope to a certain user profile or context, but this may not be possible for medical technologies designed for a particular condition, not for a particular user group. Another approach is to configure products based on user needs, for example, allowing customization of the exterior shell [34] and/or allowing configuration of the user interface. This poses a dilemma for evaluation in that as the number of possible configurations increases, the burden associated with management, evaluation, and support also increases [35,36]. On the one hand, allowing for flexibility reduces the chance of nonadoption or noncompliance; on the other hand, complexity in the product adds to the resource required to develop and maintain it.

Opportunities

One option would be to increase emphasis on postmarket assessment, such as monitoring the use of equipment in context (including self-report), alongside conducting research to understand how users are really experiencing this type of equipment. Exploratory qualitative methods have been applied in other domains to focus on the situated use of interactive devices and they are also relevant here.

For instance, diary studies [37] involve users taking note of when, where, how, and why they use their device in their everyday life. They avoid the invasiveness of observation. Autoethnography, a form of self-study, is a quick and easy way to probe the everyday use of mobile medical devices [30]. Even though there has been progress toward using exploratory methods to investigate the context of use [38], standards are lacking in their treatment of situated user experience. Situated user experience relates to the notion that the localized context is an important factor in determining how people will experience and interact with technology. As it has been shown that context influences the use of pervasive health care technologies [30], testing technology away from this context (e.g. in a laboratory) does not anticipate how well the technology will meet the needs of the user. There is therefore an opportunity for standards to support consideration of a broader range of context, given the mobile nature of devices, and use outside of clinical settings. The inclusion of exploratory qualitative methods would allow for this by probing context of use and revealing of individual differences.

Example 3: Health-Related Mobile Phone Apps

Background

Whereas personal medical technologies can provide benefits to those who require specific equipment, dedicated mobile phone apps have potential to be advantageous to almost anyone. People have access to thousands of free health-related mobile phone apps [39,40]. They range from behavior change apps supporting people who want to improve their health and well-being [3], for example, apps supporting smoking cessation (eg, SmokeFree28 [41]), or providing informed choice regarding alcohol intake (eg, Drinks Meter [42]), to apps focusing on specific conditions (eg, pain management [43]). They can be used to prevent forgetfulness (eg, medication reminder apps [44]) and general adherence support apps [45]. As people tend to keep their mobile phones close and hardly ever switch them off, health apps can provide useful functions at any time. They are always at hand to help track the on-going behavior.

Another (currently unreleased) example would be a software app to support oral contraception adherence. We are currently researching how mobile phone apps could be used to reduce unintentional nonadherence. This involves developing a medication reminder app that supports the formation of medication-taking habits and assists users as they search for the best way to embed medication taking into their daily routine; how this could be achieved is described in [44]. One of the major challenges to understanding the approach required to evaluate this type of software is the fact that it is not entirely clear whether an app should be regulated as a medical device.

Regulatory Challenges

During our research we have identified many issues concerning the certification of health-related apps. One example relates to the wording that is used to describe them. Many apps make medical claims and by doing so, could pose a serious public health issue, especially they are if ineffective or inaccurate [40]. Regulation is required and although guidance on health-related software and apps exists [46-49], it may not be clear whether such apps should be covered by regulations. Moreover, when considering the market as a whole, regulations may be bypassed and in some cases, ignored. In the European Union, in some cases, health-related apps fall under the control of the medical devices directive. In this scheme, assuming the software is not an accessory to a device, classification rules can treat medical standalone software as a comparatively low risk [49]:

"The classification rules were not written with software in mind. Due to the restrictive nature of Rules 9-11 of Annex IX to Directive 93/42/EEC, a large number of software devices therefore fall in Class I, where compliance is based on self-declaration, ie, no third party assessment."

As a result, the product can be self-certified by the organization producing it. This raises concerns about the level of peer review and testing that this type of software receives.

In other cases, software can be used in a medical context, but not for a medical purpose, and not considered to be a medical device. For example, based on similar UK Medicines and Health Care Products Regulatory Agency (MHRA) advice, an app could use an accelerometer or gyroscope to detect falls in epileptic patients and therefore be regulated as a medical device. However, if the same technology is used to measure steps or detect whether an elderly person has got up from a chair or a bed in a social care context, the regulation would not apply [47]. This is important because it impacts on the type of testing that would apply (both in terms of usability and general safety and performance requirements), as well as the approach to monitoring the device in situ.

To help determine whether an app should be treated as a medical device, the MHRA has produced the following list of keywords and phrases that if used in the app's description indicate that it should be regulated: *amplify, analysis, interpret, alarms, calculates, controls, converts, detects, diagnose, measures*, and *monitors* [47]. Take the example of an app that aims to send alerts to users to check whether they have taken their medication and records how many times they say they did not. Based on this information, it can suggest changes to the routine. Therefore, it could be said that the app *monitors* users and occasionally *alarms* when their behavior needs to be modified. Does it mean the app should be certified?

If the app is labeled as a tool for supporting *medication-taking habits*, then the answer is likely to be yes. However, if the wording changes to simply *habits*, then even though the functionality stays the same, does the answer become no? Such a small change in wording could be enough to avoid device regulations, but it might not even be needed. Some developers simply publish their health apps without worrying about regulations at all, whereas others add liability disclaimers to app descriptions [40]. Based on the US guidance relating to mobile medical applications [46], "Mobile apps that keep track of medications and provide user-configured reminders for improved medication adherence" are an example of "...mobile apps for which FDA intends to exercise enforcement discretion" [46].

Based on both US and UK frameworks, this type of app may or may not be regulated. This problem has been covered in recent UK media reports [50], where there are several examples of gray areas and products that sit on the boundary. In many respects there is not anything new about technologies that sit on the boundary between regulated and unregulated products. The concern is that the assumptions used to determine whether technology is regulated as a medical device may not reflect how the technology is actually used (eg, unregulated apps being used in a context when a regulated app is appropriate).

Of specific concern is the quality of the software code and process used during programming, which may be invisible once the software has been released. Even if there is intent to follow medical device standards, they may be difficult to realize in practice. For example, the medical device usability standard IEC 62366 [14] combines evaluation of safety and usability, which may be in opposition to each other [35]. Other standards may be insufficient when it comes to the testing of software: they may not require exhaustive testing, complete coverage, or proof that a solution is correct by construction [51]. This is evident in the number of software-related defects observed in medical device user interfaces [52].

Opportunities

Because of the intangible nature of apps and the fact that they can be easily updated in situ, assessment and classification at a single point in time may not be feasible or appropriate. Obiodu and Obiodu [40] suggest that one way to deal with the issue of certification might be producing evidence-based guidelines for designing health apps rather than trying to strictly regulate them. We agree with this point, but would see an opportunity to take this further. Rather than just relying on guidelines issued by authorities, patient groups could produce best practice guidelines for specific conditions and, by following the example of open hardware initiatives, publish them openly to encourage collaboration with other patient groups, app developers, and mobile phone manufacturers, who are already starting to release health care kits [53].

Conclusions

For medical technology, standards and regulations are needed to ensure safety, protect the public, and guarantee that products are fit for purpose. However, in the context of novel and personal medical technologies, the current approach to regulation is not only infeasible and difficult to enforce, but also work against health care innovation. Given that it is inevitable that three-dimensional-printed components, mobile devices, and apps will be used to support and deliver health care, as well as have impact on medical practice, regulators may need to rethink their approach.

Based on our work, we have presented the benefits of new technologies and personal medical devices. In many cases, growing pressure on health services makes their introduction inevitable. At the same time, we have outlined some of the regulatory challenges. For example, by allowing for rapid manufacturing of bespoke components, three-dimensional printing raises concerns regarding quality control; the standards underpinning the usability of personal mobile medical devices are not enough to guarantee the "design meets users' needs" concept; in addition, mobile phone apps may or may not be certified, depending on how a product is described. These challenges open up new possibilities and encourage new ways of thinking.

The health care domain is not the only one feeling the impact of these technologies. The situation resembles the issues with touch-screen tablets being used in the office environment. Although office work and office equipment are regulated (eg, computer workstations), the health and safety regulations are unlikely to apply to tablet computers when it is not possible to easily control how or where they are being used [54]. Rather than trying to regulate them, different, more flexible approaches are needed. For example, by shifting focus away from the

introduction of technology, and toward educating users about the implications of using it (eg, raising awareness of human factors), we allow those in different environments to make sure their technology is safe and fits their needs. Although due diligence occurs during the design of technology, continual research and review aims to tailor the properties of equipment with the needs of users and characteristics of the work environment.

The same could apply to novel and personal health care technologies. When it comes to testing for usability, we cannot predict every possible combination of user and usage before the deployment of technology. The alternative is to conduct research into how equipment is really used. We then realize that improvement will occur when a device is in situ, and this will be specific to a given context. As we can rapidly iterate a design, we can continually improve and share the benefit of this improvement. Much of the existing guidance concerning safety and usability needs updating to accommodate this approach. There also needs to be a consideration of how adequate levels of safety can be guaranteed, without making it prohibitive for small organizations to create products with relatively short life cycles. The problem with the existing approach to regulation is that historically, those producing technology would be likely to stay in business for extended periods, compared with the hobbyists and small organizations producing apps, who would rapidly develop and release technology, but then may not be in place to support it in the future. In the past, we had traceability and accountability, whereas in the present we have little recompense if something goes wrong.

We suggest ways of addressing these challenges, such as publishing documentation and making it openly available to review, therefore increasing transparency; adding situated methods to usability standards to cover people's everyday use of personal health technology, and allowing patient groups to review mobile phone apps, draft their own guidelines, and collaborate with each other and with app developers. This would help to ensure that patients' needs are met. Realizing a code of practice for app developers, such as PAS 277 [55], would help to build confidence. There is also an opportunity to educate those buying and using such technology on requirements relating to safety and usability.

If there is a need to comply with medical device software process standards (eg, IEC 62304 [56]), there have been recent developments in guidance. There are now worked examples of assessment process (ISO 33030 [57]); and support for process tailored to the safety class of the software (IEC/TR 80002-3 [58]). There are groups such as Medi SPICE [59].

This viewpoint represents a series of observations from our own research on the challenges of regulating health technology. We hope to start a discussion about the obstacles and opportunities in addressing the design of novel technologies within regulatory frameworks. Given the need to address the increasing pressures on health services, this discussion is urgently required. Future research could apply a structured methodology to review this context and a case study approach [60], to articulate practical, balanced, and proportionate approaches in line with this discussion.

Acknowledgments

We would like to thank our reviewers and other members of the team that have helped conduct the research presented in this paper. Our work is a part of the CHI + MED: Multidisciplinary Computer-Human Interaction Research for the Design and Safe Use of Interactive Medical Devices project, supported by the UK Engineering and Physical Sciences Research Council (EP/G059063/1). We would like to thank Professor Harold Thimbleby and Ms Kate Michi Ettinger for their feedback on an earlier version of this work.

Conflicts of Interest

None declared.

References

1. Ofcom. Communications Market Report 2013. 2013 Aug 01. URL: http://stakeholders.ofcom.org.uk/binaries/research/cmr/cmr13/2013_UK_CMR.pdf [accessed 2014-09-30] [WebCite Cache ID 6SyUL7SA8]
2. Roberts M. GPs 'To Prescribe Health Apps'. London, UK: BBC; 2012 Feb 22. URL: http://www.bbc.co.uk/news/health-17111092 [accessed 2014-09-30] [WebCite Cache ID 6SyST3xwb]
3. 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]
4. UK Department of Health (DoH). Ten Things You Need to Know About Long Term Conditions. London, UK: UK Department of Health (DoH); 2013. URL: http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/Healthcare/Longtermconditions/tenthingsyouneedtoknow/index.htm [accessed 2014-09-30] [WebCite Cache ID 6SyTrwLhG]
5. RAND. Health and Well-Being in the Home.: RAND; 2010. URL: http://www.rand.org/content/dam/rand/pubs/occasional_papers/2010/RAND_OP323.pdf [accessed 2014-09-30] [WebCite Cache ID 6SyTihdPE]
6. Tunstall-Pedoe H. Preventing Chronic Diseases: A Vital Investment. Geneva, Switzerland: WHO; 2005. URL: http://www.who.int/chp/chronic_disease_report/en/ [accessed 2014-09-30] [WebCite Cache ID 6SySA3DtD]

7. Clinch M. 3-D Printing Market to Grow 500% in 5 Years. Englewood Cliffs, NJ: CNBC; 2014. URL: http://www.cnbc.com/id/101542669 [accessed 2014-10-03] [WebCite Cache ID 6T2wJ0kDP]
8. Schubert C, van Langeveld CM, Donoso LA. Innovations in 3D printing: A 3D overview from optics to organs. Br J Ophthalmol 2014 Feb;98(2):159-161. [doi: 10.1136/bjophthalmol-2013-304446] [Medline: 24288392]
9. Dotz D. Open Science at Tech4Dev.: OKF Open Science Working Group; 2014. URL: http://science.okfn.org/category/research/ [accessed 2014-10-01] [WebCite Cache ID 6T0IxB0qA]
10. Vincent C. Mind the Gap: What Interactive Medical Device Manufacturers Need. London, UK: BCS; 2010. URL: http://www.chi-med.ac.uk/publicdocs/WP012.pdf [accessed 2014-10-03] [WebCite Cache ID 6T2tfGMHa]
11. Coleman C, Menikoff J, Goldner J, Dubler N. Ethics and Regulation of Research on Human Subjects. San Francisco, CA: LexisNexis; 2012. URL: http://www.lexisnexis.com/store/images/Supplements/3620_2012S.pdf [accessed 2014-10-01] [WebCite Cache ID 6T0Jr1guI]
12. European Commission (EC). Summary List of Titles and References Harmonised Standards under Directive 93/42/EEC for Medical Devices. Brussels, Belgium: European Commission; 2014. URL: http://ec.europa.eu/enterprise/policies/european-standards/harmonised-standards/medical-devices/index_en.htm [accessed 2014-10-01] [WebCite Cache ID 6SziVOw3u]
13. International Electrotechnical Commission. Medical Electrical Equipment-All Parts IEC 60601. Geneva, Switzerland: International Electrotechnical Commission; 1977. URL: http://www.webcitation.org/6YqOWpaVL [accessed 2015-06-02]
14. International Organization for Standardization. Medical Devices: Application of Usability Engineering to Medical Devices 62366:2007. Geneva, Switzerland: International Organization for Standardization; 2007. URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=38594 [accessed 2015-05-27] [WebCite Cache ID 6YqMI7wDO]
15. International Organization for Standardization. Medical Devices: Application of Risk Management to Medical Devices 14971:2007. Geneva, Switzerland: International Organization for Standardization; 2007. URL: http://www.webcitation.org/6SQqjQ7pz [accessed 2015-06-02]
16. Longhurst CA, Landa HM. Health information technology and patient safety. BMJ 2012 Feb 20;344:e1096. [doi: 10.1136/bmj.e1096] [Medline: 22349580]
17. Vincent C, Blandford A. Maintaining the Standard: Challenges in Adopting Best Practice When Designing Medical Devices and Systems. Washington, DC: American Medical Informatics Association (AMIA); 2011. URL: http://www.chi-med.ac.uk/publicdocs/WP037.pdf [accessed 2015-05-27] [WebCite Cache ID 6YqGHjxjz]
18. Robohand. Robohand. Centurion, Gauteng, South Africa: Robohand; 2014. URL: http://www.robohand.net/ [accessed 2014-09-30] [WebCite Cache ID 6SyWJdfOA]
19. Lee WM, Upadhya A, Reece PJ, Phan TG. Fabricating low cost and high performance elastomer lenses using hanging droplets. Biomed Opt Express 2014 May 01;5(5):1626-1635 [FREE Full text] [doi: 10.1364/BOE.5.001626] [Medline: 24877020]
20. Wijnen B, Hunt EJ, Anzalone GC, Pearce JM. Open-source syringe pump library. PLoS One 2014;9(9):e107216 [FREE Full text] [doi: 10.1371/journal.pone.0107216] [Medline: 25229451]
21. Niezen G. Openpump. Wales, UK: Openpump; 2014. URL: http://openpump.org/ [accessed 2014-10-02] [WebCite Cache ID 6T1manphl]
22. Wittbrodt B, Glover A, Laureto J, Anzalone G, Oppliger D, Irwin J, et al. Life-cycle economic analysis of distributed manufacturing with open-source 3-D printers. Mechatronics 2013 Sep;23(6):713-726. [doi: 10.1016/J.Mechatronics.2013.06.002]
23. Reynolds CJ, Wyatt JC. Open source, open standards, and health care information systems. J Med Internet Res 2011;13(1):e24 [FREE Full text] [doi: 10.2196/jmir.1521] [Medline: 21447469]
24. O'Kane AA, Mentis HM, Thereska E. Non-static nature of patient consent over time: Diabetic patients' changing privacy perspectives. In: Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing (ACM CSCW 2013). New York, USA: ACM; 2013 Presented at: ACM CSCW 2013; February 23-27, 2013; San Antonio, Texas, USA p. 553-562 URL: http://research.microsoft.com/pubs/179620/cscw12_helena.pdf [doi: 10.1145/2441776.2441838]
25. Diabetes UK. Type 2 Diabetes: Treatment, Food, Lifestyle: What You Need to Know. London, UK: Diabetes UK; 2013. URL: http://www.diabetes.org.uk/Global/tesco-partnership/type-2-diabetes-need-to-know-0613.pdf [accessed 2014-10-03] [WebCite Cache ID 6T2uAbx6n]
26. Clarke SF, Foster JR. A history of blood glucose meters and their role in self-monitoring of diabetes mellitus. Br J Biomed Sci 2012;69(2):83-93. [Medline: 22872934]
27. O'Kane AA, Rogers Y, Blandford A. Gaining empathy for non-routine mobile device use through autoethnography. In: Proceedings of the 32nd ACM Conference on Human Factors in Computing Systems. New York, USA: ACM; 2014 Presented at: ACM CHI 2014; April 29, 2014; Toronto, ON, Canada p. 987-990 URL: http://www.chi-med.ac.uk/publicdocs/WP228.pdf [doi: 10.1145/2556288.2557179]
28. O'Kane AA, Rogers Y, Blandford A. Concealing or revealing mobile medical devices? Designing for onstage and offstage presentation. In: Proceedings of the 33rd ACM Conference on Human Factors in Computing Systems. New York, USA: ACM; 2015 Presented at: ACM CHI 2015; April 21, 2015; Seoul, Korea p. 1689-1698 URL: http://www.chi-med.ac.uk/publicdocs/WP289.pdf [doi: 10.1145/2702123.2702453]

29. Owen T, Buchanan G, Thimbleby H. Understanding user requirements in take-home diabetes management technologies. In: Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers. London, UK: British Computer Society; 2012 Presented at: BCS-HCI 12; Sept 12, 2012; Birmingham, UK p. 268-273.
30. O'Kane AA. Using a Third-Wave HCI Approach for Researching Mobile Medical Devices. Toronto, ON, Canada: ACM; 2014. URL: https://hcihealthcarefieldwork.files.wordpress.com/2013/11/theoryhealthchi2014_submission_17.pdf [accessed 2015-05-27] [WebCite Cache ID 6YqI768Y3]
31. Zenni TRL, Flores MP, Soares EME. Functional and emotional dimensions: Method to measure through data collection instruments for new products in healthcare. In: 2013 Proceedings of PICMET'13. New York, USA: IEEE; 2013 Presented at: Technology Management in the IT-Driven Services (PICMET); Jul 28-Aug 1, 2013; San Jose, CA p. 2551-2563.
32. Vincent C, Blandford A. Designing for safety and usability: User-centered techniques in medical device design practice. Santa Monica, CA: Human Factors and Ergonomics Society; 2011 Presented at: Proceedings of the Human Factors and Ergonomics Society Annual Meeting; Sept 19-23, 2011; Las Vegas, NV p. 793-797. [doi: 10.1177/1071181311551164]
33. Law ELC, Roto V, Hassenzahl M, Vermeeren APOS, Kort J. Understanding, scoping and defining user experience: A survey approach. New York, USA: ACM CHI 2009; 2009 Presented at: Proceedings of the 27th Annual ACM Conference on Human Factors in Computing Systems; Apr 4-9, 2009; Boston, MA p. 719-728. [doi: 10.1145/1518701.1518813]
34. Cordrey K. 'Skinit' Decorative Meter Skins Customize Freestyle Blood Monitors. 2010 Sep 07. 2011 URL: http://www.trendhunter.com/trends/skinit-decorative-meter-skins [accessed 2014-09-30] [WebCite Cache ID 6SyZ8q3jD]
35. Vincent CJ, Li Y, Blandford A. Integration of human factors and ergonomics during medical device design and development: It's all about communication. Appl Ergon 2014 May;45(3):413-419. [doi: 10.1016/j.apergo.2013.05.009] [Medline: 23778022]
36. McCaffery F, O'Connor R, Coleman G. Mapping medical device standards against the CMMI for configuration management. In: Filipe J, Shishkov B, Helfert M, editors. Software and Data Technologies: First International Conference, ICSOFT 2006, Setubal, Portugal, September 11-14, 2006, Revised Selected Papers. Berlin, Germany: Springer; 2008:153-164.
37. Rieman J. The diary study: A workplace-oriented research tool to guide laboratory efforts. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. New York, USA: ACM; 1993 Presented at: ACM CHI 93; April 24-29, 1993; Amsterdam, The Netherlands p. 321-326.
38. Sharples S, Martin J, Lang A, Craven M, O'Neill S, Barnett J. Medical device design in context: A model of user–device interaction and consequences. Displays 2012 Oct;33(4-5):221-232. [doi: 10.1016/j.displa.2011.12.001]
39. Laird S. How Smartphones Are Changing Healthcare. 2012 Sep 26. URL: http://mashable.com/2012/09/26/smartphones-health-care-infographic/ [accessed 2014-09-30] [WebCite Cache ID 6SyavXFeQ]
40. Obiodu V. An empirical review of the top 500 medical apps in a European Android market. J Mob Technol Med 2012 Dec 22;1(4):22-37. [doi: 10.7309/jmtm.74]
41. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: Preliminary evaluation of SmokeFree28. J Med Internet Res 2015;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
42. National Institute for Health Research. Drinks Meter. London, UK: National Institute for Health Research; 2015. URL: http://www.drinksmeter.com/about.php [accessed 2015-03-05] [WebCite Cache ID 6WoBbYPLc]
43. Rosser BA, Eccleston C. Smartphone applications for pain management. J Telemed Telecare 2011;17(6):308-312. [doi: 10.1258/jtt.2011.101102] [Medline: 21844177]
44. Stawarz K, Cox AL, Blandford A. Don't forget your pill!: Designing effective medication reminder apps that support users' daily routines. In: Proceedings of the 32nd ACM Conference on Human Factors in Computing Systems. New York, USA: ACM; 2014 Presented at: ACM CHI 2014; April 29, 2014; Toronto, Canada p. 2269-2278 URL: http://discovery.ucl.ac.uk/1418104/1/StawarzCoxBlandford2014-reminders-submittedManuscript.pdf [doi: 10.1145/2556288.2557079]
45. Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: Potential benefits to patients and providers: Response to Aungst. J Am Pharm Assoc (2003) 2013;53(4):345. [doi: 10.1331/JAPhA.2013.13121] [Medline: 23892801]
46. Food and Drug Administration. Mobile Medical Applications. Silver Spring, MD: Food and Drug Administration; 2015. URL: http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf [accessed 2015-03-05] [WebCite Cache ID 6WoBZ9rLT]
47. Medicines and Healthcare Products Regulatory Agency (MHRA). Guidance on Medical Device Stand-Alone Software (Including Apps). London, UK: Medicines and Healthcare Products Regulatory Agency (MHRA); 2014 Aug 08. URL: http://www.mhra.gov.uk/Howweregulate/Devices/Software/index.htm [accessed 2014-09-30] [WebCite Cache ID 6SycFISSG]
48. Royal College of Physicians. Appwardly Mobile. London, UK: Royal College of Physicians; 2015. URL: https://www.rcplondon.ac.uk/commentary/appwardly-mobile [accessed 2015-03-06] [WebCite Cache ID 6WpWxpFu0]
49. Ebenezer N. The Regulation of Standalone Software as a Medical Device. London, UK: Medicines and Healthcare Products Regulatory Agency (MHRA); 2011. URL: http://www.medelec.co.uk/Presentations/9.35_DrNeilEbenezer.pdf [accessed 2014-10-01] [WebCite Cache ID 6SztnUEQA]
50. Graham F. Health Tech: When Does an App Need Regulating?. London, UK: BBC; 2014. URL: http://www.bbc.co.uk/news/business-29605951 [accessed 2015-03-05] [WebCite Cache ID 6WoBiiEP9]
51. Praxis High Integrity Systems Ltd. Correct By Construction. Bath, England: Praxis High Integrity Systems Ltd; 2010. URL: http://c2.com/cgi/wiki?CorrectByConstruction [accessed 2014-10-01] [WebCite Cache ID 6T0G2d1Bk]

52. Masci P, Oladimeji P, Curzon P, Thimbleby H. Using PVSio-web to demonstrate software issues in medical user interfaces. In: Proceedings of the 4th International Symposium on Foundations of Health Information Engineering and Systems. Berlin: Springer-Verlag; 2015 Presented at: FHIES/SEHC 2014; July 17th-18th, 2014; Washington DC URL: http://www.eecs.qmul.ac.uk/~masci/works/masci-FHIES14.pdf
53. Forbes. Apple Gives Epic And Mayo Bear Hug With HealthKit. 2014. URL: http://www.forbes.com/sites/danmunro/2014/06/03/apple-gives-epic-and-mayo-bear-hug-with-healthkit/ [accessed 2014-09-30] [WebCite Cache ID 6SyfFePdi]
54. Stawarz K, Benedyk R. Bent necks and twisted wrists: Exploring the impact of touch-screen tablets on the posture of office workers. In: Proceedings of the 27th International BCS Human Computer Interaction Conference.: British Computer Society; 2013 Presented at: BCS HCI 2013; September 9-13, 2013; London, UK URL: http://ewic.bcs.org/upload/pdf/ewic_hci13_short_paper22.pdf
55. British Standards Institution. Health and wellness apps: Quality criteria across the life cycle–Code of practice PAS 277. London, UK: British Standards Institution; 2015. URL: http://shop.bsigroup.com/upload/271432/PAS%20277%20(2015)bookmarked.pdf [accessed 2015-05-28] [WebCite Cache ID 6YrqfBbtj]
56. International Organization for Standardization. Medical Device Software: Software Life Cycle Processes IEC 62304:2006. Geneva, Switzerland: International Organization for Standardization; 2006 May 01. URL: http://www.webcitation.org/6YqNxmJHa [accessed 2015-06-02]
57. International Organization for Standardization. Documented Assessment Process ISO 33030. Geneva, Switzerland: International Organization for Standardization; 2010. URL: http://www.webcitation.org/6Yrm9zGix [accessed 2015-06-02]
58. International Organization for Standardization. Medical Device Software. Part 3: Process Reference Model of Medical Device Software Life Cycle Processes IEC 62304 IEC/TR 80002-3. Geneva, Switzerland: International Organization for Standardization; 2014. URL: http://www.webcitation.org/6YqNA4cmW [accessed 2015-06-02]
59. Medi SPICE. Medi SPICE. 2015. URL: http://medispice.ning.com/ [accessed 2015-03-05] [WebCite Cache ID 6WoBtABdB]
60. Schramm W. Notes on Case Studies of Instructional Media Projects. Stanford, CA: Institute for Communication Research, Stanford University; 1971 Dec. URL: http://files.eric.ed.gov/fulltext/ED092145.pdf [accessed 2015-05-27] [WebCite Cache ID 6YqLnlM7i]

Abbreviations

CHI+MED: Computer-Human Interaction for Medical Devices
MHRA: Medicines and Health Care Products Regulatory Agency
MOST: Michigan Tech Open Sustainability

Edited by G Eysenbach; submitted 04.10.14; peer-reviewed by R O'Connor, L Dayer; comments to author 22.02.15; revised version received 07.03.15; accepted 23.03.15; published 03.06.15.

Please cite as:
Vincent CJ, Niezen G, O'Kane AA, Stawarz K
Can Standards and Regulations Keep Up With Health Technology?
JMIR mHealth uHealth 2015;3(2):e64
URL: http://mhealth.jmir.org/2015/2/e64/
doi:10.2196/mhealth.3918
PMID:26041730

Original Paper

Validation of Physical Activity Tracking via Android Smartphones Compared to ActiGraph Accelerometer: Laboratory-Based and Free-Living Validation Studies

Eric B Hekler[1], PhD; Matthew P Buman[1], PhD; Lauren Grieco[2], PhD; Mary Rosenberger[2], PhD; Sandra J Winter[2], PhD; William Haskell[2], PhD; Abby C King[2], PhD

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

[2]Stanford University, Stanford, CA, United States

Corresponding Author:
Eric B Hekler, PhD
Arizona State University
School of Nutrition and Health Promotion
500 N. 3rd St.
Phoenix, AZ, 85003
United States
Phone: 1 6028272271
Fax: 1 6028272253
Email: ehekler@asu.edu

Abstract

Background: There is increasing interest in using smartphones as stand-alone physical activity monitors via their built-in accelerometers, but there is presently limited data on the validity of this approach.

Objective: The purpose of this work was to determine the validity and reliability of 3 Android smartphones for measuring physical activity among midlife and older adults.

Methods: A laboratory (study 1) and a free-living (study 2) protocol were conducted. In study 1, individuals engaged in prescribed activities including sedentary (eg, sitting), light (sweeping), moderate (eg, walking 3 mph on a treadmill), and vigorous (eg, jogging 5 mph on a treadmill) activity over a 2-hour period wearing both an ActiGraph and 3 Android smartphones (ie, HTC MyTouch, Google Nexus One, and Motorola Cliq). In the free-living study, individuals engaged in usual daily activities over 7 days while wearing an Android smartphone (Google Nexus One) and an ActiGraph.

Results: Study 1 included 15 participants (age: mean 55.5, SD 6.6 years; women: 56%, 8/15). Correlations between the ActiGraph and the 3 phones were strong to very strong (ρ=.77-.82). Further, after excluding bicycling and standing, cut-point derived classifications of activities yielded a high percentage of activities classified correctly according to intensity level (eg, 78%-91% by phone) that were similar to the ActiGraph's percent correctly classified (ie, 91%). Study 2 included 23 participants (age: mean 57.0, SD 6.4 years; women: 74%, 17/23). Within the free-living context, results suggested a moderate correlation (ie, ρ=.59, P<.001) between the raw ActiGraph counts/minute and the phone's raw counts/minute and a strong correlation on minutes of moderate-to-vigorous physical activity (MVPA; ie, ρ=.67, P<.001). Results from Bland-Altman plots suggested close mean absolute estimates of sedentary (mean difference=–26 min/day of sedentary behavior) and MVPA (mean difference=–1.3 min/day of MVPA) although there was large variation.

Conclusions: Overall, results suggest that an Android smartphone can provide comparable estimates of physical activity to an ActiGraph in both a laboratory-based and free-living context for estimating sedentary and MVPA and that different Android smartphones may reliably confer similar estimates.

(JMIR mHealth uHealth 2015;3(2):e36) doi:10.2196/mhealth.3505

KEYWORDS

telemedicine; cell phones; accelerometry; motor activity; validation studies

Introduction

Reduced sitting time and increased moderate-to-vigorous physical activity (MVPA) confers an array of health benefits [1-3]. Identifying cost-efficient solutions for tracking physical activity passively has become an important scientific objective [4]. Previous research on activity monitoring has focused on devices dedicated solely to this purpose, such as the ActiGraph accelerometer (ActiGraph, Fort Walton, FL, USA). Indeed, this trend of dedicated activity monitoring devices has continued with consumer devices such as the Fitbit, Jawbone UP, and Misfit Shine, among others. There is also increasing interest in improving physical activity detection through multiple sensors [5,6], but these added sensors (eg, heart rate, global positioning systems) often impact the ability to enable long-term monitoring [4] in context.

One potentially cost-efficient and low-burden mechanism for tracking daily physical activity is the smartphone [6,7]. The smartphone includes a variety of advantages that make it an excellent stand-alone device. Specifically, smartphones include a variety of sensors, such as built-in accelerometry and global positioning system (GPS), increasingly powerful computing capabilities, large data capacity, wireless connectivity to other sensors (eg, connectivity to weight scales or blood pressure monitors), and Internet access. These technical capabilities allow smartphones to track, process, and send physical activity information while also functioning as a "hub" for other health information [7,8]. Beyond these technical capabilities, smartphones often accompany individuals throughout the day and, thus, easily fit into an individual's daily routine.

At present, there is relatively little systematic research exploring how accurately physical activity can be tracked via smartphones. Some studies have explored the utility of ecological momentary assessment (EMA) via smartphones [9], activity recognition using machine learning/neural network analyses [10-13], and activity classification via an iPod Touch for older adults [14]. There remain many unresolved questions related to tracking physical activity via smartphones. For example, although there was a study examining tracking among older adults, the vast majority of research is conducted among younger cohorts. The ActiGraph is a well-validated accelerometer commonly used in epidemiological, surveillance, and intervention research to quantify physical activity [1,15-18]. As such, a comparison of physical activity estimates collected via smartphone versus ActiGraph accelerometry would provide insights into whether a common smartphone could provide estimates of physical activity with similar levels of accuracy to this common field-based assessment strategy. This would be valuable based on the large body of research linking these cut-point-based estimates of sedentary, light, and moderate-to-vigorous levels of physical activity to health outcomes [1,15-18]. These data are largely absent for more advanced machine learning techniques of activity classification; thus, comparison to cut-point-based estimates is a scientifically important question until the health linkages to the more advanced analytic techniques can be made.

The purpose of this work was to determine the validity and reliability of Android smartphones for tracking physical activity utilizing cut-point–based methods of activity classification. In particular, we sought to determine the validity of Android smartphones for categorizing physical activity into sedentary, light physical activity, and moderate-to-vigorous levels of physical activity in a laboratory setting. We also explored interdevice reliability by comparing the estimates from 3 Android phones used simultaneously in the laboratory. Finally, we sought to determine the validity of the smartphones by comparing daily estimates of physical activity between the ActiGraph and an Android smartphone in a free-living context.

Methods

Overview

We conducted both a laboratory-based and free-living study. In the laboratory-based study (study 1), individuals engaged in prescribed activities of various intensities over a 2-hour period wearing both an ActiGraph and 3 Android smartphones (ie, HTC MyTouch, Google [manufactured by HTC] Nexus One, and Motorola Cliq) all worn both on the hip and in the pocket. In the free-living study (study 2), individuals engaged in usual daily activities that were tracked over a 7-day period via an Android smartphone (Google Nexus One) and an ActiGraph.

Study 1: Laboratory Study

Participants

Participants were a convenience sample of 15 midlife and older adults living in the San Francisco Bay area in California. Midlife and older adults (ie, aged 40 or older) were chosen because they represent an understudied population for activity classification and because the algorithms developed were explicitly being developed for a smartphone-based physical activity intervention [8]. Participants were recruited via word of mouth, advertisements, and email listservs. Participants completed the Physical Activity Readiness Questionnaire (PAR-Q) to ensure that they could safely engage in physical activity [19].

Procedures

The procedures for this study were similar to previously published work conducted by the study investigators [20]. Specifically, all participants completed a written informed consent that was approved by Stanford University's Human Use Committee. Participants were asked to wear an ActiGraph GT3X+ accelerometer on their nondominant hip along with the 3 Android smartphone phones also on the nondominant hip. Participants were asked to wear clothing to the laboratory session that was suitable for exercise and had pockets.

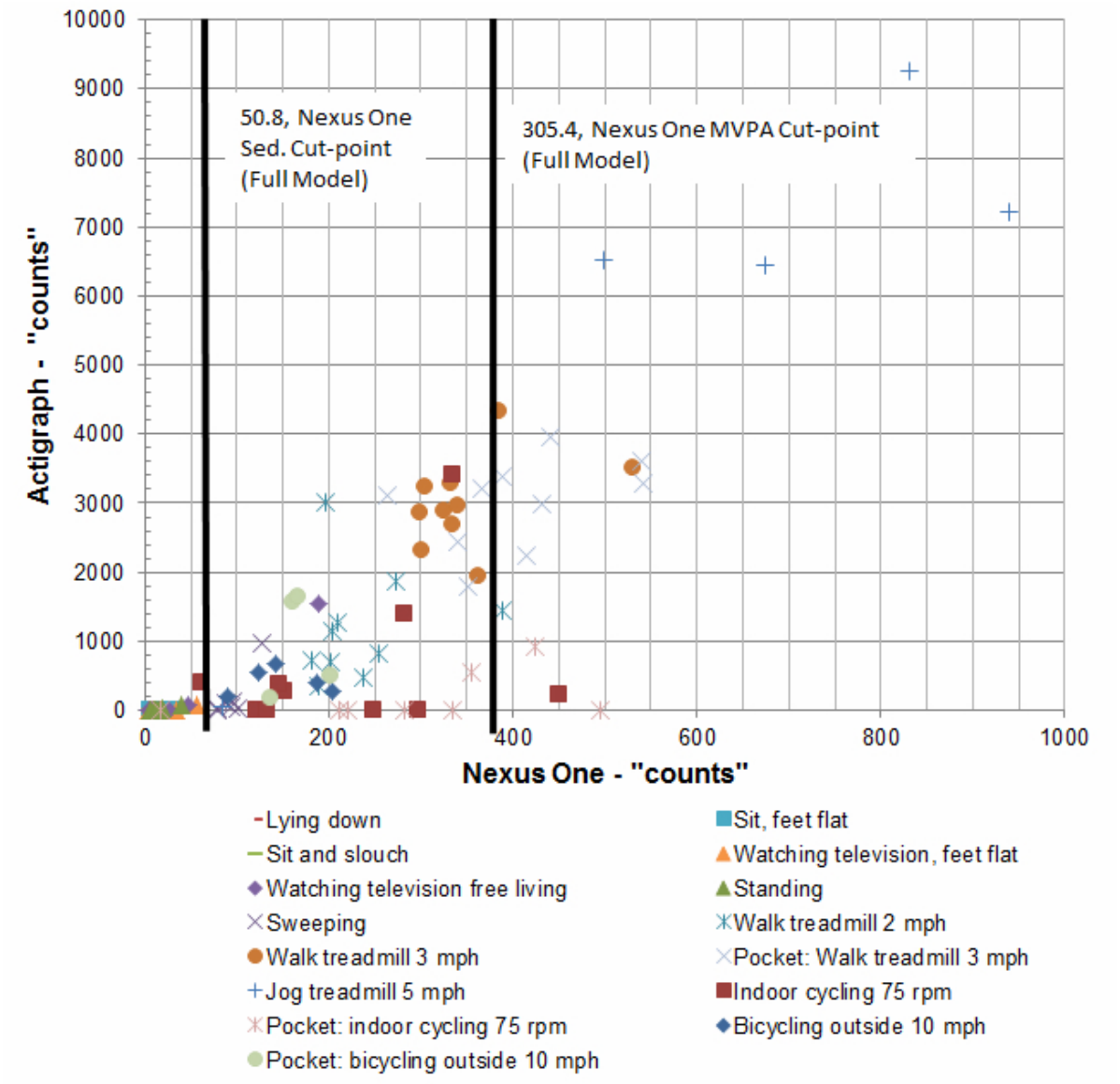
The ActiGraph GT3X+ is a small, electronic, triaxial device that is worn on the waist and measures activity "counts" (epoch set to raw for the 3 axes for this study but converted to 1-minute values with just the vertical axis for comparison with previously validated cut-point calculations) [1,21,22]. The 3 smartphones were chosen because at the time (2009) they were common Android devices and because comparison between the 3 phones would provide insights about reliability both within and between manufacturers. The Android platform was chosen because it

allowed the built-in accelerometer to run continuously and because there is increased use of the Android phones, particularly among lower income groups.

We chose to include 3 different Android smartphones to gauge the reliability and validity across manufacturers (ie, Motorola vs HTC) and within manufacturer by using different firmware versions of Android (ie, Nexus One vs MyTouch). After attaching the devices, participants were asked to engage in a series of activities while wearing the ActiGraph and the Android phones (see Figure 1 for a list of activities). Each activity was conducted for 5 minutes followed by a transition period to the next activity. The overall protocol lasted approximately 2 hours per participant. For some activities, particularly running on the treadmill at 5 mph, participants were given the option to opt out if they were not able to accomplish it safely.

A conversion was required to translate the raw values gathered from the smartphones into values similar to ActiGraph "counts." To calculate these phone-based counts, the first step was the use of a low-pass filter to account for the effects of gravity, followed by the calculation of an area under the curve measurement that represented total movement detected by the accelerometers. This technique is commensurate with previously published work [20].

Figure 1. Phone to ActiGraph comparison across activities (laboratory study).



Data Processing

Standard data processing techniques for calculating cut-point-derived estimates of sedentary, light physical activity, and moderate/vigorous levels of physical activity were used for processing the ActiGraph data based on previously published work on meaningful cut-points among older adults for sedentary (<100 counts) and moderate-to-vigorous intensity physical activity (>1951) [1]. These included setting the sampling rate for the ActiGraph to 80 Hz and gathering raw data across the 3 axes, which was then converted into counts using ActiLife 6.1 software (ActiGraph, Pensacola, FL, USA) based on the vertical axis. For the phone data, a custom app was developed that allowed for raw acceleration values similar to ActiGraph counts

to be attained. For all phones, the custom app sampled at the maximum rate allowed for each phone (ie, 20 Hz for the HTC MyTouch and Google Nexus One; 80 Hz for the Motorola Cliq).

Statistical Analyses

We first calculated an intraclass coefficient (ICC) reflecting the effect of individual variability on observed counts. Our analyses found that the within-subject variance was effectively zero across all possible comparisons using a leave-one-out cross-validation method for each mixed model analysis described subsequently. Based on this and that the counts were not normally distributed, Spearman rank order correlations were deemed an acceptable method for determining these associations. Mixed model analyses [23] were conducted to create a regression equation comparing each phone to the ActiGraph. The regression equations were validated utilizing a leave-one-out cross-validation procedure for each subject's activity as has been used in previous research [21]. Because the cut-point-based classification via the ActiGraph is known to have difficulties with properly estimating cycling activities and standing [24], and because phones worn in the pocket would likely affect the phone's counts, a variety of datasets with different filtering criterion were explored (eg, hip-only data, no biking and standing activities included).

As previously mentioned, we used fairly standardized cut-point values for ActiGraph counts of sedentary physical activity (ie, ActiGraph counts <100) [1,25] and MVPA (ie, ActiGraph counts >1952 [26]) because they have been validated among older adults specifically [1]. We utilized our regression equation (ie, phone cut-point values= beta*[imputed ActiGraph cut-point values] + intercept) to convert the validated ActiGraph cut-points into Android phone cut-points for the phone counts. We then utilized these derived phone-based cut-points to then label each minute of measured data via the phones as sedentary, light physical activity, or moderate-to-vigorous levels of physical activity.

We used standard conventions for labeling the strength of our Spearman correlations (ie, very weak=0-.19, weak=.2-.39, moderate=.4-.59, strong=.6-.79, and very strong=.8-1.0 [27]). Within the laboratory study, we had known activities with known metabolic equivalent (MET) values based on the Compendium of Physical Activities [28]. We calculated the percentage of total observations each phone correctly classified each activity as sedentary, light, or moderate-to-vigorous. To support interpretation of these percentages, we utilized the percent correctly classified by the ActiGraph minus 5% as the range for acceptable percent classification, similar to previous work [15]. For example, if the ActiGraph correctly classified 70% of total activities, the comparable range for the phones would be 65% or greater.

Study 2: Free-Living Study

Participants

Participants were a subsample from the previously reported Mobile Interventions for Lifestyle Exercise and eating at Stanford (MILES) study and, thus, recruitment procedures have been reported previously [8]. In brief, the target population consisted of community-dwelling adults in the San Francisco Bay area aged 45 years and older who were insufficiently physically active (ie, engaged in less than 60 minutes of self-reported MVPA per week), self-reported typically sitting for 10 or more hours per day, were able to participate safely in a physical activity program based on the PAR-Q [19], and were currently using a mobile phone but not a smartphone. In addition, participants were excluded if they did not provide sufficient data to complete the analyses (eg, insufficient wear time of the phone or ActiGraph; see data processing described subsequently).

Procedures

For this validation study, only the baseline phase (ie, the first week when no intervention was provided) from the MILES intervention study was used. Participants were requested to continue with their normal physical activity during the baseline phase. All participants (N=23) were provided a Nexus One smartphone equipped with a custom app for tracking physical activity via the built-in accelerometer and also wore an ActiGraph (GT3X+), utilizing standard quality control procedures for the ActiGraph [29]. Participants wore the ActiGraph accelerometer for at least 7 days (our request was 7 days, but due to scheduling issues some individuals wore the ActiGraph longer). Participants were asked to wear the ActiGraph and smartphone at the same times during waking hours. Participants wore the ActiGraph on their hip and were allowed to wear the smartphone either on the hip or in their pocket. After the baseline week, participants reported where they wore the phone and exploratory analyses suggested that placement did not influence any of the results (which were in-line with our laboratory-based study as described subsequently).

Data Processing

Data compliance and cleaning procedures for the ActiGraph and smartphone were consistent with other large-scale cross-sectional accelerometer studies [25,30], such that (1) valid hours of data consisted of no more than 60 consecutive "zero" values (interpreted as nonwear time) and (2) a valid day was defined as at least 10 valid hours/day for both the Nexus One phone and the ActiGraph. We included this 10-hour stipulation for both phones to ensure comparable wear time (note: in separate analyses not reported, there were no significant differences in wear time between the ActiGraph and Nexus One phone and individuals did wear them at the same times throughout the day). We did not include a stipulation on the number of valid days an individual must complete to be part of the study. Instead, we included any day that included both valid ActiGraph and smartphone accelerometry data. The same cut-points that were used in the laboratory-based study were used in the free-living study for the ActiGraph [26,31] and the phone cut-points were based on the regression equations calculated in the laboratory-based study.

Statistical Analyses

Although ICC were originally used, we again found minimal within-person variation. As such, Spearman rank order correlations were calculated between the ActiGraph and smartphone. In addition, Bland-Altman plots [32] were created

to compare the difference between the estimates of minutes in each activity intensity per day between the ActiGraph and the smartphone utilizing similar procedures reported previously [29]. As is convention with Bland-Altman plots, a criterion level of agreement that would be expected is required for proper interpretation of the mean difference and variance calculation. We assumed that a mean difference of less than 10 minutes constituted agreement for MVPA based on national guideline recommendations suggesting clinically meaningful activity occurring past the 10-minute threshold [3]. There is no consensus on what "meaningful" differences in sedentary or light activity might be. Nonetheless, based on our own previous work [1], we postulated that mean differences greater than 60 minutes for sedentary and light activity would suggest poor agreement. For this study, we did not have a "gold-standard" measure of sedentary, light physical activity, and moderate-to-vigorous levels of physical activity but instead a well-validated field measure, the ActiGraph. Without a gold-standard metric, it is hard to determine what the "right" answer should be. This is important to consider with regard to interpretation of the confidence intervals because wide confidence intervals could occur based on error from both the ActiGraph and the smartphone. As such, although we report confidence intervals in our Bland-Altman plots, we suggest caution in overinterpretation of these confidence intervals.

Results

Study 1: Laboratory Study

There were a total of 15 participants (age: mean 55.5 years, SD 6.6, range 43-65; women: 56%, 8/15; education: mean 16.25 years, SD 1.6) who participated in study 1. A list of the activities in the laboratory study can be found in Figure 1. Table 1 reports Spearman rank order correlations between the ActiGraph counts and the 3 phones. Overall, results suggest strong correlations between the ActiGraph and the 3 phones and between the 3 phones (ρ=.90). Figure 1 is a visual representation of all activities with color-coded labeling to identify the classification of each activity as sedentary, light physical activity, or MVPA.

Table 1. Spearman rank order correlations[a] (ρ) between raw ActiGraph and raw phone counts for the laboratory study (N=15).

Statistical model by activity monitoring device[b]	Cliq		MyTouch		Nexus One	
	ρ	*P*	ρ	*P*	ρ	*P*
Full model						
ActiGraph	.77	<.001	.82	<.001	.80	<.001
Cliq			.95	<.001	.90	<.001
MyTouch					.90	<.001
No bike & standing						
ActiGraph	.85	<.001	.89	<.001	.83	<.001
Cliq			.93	<.001	.88	<.001
MyTouch					.87	<.001
Hip only						
ActiGraph	.82	<.001	.86	<.001	.85	<.001
Cliq			.93	<.001	.95	<.001
MyTouch					.93	<.001
Hip only and no bike & standing						
ActiGraph	.81	<.001	.85	<.001	.83	<.001
Cliq			.92	<.001	.92	<.001
MyTouch					.89	<.001

[a] The Spearman correlations are between counts/min derived for the ActiGraph and the 3 Android smartphones (Motorola Cliq, HTC MyTouch, and Google/HTC Nexus One).

[b] The different models correspond to different filters (ie, no bike & standing excludes bicycling and standing; hip-only excludes measures whereby the phones were in the pocket).

Table 2 reports results of the mixed model regression equations for each cut-point estimate. The betas and intercepts were the values used to calculate the phone-based cut-points listed when the ActiGraph cut-points (ie, <100 and >1952) were imputed into the equation. Figure 2 provides estimates of the mean differences between the predicted values from the phones and the actual ActiGraph counts across phones and across different modeling datasets (eg, the full dataset, to the dataset that only included hip data and excluded biking/standing). Overall, there was some instability in the overall estimate depending on the observations included/excluded from the models using the leave-one-out technique based on the large root mean standard errors but that these differences were not greatly impacted by the specific phone used or the filtering strategies. Overall, the models were improved if data were aggregated across all phones as opposed to using phone-specific estimates. For the remainder

of the paper, we only refer to the aggregated estimates of cut-points across phones as these appeared most stable and we continue to report both the "full model" and "no bike and standing" models because there appeared to be improved model fit using this filter but no improved model fit when excluding measurements from the phones while worn in the pocket.

Table 2. Regression equations and cut-points for sedentary and moderate-to-vigorous levels of physical activity in the laboratory.

Phone	Intercept[a]	b[a]	N	Sedentary cut-point[b]	MVPA cut-point[c]
Full data					
Moto Cliq	–260.34	6.73	169	53.54	328.72
HTC MyTouch	–304.62	7.96	169	50.82	283.46
Google Nexus One	–247.00	7.15	169	48.56	307.76
All phones mean	–270.65	7.28		50.92	305.36
No bike & standing					
Moto Cliq	–182.81	7.63	133	37.06	279.77
HTC MyTouch	–205.17	8.67	133	35.20	248.86
Goo/HTC Nexus One	–171.66	8.14	133	33.36	260.75
All phones mean	–186.55	8.15		35.17	262.47

[a] The betas and intercepts were developed as an aggregation of the results of the leave-one-out technique (ie, averaging the beta and intercept estimates from all models).

[b] The cut-point value imputed into the regression equation for the sedentary cut-point was <100.

[c] The cut-point value imputed into the regression equation for the MVPA cut-point was >1951.

Figure 2 shows the mean difference between the predicted value of an ActiGraph count compared to the actual ActiGraph count (ie, predicted count – actual count). This mean difference across different models allows for an estimate of differences across models when using the leave-one-out strategy of model building. Higher mean differences suggest less stable regression models that are more influenced by individual observations in the model. The error bars represent the root mean standard error across all models run utilizing a leave-one-out procedure, which further provides insights on the interpretability of the models. High root mean standard errors suggest large variation in estimates during the leave-one-out procedure, but also provide an estimate of "meaningful" mean difference estimates across the various regression models. Results suggest that the mean differences observed did not differ by phone or data-filtering strategy used, but that there was large variation of impact across observations, thus justifying the use of the leave-one-out method for creating a more stable regression model.

Table 3 reports the percentage of time each device accurately classified an activity into its corresponding MET classification (ie, sedentary, light physical activity, or MVPA) [28]. Overall, results suggest that the phones correctly classified activities at nearly the same level as the ActiGraph. For example, when excluding behaviors from the model that are known to be poorly classified using cut-point estimates (ie, bicycling and standing), the ActiGraph correctly classified 91% (108/119) of all activities across all participants. In comparison, the phones correct classification levels ranged from 78% (73/93; MyTouch, Full Model) to 91% (85/93; Nexus One, no bike and standing model). Not surprisingly, cycling activities and standing exhibited the poorest estimated agreement (see Table 3). Interestingly, the phones actually did better at correctly classifying the light intensity activity of sweeping. Again, Figure 1 provides a visual summary of these classifications on a labeled scatterplot for the Nexus One.

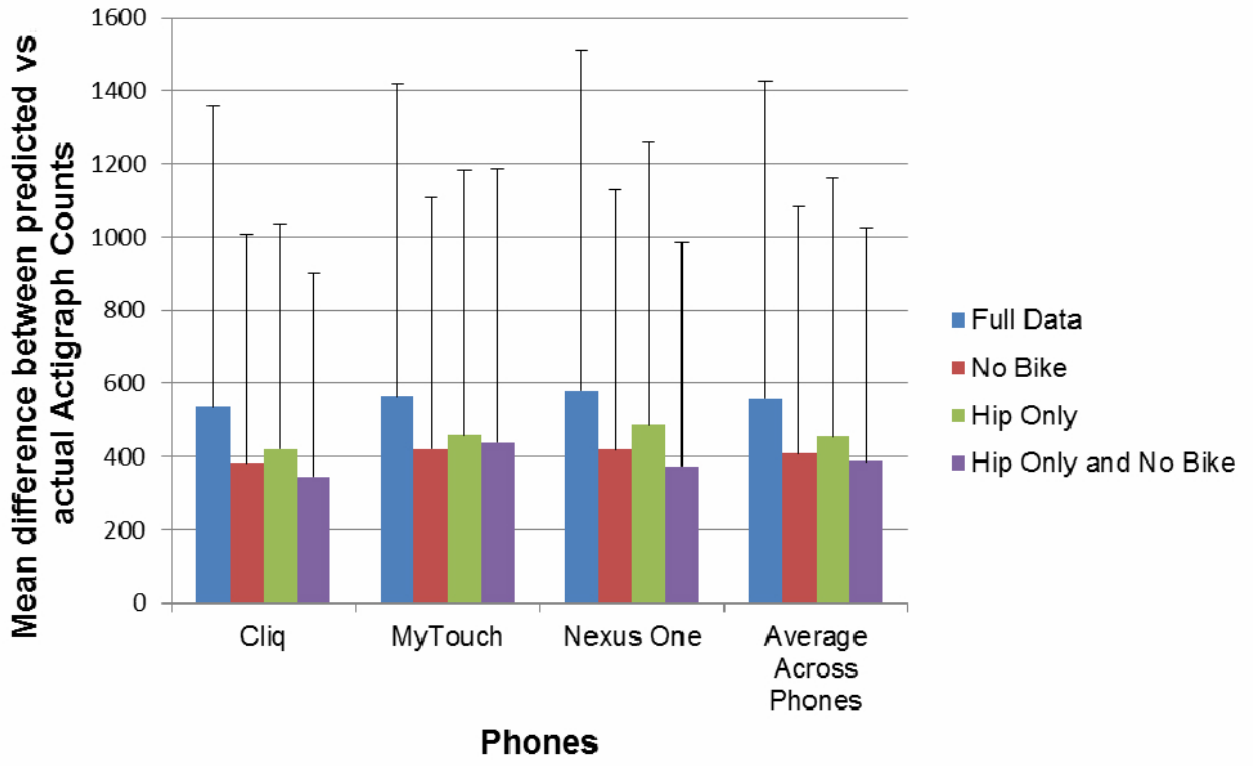
Table 3. Correct classification of activity intensity level for each device.

Activity	Placement	Correct classification,[a] %						
		Nexus One		MyTouch		Cliq		ActiGraph
		Full	No bike & stand[b]	Full	No bike & stand[b]	Full	No bike & stand[b]	N/A
Overall		69%	73%	59%	60%	63%	65%	64%
Overall excluding behaviors[b]		90%	91%	78%	80%	83%	83%	91%
Bicycling outside 10 mph	Hip	0%	0%	0%	0%	0%	0%	0%
Bicycling outside 10 mph	Pocket	0%	0%	0%	0%	0%	0%	0%
Cycling indoors 75 rpm	Hip	18%	36%	0%	0%	0%	9%	7%
Cycling indoors 75 rpm	Pocket	50%	63%	63%	63%	63%	88%	0%
Lying down	Hip	91%	91%	82%	82%	82%	82%	100%
Sitting while slouching	Hip	100%	100%	89%	89%	89%	89%	100%
Sitting with back straight	Hip	100%	100%	80%	80%	90%	90%	100%
Television (free-movement)	Hip	91%	82%	82%	73%	82%	73%	93%
Television (sitting straight)	Hip	90%	90%	80%	80%	80%	80%	100%
Standing Straight	Hip	9%	18%	0%	0%	9%	9%	0%
Sweeping	Hip	100%	100%	100%	100%	100%	100%	50%
Treadmill 2 mph	Hip	90%	80%	90%	90%	90%	90%	92%
Treadmill 3 mph	Hip	80%	100%	20%	40%	80%	90%	92%
Treadmill 3 mph	Pocket	80%	90%	90%	90%	70%	70%	91%
Treadmill 5 mph	Hip	80%	80%	80%	80%	60%	60%	80%

[a] These values are the percentage of times the activity was correctly categorized according to its physical activity intensity level (ie, sedentary, light, or moderate-to-vigorous intensity physical activity) by each of the 4 devices. For this work, we explored correct classification both using different phones and via different cut-point algorithms based on different regression models. The cut-points used here were the average cut-point estimates across all the phones for both the full model cut-points (ie, <50.92 for sedentary and >305.36 for MVPA) and the model generated when biking and standing was excluded when creating the cut-point estimates (ie, “no bike & standing” model cut-points were <35.17 for sedentary and >262.47 for MVPA). For the full model, N=132.

[b] This overall estimate of correct classification excluded the following behaviors that are known to be problematic for classifying using a cut-point strategy: bicycling outdoors, indoor cycling, and standing still (N=93).

Figure 2. Comparison of the stability of different regression model estimates (laboratory study). The error bars represent the root mean standard error across all models run utilizing a leave one out procedure.



Study 2: Free-Living Study

There were 23 participants who had acceptable data for comparing the phone accelerometer to the ActiGraph (age: mean 57.0 years, SD 6.4, range 45-69; women: 74%, 17/23; body mass index: mean 29.5, SD 5.9, range: 20.8-40.9; white: 70%, 16/23; Asian/Asian American: 22%, 5/23; bachelor's degree or higher: 79%, 18/23; married: 44%, 10/23; working full time: 64%, 14/23). On average, participants had approximately 5 days of valid wear time for both the ActiGraph and smartphone accelerometry data (days per participant: mean 4.8, SD 2.1; total days=107). Table 4 reports correlations between the estimated number of minutes engaged in sedentary, light physical activity, and moderate-to-vigorous levels of physical activity based on the Nexus One phone using the full model cut-points derived from study 1 compared to the ActiGraph (listed in Figure 1). Overall, results suggested moderate to strong correlations between the direct estimates for sedentary physical activity and MVPA (eg, raw count comparisons: ρ=.44, *P*<.001; sedentary: ρ=44, *P*<.001; MVPA: ρ=.67, *P*<.001) and weak correlations for light physical activity (ρ=.38, *P*<.001). These correlations were nearly the same as the no bike and standing correlations (eg, raw count comparisons: ρ=.35, *P*<.001; sedentary: ρ=.44, *P*<.001; light: ρ=.34, *P*<.001; MVPA: ρ=.68, *P*<.001).

Table 4. Free-living Spearman rank correlations between ActiGraph and NexusOne smartphone.

Actigraph	Smartphone[a]							
	Raw count		Sedentary		Light		MVPA	
	ρ	*P*	ρ	*P*	ρ	*P*	ρ	*P*
Raw count	.59	<.001	–.22	.02	.32	<.001	.57	<.001
Sedentary	–.34	<.001	.44	<.001	.11	.27	.00	.98
Light	.14	.16	–.13	.20	.38	<.001	–.07	.49
MVPA	.54	<.001	–.21	.03	.06	.53	.67	<.001

[a] Smartphone estimates of min/day in each category are based on the "full" model average cut-points that were derived from study 1.

Figures 3-5 report Bland-Altman plots comparing the ActiGraph and smartphones utilizing the full model cut-point estimates. Comparison of the plots suggest good absolute mean-level differences in sedentary and MVPA activity minutes/day estimates (sedentary: mean difference=–26.0 min/day, 95% CI –279.5 to 227.6; MVPA: mean difference=–1.3 min/day, 95% CI –38.4 to 35.8). Absolute mean-level differences for light physical activity were outside of the acceptable range (mean difference=–111.2 min/day, 95% CI –285.8 to –63.5). Although not shown, we did create Bland-Altman plots for the no bike and standing models and consistently found poorer absolute estimates suggesting that for absolute estimates in a free-living

context, the full model cut-points listed in Table 2 (ie, <50.92 for sedentary and >305.36 for MVPA) were superior. Based on the lack of a gold standard in this study, the 95% confidence intervals are not as easily interpreted because the error likely comes from measurement issues with both devices. As such, they are provided more for broader context but should be interpreted with caution.

Figure 3. Bland-Altman plot comparing estimated minutes of MVPA per day between the phone and ActiGraph, full model (free-living study).

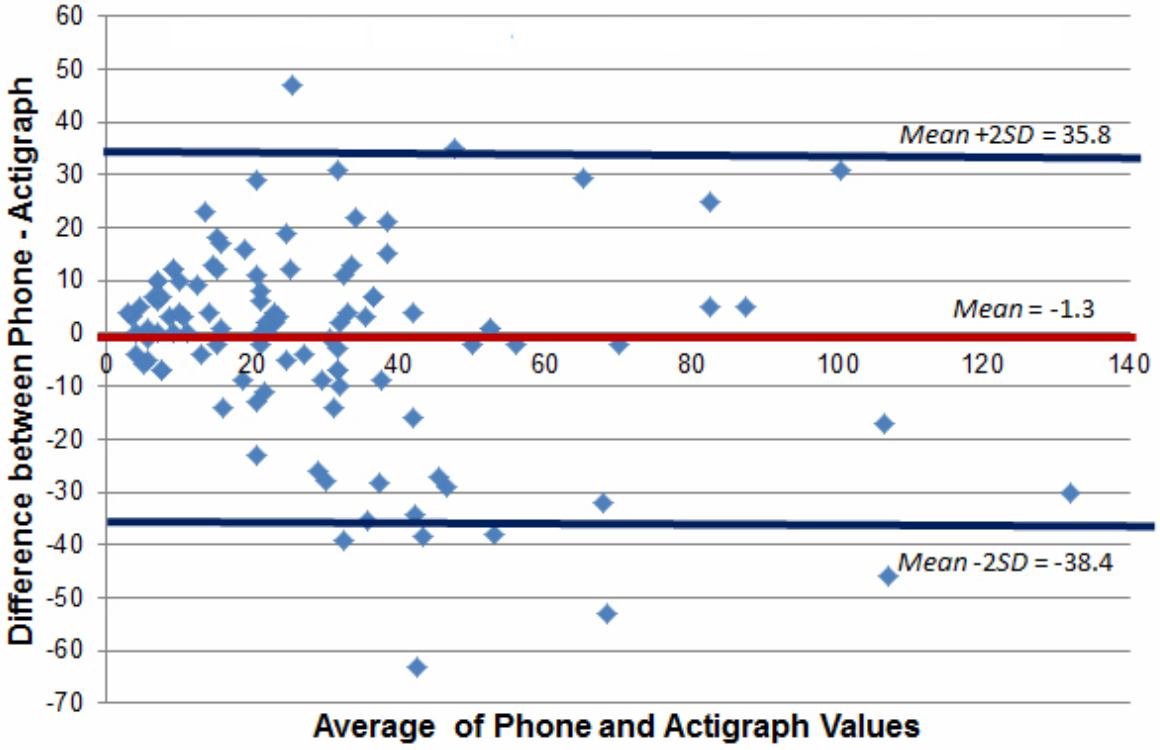


Figure 4. Bland-Altman plot comparing estimated minutes of light activity per day between the phone and ActiGraph, full model (free-living study).

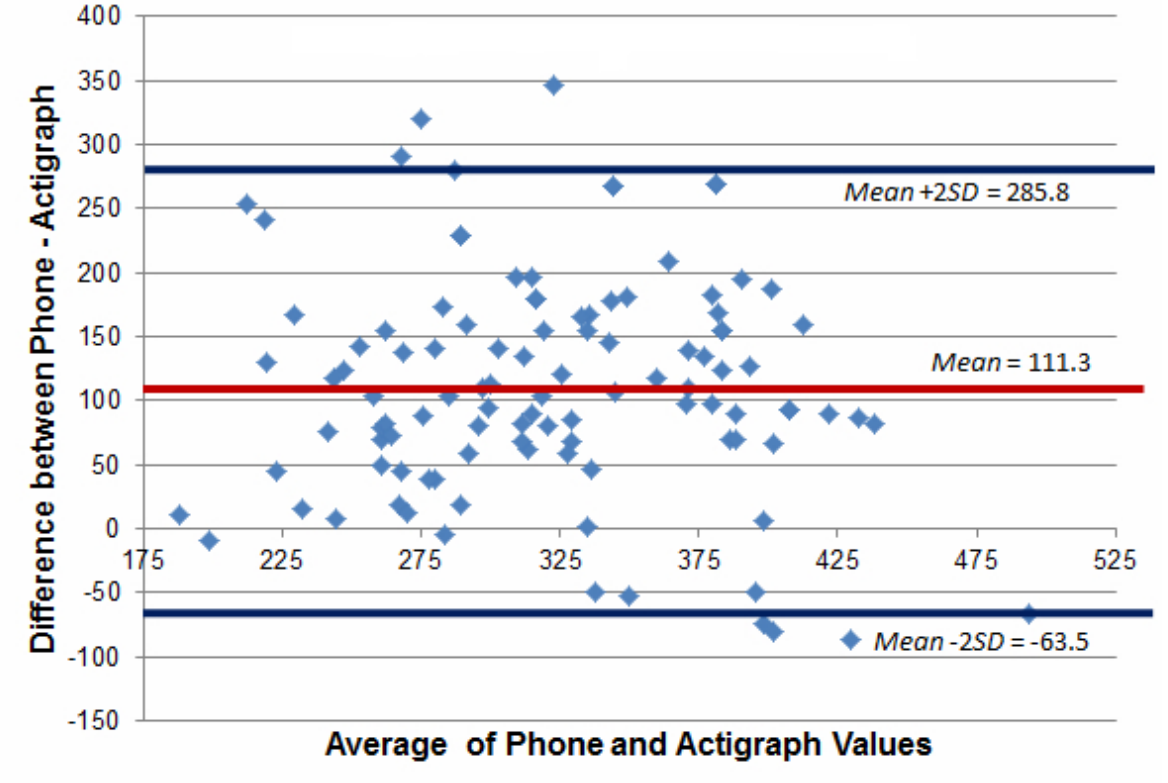
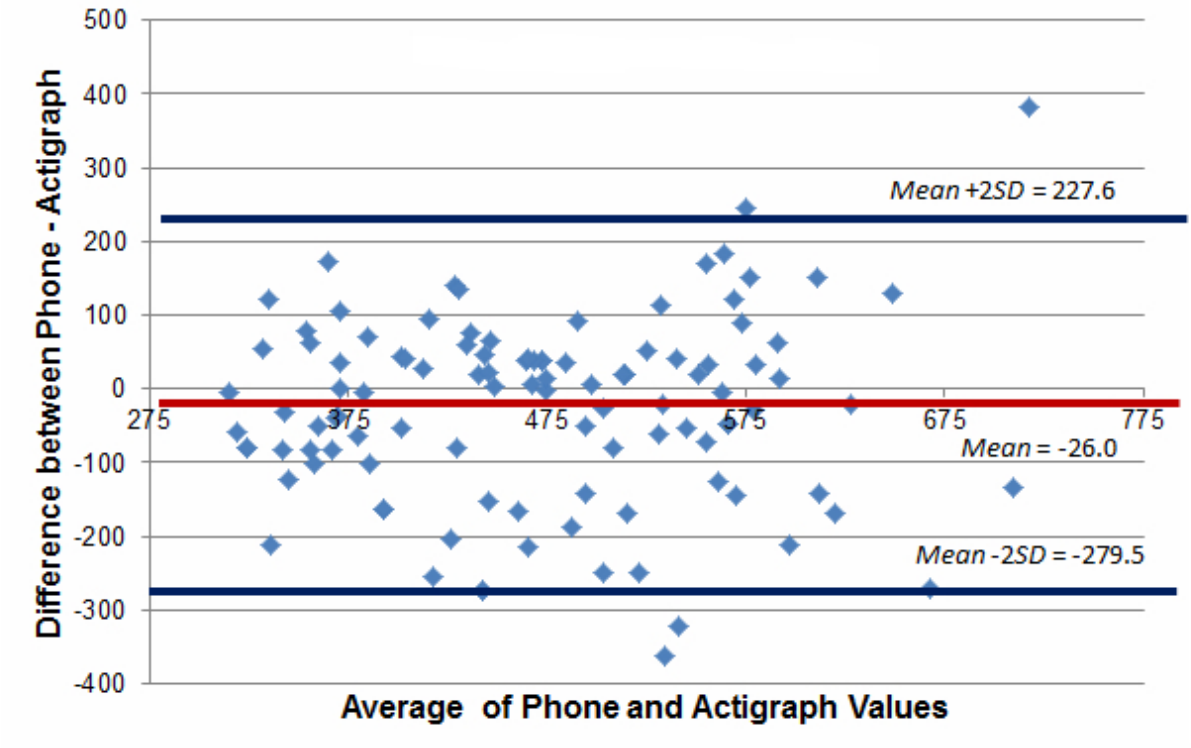


Figure 5. Bland-Altman plot comparing estimated minutes of sedentary behavior per day between the phone and ActiGraph, full model (free-living study).



Discussion

Principal Findings

The purpose of this work was to determine the validity and reliability across different Android smartphones for tracking physical activity among midlife and older adults. Overall, results indicated (1) Android smartphone raw counts are strongly correlated to ActiGraph counts in a laboratory-based and free-living setting suggesting the accelerometers provide similar estimates regardless of the activity classification algorithm used, (2) Android smartphone raw counts were strongly correlated with one another suggesting that different Android phones reliably provide similar estimates before any algorithm, (3) placement in the pocket versus on the hip both in a laboratory-based and free-living context did not detrimentally impact estimates suggesting that phones can reliably be worn in either location, and (4) absolute classifications of activity intensity between the phone and ActiGraph were comparable both in-laboratory (based on similar correct classification percentages of known activities) and free-living context (based on absolute mean difference estimates via Bland-Altman plots) for sedentary and MVPA estimates but not light physical activity.

Results indicating that the phone accelerometers and ActiGraph are strongly correlated have important implications for not only cut-point-based activity classification strategies, but also any classification strategy utilizing an accelerometer. In particular, because the effective raw signals are providing similar information, it is quite plausible that similar algorithm strategies can be used between devices with similar effects, assuming some degree of calibration. This is important because it suggests other activity recognition strategies, such as machine learning/neural network analyses [10-13], can likely provide similar estimates from the raw accelerometer signals across phones. This also increases confidence that estimates between Android phones from apps that utilize the accelerometer for activity classification will likely be comparable across phones.

Results suggesting phone placement did not impact sedentary and MVPA estimates have important implications for usability. Specifically, we found an almost even split in preferences related to wearing the phone, with some desiring to wear it on the hip and others desiring to keep it in their pockets. A phone will likely be worn if it fits into a person's daily life. As such, allowing individuals to wear the device in whatever way they so desire increases the likelihood of continuing to gather the data in the long-term (as we did in our MILES intervention trial [8]). Our results suggest that acceptable activity estimates can be acquired either on the hip or in the pocket. Further work should continue to explore ways to further improve "wearability" of a system. This interest is evident in the commercial sector with devices such as the Fitbit Flex, Jawbone UP, or Misfit Shine, which are all wrist-worn activity monitors.

Finally, our results suggesting that the phones gave similar absolute estimates of sedentary and MVPA activities to an ActiGraph in both in-laboratory and free-living studies has important implications for epidemiologic and intervention research. At present, there is far more data supporting the linkage between cut-point-derived estimates of physical activity with health outcomes compared to the other data analytic techniques [1,15,33]. Based on this, although it is likely the machine learning-based classification techniques will eventually provide better methods for classifying physical activity, at present, there is still value in focusing on validating the smartphone compared to the classical cut-point-derived

estimates from the ActiGraph. Our results suggest that the phones can provide an acceptable alternative to an ActiGraph for classifying sedentary and MVPA, but not light physical activity.

Limitations

The ActiGraph was utilized as our primary comparison metric. Although the ActiGraph is acceptable for field-based research, it is not a gold-standard activity monitoring device. This is particularly important to be mindful of when interpreting the Bland-Altman plots because the Bland-Altman plots are designed to support comparison of a new measure to a previous gold-standard metric. The Bland-Altman plots revealed good mean-level differences but poor confidence interval estimates between the smartphone and ActiGraph. Because the ActiGraph is not a gold-standard measure, it is quite plausible that the high confidence intervals can be attributed not only to variance from the smartphone but also the ActiGraph. As such, the confidence intervals from the Bland-Altman plots need to be interpreted with caution. Further, the ActiGraph cut-points utilized incorporated only the vertical axis whereas the phones used all 3 axes, thus establishing another potential source of error between the two.

A second limitation of our work is that we utilized a cut-point-based strategy for activity classification rather than newer strategies such as neural network analysis or machine learning techniques [10-13]. Although a cut-point strategy does not afford the level of precision that these newer techniques do, cut-point estimates are still pragmatic to use in both epidemiologic and intervention research. In particular, the national guidelines for physical activity do not focus on specific behaviors, but instead the broad classes of moderate and vigorous intensity physical activity that we can classify with cut-point algorithms [3]. As such, intensity classification has utility for physical activity recommendations because it more directly corresponds with national guidelines. In addition, there is also much more work linking cut-point-based estimates of physical activity to health outcomes compared with the more precise activity classification techniques [1,33]. As such, the regression equations generated provide a strategy for establishing a linkage from estimates derived from smartphones to those health outcome linkages made via previous work that found associations with ActiGraph estimates via cut-point algorithms and health outcomes.

Another limitation is the brands of the smartphones we studied. In 2009, HTC and Motorola were among the most important manufacturers of Android smartphones, but at this point Samsung has become the primary Android phone manufacturer. Unfortunately, our study cannot provide any insights on the quality of Samsung devices. Further, although the data were comparable across our phones, it is impossible to determine if the phone estimates would remain comparable in newer phones. That said, based on the plethora of different Android phones available, our data at least increases confidence that different manufacturers and running different firmware versions of Android may provide similar estimates of physical activity based on the accelerometry.

Strengths of this research were that data are available from both an in-laboratory and free-living context. Further, the samples used to create the estimates are in-line with other physical activity assessment/validation protocols. Finally, validation occurred via direct observation of activities in laboratory.

Conclusions

Overall, our results suggest that an Android smartphone appears to be an acceptable alternative for estimating sedentary and moderate-to-vigorous intensity physical activity to an ActiGraph accelerometer in both a laboratory-based and free-living context. This suggests that smartphones might be an effective mechanism, by themselves, for tracking physical activity. Future work should explore other potential analytic techniques (eg, machine learning) for further improving the classification.

Acknowledgments

This study was funded, in part, by a grant from the National Heart, Lung, and Blood Institute (NHLBI-1RC1 HL099340, PI-King).

Conflicts of Interest

None declared.

References

1. Buman MP, Hekler EB, Haskell WL, Pruitt L, Conway TL, Cain KL, et al. Objective light-intensity physical activity associations with rated health in older adults. Am J Epidemiol 2010 Nov 15;172(10):1155-1165 [FREE Full text] [doi: 10.1093/aje/kwq249] [Medline: 20843864]
2. Owen N, Bauman A, Brown W. Too much sitting: a novel and important predictor of chronic disease risk? Br J Sports Med 2009 Feb;43(2):81-83. [doi: 10.1136/bjsm.2008.055269] [Medline: 19050003]
3. Physical Activity Guidelines Advisory Committee. Physical Activity Guidelines Advisory Committee Report. Washington, DC: US Department of Health and Human Services; 2008.
4. Hekler EB, Klasnja P, Traver V, Hendriks M. Realizing effective behavioral management of health: the metamorphosis of behavioral science methods. IEEE Pulse 2013 Sep;4(5):29-34. [doi: 10.1109/MPUL.2013.2271681] [Medline: 24056791]
5. Liu S, Gao RX, Freedson PS. Computational methods for estimating energy expenditure in human physical activities. Med Sci Sports Exerc 2012 Nov;44(11):2138-2146 [FREE Full text] [doi: 10.1249/MSS.0b013e31825e825a] [Medline: 22617402]

6. Intille SS, Lester J, Sallis JF, Duncan G. New horizons in sensor development. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S24-S31 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399c7d] [Medline: 22157771]
7. Kranz M, Möller A, Hammerla N, Diewald S, Plötz T, Olivier P, et al. The mobile fitness coach: towards individualized skill assessment using personalized mobile devices. Pervasive and Mobile Computing 2013 Apr;9(2):203-215. [doi: 10.1016/j.pmcj.2012.06.002]
8. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
9. 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]
10. Lau S, David K. Movement recognition using the accelerometer in smartphones. Future Network and Mobile Summit' 2010:10.
11. Saponas T, Lester J, Froehlich J, Fogarty J, Landay J. iLearn on the iPhone: real-time human activity classification on commodity mobile phones. University of Washington CSE Tech Report UW-CSE-08-04-02 2008:1-4. [doi: 10.1145/1738826.1738911]
12. Kwapisz JR, Weiss GM, Moore SA. Activity recognition using cell phone accelerometers. SIGKDD Explor Newsl 2011 Mar 31;12(2):74-82. [doi: 10.1145/1964897.1964918]
13. Győrbíró N, Fábián Á, Hományi G. An activity recognition system for mobile phones. Mobile Netw Appl 2008 Nov 7;14(1):82-91. [doi: 10.1007/s11036-008-0112-y]
14. Vankipuram M, McMahon S, Fleury J. ReadySteady: app for accelerometer-based activity monitoring and wellness-motivation feedback system for older adults. AMIA Annu Symp Proc 2012;2012:931-939 [FREE Full text] [Medline: 23304368]
15. Copeland JL, Esliger DW. Accelerometer assessment of physical activity in active, healthy older adults. J Aging Phys Act 2009 Jan;17(1):17-30. [Medline: 19299836]
16. Freedson P, Bowles HR, Troiano R, Haskell W. Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S1-S4 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399b7e] [Medline: 22157769]
17. Freedson PS, John D. Comment on "estimating activity and sedentary behavior from an accelerometer on the hip and wrist". Med Sci Sports Exerc 2013 May;45(5):962-963. [doi: 10.1249/MSS.0b013e31827f024d] [Medline: 23594509]
18. Tudor-Locke C, Camhi SM, Troiano RP. A catalog of rules, variables, and definitions applied to accelerometer data in the National Health and Nutrition Examination Survey, 2003-2006. Prev Chronic Dis 2012;9:E113 [FREE Full text] [Medline: 22698174]
19. Cardinal BJ, Esters J, Cardinal MK. Evaluation of the revised physical activity readiness questionnaire in older adults. Med Sci Sports Exerc 1996 Apr;28(4):468-472. [Medline: 8778552]
20. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille S. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. Med Sci Sports Exerc 2013 May;45(5):964-975 [FREE Full text] [doi: 10.1249/MSS.0b013e31827f0d9c] [Medline: 23247702]
21. Welk GJ. Use of accelerometry-based activity monitors to assess physical activity. In: Welk GJ, editor. Physical Activity Assessments for Health-Related Research. Chicago, IL: Human Kinetics, Inc; 2002:124-141.
22. Pruitt LA, Glynn NW, King AC, Guralnik JM, Aiken EK, Miller G, et al. Use of accelerometry to measure physical activity in older adults at risk for mobility disability. J Aging Phys Act 2008 Oct;16(4):416-434 [FREE Full text] [Medline: 19033603]
23. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. New York: Oxford University Press; 2003.
24. Crouter SE, Kuffel E, Haas JD, Frongillo EA, Bassett DR. Refined two-regression model for the ActiGraph accelerometer. Med Sci Sports Exerc 2010 May;42(5):1029-1037 [FREE Full text] [doi: 10.1249/MSS.0b013e3181c37458] [Medline: 20400882]
25. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
26. 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]
27. Martella RC, Nelson JR, Morgan RL, Marchand-Martella NE. Understanding and Interpreting Educational Research. New York: Guilford Press; 2013.
28. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S498-S504. [Medline: 10993420]
29. Hekler EB, Buman MP, Haskell WL, Conway TL, Cain KL, Sallis JF, et al. Reliability and validity of CHAMPS self-reported sedentary-to-vigorous intensity physical activity in older adults. J Phys Act Health 2012 Feb;9(2):225-236. [Medline: 22368222]

30. Sallis JF, Saelens BE, Frank LD, Conway TL, Slymen DJ, Cain KL, et al. Neighborhood built environment and income: examining multiple health outcomes. Soc Sci Med 2009 Apr;68(7):1285-1293 [FREE Full text] [doi: 10.1016/j.socscimed.2009.01.017] [Medline: 19232809]
31. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol 2008 Apr 1;167(7):875-881 [FREE Full text] [doi: 10.1093/aje/kwm390] [Medline: 18303006]
32. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
33. Loprinzi PD, Lee H, Cardinal BJ, Crespo CJ, Andersen RE, Smit E. The relationship of actigraph accelerometer cut-points for estimating physical activity with selected health outcomes: results from NHANES 2003-06. Res Q Exerc Sport 2012 Sep;83(3):422-430. [doi: 10.1080/02701367.2012.10599877] [Medline: 22978192]

Abbreviations

EMA: ecological momentary assessment
GPS: global positioning system
MET: metabolic equivalent
MILES: Mobile Interventions for Lifestyle Exercise and eating at Stanford
PAR-Q: Physical Activity Readiness Questionnaire
MVPA: Moderate-to-Vigorous Physical Activity
ICC: intraclass coefficient

Edited by G Eysenbach; submitted 22.05.14; peer-reviewed by G Norman, S Weegen; comments to author 25.07.14; revised version received 02.09.14; accepted 19.10.14; published 15.04.15.

Please cite as:
Hekler EB, Buman MP, Grieco L, Rosenberger M, Winter SJ, Haskell W, King AC
Validation of Physical Activity Tracking via Android Smartphones Compared to ActiGraph Accelerometer: Laboratory-Based and Free-Living Validation Studies
JMIR mHealth uHealth 2015;3(2):e36
URL: http://mhealth.jmir.org/2015/2/e36/
doi:10.2196/mhealth.3505
PMID:25881662

Short Paper

A New Mobile Phone-Based Tool for Assessing Energy and Certain Food Intakes in Young Children: A Validation Study

Hanna Henriksson[1,2], BSc; Stephanie E Bonn[3], PhD; Anna Bergström[4], PhD; Katarina Bälter[3], PhD; Olle Bälter[5], PhD; Christine Delisle[2], MSc; Elisabet Forsum[1], PhD; Marie Löf[1,2], PhD

[1]Linköping University, Department of Clinical and Experimental Medicine, Linköping, Sweden

[2]Karolinska Institutet, Department of Biosciences and Nutrition, Huddinge, Sweden

[3]Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Stockholm, Sweden

[4]Karolinska Institutet, Institute of Enviromental Medicine, Stockholm, Sweden

[5]Royal Institute of Technology, School of Computer Science and Communication, Stockholm, Sweden

Corresponding Author:
Marie Löf, PhD
Karolinska Institutet
Department of Biosciences and Nutrition
Novum
Huddinge, 14183
Sweden
Phone: 46 734426417
Fax: 46 8 31 11 01
Email: marie.lof@ki.se

Abstract

Background: Childhood obesity is an increasing health problem globally. Obesity may be established already at pre-school age. Further research in this area requires accurate and easy-to-use methods for assessing the intake of energy and foods. Traditional methods have limited accuracy, and place large demands on the study participants and researchers. Mobile phones offer possibilities for methodological advancements in this area since they are readily available, enable instant digitalization of collected data, and also contain a camera to photograph pre- and post-meal food items. We have recently developed a new tool for assessing energy and food intake in children using mobile phones called the Tool for Energy Balance in Children (TECH).

Objective: The main aims of our study are to (1) compare energy intake by means of TECH with total energy expenditure (TEE) measured using a criterion method, the doubly labeled water (DLW) method, and (2) to compare intakes of fruits and berries, vegetables, juice, and sweetened beverages assessed by means of TECH with intakes obtained using a Web-based food frequency questionnaire (KidMeal-Q) in 3 year olds.

Methods: In this study, 30 Swedish 3 year olds were included. Energy intake using TECH was compared to TEE measured using the DLW method. Intakes of vegetables, fruits and berries, juice, as well as sweetened beverages were assessed using TECH and compared to the corresponding intakes assessed using KidMeal-Q. Wilcoxon matched pairs test, Spearman rank order correlations, and the Bland-Altman procedure were applied.

Results: The mean energy intake, assessed by TECH, was 5400 kJ/24h (SD 1500). This value was not significantly different (P=.23) from TEE (5070 kJ/24h, SD 600). However, the limits of agreement (2 standard deviations) in the Bland-Altman plot for energy intake estimated using TECH compared to TEE were wide (2990 kJ/24h), and TECH overestimated high and underestimated low energy intakes. The Bland-Altman plots for foods showed similar patterns. The mean intakes of vegetables, fruits and berries, juice, and sweetened beverages estimated using TECH were not significantly different from the corresponding intakes estimated using KidMeal-Q. Moderate but statistically significant correlations (ρ=.42-.46, P=.01-.02) between TECH and KidMeal-Q were observed for intakes of vegetables, fruits and berries, and juice, but not for sweetened beverages.

Conclusion: We found that one day of recordings using TECH was not able to accurately estimate intakes of energy or certain foods in 3 year old children.

(JMIR mHealth uHealth 2015;3(2):e38) doi:10.2196/mhealth.3670

KEYWORDS

cell phone; digital camera; food intake; energy intake; child; DLW; FFQ

Introduction

According to the World Health Organization (WHO), childhood obesity is one of the most serious public health challenges of the 21st century [1]. Therefore, there is growing interest for interventional and observational studies already in pre-school children (2-6 years) [2]. However, such studies are difficult to conduct since traditional dietary assessment methods have limited accuracy, and involve excessive effort for caretakers and researchers [3]. Mobile phones open new possibilities since they are readily available, enable instant data digitalization, and contain a camera for photographing pre- and post-meal food items. Photos using digital cameras have shown potential for assessing dietary intake in both adults [4-6] and children [7-9]. We have developed a new tool for assessing energy and food intake using mobile phones, the Tool for Energy Balance in Children (TECH). The aims of this pilot study in healthy 3 year olds were (1) to compare energy intake by means of TECH with total energy expenditure (TEE) measured using the doubly labeled water (DLW) method, and (2) to compare intakes of fruits and berries, vegetables, juice, and sweetened beverages assessed by means of TECH with corresponding intakes obtained using the Web-based KidMeal questionnaire (KidMeal-Q), previously validated against a 7-day food record. The selected foods were considered as relevant markers for good (fruits and berries, vegetables, and juice) and bad dietary habits (sweetened beverages).

Methods

Recruitment and Protocol

Healthy, 3 year old Swedish children (N=30) were recruited in 2010-2011 [10-12]. Their mean TEE was measured for 14 days using the DLW method [11,12]. During this period, the children's intakes of foods and drinks were assessed using TECH, and parents completed the KidMeal-Q. A complete data collection was obtained from 30 children. The mean change in body weight between day 1 and 14 was -0.008 kg (SD 0.317). All children originated from a well-educated, middle-income population. The study was approved by the Research Ethics Committee, Linköping, Sweden. Written informed consent was obtained from all parents.

Tool for Energy Balance in Children (TECH)

Parents and other caretakers were instructed to take pre- and post-meal photographs of all the food items and beverages consumed by their child during one 24-hour period using a mobile phone provided for the study. At each meal, they also answered 6-7 questions regarding the type of milk, butter/margarine/oil, meat, bread, cereal, and sauce using a JAVA-based questionnaire installed on the mobile phone (see Figure 1). The parents were instructed to photograph the meals from three angles, to use tableware provided specifically for the study, and to place a matchbox in each image. Volumes of foods were assessed from images using known sizes of the tableware and matchbox identifications by means of the software Paint (Microsoft, version 6.1), and converted into weight by being multiplied by the appropriate weight per volume [13]. The energy intake was calculated from intakes of foods and drinks through linkage to the Swedish Food Database [14].

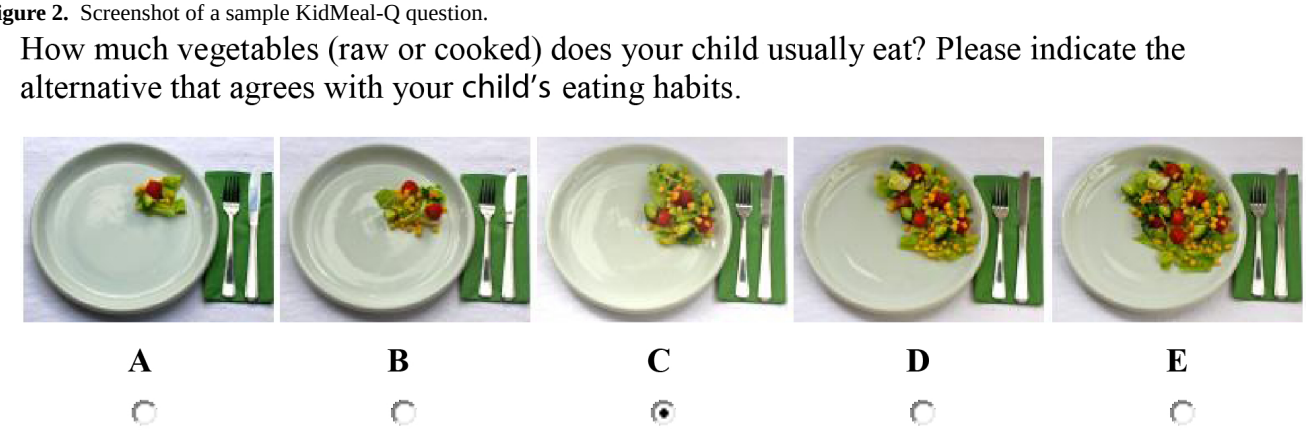
Figure 1. Screenshot of a TECH question and picture.

KidMeal-Q

KidMeal-Q was developed in 2008 for the LifeGene study [15]. It is an online, meal-based food frequency questionnaire designed for children aged 3-5 years for assessing dietary intake during previous months, and covers 42-86 food items, drinks, and dishes, depending on the number of follow-up questions (see Figure 2). For each child, we converted the reported frequency for vegetables into daily intakes by multiplying them by the reported portion sizes (using six pictures). KidMeal-Q does not provide any portion sizes for juice, fruits, berries, or sweetened beverages, and thus standard portion sizes [14] were used to convert reported frequencies into daily intakes.

KidMeal-Q was validated against a 7-day weighed food record in 23 healthy Swedish children with a mean age of 4.6 years (SD 1.5), weight of 18.4 kg (SD 3.7), and height of 1.09 m (SD 0.11) (data to be published). In that study, correlation coefficients between intakes of vegetables, fruits, and juice to sweetened beverages assessed using KidMeal-Q and food record estimates were .45 (P=.03), .59 (P=.003), and .53 (P<.001), respectively. These correlation coefficients are similar to those reported for adults when comparing food-frequency questionnaires with food records [16,17]. Therefore, although not an established reference method, we consider KidMeal-Q to be an appropriate reference in this first evaluation of TECH.

Figure 2. Screenshot of a sample KidMeal-Q question.



Statistical Analyses

Values are given as means (SD). Significant differences between mean values were identified using the Wilcoxon matched pairs test. Correlation analyses were performed using Pearson or Spearman rank order correlations. The Bland-Altman procedure [18] was used to assess the agreement between methods. Analyses were performed using Statistica Software, version 10 (STAT SOFT, Scandinavia AB, Uppsala, Sweden).

Results

The mean age of the children in the study was 3.00 years (SD 0.04), with a mean weight of 15.4 kg (SD 1.6), height of 0.96 m (SD 0.03), and body mass index (BMI) of 16.6 kg/m^2(SD 1.2). Five children were classified as overweight, and none as obese [19].

The mean energy intake, assessed using TECH (5400 kJ/24h, SD 1500) was not significantly different (P=.23) from TEE (5070 kJ/24h, SD 600). Figure 3 shows the Bland-Altman plot for energy intake compared to TEE. The limits of agreement were wide, and TECH overestimated high energy intakes and underestimated low energy intakes.

The mean daily intakes of fruits and berries, vegetables, juice, and sweetened beverages using TECH and KidMeal-Q are shown in Table 1. No significant differences between the two methods were observed. When comparing intakes of fruits and berries, vegetables, juice, and sweetened beverages using TECH with the corresponding KidMeal-Q estimates, the Bland-Altman plots were similar to the corresponding plot for energy intake (ie, they showed large limits of agreement and trends toward an overestimation of high intakes and an underestimation of low intakes, figures not shown). Furthermore, significant correlations between the two methods were observed for intakes of fruits and berries, vegetables, and juice (Table 2).

Table 1. Mean daily food intake estimated by means of TECH andKidMeal-Q (N=30).

	TECH[a]		KidMeal-Q[b]		
Food group	Intake (g/day[c]), mean (SD)	Range (g/day), min-max	Intake (g/day), mean (SD)	Range (g/day), min-max	P [d]
Fruits and berries	89 (110)	0-309	88 (31)	37-125	.636
Vegetables	40 (52)	0-255	40 (36)	0-120	.568
Juice	50 (79)	0-240	59 (64)	0-300	.243
Sweetened beverages	95 (120)	0-320	36 (30)	0-96	.061

[a]Tool for Energy Balance in Children

[b]KidMeal Questionnaire

[c]grams/day

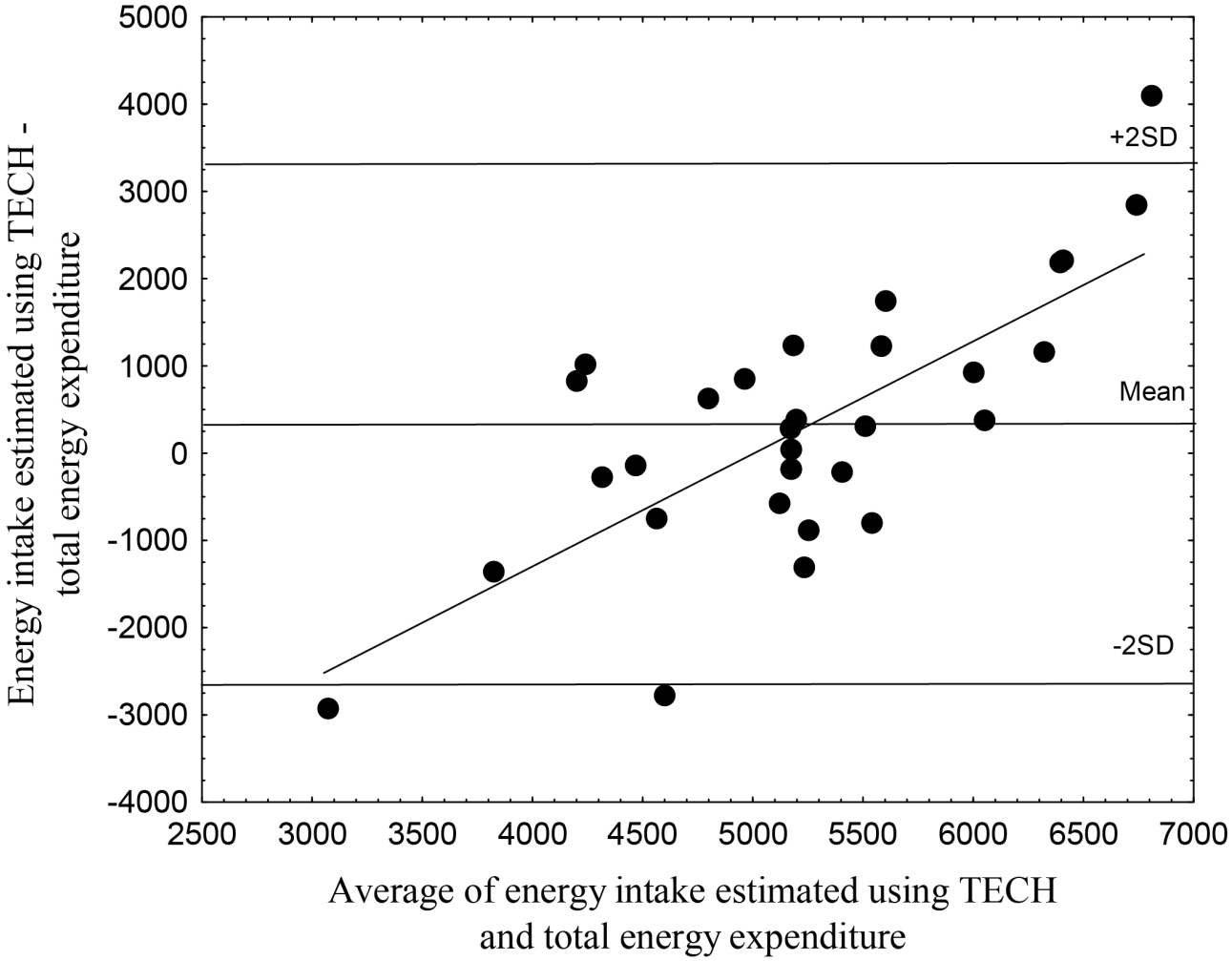
[d]P value for difference between daily intake estimated by means of TECH and KidMeal-Q using the Wilcoxon matched pairs test

Table 2. Correlation coefficients between food intakes (grams/day) estimated using TECH and KidMeal-Q (N=30).

Food group	ρ [a]	*P*
Fruits and berries	.46	.010
Vegetables	.43	.020
Juice	.42	.020
Sweetened beverages	.15	.424

[a]Spearman rank order correlation

Figure 3. A Bland-Altman plot comparing energy intake estimated using TECH and total energy expenditure measured using the doubly labeled water method in 30 healthy 3 year old children. The mean difference between the methods was 330 kJ/24h with limits of agreement (2SD) of 2990 kJ/24h [16]. The regression equation was y=1.27(x−6312.67) (*r*=.73, *P*<.001).



Discussion

Principal Findings

This is the first study that has evaluated energy intake assessed using mobile phones versus the DLW method in preschoolers. The average energy intake estimated using TECH was not significantly different from the average TEE (mean difference +7%). However, the limits of agreement in the Bland-Altman plot were wide, indicating low accuracy for TECH in estimating energy intake for individuals. Furthermore, we observed a bias where high energy intakes were overestimated while low energy intakes were underestimated. Although we only used TECH for one day, the mean difference and our limits of agreement for energy are comparable to previous validation studies for established dietary methods in children aged 3-6 years [20-25]. In our study, part of the inaccuracy for individuals may be the result of TECH being applied for only one day.

There is no reference method for intakes of foods and drinks. Hence, as commonly applied, we compared TECH with another dietary method to evaluate its relative accuracy for foods. Despite that TECH was only applied for one day, which may not represent habitual intakes, we obtained correlations for fruits and berries, vegetables, and juice with KidMeal-Q (which assesses habitual intakes in past months) that were of similar magnitude as those commonly reported when comparing two dietary methods [16,17]. Although the accuracy in individual

children was low, the average intakes of fruits and berries, vegetables, and juice assessed using TECH were comparable to corresponding figures with KidMeal-Q. However, our results for foods need confirmation due to the different assessment periods.

Study Strengths and Limitations

The major strength of our study is that we compared energy intake to TEE using the DLW method, which is the gold standard when validating reports of energy intake [3,26]. This is superior to using another dietary method as a reference since all such methods are well known to be associated with systematic errors [3,27]. The DLW method can be used as a reference since energy intake and TEE should be equal for subjects in energy balance during the measurement period. This criteria is valid both in normal-weight and overweight subjects, and was fulfilled since our children were weight stable through the 14-day-period, and since the energy content of retained tissue corresponds to only approximately 1% of energy intake at this age [28,29]. A limitation is that we compared one-day TECH data with mean TEE data from 14 days. However, this has unlikely influenced our results since the day-to-day variation in TEE is small [30,31].

The major limitation of this pilot study is that we applied TECH during only one day. The reason for this is that this first evaluation of TECH was conducted within an on-going study and we did not want to affect the parents' participation in the original study by adding more assessment days. Furthermore, our sample size was small and participating families represented a selected group, which may limit generalizability. However, since the accuracy of using TECH for one day was comparable to established dietary methods, future research should evaluate if the accuracy for TECH can be improved with more days. In addition, identification of the underlying reasons for the observed bias in TECH could be the topic for future studies.

Conclusion

In conclusion, one day of recordings using TECH is not able to accurately estimate intakes of energy or certain foods in 3 year old children.

Acknowledgments

The authors thank all parents and children who participated in the study, Katrin Bergström (nutritionist) for analyzing the food images, as well as Johan Cederlund and Jan Fjellström for help with the mobile phone questionnaire. This work was supported by grants from the Swedish Research Council (15402, 2008-4322), AFA Insurance, the Medical Faculty at Linköping University, the County Council of Östergötland, the Crown Princess Lovisa Foundation, the Swedish Research Council for Health, Working Life and Welfare, and the Swedish Nutrition Foundation. The funders had no role in the design, analysis, or writing of this article.

Authors' Contributions

ML, HH, EF, AB, SB, and KB designed the study. KB, AB, and SB designed KidMeal-Q. HH was responsible for the recruitment of children, data collection, and statistical analyses. SB performed data collection and analyses of KidMeal-Q. OB analyzed KidMeal-Q. CD contributed to the analysis of the food images. HH drafted the manuscript. ML, EF, KB, OB, CD, and AB and SB reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. World Health Organization (2014). World Health Organization website URL: http://www.who.int/dietphysicalactivity/childhood/en/ [accessed 2014-11-28] [WebCite Cache ID 6UQHN6twc]
2. Hesketh KD, Campbell KJ. Interventions to prevent obesity in 0-5 year olds: an updated systematic review of the literature. Obesity (Silver Spring) 2010 Feb;18 Suppl 1:S27-S35. [doi: 10.1038/oby.2009.429] [Medline: 20107458]
3. Burrows TL, Martin RJ, Collins CE. A systematic review of the validity of dietary assessment methods in children when compared with the method of doubly labeled water. J Am Diet Assoc 2010 Oct;110(10):1501-1510. [doi: 10.1016/j.jada.2010.07.008] [Medline: 20869489]
4. Zhu F, Bosch M, Woo I, Kim S, Boushey CJ, Ebert DS, et al. The Use of Mobile Devices in Aiding Dietary Assessment and Evaluation. IEEE J Sel Top Signal Process 2010 Aug;4(4):756-766 [FREE Full text] [doi: 10.1109/JSTSP.2010.2051471] [Medline: 20862266]
5. Martin CK, Nicklas T, Gunturk B, Correa JB, Allen HR, Champagne C. Measuring food intake with digital photography. J Hum Nutr Diet 2014 Jan;27 Suppl 1:72-81 [FREE Full text] [doi: 10.1111/jhn.12014] [Medline: 23848588]
6. Wang DH, Kogashiwa M, Kira S. Development of a new instrument for evaluating individuals' dietary intakes. J Am Diet Assoc 2006 Oct;106(10):1588-1593. [doi: 10.1016/j.jada.2006.07.004] [Medline: 17000191]
7. Nicklas TA, O'Neil CE, Stuff J, Goodell LS, Liu Y, Martin CK. Validity and feasibility of a digital diet estimation method for use with preschool children: a pilot study. J Nutr Educ Behav 2012;44(6):618-623 [FREE Full text] [doi: 10.1016/j.jneb.2011.12.001] [Medline: 22727939]

8. Small L, Sidora-Arcoleo K, Vaughan L, Creed-Capsel J, Chung K, Stevens C. Validity and Reliability of Photographic Diet Diaries for Assessing Dietary Intake Among Young Children. ICAN: Infant, Child, & Adolescent Nutrition 2009 Feb 01;1(1):27-36. [doi: 10.1177/1941406408330360]
9. Mitchell SA, Miles CL, Brennan L, Matthews J. Reliability of the School Food Checklist for in-school audits and photograph analysis of children's packed lunches. J Hum Nutr Diet 2010 Feb;23(1):48-53. [doi: 10.1111/j.1365-277X.2009.00996.x] [Medline: 19788706]
10. Eriksson B, Löf M, Forsum E. Body composition in full-term healthy infants measured with air displacement plethysmography at 1 and 12 weeks of age. Acta Paediatr 2010 Apr;99(4):563-568. [doi: 10.1111/j.1651-2227.2009.01665.x] [Medline: 20064135]
11. Eriksson B, Henriksson H, Löf M, Hannestad U, Forsum E. Body-composition development during early childhood and energy expenditure in response to physical activity in 1.5-y-old children. Am J Clin Nutr 2012 Sep;96(3):567-573 [FREE Full text] [doi: 10.3945/ajcn.111.022020] [Medline: 22836033]
12. Henriksson H, Forsum E, Löf M. Evaluation of Actiheart and a 7 d activity diary for estimating free-living total and activity energy expenditure using criterion methods in 1·5- and 3-year-old children. Br J Nutr 2014 May 28;111(10):1830-1840. [doi: 10.1017/S0007114513004406] [Medline: 24503109]
13. National Food Administration. Weights for foods and dishes. Uppsala, Sweden: National Food Administration; 1992.
14. National Food Administration (2014). Swedish Food Database URL: http://www.slv.se/en-gb/ [accessed 2014-11-28] [WebCite Cache ID 6UQHUQJjm]
15. Almqvist C, Adami HO, Franks PW, Groop L, Ingelsson E, Kere J, et al. LifeGene--a large prospective population-based study of global relevance. Eur J Epidemiol 2011 Jan;26(1):67-77. [doi: 10.1007/s10654-010-9521-x] [Medline: 21104112]
16. Khani BR, Ye W, Terry P, Wolk A. Reproducibility and validity of major dietary patterns among Swedish women assessed with a food-frequency questionnaire. J Nutr 2004 Jun;134(6):1541-1545 [FREE Full text] [Medline: 15173426]
17. Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, et al. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. Am J Clin Nutr 1999 Feb;69(2):243-249 [FREE Full text] [Medline: 9989687]
18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
19. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ 2000 May 6;320(7244):1240-1243 [FREE Full text] [Medline: 10797032]
20. Börnhorst C, Bel-Serrat S, Pigeot I, Huybrechts I, Ottavaere C, Sioen I, et al. Validity of 24-h recalls in (pre-)school aged children: comparison of proxy-reported energy intakes with measured energy expenditure. Clin Nutr 2014 Feb;33(1):79-84. [doi: 10.1016/j.clnu.2013.03.018] [Medline: 23622780]
21. Collins CE, Burrows TL, Truby H, Morgan PJ, Wright IR, Davies PSW, et al. Comparison of energy intake in toddlers assessed by food frequency questionnaire and total energy expenditure measured by the doubly labeled water method. J Acad Nutr Diet 2013 Mar;113(3):459-463. [doi: 10.1016/j.jand.2012.09.021] [Medline: 23317500]
22. Davies PS, Coward WA, Gregory J, White A, Mills A. Total energy expenditure and energy intake in the pre-school child: a comparison. Br J Nutr 1994 Jul;72(1):13-20. [Medline: 7918322]
23. Kaskoun MC, Johnson RK, Goran MI. Comparison of energy intake by semiquantitative food-frequency questionnaire with total energy expenditure by the doubly labeled water method in young children. Am J Clin Nutr 1994 Jul;60(1):43-47 [FREE Full text] [Medline: 8017336]
24. Livingstone MB, Prentice AM, Coward WA, Strain JJ, Black AE, Davies PS, et al. Validation of estimates of energy intake by weighed dietary record and diet history in children and adolescents. Am J Clin Nutr 1992 Jul;56(1):29-35 [FREE Full text] [Medline: 1609757]
25. Reilly JJ, Montgomery C, Jackson D, MacRitchie J, Armstrong J. Energy intake by multiple pass 24 h recall and total energy expenditure: a comparison in a representative sample of 3-4-year-olds. Br J Nutr 2001 Nov;86(5):601-605. [Medline: 11737958]
26. Schoeller DA, Schoeller DA. Validation of habitual energy intake. Public Health Nutr 2002 Dec;5(6A):883-888. [doi: 10.1079/PHN2002378] [Medline: 12633511]
27. Hill RJ, Davies PS. The validity of self-reported energy intake as determined using the doubly labelled water technique. Br J Nutr 2001 Apr;85(4):415-430. [Medline: 11348556]
28. Fomon SJ, Haschke F, Ziegler EE, Nelson SE. Body composition of reference children from birth to age 10 years. Am J Clin Nutr 1982 May;35(5 Suppl):1169-1175 [FREE Full text] [Medline: 7081099]
29. Roberts SB, Young VR. Energy costs of fat and protein deposition in the human infant. Am J Clin Nutr 1988 Oct;48(4):951-955 [FREE Full text] [Medline: 3421204]
30. Coward WA, Cole TJ. The Doubly Labeled Water Method for the Measurement of Energy Expenditure in Humans: Risks and Benefits. In: New Techniques in Nutritional Research. 9th edition. San Diego: Academic Press; 1991:139-176.
31. Speakman JR. Doubly labelled water. Theory and practice. 1st edition. London, UK: Chapman & Hall; 1997.

Abbreviations

TECH: Tool for Energy Balance in Children
DLW: Doubly labeled water
TEE: Total energy expenditure
KidMeal-Q: KidMeal questionnaire

Edited by G Eysenbach; submitted 04.07.14; peer-reviewed by K Westerterp, L Zepeda; comments to author 19.09.14; revised version received 08.11.14; accepted 25.02.15; published 24.04.15.

Please cite as:
Henriksson H, Bonn SE, Bergström A, Bälter K, Bälter O, Delisle C, Forsum E, Löf M
A New Mobile Phone-Based Tool for Assessing Energy and Certain Food Intakes in Young Children: A Validation Study
JMIR mHealth uHealth 2015;3(2):e38
URL: http://mhealth.jmir.org/2015/2/e38/
doi:10.2196/mhealth.3670
PMID:25910494

Original Paper

Feasibility and Acceptability of Smartphone-Based Ecological Momentary Assessment of Alcohol Use Among African American Men Who Have Sex With Men in Baltimore

Cui Yang[1*], PhD; Beth Linas[2*], PhD; Gregory Kirk[2], MD, PhD; Robert Bollinger[3], MD; Larry Chang[3], MD; Geetanjali Chander[3], MD; Daniel Siconolfi[1], MPH; Sharif Braxton[1], BA; Abby Rudolph[4], PhD; Carl Latkin[1], PhD

[1]Johns Hopkins School of Public Health, Department of Health, Behavior and Society, Baltimore, MD, United States

[2]Johns Hopkins School of Public Health, Department of Epidemiology, Baltimore, MD, United States

[3]Johns Hopkins School of Medicine, Baltimore, MD, United States

[4]Pacific Institute for Research and Evaluation, Beltsville, MD, United States

[*]these authors contributed equally

Corresponding Author:
Cui Yang, PhD
Johns Hopkins School of Public Health
Department of Health, Behavior and Society
2213 McElderry St. 2n FL
Baltimore, MD, 21205
United States
Phone: 1 4105025368
Fax: 1 4105025385
Email: cyang29@jhu.edu

Abstract

Background: Alcohol use is a risk factor for the acquisition of human immunodeficiency virus (HIV) among African American men who have sex with men (MSM). Mobile phone-based ecological momentary assessments (EMA) could minimize bias due to retrospective recall and thus provide a better understanding of the social and structural context of alcohol use and its relationship with HIV-related risk behaviors in this population as well as other highly stigmatized populations.

Objective: We describe the study design and the implementation, feasibility, reactivity, and acceptability of an EMA study of alcohol use and HIV-related behaviors among African American MSM in Baltimore.

Methods: Participants were recruited through flyers and word-of-mouth in Baltimore from September 2013 to November 2014. Each participant was loaned an Android smartphone and instructed to respond to multiple prompts from the mobile app for 4 weeks. Data were collected through (1) random prompts delivered three times daily assessing participants' location, activity, mood, and social context, (2) daily prompts capturing drinking and sex events occurring in the past 24 hours, and (3) event-contingent responses collecting participants' self-reported episodes of drinking.

Results: A total of 16 participants enrolled in the study. The current analyses focused on 15 participants who completed at least 24 days of follow-up (mean follow-up time 29 days; range 24-35 days). Study participants (N=15) were a median 38 years of age (range 27-62 years) with low levels of income and educational attainment. Ten individuals self-reported living with HIV/AIDS, over half reported drinking alcohol at least 2-3 times a week, and a third reported binge drinking (ie, 6 or more drinks on one occasion) on a weekly basis. Based on the Alcohol Use Disorders Identification Test (AUDIT) score, nearly half were classified as hazardous drinkers (score 8-15) and a fifth were likely dependent (score ≥16). A total of 140 participant-initiated events were reported, and 75% of 1308 random prompts and 81% of 436 daily prompts delivered were answered. Of seven devices used during the study, five were reported lost by participants. We did not observe strong reactivity effects, and self-reported acceptability to study procedures was uniformly favorable.

Conclusions: This study provides evidence to support the feasibility and acceptability of using EMA methods for collecting data on alcohol use among African American men who have sex with men living in urban settings. These data provide the basis for future studies of EMA-informed mHealth interventions to promote the reduction of substance use and HIV risk-taking behaviors among African American MSM living in urban settings.

(JMIR mHealth uHealth 2015;3(2):e67) doi:10.2196/mhealth.4344

KEYWORDS

ecological momentary assessment (EMA); alcohol use; HIV; African American; men who have sex with men (MSM)

Introduction

Epidemiological data suggest the highest rates of human immunodeficiency virus (HIV) infection in the United States are among African American men who have sex with men (MSM) [1]. The 2012 National HIV Behavioral Surveillance survey demonstrated that among HIV-infected MSM in Baltimore, 48% were African American and less than 20% were white. Alcohol use and its impact on HIV transmission and treatment are major public health burdens in many parts of the world. Reviews indicate that alcohol consumption is associated with HIV incidence [2]; furthermore, alcohol is a potential cause of poorer HIV outcomes [2] and is associated with lower adherence to HIV medications [3]. African American MSM report significantly more drinks per drinking day compared to MSM from other race/ethnicity groups [4]. Previous research in Baltimore found that, based on Alcohol Use Disorders Identification Test (AUDIT) scores, 22% of African American MSM were in the hazardous category (ie, AUDIT score 8-15) and 21% in the high risk/likely dependent category (AUDIT score ≥16) [5].

Most research on substance use (alcohol and illicit drugs) and HIV has relied on self-reported information over a period of time, which varies by study. A major limitation of this method of measurement is that it is affected by recall bias (reliability and accuracy), and it may not be context specific [6]. Information recall is affected by heuristics used in memory search and reconstruction, which can systematically bias participant responses [7,8]. Imprecise or inaccurate information can impede the advancement of knowledge regarding alcohol use and risky sex behaviors among highly stigmatized populations [9]. There is a need to improve the methodologies for behavioral data collection and to obtain a more detailed understanding of the relationship between alcohol use and HIV risk behaviors among key populations, especially African American MSM.

Mobile health (mHealth) opens new avenues for research on substance use and HIV, as ubiquitous technology allows for more frequent and close to real-time collection of behaviors, locations, and physiologies [10]. Ecological Momentary Assessment (EMA), an mHealth method that utilizes mobile technologies, such as a personal digital assistant (PDA) or mobile phone, allows participants to record their daily activities on the device in real time [11]. EMA minimizes biases, specifically recall bias, by requiring participants to immediately respond to random prompts or record specific events on a daily basis in their natural environments [12]. EMA is an especially suitable tool to study health behaviors, such as alcohol use, which are discrete and episodic behaviors, and is ideal for event-contingent recording [13]. Prior research suggests that MSM may use alcohol to cope with internal experiences (eg, stress associated with internalized homophobia) [14] and situational stimuli and cues (eg, social pressure to use), making EMA an excellent method to capture these fleeting states. Empirical data from previous research of EMA have shown a high compliance rate among diverse populations, including homeless persons with crack-cocaine addiction [15], heroin and cocaine users in treatment [16], ecstasy users who also engaged in use of alcohol, marijuana, cocaine and hallucinogens [17], and social drinkers [18]. Evidence has also shown that intoxicated participants could enter data accurately on a mobile device [13]. Finally, in studies assessing reactivity of substance-use recording in EMA, the possibility that the repeated assessments may affect the behavior under study and thus distort the findings, have not indicated strong reactivity effects [19,20].

While EMA methods have been used successfully, there have been few studies that employ these methods with African American MSM in everyday, community-dwelling, non-treatment settings. In response to this limitation, we developed EMA methods for near real-time characterization of alcohol use in individuals' natural environments. In this paper, we characterize implementation barriers and examine the feasibility, acceptability, and reactivity of using intensive EMA methods among African American MSM living in urban settings.

Methods

Study Participants

Study participants were recruited through flyers and word-of-mouth in Baltimore, Maryland. Flyers were placed at the front desk of the research facility, where several other substance use and HIV projects were conducted. Inclusion criteria for this study were (1) at least 18 years of age, (2) self-reported African American race/ethnicity, (3) self-reported male sex, (4) self-reported having had sex with a male in the prior 30 days, (5) self-reported drinking alcohol at least once a week in the prior 30 days, (6) reported living within the Baltimore metropolitan area, and (7) able to understand and follow directions on how to use the mobile phone, as assessed by study staff in a one-on-one orientation session.

Study Procedures

Recruitment/Enrollment

Between September 2013 and November 2014, we screened 25 individuals via phone or in-person by a trained study coordinator or principal investigator. Of those screened, 17 were eligible to participate in our study and 16 participants provided informed verbal consent to participate in the study. Participants were enrolled in 6 waves of data collection (2-3 participants/wave). Participants first completed a baseline interview at the research facility. They were then loaned a mobile phone (Samsung Galaxy S4), and the mobile phone was reused in each wave. Once a phone was lost, a replacement phone was acquired. To protect participants' privacy, mobile phones were set back to factory setting and therefore all personal information was deleted. Before leaving the study office and beginning mobile

data collection in the field, study participants were trained on how to use the device and app by the study coordinator and all participants were required to show their understanding of the mobile app by running a "demo" questionnaire.

Audio-Computer Assisted Standardized Interview Surveys

At baseline, all participants completed an audio computer-assisted self-interview (ACASI), which assessed sociodemographics (eg, age, education, employment, income, and homelessness), drug and sex behaviors (eg, self-reported alcohol, tobacco, and illicit drug use and number of sex partners), clinical diagnoses (eg, self-reported HIV status), and prior experience with mobile technology. Additional data collected at baseline included depressive symptoms assessed using the Center for Epidemiologic Studies Depression Scale (CES-D), and past week and hazardous drinking and likely alcohol dependence assessed via the AUDIT score (respective cut-offs: ≥8 and ≥16). Participants returned to the research center after 1 and 4 weeks of follow-up to complete additional ACASI surveys that collected information about their behaviors during the prior week (1-week assessment) or during the prior 30 days (4-week assessment). One of the goals of the ACASI surveys was to compare aggregated responses with real-time responses from the EMA data collection.

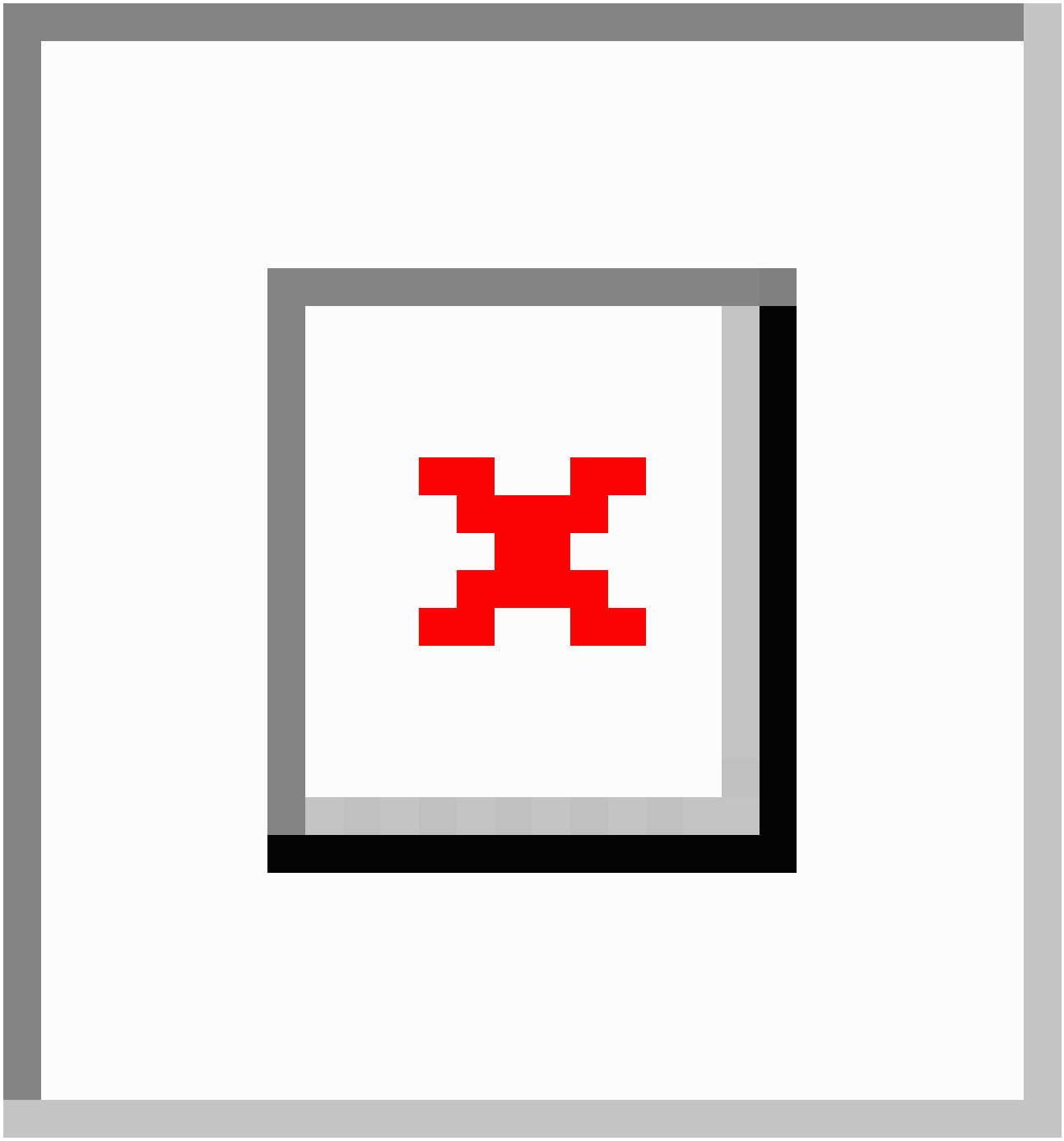
Mobile App

The mobile app used in this study, emocha, was created by the Center for Clinical Global Health Education at the Johns Hopkins School of Medicine. EMA surveys used in this study were adapted from prior research conducted by collaborators with drug using populations in Baltimore [21]. The following three types of EMA prompts were used in this study:

1. Random prompts: Three times daily, emocha sent an alert to the participant's phone between 10 a.m. and 10 p.m. These random prompts asked about participants' immediate mood, surroundings, and potential environmental cues that may trigger alcohol use. Participants answered no more than 46 questions in each of the random prompt surveys (including skip patterns).
2. Daily prompts: One daily alert at 9 a.m. was sent to participant's phone. This survey consisted of questions that summarized activities during the previous 24 hours, including alcohol, other substance use, and sexual activities. The daily survey contained up to 98 questions, depending the number of alcohol drinks and number of sex partners reported during the past 24 hours.
3. Event-contingent entries: Participants were instructed to initiate an electronic entry every time they finished one episode of drinking, which was defined as a cluster of alcohol use in one sitting. Event-contingent surveys asked up to 40 questions concerning the location, alcohol use expectancy, drinking partner, types, and amount of alcohol, co-use of other substances, and current mood and stress level (see Figure 1).

Participants had 30 minutes after being prompted to complete surveys for both the random and daily prompts. After 30 minutes, the prompt was considered missed. Each day at 10 p.m., mocha uploaded the encrypted EMA data to a secure server and removed the data from the device.

Figure 1. Screenshot of emocha event-contingent survey.



Qualitative Assessment

During the week 1 and week 4 visits, participants also completed 10-minute, semistructured interviews to provide feedback on their experience using the mobile phone and the emocha app. The qualitative interview guides had predefined main themes, but these guides were meant to be dynamic and allowed for new topics to emerge during the course of the interview. Important topics included (1) satisfaction and challenges with the study, (2) suggestions for future studies, and (3) ways the mobile phone may facilitate reducing alcohol use and promoting HIV risk reduction.

Participants received remuneration for attendance at study visits (Table 1), for providing adequate responses to weekly random and daily prompts, and for returning devices upon study completion. Participants were informed at enrollment that loss of two study devices would result in their dismissal from the study.

Table 1. Participant reimbursement/visit schedule (in USD).

	Intake	Week 1	Week 2	Week 3	Week 4	Total
Baseline visit	$10					
ACASI	$20	$20			$20	
EMA[a]		$25/$50	$25/$50	$25/$50	$25/$50	
Close out					$10	
Smartphone return[b]					$50/$100	
Total	$30	$45-$70	$25-$50	$25-$50	$105-$180	$230-$380

[a]Participants were paid US $50 every week for answering 80% of their alarms or US $25 every week for answering 60% of their alarms. They received no bonus for answering less than 60% of their alarms or if their phone was uncharged.

[b]Participants received US $100 at close out for returning their original phone or US $50 for returning their replacement. They would be excused from the proposed study if they lost their replacement phone.

The Institutional Review Board at Johns Hopkins University Bloomberg School of Public Health approved the study protocol and a Certificate of Confidentiality was obtained through the National Institute of Allergy and Infectious Diseases.

Data Analysis

Descriptive statistics were used to examine characteristics of participants and study compliance (eg, days of follow-up, random and daily prompt response rates, EMA survey completion time, and device loss rate). Feasibility was assessed through participant retention, days of follow-up, device loss rate, response rates to EMA surveys, and amount of time needed to complete each EMA survey. We also assessed the number of questions completed in each survey.

Reactivity analysis was conducted by examining any correlation between day of study and total number of drinks reported in daily surveys. Repeated measures were accounted for in linear regression models using Generalized Estimating Equations [22]. We summed the number of drinks each individual reported consuming per day in the daily prompts and plotted this against the day of study. We used a non-parametric lowess curve, which is able to show a relationship between variables and any trends that may exist in the data. All quantitative analyses were performed using Stata version 13.0.

For the qualitative evaluation, we identified core consistencies and meanings in the data through careful repeated reading of interview texts. We labeled sections of text based on themes and particular domains of interest related to feasibility and acceptance. Results were summarized by main themes and reviewed by investigators continually throughout the study, with the goal of identifying strategies to refine the emocha design in preparation for a subsequent data collection.

Results

Baseline Characteristics of Participants

Of the 16 participants enrolled, one participant was lost to follow-up 1 day after the baseline visit. The current analyses focused on 15 participants who completed at least 24 days of follow-up (median 29, interquartile range 27-31). Of these 15 participants, 5 saw the study flyers and 10 heard about our study from other people. The baseline characteristics of 15 participants were summarized in the Table 2.

Table 2. Baseline characteristics of participants (N=15).

Characteristics	n (%)
Age, median (IQR)	32 (29-45)
At least grade 12 or GED education	13 (87)
Full/part time job	2 (13)
<US $10,000 income (last year)	8 (53)
Homeless (past 6 months)	4 (27)
Arrested (past 6 months)	2 (13)
HIV positive (self-report)	10 (67)
CES-D score, median (IQR)	32 (21-44)
Depressive symptoms (CES-D>20)	12 (80)
Frequency of cigarette use (past 30 days)	
Never	3 (20)
Once a week	1 (7)
A few times a week	2 (13)
Every day	9 (60)
Have smoked crack/cocaine/heroin/inject drugs to get high (past 3 months)	4 (27)
Often or always smoke marijuana while drinking alcohol	7 (47)
Frequency of alcohol use	
Monthly or less	1 (6)
2-4 times a month	6 (40)
2-3 times a week	4 (27)
4 or more times a week	4 (27)
Frequency of binge drinking at least weekly	5 (33)
AUDIT score, median (IQR)	9 (6-14)
Hazardous drinker (AUDIT score: 8-15)	7 (47)
Probable alcohol dependence (AUDIT score≥16)	2 (20)
Number of sex partners (past 30 days), median (IQR)	2 (1-5)
Have owned a mobile phone (past 6 months)	14 (93)
Currently using a smartphone	11 (79)

Feasibility Assessment

Overall, 15 participants provided 436 days of observation (mean 29 days; Table 3). Seven participants enrolled in the study over 4 weeks (ie, 28 days) due to scheduling. Of the 15 participants, 2 completed 4 weeks of EMA entries but failed to return to the clinic for the 4-week visit and to return the phone. Two more participants were unable to complete week 4 of EMA surveys due to phone loss (n=1) and incarceration (n=1); however, both men completed the last clinic visit. A total of seven phones were issued to participants, and phones were re-used for multiple waves of data collection. At the end of the study, five phones were either reported lost by participants (n=2) or were unable to be retrieved by study staff due to loss to follow up with participants (n=3). One participant reported losing his phone on the last day of follow-up. The last phone was lost during the fourth week of EMA data collection, but the participant did not report the phone as lost until the last day of the study.

Table 3. Days of follow-up and device loss (participants were enrolled in 6 waves of data collection [2-3 participants/wave]. Phone was reused in each wave. Once a phone was lost, a replacement phone was acquired).

	Baseline	Week 1	Week 2	Week 3	Week 4		Days	Device
Participant 1	X	X	X	X	X		29	lost
Participant 2	X	X	X	X	X	X	30	returned
Participant 3	X	X	X	X	X	X	29	returned
Participant 4	X	X	X	X	X	X	27	returned
Participant 5	X	X	X	X	X	X	31	returned
Participant 6	X	X	X	X	X		28	lost
Participant 7	X	X	X	X	X	X	27	returned
Participant 8	X	X	X	X	X	X	31	returned
Participant 9[a]	X						1	lost
Participant 10	X	X	X	X	X	X	27	returned
Participant 11	X	X	X	X	X	X	35	returned
Participant 12	X	X	X	X	X	X	32	returned
Participant 13	X	X	X	X		X	24	lost
Participant 14	X	X	X	X		X	24	returned
Participant 15	X	X	X	X	X	X	35	returned
Participant 16	X	X	X	X	X	X	27	lost

[a]Excluded from the current analyses.

Response rates, time to complete EMA survey and number of questions completed over the 4-week study period are summarized in Table 4. A total of 436 daily prompts, which were initiated at 9 a.m. every day, were sent to participant's mobile phones. In all, 352 daily prompts were completed resulting in an overall compliance rate of 80.7%. The compliance rate of daily survey completion ranged from 62.5% to 100% among all participants. Table 4 describes the peak compliance rate in week 3 (92.4%) followed by a drop in week 4 (63.6%).

A total of 1308 random prompts were sent to participants' mobile phones over follow-up and 968 were completed. This represents an overall compliance rate of 74%, translating to an average of 2.22 random-prompt responses per day per person. Table 4 shows the compliance rate per week as steady in the first 3 weeks, followed by a drop-off in week 4. Among all participants, the compliance rate of random prompts ranged from 48.1% to 98.9%.

The 15 participants of the current study reported a total of 140 drinking events over follow-up through emocha. The average number of self-reported drinking events per person was 9, ranging from of 2 to 32 reports over 4 weeks of follow-up. Of note, 40% of drinking events were reported in week 1 and the number reported per week decreased at each week over the course of the study.

We assessed completion time as the number of minutes elapsed from initiation of each survey to synchronization with the server. The average time to finish the daily survey was 1.43 minutes, the average time to complete the random survey was 1.15 minutes, and the average time needed to complete the event survey was 1.52 minutes. In both daily and random surveys, we observed a trend of a learning curve so that initially participants took longer to complete the survey. In week 4, it took the participants an average of 1 minute to finish each survey. Taken together, the amount of time it took to complete one daily survey and three random surveys averaged 4.88 minutes per day plus any additional time to fill out event surveys if drinking occurred. We also assessed the number of questions participants completed in different the types of surveys per week over 4 weeks. There were no significant changes over time in terms of number of completed questions in both random and event surveys. Although the change was statistically significant (P=.006) in the daily survey, the magnitude of change was from 19.33 in week 1, 15.10 in week 2, 15.21 in week 3, and 15.91 in week 4.

Table 4. Response rates, time to complete EMA survey, and number of questions completed.

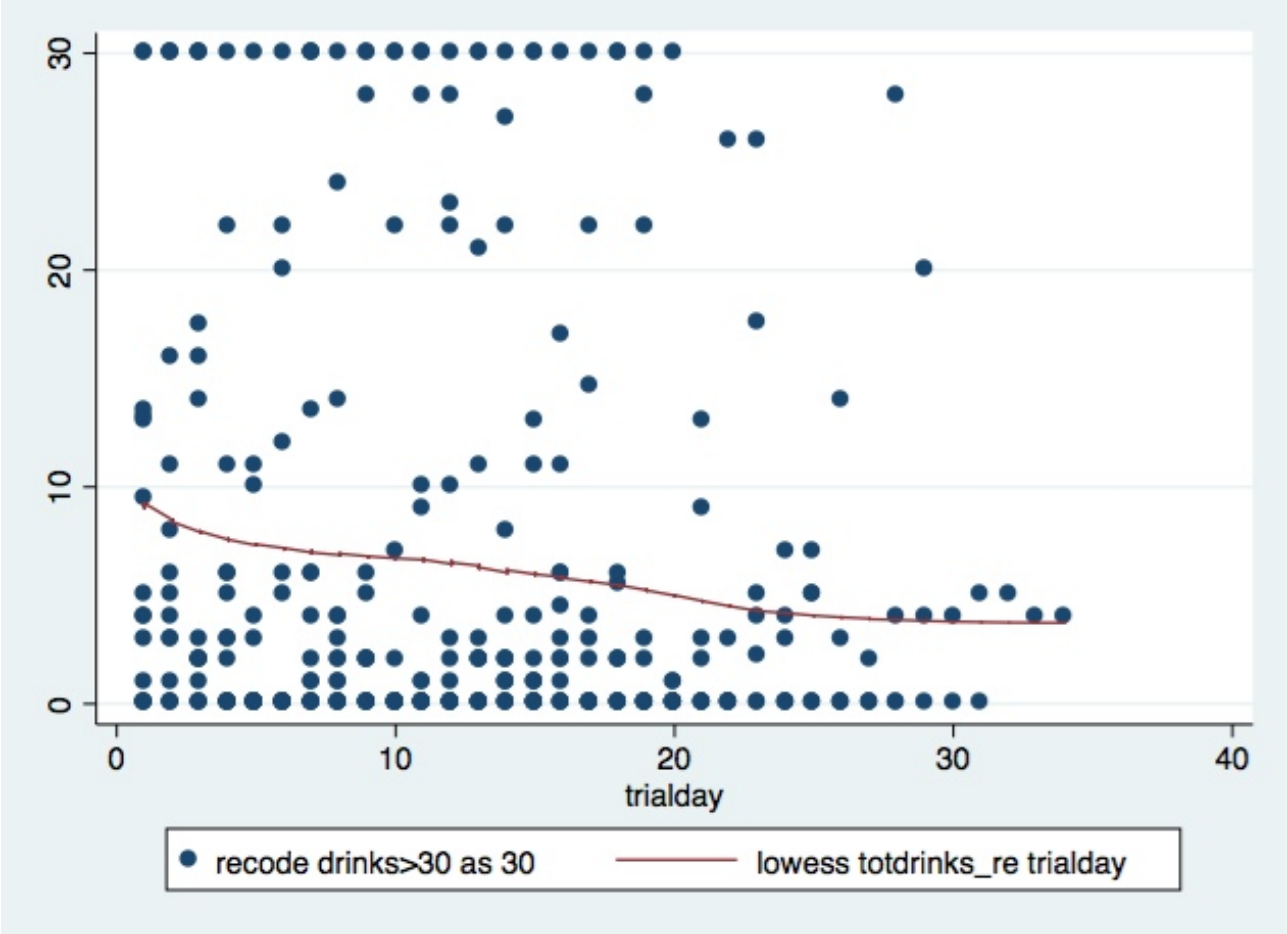
	Average response rate			Time to finish EMA survey in minutes, median (IQR)			Number of questions completed in each survey, mean (SD)		
	Daily survey	Random survey	Event, n	Daily survey	Random survey[a]	Event survey	Daily survey	Random survey	Event survey
Overall	80.7%	74%	140	1.43 (0.91-2.53)	1.15 (0.83-1.60)	1.52 (1.15-2.10)	16.36 (9.24)	18.06 (3.15)	18.92 (1.04)
Week 1	85.7%	81.6%	58	2.64 (1.57-3.68)	1.48 (1.11-1.99)	1.97 (1.55-2.78)	19.33 (12.01)	17.96 (3.14)	19.14 (1.32)
Week 1	83.8%	84.8%	28	1.28 (0.89-2.19)	1.17 (0.83-1.71)	1.18 (1.06-1.58)	15.10 (8.98)	18.24 (3.37)	18.75 (0.79)
Week 3	92.4%	80.9%	32	1.19 (0.84-1.98)	1.06 (0.79-1.35)	1.29 (1.11-1.90)	15.21 (7.03)	17.75 (2.82)	18.75 (0.62)
Week 4	63.6%	52.1%	22	1.05 (0.81-1.84)	0.95 (0.70-1.24)	1.16 (0.89-1.55)	15.91 (7.73)	18.33 (3.24)	18.82 (0.91)
P value				<.001	.28	<.001	.006	.18	.23

[a]Data only available for ID8-ID16.

Reactivity Assessment

The median number of drinks per day per person was 2 (IQR 0-6). Cases with number of drinks per day over 30, which represented 9% of all cases, were recoded as 30 for further analysis. The correlation between number of drinks per day and days of study was -.015 (*P*=.01). We plotted the number of drinks per day against the day of study, and each dot represents each individual's self-reported number of drinks. As shown in Figure 2, there was a trend towards decreasing alcohol use over the course of the study, and it flattened out after 25 days.

Figure 2. Trend in reporting number of drinks in daily survey.



Acceptability Assessment

As seen in Table 5, most participants reported that the phones were “easy or very easy” to use and that the reporting burden was “just right or not enough”. Comprehension of survey questions was also high (85%-94% reported “most” or “all made sense”). The majority (87%-93%) reported being mostly or extremely confident that their privacy would be protected. Finally, 20%-31% of participants indicated that answering questions about drinking made them want to drink less.

Table 5. Acceptability survey.

	Week 1 (n=15), n (%)	Week 4 (n=13), n (%)
In general, how easy is it to use the smartphone?		
Very easy	13 (87)	13 (100)
Easy	2 (13)	0
Difficult	0	0
Very difficult	0	0
What do you think about the number of times that your alarm goes off every day?		
Not enough	1 (6)	1 (8)
Just right	14 (94)	12 (92)
A little too much	0	0
Too much	0	0
Do the questions on the phone make sense to you?		
None of the questions make sense to me	0	0
Some of the questions do not make sense to me	1 (6)	2 (15)
Yes, most of them make sense to me	4 (27)	5 (39)
Yes, all of them make sense to me	10 (67)	6 (46)
Do you feel comfortable carrying the smartphone?		
Extremely comfortable	14 (93)	10 (77)
Mostly comfortable	0	2 (15)
Somewhat comfortable	1 (7)	1 (8)
Not too comfortable	0	0
Not comfortable at all	0	0
Do you feel confident that the information collected will only be seen by researchers and not used against you?		
Extremely confident	12 (80)	11 (86)
Mostly confident	1 (7)	1 (7)
Somewhat confident	2 (13)	1 (7)
Not too confident	0	0
Not confident at all	0	0
Do you feel that the size of the device is:		
Too small	1 (7)	2 (15)
A good size	14 (93)	11 (85)
Too big	0	0
Does answering questions about drinking make you want to drink more, less, or about the same?		
More	1 (7)	1 (7)
Less	3 (20)	4 (31)
The same	11 (73)	8 (62)

Qualitative Assessment

In qualitative interviews, participants provided positive feedback concerning the study methods. Familiarity with the technology seemed to have helped participants navigate the study as participants stated "I have the same phone, so I like it" (Participant 8) and "easy to use, similar to my own phone" (Participant 13). Participants enjoyed the technology as part of the research, as one participant stated "fun to answer the questions on phone, much easier" (Participant 5). Participants felt the mobile technology may have even increased their engagement in the study, as participants stated "fun study which makes people more engaged and willing to participate" (Participant 13) and "good way to answer questions without coming to a clinic, it is cool" (Participant 14). One participant mentioned "The questions also make you more cognizant of your surroundings and habits (anxiety level, seeing things happen, amount of drinking)" (Participant 13).

Participants also expressed concerns and provided suggestions for the study procedures and future studies. For participants reporting that they already owned a smartphone, the study phone may have been a burden. As one participant stated, "two phones are too much (to carry)" (Participant 8). Another participant suggested that in future studies, study staff should "install apps (emocha) on my phone" (Participant 15). Participants also reported, "the battery life dies fast" (Participant 14), "volume is too low to hear the alarms and snooze time should be longer" (Participant 7), "[the apps] need to be more personalized" (Participant 13).

Participants expressed their interest in keeping the study phone as they stated "it [study phone] can't be kept" (Participant 12) and "had to give it [study phone] back" (Participant 15). Participants also suggested extending the hours of data collection to better capture times when people are more likely to be out drinking. For example, one participant suggested that we "increase the time on weekends to around 1 or 2 [am]. That's when the clubs close and if you do that, you would get some good data" (Participant 13), and another suggested that it "May be a good idea to have the alarm period extended to 11-12, people may be getting ready for parties" (Participant 14). In addition to data collection, participants felt the "smartphone can be a good tool to deliver health messages" (Participant 5).

Discussion

Principal Findings

To our knowledge, this is the first study to evaluate the use of a mobile app–based EMA to prospectively capture alcohol use among African American men who have sex with men living in urban settings. Despite challenges, this study provides evidence to support the feasibility and acceptability of using EMA methods for collecting data on alcohol use in this population.

Given the highly demanding protocol in EMA, we were particularly sensitive to participation burden. Our study protocol included 3 random prompts and 1 daily prompt per day, which represents a lower or moderate participant burden as compared to previous EMA studies in substance using populations [16,21]. In both daily and random surveys, we observed a trend of a learning curve, in that participants initially took longer to complete surveys. Overall, our data demonstrate that participants spent on average less than 5 minutes per day to compete mandatory EMA surveys. Event-driven surveys required an additional 1-2 minutes per drinking episode reported. The quality of EMA data depends heavily on participants' compliance to prompts and timeliness of recording episodes of the desired event (ie, alcohol use). Noncompliance leads not only to missing data but can even introduce bias in the data collected [12]. In the current study, participants answered 74% of random prompts and 80.7% of daily prompts, which is comparable to response rates reported in previous EMA studies (50%-90%) [16,23,24]. This finding is consistent with what participants reported in the acceptability survey, as they were very clear in reporting that study procedures were not overly burdensome. Overall, findings from the current study represent a moderate burden to participants enrolled in the study.

Assessing compliance through the recorded events (eg, alcohol use) is much more challenging. There is often no way to independently assess or verify whether participants failed to report the events that have actually occurred. The idea of providing incentives for reporting substance use events is debatable and needs further evaluation. In the current study, 40% of drinking events were reported in week 1. Although the reactivity analysis found a significant decrease in reporting of number of drinks per day, the magnitude of decrease was minor (-.015). Taken together, underreporting of alcohol events through event-contingent surveys is expected in the current study. More research should explore good participant management procedures that can yield high compliance [13], such as a regular reminder to participants to report their drinking events when they have an onsite visit or through text messages to their phones. Future studies could consider using biochemical markers of alcohol, such as transdermal alcohol sensors as a way to objectively validate self-reported alcohol use (and compliance).

One of the challenges associated with EMA is exhaustion of participants within the study period due to the highly demanding research protocols, which can diminish the level of participation. In our analyses of weekly response rates, we did find some evidence of exhaustion, as the response rates to both daily and random prompts dropped significantly in week 4. These results may signify that further examination of the assessment windows, such as shorter follow-up are necessary. Additionally, we had significant loss to follow-up, which mostly occurred in week 4. In the future, studies should provide better monitoring to seek a better understanding of exhaustion with similar populations. We were able to find out that one participant was not be able to complete EMA in week 4 due to a brief incarceration. African American MSM living in urban settings may experience unique social and structural challenges, such unemployment, low-income status, incarceration, and community violence. These sociostructural factors may operate independently or together in a dynamic fashion to create a context in which they experience challenges or inabilities to engage in prevention and treatment programs [25]. Impacts of future mHealth research

and programs will come from a better understanding of broader social contexts where mHealth is implemented.

Another challenge is related to an issue that concerns all EMA studies of substance use, namely the hours of coverage for EMA assessments [13]. In the current study, EMA assessments occurred between 10 a.m. to 10 p.m., in order to avoid alarming participants when they are asleep. However, as some participants suggested, these times of day may not be representative of hours in which drinking occurs, particularly as alcohol consumption tends to occur later in the night, and mood, activity, and social settings vary by time (eg, weekdays vs weekends). Thus, it is important for future studies to assess the full range of an individual participant's waking hours [26] and to possibly provide personalized hours of coverage of EMA for each participant. In Epstein et al's research with cocaine- and heroin-abusing outpatients who were being treated with methadone, typical waking hours were programmed on a weekly basis when participants were issued the device [16]. With technology development, personalized EMA could be executed remotely.

Device loss posed a major challenge in our study, as five of seven devices issued to participants were lost. Participants were informed at enrollment that they would be dismissed from the study after losing two devices. There is a concern that when using mobile devices among impoverished populations, participants might sell the devices [21]. Therefore, we provided an incentive (US $100) for returning the devices upon completion of the study. However, we later realized that the street value for the Samsung Galaxy 4 can be more than the incentive. Several participants expressed their disappointment at returning the study phones in qualitative interviews. Future studies may consider the option of installing the mobile app on participants' own phones or letting participants keep the study phone instead of monetary incentive or using a less well-known brand of mobile phone. Additionally, future research could utilize the remote inactivation feature if mobile phones are stolen or misplaced.

Limitations

This is a preliminary feasibility study, and so several limitations need to be addressed in future research. Given the small sample size, the current study does not have enough statistical power to detect the significant differences in sociodemographic, behavior, or clinical characteristics between participants with higher response rates and those with lower response rates. Research involving larger samples is needed to explore various factors associated with variation in compliance rates that can be used for targeted EMA training to enhance compliance. Despite the less restrictive inclusion criteria of our study, we were able to enroll participants with varied alcohol use, including 7 hazardous drinkers and 3 likely alcohol dependent. Our limited sample size, however, may not extend to problematic alcohol users with more intense alcohol use patterns. Our study confirmed previous findings that did not demonstrate strong reactivity from EMA assessment of substance use [20]; however, future research should generate more rigorous evidence.

Conclusions

In conclusion, findings from our study demonstrate that EMA methods are feasible and acceptable approaches for data collection among African American men who have sex with men. Eliminating health disparities and reversing HIV epidemic trends will require innovative combination prevention approaches to reduce high-risk behaviors, including substance use, expanded HIV testing, and increased linkage to and retention in care. The high ownership of mobile phones among minority MSM may provide a promising platform for data collection and the delivery of substance use and HIV risk reduction messages to this hard-to-reach population [27]. Using EMA data can identify individualized sets of triggers to be used to further tailor "ecological momentary intervention (EMI)" content and delivery [28]. These methods could reinforce the systematic use of prevention or treatment components in real-world settings.

Acknowledgments

This research has been supported by National Institute of Alcohol Abuse and Alcoholism (K99/R00AA020782), and the Johns Hopkins Center for AIDS Research (1P30AI094189).

Conflicts of Interest

LC and RB are consultants to, minority equity holders in, and entitled to royalties from emocha Mobile Health, Inc. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflict of interest policies.

References

1. Centers for Disease Control and Prevention (CDC). HIV/AIDS Surveillance Report, 2007; Vol 19. 2009. URL: http://www.cdc.gov/hiv/topics/surveillance/resources/reports/ [accessed 2015-06-10] [WebCite Cache ID 6ZC2dWZU4]
2. Shuper PA, Neuman M, Kanteres F, Baliunas D, Joharchi N, Rehm J. Causal considerations on alcohol and HIV/AIDS--a systematic review. Alcohol Alcohol 2010;45(2):159-166 [FREE Full text] [doi: 10.1093/alcalc/agp091] [Medline: 20061510]
3. Hendershot CS, Stoner SA, Pantalone DW, Simoni JM. Alcohol use and antiretroviral adherence: review and meta-analysis. J Acquir Immune Defic Syndr 2009 Oct 1;52(2):180-202 [FREE Full text] [doi: 10.1097/QAI.0b013e3181b18b6e] [Medline: 19668086]
4. Irwin T, Morgenstern J. Drug-use patterns among men who have sex with men presenting for alcohol treatment: differences in ethnic and sexual identity. J Urban Health 2005;82:127-133.

5. Tobin K, Davey-Rothwell M, Yang C, Siconolfi D, Latkin C. An examination of associations between social norms and risky alcohol use among African American men who have sex with men. Drug Alcohol Depend 2014 Jan 1;134:218-221. [doi: 10.1016/j.drugalcdep.2013.10.002] [Medline: 24210770]
6. Jaccard J, McDonald R, Wan CK, Guilamo-Ramos V, Dittus P, Quinlan S. Recalling sexual partners: the accuracy of self-reports. J Health Psychol 2004 Nov;9(6):699-712. [doi: 10.1177/1359105304045354] [Medline: 15367750]
7. Stone A, Shiffman S, DeVries M. Rethinking our self-report assessment methodologies: An argument for collecting ecological valid, momentary measurements. In: Kahneman D, Diener E, Schwarz N, editors. Well-being: The foundations of hedonic psychology. New York: Russell Sage Foundation; 1999:26-39.
8. Smyth J, Stone A. Ecological momentary assessment research in behavioral medicine. J Happiness Stud 2003;4:35-52.
9. Woolf SE, Maisto SA. Alcohol use and risk of HIV infection among men who have sex with men. AIDS Behav 2009 Aug;13(4):757-782. [doi: 10.1007/s10461-007-9354-0] [Medline: 18236149]
10. Collins F. Scientific American. 2012. The Real Promise of Mobile Health Apps URL: https://www.scientificamerican.com/article/real-promise-mobile-health-apps/ [accessed 2015-06-10] [WebCite Cache ID 6ZC313mgW]
11. Stone A, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. Annals of Behavioral Medicine 1994;16:199-202.
12. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [Medline: 12173681]
13. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]
14. Amadio D, Chung Y. Internalized homophobia and substance use among lesbian, gay, and bisexual persons. Journal of Gay and Lesbian Social Services 2006;17:83-101.
15. Freedman MJ, Lester KM, McNamara C, Milby JB, Schumacher JE. Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. J Subst Abuse Treat 2006 Mar;30(2):105-111. [doi: 10.1016/j.jsat.2005.10.005] [Medline: 16490673]
16. Epstein DH, Willner-Reid J, Vahabzadeh M, Mezghanni M, Lin J, Preston KL. Real-time electronic diary reports of cue exposure and mood in the hours before cocaine and heroin craving and use. Arch Gen Psychiatry 2009 Jan;66(1):88-94 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2008.509] [Medline: 19124692]
17. Hopper JW, Su Z, Looby AR, Ryan ET, Penetar DM, Palmer CM, et al. Incidence and patterns of polydrug use and craving for ecstasy in regular ecstasy users: an ecological momentary assessment study. Drug Alcohol Depend 2006 Dec 1;85(3):221-235. [doi: 10.1016/j.drugalcdep.2006.04.012] [Medline: 16730923]
18. Collins RL, Kashdan TB, Gollnisch G. The feasibility of using cellular phones to collect ecological momentary assessment data: application to alcohol consumption. Exp Clin Psychopharmacol 2003 Feb;11(1):73-78. [Medline: 12622345]
19. Simpson TL, Kivlahan DR, Bush KR, McFall ME. Telephone self-monitoring among alcohol use disorder patients in early recovery: a randomized study of feasibility and measurement reactivity. Drug Alcohol Depend 2005 Aug 1;79(2):241-250. [doi: 10.1016/j.drugalcdep.2005.02.001] [Medline: 16002033]
20. Hufford MR, Shields AL, Shiffman S, Paty J, Balabanis M. Reactivity to ecological momentary assessment: an example using undergraduate problem drinkers. Psychol Addict Behav 2002 Sep;16(3):205-211. [Medline: 12236455]
21. Kirk G, Linas B, Westergaard R, Piggott D, Bollinger R, Chang L, et al. The exposure assessment in current time study: implementation, feasibility, and acceptability of real-time data collection in a community cohort of illicit drug users. AIDS Res Treat 2013;2013:594671 [FREE Full text] [doi: 10.1155/2013/594671] [Medline: 24307943]
22. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986 Mar;42(1):121-130. [Medline: 3719049]
23. Todd M, Armeli S, Tennen H. Interpersonal problems and negative mood as predictors of within-day time to drinking. Psychol Addict Behav 2009 Jun;23(2):205-215 [FREE Full text] [doi: 10.1037/a0014792] [Medline: 19586137]
24. Otsuki M, Tinsley BJ, Chao RK, Unger JB. An ecological perspective on smoking among Asian American college students: the roles of social smoking and smoking motives. Psychol Addict Behav 2008 Dec;22(4):514-523. [doi: 10.1037/a0012964] [Medline: 19071976]
25. Diaz T, Chu SY, Sorvillo F, Mokotoff E, Davidson AJ, Samuel MC, et al. Differences in participation in experimental drug trials among persons with AIDS. J Acquir Immune Defic Syndr Hum Retrovirol 1995 Dec 15;10(5):562-568. [Medline: 8548336]
26. McCarthy DE, Piasecki TM, Fiore MC, Baker TB. Life before and after quitting smoking: an electronic diary study. J Abnorm Psychol 2006 Aug;115(3):454-466. [doi: 10.1037/0021-843X.115.3.454] [Medline: 16866586]
27. Muessig KE, Pike EC, Fowler B, LeGrand S, Parsons JT, Bull SS, et al. Putting prevention in their pockets: developing mobile phone-based HIV interventions for black men who have sex with men. AIDS Patient Care STDS 2013 Apr;27(4):211-222 [FREE Full text] [doi: 10.1089/apc.2012.0404] [Medline: 23565925]
28. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]

Abbreviations

ACASI: audio computer-assisted self-interview
EMA: ecologic momentary assessment
emocha: electronic Mobile Comprehensive Health App
HIV: human immunodeficiency virus
MSM: men who have sex with men

Edited by G Eysenbach; submitted 12.02.15; peer-reviewed by L Hightow-Weidman, D Epstein, S Legrand; comments to author 03.03.15; revised version received 15.03.15; accepted 03.04.15; published 17.06.15.

Please cite as:
Yang C, Linas B, Kirk G, Bollinger R, Chang L, Chander G, Siconolfi D, Braxton S, Rudolph A, Latkin C
Feasibility and Acceptability of Smartphone-Based Ecological Momentary Assessment of Alcohol Use Among African American Men Who Have Sex With Men in Baltimore
JMIR mHealth uHealth 2015;3(2):e67
URL: http://mhealth.jmir.org/2015/2/e67/
doi:10.2196/mhealth.4344
PMID:26085078

Original Paper

Usability and Feasibility of PIERS on the Move: An mHealth App for Pre-Eclampsia Triage

Joanne Lim[1], MAS; Garth Cloete[2], MSc; Dustin T Dunsmuir[1], MSc; Beth A Payne[3], PhD; Cornie Scheffer[2†], PhD; Peter von Dadelszen[3], MBChB, D Phil; Guy A Dumont[4], PhD; J Mark Ansermino[1], MBBCh

[1]Department of Anesthesiology, Pharmacology and Therapeutics, The University of British Columbia, Vancouver, BC, Canada

[2]Department of Mechanical and Mechatronic Engineering, Stellenbosch University, Stellenbosch, South Africa

[3]Department of Obstetrics and Gynaecology, The University of British Columbia, Vancouver, BC, Canada

[4]Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada

[†]deceased

Corresponding Author:
Joanne Lim, MAS
Department of Anesthesiology, Pharmacology and Therapeutics
The University of British Columbia
Clinical Support Building
950 W. 28th Avenue
Vancouver, BC, V5Z 4H4
Canada
Phone: 1 604 875 2000 ext 6669
Fax: 1 604 875 2668
Email: jlim2@cw.bc.ca

Abstract

Background: Pre-eclampsia is one of the leading causes of maternal death and morbidity in low-resource countries due to delays in case identification and a shortage of health workers trained to manage the disorder. Pre-eclampsia Integrated Estimate of RiSk (PIERS) on the Move (PotM) is a low cost, easy-to-use, mobile health (mHealth) platform that has been created to aid health workers in making decisions around the management of hypertensive pregnant women. PotM combines two previously successful innovations into a mHealth app: the miniPIERS risk assessment model and the Phone Oximeter.

Objective: The aim of this study was to assess the usability of PotM (with mid-level health workers) for iteratively refining the system.

Methods: Development of the PotM user interface involved usability testing with target end-users in South Africa. Users were asked to complete clinical scenario tasks, speaking aloud to give feedback on the interface and then complete a questionnaire. The tool was then evaluated in a pilot clinical evaluation in Tygerberg Hospital, Cape Town.

Results: After ethical approval and informed consent, 37 nurses and midwives evaluated the tool. During Study 1, major issues in the functionality of the touch-screen keyboard and date scroll wheels were identified (total errors n=212); during Study 2 major improvements in navigation of the app were suggested (total errors n=144). Overall, users felt the app was usable using the Computer Systems Usability Questionnaire; median (range) values for Study 1 = 2 (1-6) and Study 2 = 1 (1-7). To demonstrate feasibility, PotM was used by one research nurse for the pilot clinical study. In total, more than 500 evaluations were performed on more than 200 patients. The median (interquartile range) time to complete an evaluation was 4 min 55 sec (3 min 25 sec to 6 min 56 sec).

Conclusions: By including target end-users in the design and evaluation of PotM, we have developed an app that can be easily integrated into health care settings in low- and middle-income countries. Usability problems were often related to mobile phone features (eg, scroll wheels, touch screen use). Larger scale evaluation of the clinical impact of this tool is underway.

(JMIR mHealth uHealth 2015;3(2):e37) doi:10.2196/mhealth.3942

KEYWORDS

pulse oximetry; mHealth app; predictive model; usability analysis; design methodology

Introduction

Background

Pre-eclampsia is generally defined as the onset of proteinuric gestational hypertension [1,2]. Pre-eclampsia remains a leading cause of maternal morbidity and mortality, particularly in low- and middle-income countries (LMICs), in which more than 99% of pre-eclampsia related maternal deaths occur [3]. In LMIC settings, the difficulty in managing women with pre-eclampsia is made greater by delays in identification of women with pre-eclampsia and a lack of adequately skilled maternity care providers [4]. As such, clinical tools for the identification and monitoring of these pregnancies are required [1,5].

PIERS on the Move (PotM) [6,7] is a low-cost, easy-to-use, mobile health (mHealth) app that has been developed to aid frontline health workers in making decisions around management of women with pre-eclampsia. PotM is based on a decision model that combines accurate risk prediction of maternal adverse outcomes associated with pre-eclampsia (miniPIERS [Pre-eclampsia Integrated Estimate of RiSk]) with World Health Organization (WHO) recommendations for the management of pre-eclampsia [8].

The PotM app guides the health worker through a standardized process of antenatal assessment, including measurement of blood pressure, dipstick proteinuria and symptoms. The standardization of care resulting from this guided process and addition of blood pressure measurement as routine practice, which is critical to diagnosis of hypertensive disorders in pregnancy, represents a significant potential to improve clinical care in low-resourced settings. In addition, the PotM app allows integrated measurement of oxygen saturation (SpO_2), a vital sign previously demonstrated to have significant association with risk of adverse maternal health outcomes in women with pre-eclampsia [9]. The PotM app uses an integrated pulse oximetry sensor (Phone Oximeter) to measure SpO_2 [10]. The interface and data collection techniques of PotM were designed and developed for use by trained frontline health workers, nurses, and midwives in Africa and South Asia where the burden of pre-eclampsia is greatest. To ensure a user centric design and development process we undertook two usability studies in South Africa; one at Tygerberg Hospital, Cape Town and the second at Frere Maternity Hospital, East London.

Previously, we have described the motivation, design, and technical development of two versions of the PotM mobile app for the diagnosis and management of pregnant women with pre-eclampsia [7]. The purpose of this manuscript is to describe the feasibility and usability evaluation process of the development of the first version of these apps. The usability studies allowed for iterative improvement of the app into a version that is simple, intuitive, and easy-to-use.

The miniPIERS Model

The miniPIERS model [6] was developed to reduce adverse pregnancy outcomes by providing community-based health workers in low-resourced settings with an evidence-based and low-cost tool to improve diagnosis and initial management of pre-eclampsia. The model uses the demographics (gestational age at presentation), clinical signs (blood pressure and dipstick proteinuria), and symptoms (chest pain or dyspnoea, headache or visual disturbances, vaginal bleeding with abdominal pain) that are more readily available in low resourced settings. Using a threshold of miniPIERS-calculated predicted probability of ≥25% defines a high-risk population with 85.5% accuracy [6]. The risk of developing a complication of pre-eclampsia is more than 30 times greater when the SpO_2 is <93% [11].

The Phone Oximeter

The other component integrated into the PotM system is the Phone Oximeter [10], which consists of a mobile phone app that guides users to measure accurate instantaneous oxygen saturation from a connected pulse oximeter. The Phone Oximeter has been optimized for use in the PotM system as an easy and efficient method for health care providers to perform one-minute spot-checks of SpO_2 [12].

mHealth Tools

The PotM app uses the massive health care delivery opportunity offered by mobile devices and networks in the developing world. Initial evaluations of mHealth programs to improve maternal health in low-resourced settings have been positive with demonstrated improvements in women's levels of access of care [13]. mHealth helps to overcome access limiting factors, such as distance to services, social marginalization, and the paucity of skilled medical personnel and finances. Mobile devices are ideal for improving health care delivery. Their popularity and near-ubiquity enables delivery of interventions to large numbers of people, while their mobility allows advanced mHealth apps to be available at any time and place. Open Data Kit (ODK) is an example of a popular mhealth information system that is freely available for download. Evaluation of ODK has shown that mHealth tools can reduce health care costs while increasing the quality and efficiency of care [14].

A study into the usability factors of four different mobile devices for accessing health care information showed that the specific device used for the app played a major role in user satisfaction as well as efficiency [15]. The conclusion was that the future of mHealth app design should place particular importance on interface quality. There are other important considerations when designing an mHealth tool, such as data security and data synchronization. These were previously described for PotM by Dunsmuir et al [7].

Design Considerations for the Usability of PIERS on the Move

System Constraints

The aim was to develop and implement an app for use in low-resourced settings where no formal health care system is available and electricity or a reliable cellular data connection is nonexistent. Therefore, the app was required to make onsite treatment recommendations. Synchronization of the data with the central server was to occur when cellular or wireless connectivity was available.

The primary target user group was frontline care providers (nurses, midwives, community health workers) who might provide care for pregnant women in the community in rural

health care facilities or in a hospital setting. The level of education of these users can vary significantly and the system was kept as simple as possible while still being comprehensive to ensure usability at all levels of intended use.

Usability Testing Objectives

The usability test objectives were to: (1) exercise the app under controlled test conditions with representative users; (2) establish baseline user performance and user-satisfaction levels of the user interface; (3) determine design inconsistencies and usability problem areas within the user interface and content areas.

Potential sources of error may include: (1) navigation errors–failure to locate functions, excessive keystrokes to complete a function, failure to follow recommended screen flow; (2) presentation errors–failure to locate and properly act upon desired information in screens, selection errors due to labeling ambiguities; (3) control usage problems–improper entry field usage.

Methods

System Design

The PotM full system consists of a client app running on the user's mobile phone [7], a REDCap [16] database server running a web data collection app, and a web-interface that allows users or supervisors to enter additional follow-up information on patients. The client app was the focus of the usability studies.

PotM was developed using the *LambdaNative* framework, which allows rapid prototyping of the developed apps to run on either Google Android or Apple iOS operating systems [17].

Hardware Specifications

For the usability studies, the app was installed on iPod Touch (4^{th} generation model; Apple Inc, Cupertino, California, United States) and iPhone 3GS (A1303 model; Apple Inc) devices. The device was hardwired via a serial connection to the dock connector of a US Food and Drug Administration certified 16-bit OEM NoninXpod (Nonin Medical Inc, Plymouth, Minnesota, United States) pulse oximeter processing module. The Xpod pulse oximeter module provides the photoplethysmograph (PPG) waveform and the processed trend values for the SpO_2 and heart rate (HR).

User Interface

PotM supports the ability to input, view, and edit the patient history, clinical information, and current symptoms. The app also allows for the viewing of past evaluations and the monitoring and viewing of prescribed medications. All entries are logged with timestamps to ensure data consistency. The expected date of delivery is calculated from ultrasound measurements, last menstrual period or fundal height, in order of reliability when available.

The user interface enforces one minute measurements of oxygen saturation from the pulse oximeter using a color-coded signal quality indicator and progress bar, repeated measurements of blood pressure (2 measurements if consistent or 3 if inconsistent) and dipstick urine protein measurement [7].

Usability Evaluation

The evaluation of the basic interface functionality, workflow, and navigation was conducted in a series of participatory design groups that included investigators, research staff, and potential end-users. From these design groups, we created the initial prototypes for formal usability evaluation. Two usability studies were performed with the potential end-users. Each step in the development process used the findings of the previous, thus iteratively improving on the design and features available in the app: usability study 1: evaluation by advanced midwifery students at Tygerberg Hospital (Cape Town, South Africa); usability study 2: evaluation of the next iteration by maternal nursing staff at Frere Maternity Hospital (East London, South Africa).

After institutional ethics approval, subject participants were recruited at each of the sites. All nurses and midwives involved in maternity care at the participating sites were invited to participate and the final selection of participants was based on availability on the day of study. Our general pool of potential subjects was chosen as they would be representative of the future end-users of the app; nurses and midwives. Participation was voluntary and participants provided written informed consent. A minimum of 10 participants were selected for each usability evaluation; the suggested sample size for usability testing [18].

The evaluation of the device was performed in a quiet environment (closed room with no distractions). Participants completed a demographic questionnaire (gender, age range, use of mobile phone/personal computer). A facilitator was seated next to the participant and directed the interaction with the hardware and app. All the sessions were videotaped with the camera positioned such that only the hands of the participant and the devices were visible. The facilitator introduced the list of tasks to be completed. Participants were instructed to delay exploratory behavior outside the task flow until after the specific tasks were completed. Time-on-task measurement began when the participant started each of the specified tasks and at the end of each task; an observer recorded the duration of each task.

The facilitator instructed the participant to think aloud [19] during the task and encouraged expression of thought processes during the task performance. The observer entered the user behavior, user comments, and system actions on paper-based data logging forms.

A post-task questionnaire and interview was used to elaborate on the task session. A computer system usability questionnaire (CSUQ) [20] was completed by the participant at the end of the session. Each question was evaluated on a scale of 1 to 7, where 1 and 7 respectively correspond to "strongly agree" and "strongly disagree".

Metrics

Scenario completion success rates, time-to-completion of the tasks, specific user comments, error rates, and subjective evaluations were collected for each task.

Each task emulated a typical step in using the PotM platform and required that the participant obtain or input specific data. The tasks were considered complete when the participant

indicated that the task's specific goal was obtained (whether successfully or unsuccessfully) or the participant requested and received sufficient guidance to score the scenario as either "help required" or "critical error".

Usability Tasks and Evaluation

The usability tasks (provided as a paper script) (Textbox 1) were derived from test scenarios developed from the specific functionality of the system as well as with the assistance of subject-matter experts. Due to the range and extent of functionality provided in the app, and the time limitations of the participants, the tasks consisted of the most common and more complex of the available functions.

Textbox 1. Usability tasks performed.

1. Start the PIERS on the Move app
2. Log into the system
3. Enter contact details
4. Enter patient characteristics
5. Enter clinical history
6. Start a new evaluation
7. Enter symptoms
8. Enter blood pressure
9. Record pulse oximetry and heart rate
10. Enter patient outcomes
11. Record management advice provided
12. Enter medications
13. End evaluation
14. Search for the existing patient
15. Enter new patient data as provided
16. Log out of the system

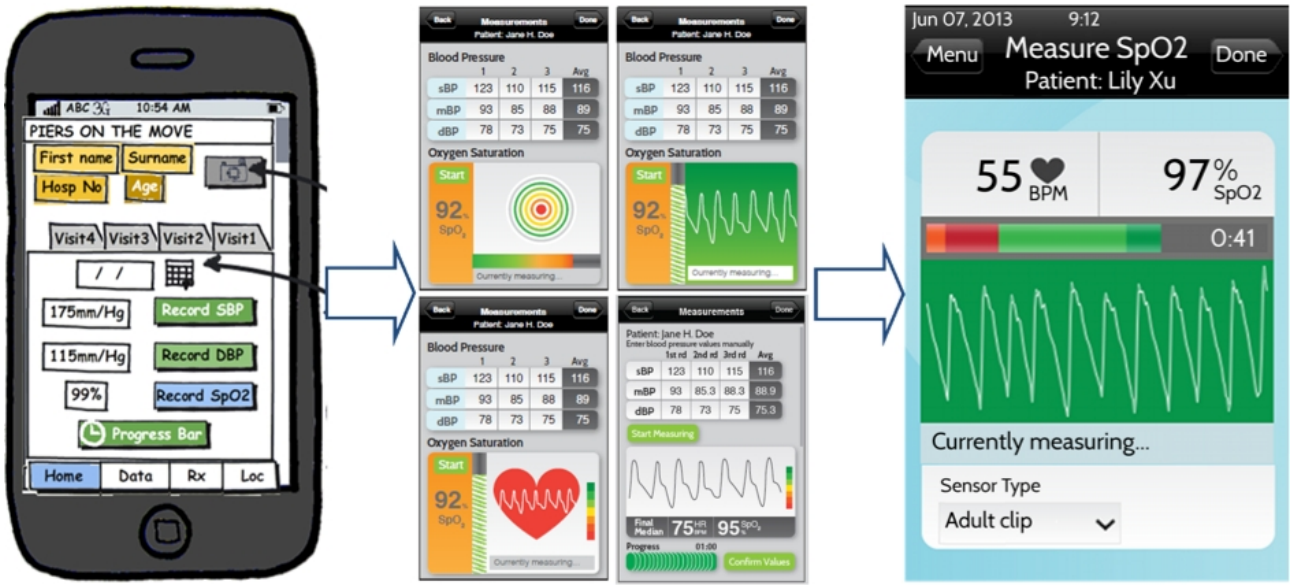
Results

Early Usability

Participatory sessions involved the generation of a series of wireframe mock-ups using Balsamiq (Balsamiq Solutions, LLC, Sacramento, California, United States) with the basic input variables and forms for each module (Figure 1). Once the mock-ups had been evaluated and optimized, full technical specifications and a final prototype were created (Figure 1). Following additional input from the design team and users, the prototype was implemented in the software framework and tested to ensure it met the required specifications.

The majority of participants were between 31 and 50 years of age (31 of 37 subjects) and were frequent users of cellphones with a minority using mobile phones. There were fewer participants in Study 1 than Study 2 who had never used a mobile phone (n=2 and n=8) or personal computer (n=0 and n=5). All of the participants were female.

Figure 1. Evolution of the pulse oximetry module.



Usability Study 1 – Tygerberg Hospital

Fifteen volunteers participated in the first study. There were a total of 212 errors in Study 1 (Figure 2).

Errors caused by the on-screen keypad and date wheels were the most common complaint. Participant feedback on the use of the keypad were directed to the layout of the keys and included: “keys are too small”, and “wrong button keeps getting pressed”. Whereas feedback concerning the date wheels were typically physics-based and included: “wheels are too sensitive”, and “I don’t understand how to change the numbers on the wheels”. Based on this feedback, both the keypad and date wheels for the app were redesigned. The keypad was enlarged to make all keys easier to press (Figure 3).

In addition, the interaction technique of the date wheels was adjusted to make the date selection easier. Originally, it was only possible to change the value on a date wheel by dragging your finger up or down and the speed of this drag determined how quickly it rotated. The speed of the rotation was faster than the actual drag in order to reach values faster and without multiple dragging of the finger. Once the finger was lifted, the rotation of the wheel stopped. This was found to be confusing and the interaction technique was redesigned. In the redesigned date wheel, the wheel turns at the same speed as the drag but when the finger is lifted, it keeps turning and then gradually slows to a stop. This enables the user to use a quick flicking motion to start the wheel spinning and they can potentially stop it early by touching it again. This more closely matches the physics of a real world spinning wheel.

Figure 2. Number of errors obtained during Study 1 & Study 2.

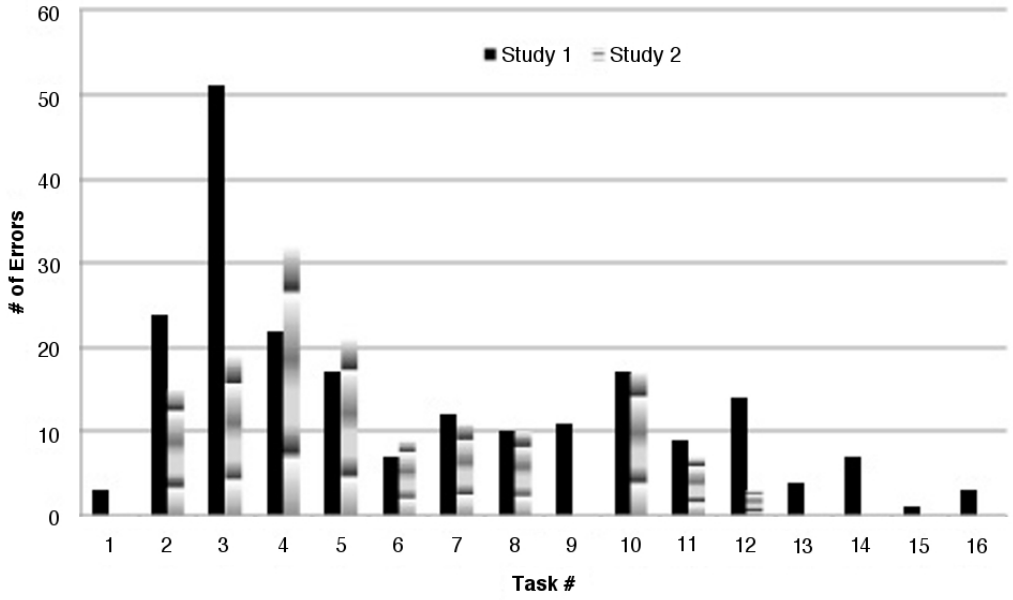
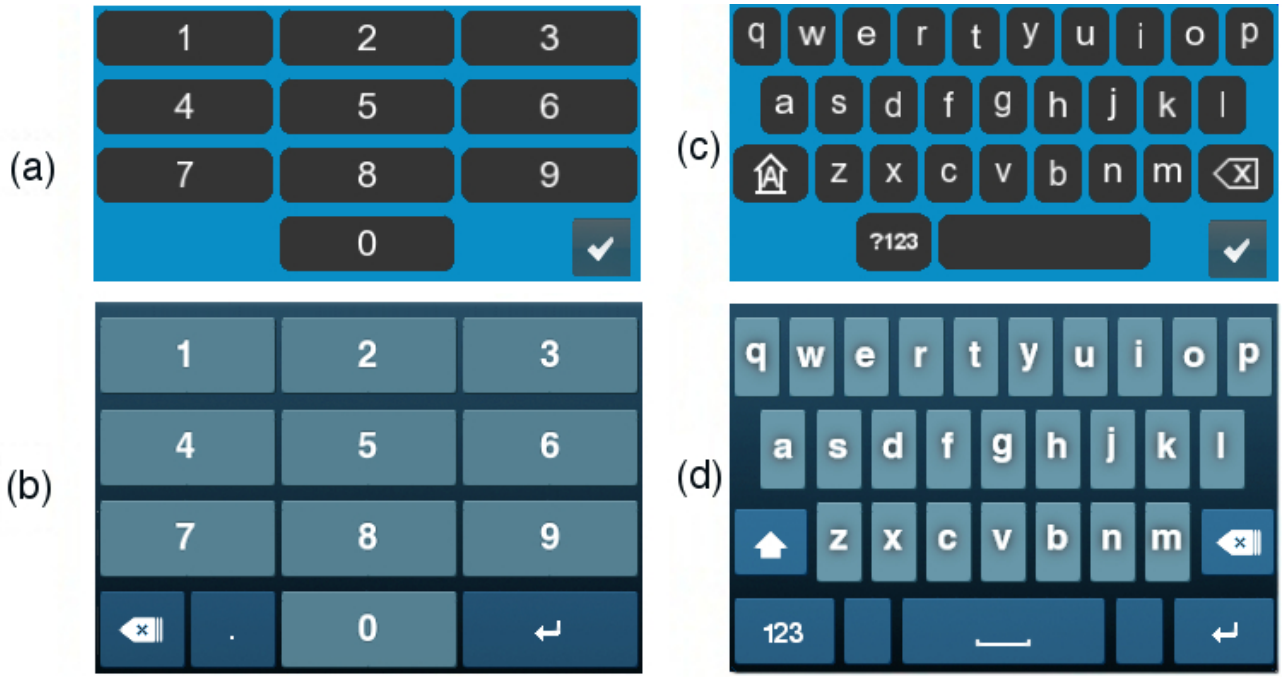


Figure 3. Redesign of the on-screen keypads: (a) & (b) two evolutions of the number pads; (c) & (d) two evolutions of the keypad.



Usability Study 2 – Frere Maternity

Evaluations were done by 22 nurses and midwives working at Frere Maternity Hospital (East London, South Africa).

The majority of errors were due to the learning curve required to master the use of a mobile phone or touch screen device for the first time. Typical errors included “why are the keyboard keys in that layout”, or “how do I move up and down on a touch screen”. These errors were quickly corrected once the participant was shown how to perform the function.

Errors were also evaluated in relation to the specific tasks they affected. A summary of the errors for Study 1 and Study 2 can be found in Table 1.

One recurring error was that the date wheel was “too difficult to use” (as seen in the increase in errors compared to Study 1 for Task 4). This occurred frequently even after redesign of the interaction technique of the wheel. This was mainly due to the participants (who are also our target end-users) in rural settings not being familiar with operating smart devices with touch screens and rotating wheel date selectors. The date wheels were subsequently removed, and replaced by up and down arrows to select day, month, and year. The arrow buttons can be held down to enable faster cycling of values, although this cycling is at a set speed. The arrow buttons do not enable as quick date selection as the wheels, but they were easier to understand for new mobile phone users.

Table 1. Errors in Study 1 and Study 2: Navigation errors–failure to locate functions, excessive keystrokes to complete a function, failure to follow recommended screen flow; Presentation errors–failure to locate and properly act upon desired information in screens, selection errors due to labeling ambiguities; Control usage problems–improper entry field usage.

Errors	Study 1	Study 2
Navigation errors	36	5
Could not locate or use navigation button	36	5
Presentation errors	84	45
Could not locate data entry feature–missed it	57	33
Feature did not appear how user expected it to	15	5
Reason for feature not understood	12	7
Control usage errors	92	92
Had trouble using data entry feature	85	80
Entered data in wrong location	5	10
Data entry was not done as user would like it to be	2	2
Program crashed	0	2
Total errors	212	144

CSUQ Results

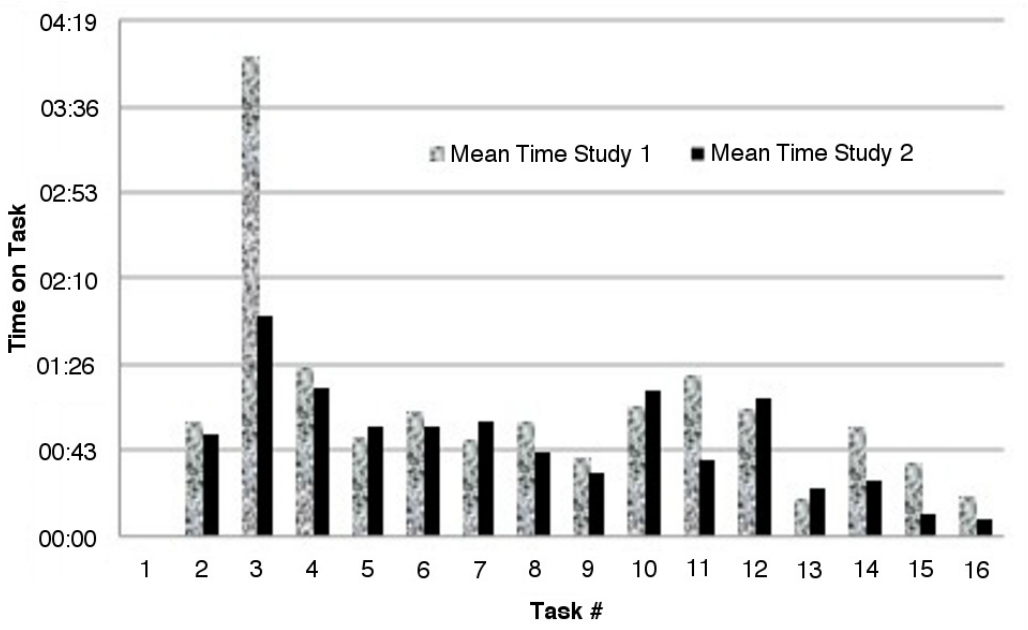
The overall satisfaction with the use of the PotM app as measured by the CSUQ was good (lower values are better) and improved in the subsequent study; Study 1 and Study 2 median (range) values are 2 (1–6) and 1 (1–7), respectively. For example, CSUQ statement "I can effectively complete my work using this interface" had a mean of 3.7 in Study 1 and 1.9 in Study 2. "I feel comfortable using this interface" had a mean of 2.7 in Study 1, and 1.6 in Study 2.

Time on Task

There were marked improvements in almost all of the task times with subsequent iterations. Although the average technology proficiencies of the participants in Study 2 were lower than those in Study 1, the participants of Study 2 demonstrated a 25% improvement in scenario completion time (Study 1 had a mean of 37 min 41 sec [SD: 9 min 7 sec], Study 2 had a mean of 30 min 07 sec [SD: 6 min 25 sec]) (Figure 4).

The largest decrease in time taken was seen in *Task 3: Enter Contact Details*, which was the biggest use of a keypad for entering (non-numeric) text. The increase in time taken to complete *Task 7: Enter Symptoms* and *Task 13: End Evaluation* was due to the additional functionality added on the screens at this stage of the study. For Task 7, the Symptoms screen had been redesigned to be more ergonomic and to produce an error message if the form was not fully complete. For Task 13 a detailed summary of why the specific treatment recommendation was given at the end of each evaluation.

Figure 4. Comparison of Study 1 & 2 time-on-tasks.



Clinical Trial

The feasibility of using the app clinically was evaluated by considering the time taken for data entry, the amount of missing data, the number of incomplete observations, and subjective feedback from a research nurse who performed all of the observations. The clinical study consisted of collecting patient data using the PotM app using a research nurse at Tygerberg Hospital over the period of November 2012-December 2013. Two hundred two women were recruited and enrolled in the study. The nurse performed multiple evaluations of each patient presenting with pre-eclampsia who met the necessary study inclusion criteria. The nurse was blinded to the recommendations given by the app.

Two secondary objectives of the clinical study were to assess the predictive ability of the miniPIERS model in this new cohort and to test if updating the miniPIERS model to include SpO_2, a variable that had not been available in the original study, improved predictive ability. Blinding was necessary to ensure clinical care was not affected by the recommendations of the model, which would prevent us from properly meeting these secondary objectives. Results of the miniPIERS model evaluation and updating have been published separately [9].

The average time taken to perform an evaluation (equivalent to Tasks 6-12 in the usability studies) was collected from data logs stored automatically by the PotM platform. In total more than 500 evaluations were performed on more than 200 patients. The median (interquartile range) time for an evaluation was 4min 55 sec (3 min 25 sec to 6 min 56 sec).

Discussion

Principal Findings

We conducted usability tests to optimize the design of an integrated mHealth tool for use by nurses and health care providers to evaluate the risk of developing adverse outcomes of pre-eclampsia. The PotM system integrated the Phone Oximeter and the miniPIERS predictive score and facilitated the collection of patient demographic information, symptoms, and clinical observations with a high degree of completeness in an average of less than 5 minutes in a clinical setting.

Nurses and midwives who participated in our study rated the usability high for the integration of these technologies.

The ease of use of mHealth tools is vitally important to the perception and uptake by the community health workers [21]. It would seem the largest barrier during the testing was the general unfamiliarness of using mobile devices with touch screen technology, and the associated functionality (eg, scroll wheels, on screen keyboards). It should be noted that these typical features in most mobile phones actually seem to be common problems for new users and were the cause of the usability problems. Once these features were explained and/or modified, users were able to complete the functions and tasks much more quickly. In Study 2 and then again for the clinical study, these common data entry features were improved and task times decreased. Even though the user subjects in Study 1 had a greater familiarity with mobile phone use, the CSUQ results demonstrate an overall improvement in usability even with less experienced mobile phone users in Study 2. This may have been caused by the improvements we made to the app between the two studies. Overall, the users we worked with were satisifed with the app and thought it would help their fieldwork. The use of the Phone Oximeter was of particular interest to the midwives as that device could be used in many different clincial settings.

Limitations

The results should be interpreted with caution as we did not test the device with front line health workers in rural setting who may have very limited education or access to technology. The nurses and midwives in our test population were relatively well-educated. Furthermore, the clinical feasibility study involved a single health care professional using PotM at one health care center, which may limit the generalizability of the findings.

Future Plans

In the Community Level Interventions for Pre-Eclampsia (CLIP) trial [22], community health care workers (CHWs) are using a version of the PotM app during their regular antenatal and postpartum visits to pregnant women. More than 500 CHWs will use the app to assess over 40,000 pregnant women throughout their pregnancies over the next two years. The CLIP PotM version is simplified for ease of use by those who have less medical training than the users of the original version of PotM. Feedback during the training of these CHWs and during the trial itself has led to further updates to the app. For example, we are displaying the PotM recommendations and record whether the pregnant woman accepted each recommendation. The recommendation summary page contains a short description and image for each appropriate recommendation such as “Transport to the hospital within 4 hours” with a pictogram of an ambulance and hospital. Each pair of image and text was meant to be clicked to go to the page containing the accepted or rejected checkbox (and reasons for rejection). Unfortunately this step was often skipped. Thus we redesigned the summary page to make each recommendation description text appear on a button, making it obvious that it could be clicked and adding a popup to warn users if they try to advance without accepting/rejecting each recommendation. Changes like this are easy to implement and can lead to much more accurate and thorough use of the app.

Our future plans include integrating a full health record for pregnant women during antenatal, intrapartum, and postnatal care including the identification and management of postpartum hemorrhage.

Finally, the Audio Phone Oximeter [23], which consists of a pulse oximeter sensor connected directly to the audio port of a phone, is an innovative solution to pulse oximetry on a mobile device. This eliminates the pulse oximeter processing module, which is expensive hardware that performs the processing of the pulse oximeter signals. Instead, the processing required to extract the HR and SpO_2 from the PPG waveform will be done directly on the mobile device. Additional sensors such as a semi-automated blood pressure cuff and a temperature sensor, which also connect to the audio port of any mobile device, are being developed.

Acknowledgments

The authors would like to thank the volunteer participants in Cape Town and East London, and Dr Justus Hofmeyr, Sister Naomi Mwesigye, and Sister Doreen Mugendi M'Rithaa for helping us to recruit volunteers. Thank you to Sister Erika van Papendorp for subject recruitment and data collection for the clinical pilot study. Thank you to Walter Karlen for technical and study assistance.

This work was supported by Stellenbosch University and the University of British Columbia. The project received grant funding from Saving Lives at Birth - Grand Challenges Canada (Grant# 0038-03-01-01-01) and the Bill & Melinda Gates Foundation (Grant # OPP1017337). Peter von Dadelszen and Mark Ansermino receive salary support from the Child & Family Research Institute. The study sponsors had no involvement in study design, data collection analysis, interpretation, writing the manuscript, and decision on where to submit the manuscript for publication.

Conflicts of Interest

None declared.

References

1. von Dadelszen P, Payne B, Li J, Ansermino JM, Broughton PF, Côté AM, et al. Prediction of adverse maternal outcomes in pre-eclampsia: development and validation of the fullPIERS model. Lancet 2011 Jan 15;377(9761):219-227. [doi: 10.1016/S0140-6736(10)61351-7] [Medline: 21185591]
2. Steegers Eric A P, von Dadelszen P, Duvekot JJ, Pijnenborg R. Pre-eclampsia. Lancet 2010 Aug 21;376(9741):631-644. [doi: 10.1016/S0140-6736(10)60279-6] [Medline: 20598363]
3. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, de Vaccaro Karen Courville, Des Jarlais Don C, Fowkes Francis Gerry R, Narayan K M Venkat, De León Felipe Rodriguez, Murray Christopher J L, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2095-2128. [doi: 10.1016/S0140-6736(12)61728-0] [Medline: 23245604]
4. Firoz T, Sanghvi H, Merialdi M, von Dadelszen P. Pre-eclampsia in low and middle income countries. Best Pract Res Clin Obstet Gynaecol 2011 Aug;25(4):537-548. [doi: 10.1016/j.bpobgyn.2011.04.002] [Medline: 21592865]
5. Lincetto O, Mothebesoane-Anoh S, Gomez P, Munjanja S. Lawn J, Kerber K. Antenatal Care URL: http://www.who.int/pmnch/media/publications/oanfullreport.pdf?ua=1 [accessed 2014-10-08] [WebCite Cache ID 6TB0TRBC2]
6. Payne BA, Hutcheon JA, Ansermino JM, Hall DR, Bhutta ZA, Bhutta SZ, miniPIERS Study Working Group. A risk prediction model for the assessment and triage of women with hypertensive disorders of pregnancy in low-resourced settings: the miniPIERS (Pre-eclampsia Integrated Estimate of RiSk) multi-country prospective cohort study. PLoS Med 2014 Jan;11(1):e1001589 [FREE Full text] [doi: 10.1371/journal.pmed.1001589] [Medline: 24465185]
7. Dunsmuir DT, Payne BA, Cloete G, Petersen CL, Görges M, Lim J, et al. Development of mHealth applications for pre-eclampsia triage. IEEE J Biomed Health Inform 2014 Nov;18(6):1857-1864. [doi: 10.1109/JBHI.2014.2301156] [Medline: 25375683]
8. World HO, Dept. of Maternal, Newborn, Child and Adolescent Health, Dept. of Reproductive Health and Research, Dept URL: http://www.who.int/reproductivehealth/publications/maternal_perinatal_health/9789241548335/en/ [accessed 2014-10-08] [WebCite Cache ID 6TB0h28SD]
9. Payne BA, Hutcheon JA, Dunsmuir D, Cloete G, Dumont G, Hall D, et al. Assessing the incremental value of blood oxygen saturation (SpO2) in the miniPIERS (Pre-eclampsia Integrated Estimate of RiSk) risk prediction model. J Obstet Gynaecol Can 2015 2015;37(1):16-24 [FREE Full text]
10. Karlen W, Dumont G, Petersen C, Gow J, Lim J, Sleiman J, et al. Human-centered Phone Oximeter interface design for the operating room. In: Traver V, Fred A, Filipe J, Gamboa H. editors. HEALTHINF 2011 - Proceedings of the International Conference on Health Informatics. Rome, Italy: SciTePress; 2011 Presented at: International Conference on Health Informatics; January 26-29, 2011; Rome, Italy p. 433-438. [doi: 10.5220/0003335204330438]
11. Millman AL, Payne B, Qu Z, Douglas MJ, Hutcheon JA, Lee T, et al. Oxygen saturation as a predictor of adverse maternal outcomes in women with preeclampsia. J Obstet Gynaecol Can 2011 Jul;33(7):705-714. [Medline: 21749746]
12. Dunsmuir D, Petersen C, Karlen W, Lim J, Dumont GA, Ansermino JM. The Phone Oximeter for mobile spot-check. In: Anest Analg(S2); S. 2012 Presented at: Proceedings of the 2012 Society for Technology in Anesthesia Annual Meeting; January 18-21, 2012; Palm Beach, FL, United State p. 26. [doi: 10.1213/01.ane.0000418552.16222.39]
13. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: a systematic review of the literature. PLoS One 2013;8(6):e65772 [FREE Full text] [doi: 10.1371/journal.pone.0065772] [Medline: 23776544]
14. Macharia P, Muluve E, Lizcano J, Cleland C, Cherutich P, Kurth A. Open Data Kit, A solution implementing a mobile health information system to enhance data management in public health. In: Cunningham P, Cunningham M, editors. 2013 Presented at: Proceedings of the IST-Africa 2013 Conference Exhibition; 29-31 May 2013; Nairobi, Kenya.
15. Sheehan B, Lee Y, Rodriguez M, Tiase V, Schnall R. A comparison of usability factors of four mobile devices for accessing healthcare information by adolescents. Appl Clin Inform 2012;3(4):356-366 [FREE Full text] [doi: 10.4338/ACI-2012-06-RA-0021] [Medline: 23227134]
16. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
17. Petersen CL, Görges M, Dunsmuir D, Ansermino JM, Dumont GA. Functional programming of mHealth applications. In: Proceedings of the 18th ACM SIGPLAN International Conference on Functional Programming. Boston, USA; 2013 Presented at: ACM SIGPLAN International Conference on Functional Programming; 2013; New York p. 15-27. [doi: 10.1145/2500365.2500615]
18. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. New York: Association for Computing Machinery; 1993 Presented at: Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems; Apr 24-29; Amsterdam, The Netherlands. [doi: 10.1145/169059.169166]
19. Kushniruk AW, Triola MM, Borycki EM, Stein B, Kannry JL. Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. Int J Med Inform 2005 Aug;74(7-8):519-526. [doi: 10.1016/j.ijmedinf.2005.01.003] [Medline: 16043081]
20. Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]

21. Palazuelos D, Diallo AB, Palazuelos L, Carlile N, Payne JD, Franke MF. User Perceptions of an mHealth Medicine Dosing Tool for Community Health Workers. JMIR Mhealth Uhealth 2013;1(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2459] [Medline: 25100670]
22. University of British Columbia. Community level interventions for pre-eclampsia (CLIP). 2014 URL: http://clinicaltrials.gov/ct2/show/NCT01911494 [accessed 2014-10-08] [WebCite Cache ID 6TB0oojLU]
23. Petersen CL, Gan H, MacInnis MJ, Dumont GA, Ansermino JM. Ultra-low-cost clinical pulse oximetry. Conf Proc IEEE Eng Med Biol Soc 2013;2013:2874-2877. [doi: 10.1109/EMBC.2013.6610140] [Medline: 24110327]

Abbreviations

CLIP: Community Level Interventions for Pre-Eclampsia
CSUQ: computer system usability questionnaire
HR: heart rate
LMIC: low and middle income country
mHealth: mobile health
miniPIERS: mini Pre-eclampsia Integrated Estimate of RiSk
PPG: photoplethysmogram
PotM: PIERS on the Move
SD: standard deviation
SpO2: oxygen saturation
WHO: World Health Organization

Edited by G Eysenbach; submitted 13.10.14; peer-reviewed by M Johnston, S Maingot, D Pan, YK Akpak; comments to author 03.11.14; revised version received 09.12.14; accepted 15.12.14; published 17.04.15.

Please cite as:
Lim J, Cloete G, Dunsmuir DT, Payne BA, Scheffer C, von Dadelszen P, Dumont GA, Ansermino JM
Usability and Feasibility of PIERS on the Move: An mHealth App for Pre-Eclampsia Triage
JMIR mHealth uHealth 2015;3(2):e37
URL: http://mhealth.jmir.org/2015/2/e37/
doi:10.2196/mhealth.3942
PMID:25887292

Original Paper

Adoption and Usage of mHealth Technology on Quality and Experience of Care Provided by Frontline Workers: Observations From Rural India

Sangya Kaphle[1], MA; Sharad Chaturvedi[2]; Indrajit Chaudhuri[2]; Ram Krishnan[2]; Neal Lesh[3], MPH, PhD

[1]Dimagi Software Innovations, Nairobi, Kenya

[2]CARE India, Patna, India

[3]Dimagi Software Innovations, Cambridge, MA, United States

Corresponding Author:
Sangya Kaphle, MA
Dimagi Software Innovations
Taarifa Road, Parklands
Nairobi,
Kenya
Phone: 254 733962903
Fax: 254 733962903
Email: sangyakaphle@gmail.com

Abstract

Background: mHealth apps are deployed with the aim of improving access, quality, and experience of health care. It is possible that any mHealth intervention can yield differential impacts for different types of users. Mediating and determining factors, including personal and socioeconomic factors, affect technology adoption, the way health workers leverage and use the technology, and subsequently the quality and experience of care they provide.

Objective: To develop a framework to assess whether mHealth platforms affect the quality and experience of care provided by frontline workers, and whether these effects on quality and experience are different depending on the level of technology adoption and individual characteristics of the health worker. Literacy, education, age, and previous mobile experience are identified as individual factors that affect technology adoption and use, as well as factors that affect the quality and experience of care directly and via the technology.

Methods: Formative research was conducted with 15 community health workers (CHWs) using CommCare, an mHealth app for maternal and newborn care, in Bihar, India. CHWs were first classified on the level of CommCare adoption using data from CommCareHQ and were then shadowed on home visits to evaluate their levels of technology proficiency, and the quality and experience of care provided. Regression techniques were employed to test the relationships. Out of all the CHWs, 2 of them refused to participate in the home visits, however, we did have information on their levels of technology adoption and background characteristics, which were included in the analysis as relevant.

Results: Level of technology adoption was important for both quality and experience of care. The quality score for high users of CommCare was higher by 33.4% (P=.04), on average, compared to low users of CommCare. Those who scored higher on CommCare proficiency also provided significantly higher quality and experience of care, where an additional point in CommCare proficiency score increased the quality score by around half a point (0.541, P=.07), and experience score by around a third of a point (0.308, P=.03). Age affected CommCare user type negatively, with an increase in age increasing the likelihood of belonging to a lower category of CommCare adoption (-0.105, P=.08). Other individual characteristics did not affect adoption or the predicted values estimating the relationship between adoption and quality and experience of care, although illiteracy was able to affect the relationship negatively.

Conclusions: mHealth technology adoption by frontline workers can positively impact the quality and experience of care they provide. Individual characteristics, especially literacy and age, can be important elements affecting technology adoption and the way users leverage the technology for their work. Our formative study provides informed hypotheses and methods for further research.

(JMIR mHealth uHealth 2015;3(2):e61) doi:10.2196/mhealth.4047

KEYWORDS

mHealth; technology adoption; community health workers; CommCare

Introduction

Background

Mobile health, or mHealth, platforms are currently in use in various programs around the world to facilitate health care delivery where frontline health workers play an important role in providing health services in resource-poor settings. While the number of programs using mobile technologies in health care is increasing globally, there exists a significant gap in knowledge regarding its impact on health outcomes, as well as intermediary factors like access, quality, and experience [1,2].

Increased access, quality, and experience of care are known to contribute toward improved health outcomes. Traditionally, the impact of any health intervention is assessed by investigating changes in the relevant health outcomes that the intervention is targeting. Improvements in quality, access, or experience are rarely the focus of study, and how and whether the intervention is leading to improvements in these indicators is usually not investigated in detail.

Regardless of their impact on health outcomes, improvements in quality and experience of care are legitimate goals for any health intervention. Not only are they the intermediate outcomes for better health from an equity perspective, they can be seen as end goals of health interventions, whereby everyone has access to a high quality and experience of health care [3]. Mobile health interventions can target all three of these factors—access, quality, and experience—leading to improvements in health outcomes [2].

It is possible that any mHealth intervention can yield differential impacts for different types of users. The quality and experience of care provided by community health workers (CHWs) can depend on individual characteristics like literacy, education, or age; social factors like the perception of CHWs by their target communities; or health system factors like receiving adequate support and essential materials needed to do their job [4]. In the case of mHealth, novel individual factors can also come into play, including those that affect technology adoption, such as perception of technology relevance and self-efficacy in utilizing the tools [5]. Mediating and determining factors include personal and socioeconomic factors affecting technology adoption [5], the way the CHWs leverage and use the technology, and subsequently the quality and experience of care they provide. Additionally, adoption and usage of the technology can directly impact the quality and experience of care provided by CHWs, while personal factors can also have a direct impact on the quality and experience of care.

With this in mind, Dimagi, in collaboration with CARE India, conducted formative research in Bihar, India, to better understand how the use of mHealth platforms affect the quality and experience of care provided by different types of CHWs. Our aim is to develop a framework to analyze the effects of mHealth technology adoption on the quality and experience of care, and to provide greater understanding of the personal factors affecting (1) mHealth technology adoption and (2) CHWs' use of the technology to support their work.

Research Objective

Our objective is to develop a framework to assess whether mHealth platforms affect the quality and experience of maternal and newborn care provided by CHWs, and whether these effects on quality and experience are different depending on the level of technology adoption and individual characteristics of the CHWs. We identified literacy, education, age, and previous mobile technology experience as individual factors that affect technology adoption and use, as well as factors that affect the quality and experience of care via the technology. Literacy and education were also identified as potential factors which can directly affect the quality and experience of care provided. Our formative work also offers some insight into factors that may affect technology adoption, and provides some field observations for comparisons between quality and experience of care provided by those using and not using the technology.

We expect to find better quality and experience of care among CHWs with higher levels of adoption and usage of mHealth tools. We also expect different effects on quality and experience based on individual characteristics, in addition to the level of technology adoption and usage.

The Setting

The study was conducted in Saharsa district in Bihar, India. Bihar is one of the more underdeveloped states in India. Most socioeconomic and mother-and-child health indicators in Bihar are considerably lower than the national average, including per-capita income, public expenditure on health, literacy rate, immunization rate among pregnant women and newborn children, institutional delivery rate, and malnutrition among children [6]. Nearly half of the population in Bihar is living under the poverty line [6]. Table 1 presents data on some socioeconomic and health indicators, as well as health infrastructure and human resource availability in Saharsa and Bihar. The low level of socioeconomic indicators adversely affects the health status of, and utilization of health services by, the population [6].

Table 1. Socioeconomic, health, and health infrastructure indicators for Saharsa and Bihar.

Indicators	Saharsa, n or %	Bihar, n or %
Socioeconomic indicators		
Total literacy [7], %	53.20	61.80
Male literacy, %	63.56	71.20
Female literacy, %	41.68	46.40
Total population [7], n	1,900,661	104,099,452
Urban population, %	8.24	11.29
Rural population, %	91.76	88.71
Health indicators		
Crude birth rate (number of live births in reference period/mid-year population x 1000) [8]	31.2	26.1
Crude death rate (number of deaths in reference period/mid-year population x 1000) [8]	7.4	6.8
Infant mortality rate (number of infant deaths [less than 1 year of age]/number of live births during reference period x 1000) [8]	55	48
Neonatal mortality rate (number of infants dying before 29 days per 1000 live births) [8]	37	32
Postneonatal mortality rate (infants dying between 29 days and 1 year per 1000 live births) [8]	18	16
Under 5 mortality rate (per 1000 live births) [8]	82	70
Maternal mortality ratio (maternal deaths per 1000 live births) [8]	33	30
Institutional deliveries, %	33.5 [9]	22.0 [10]
Full immunization in children, %	52.4 [9]	39.8 [11]
Health infrastructure and human resources [10]		
Number of doctors, n	53	N/A[a]
Number of Auxiliary Nurse Midwives (ANMs), n	225	N/A
Number of c (ASHAs), n	1242	N/A
Number of Aganwadi Workers (AWWs), n	1367	N/A
District hospitals, n	1	N/A
Referral hospitals, n	0	N/A
Primary health centers (PHCs), n	10	N/A
Additional primary health centers (APHCs), n	15	N/A
Health sub-centers (HSCs), n	152	N/A
Blood banks, n	1	N/A

[a]Not applicable (N/A). The data for Bihar were unavailable.

Figure 1 describes the health system structure in place in Bihar's districts including Saharsa. Each state in India has its own health care delivery system. The backbone of the system is a three-tier delivery system comprised of a tier one health sub-center (HSC), a tier two primary health center (PHC) and community health center (CHC), and a tier three district hospital. The Integrated Child Development Services (ICDS) provides immunization, health checkups, referrals, nutrition, health education, and preschool education to children below 6 years of age and women of reproductive age at the village level. Aganwadi Workers (AWWs) are at the center of the ICDS. AWWs run an Aganwadi center in each village, where the ICDS services are available to the population. However, the availability of health infrastructure and resources is still inadequate and the quality of services is poor. In order to improve the quality and access to services, especially in rural areas, the Ministry of Health and Family Welfare introduced the National Rural Health Mission (NRHM).

India announced and started implementing the NRHM with the goal of improving public health outcomes through community-driven approaches in 2005 [12]. The NRHM aims to improve access, affordability, accountability, and effectiveness of health care facilities available to the poor and vulnerable segments of the population. As part of the mission to bridge the gap in rural health services, the mission has created a cadre of community health workers, or Auxiliary Nurse Midwives (ASHAs), who are tasked with providing maternal and child health services to the communities that they are a part

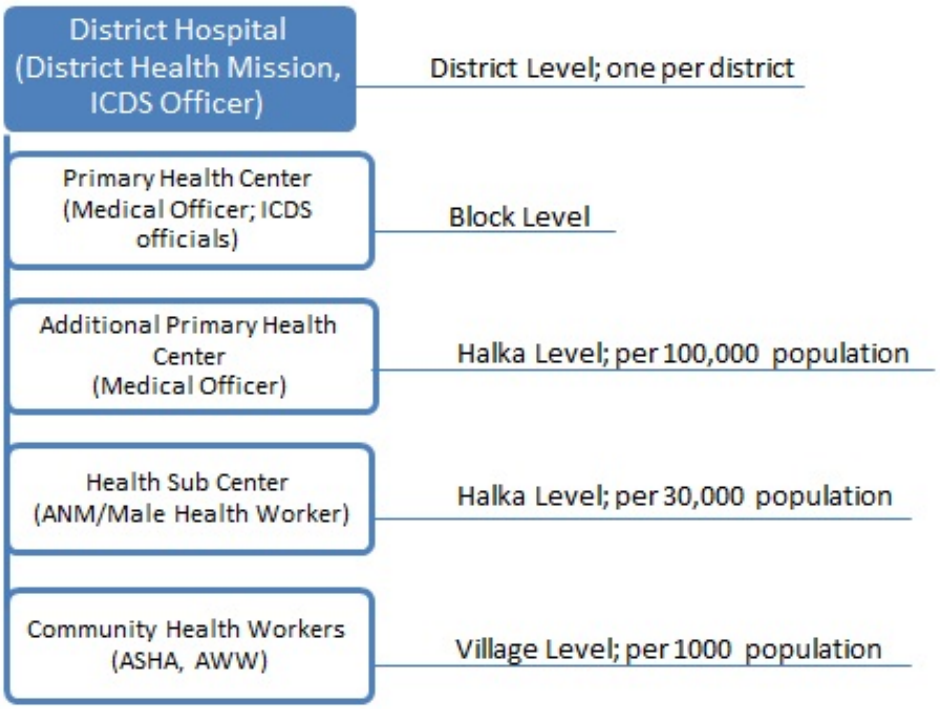
of. The ASHA program is a cornerstone of the NRHM, and it involves selecting, training, and supporting a locally recruited community-based health worker and change agent for every 1000 individuals in the community. The primary role of ASHAs is to create awareness and behavior change in health practices and improve utilization and accountability of the existing health systems, leading to stronger primary health care systems and services. ASHAs are trained to provide basic care, health information, and guidance, and to make referrals when appropriate. Community-led initiatives have been strikingly successful in improving health outcomes, and actions taken by households and families can prevent over 30% of child deaths [12].

CHWs, like ASHAs, have a vital role to play when it comes to influencing household and family choices that contribute to better health. In Bihar, where literacy levels are low, a majority of the ASHAs are from poor and low-literacy backgrounds [12]. ASHA training and support often involves content that requires more than basic literacy to grasp, which affects their effectiveness and performance. Working in remote, isolated settings, they can suffer from low morale and motivation. Capacity building using innovative training techniques suited for low-literacy adult learning, and equipping ASHAs with the resources to impart the knowledge, such as pictorial materials, radio access, or mHealth platforms, are some interesting strategies which are currently being deployed or being tested for their effectiveness to improve ASHA knowledge and skills [12]. ASHAs are complimented by AWWs, both of whom are supervised by an Auxiliary Nurse Midwife (ANM). The ANM runs village-level camps for maternal and child health services, which includes immunization of mothers and children, tracking nutritional status and growth, providing antenatal and postnatal checkups, and making referrals to the facilities when necessary [13].

There are some important challenges that need to be addressed in the health sector in Bihar. There are substantial gaps in health infrastructure, including primary health centers and community health centers. There are also gaps in the essentials required for effective functioning of the health facilities, including in drugs, consumables, equipment, and manpower [12]. Immunization coverage is low, there are high levels of malnutrition in children and mothers, and fertility rates are high [12]. Low quality of care provided at the district level down to the community levels, perpetuated by a lack of technical knowledge and skills, as well as gaps in health infrastructure and essentials, is also an important gap that needs to be addressed. Research on the challenges faced by ASHAs has identified a lack of support from PHC staff, a lack of adequate training, unclear incentives policy, and poor clarity in how to collaborate with the ANM and AWW as the main barriers to improving the quality of services they provide [14].

CARE India, in collaboration with the Ministry of Health and Ministry of Social Welfare, with the support of the Bill and Melinda Gates Foundation, is implementing CommCare, an mHealth platform targeting maternal and newborn care, with 600 ASHAs and AWWs in four blocks of the Saharsa district in Bihar. The deployment is part of a randomized controlled trial (RCT), which is implemented by CARE India with other consortium partners, and which will be evaluated by Mathematica Policy Research. The RCT compares health outcomes in different catchment areas where CHWs are (1) using CommCare or (2) using paper-based job aids. While our study is not associated with the broader RCT, the design of the intervention provides an ideal setting to better understand how mHealth platforms affect the quality and experience of home visits by CHWs. Both groups of ASHAs received content and capacity-building training facilitated by CARE at a fixed platform through an ANM, which means that CommCare ASHAs did not systematically receive more supervision and support than non-CommCare ASHAs. Additionally, the ASHAs using CommCare were randomly selected from the four blocks, which means that there was no bias to consider during the selection of our sample, which was drawn from the RCT's sample of ASHAs using CommCare. We are able to observe differences in the performances of CHW's using mHealth and those not using the technology and be more certain that any differences in visit quality and experience can be attributed to CommCare.

Figure 1. Public health system structure in India. The manpower available at each tier and administrative level of each type of health facility are shown. Each tier acts as a referral unit for the tier below. The district hospital services the facilities below with necessary support, resources, and essentials. Halka is a collection of villages, and Block is a collection of Halka.



Ethical Considerations

We have taken all measures possible to ensure that our study follows research governance and ethical protocol necessary for such research. The study benefits the participating ASHAs and the larger community where they work by understanding whether the mHealth platforms they are already using improve the quality and experience of the health services they provide to their communities. All ASHAs were informed about the nature of the study, and their consent was required prior to the researchers accompanying them on any home visits. Since our research does not involve patients or patient outcomes, it did not require any approvals from the Internal Review Board (IRB). Conflicts of interest that could arise due to the researchers having been employed by either Dimagi, who makes CommCare, or CARE India, who is implementing CommCare in Saharsa, have been mitigated by taking an unbiased view of CommCare in the study.

Analytical Framework

mHealth Technology Adoption and Improvements in Quality and Experience of Care

The level of technology adoption and usage can significantly affect the quality and experience of care provided by CHWs using mHealth platforms. CHWs are using CommCare to record and track pregnancies, newborns, and children up to 2 years of age. CHWs use CommCare to register pregnancies, provide counseling to the mother and families on safe pregnancies and newborn care, register births and deaths, and track immunization histories for both mother and child. CommCare provides decision support; a reproductive health checklist to ensure comprehensive care; multimedia, including images, audio, and video to enhance behavioral change communication; and features to aid with work planning and scheduling. Data from CommCare is reported to a central database, which helps supervisors provide targeted supervision to the CHW. These features aid the CHW in providing better access, quality, and experience of care to her clients [2,15] by addressing the gaps in medical information and skills [1], increasing adherence to protocol and guidelines [16], and improving engagement of the client with the rich multimedia components of the app to help with behavior change [17].

Factors Influencing Technology Adoption

Rogers' Diffusion Model identifies different users that adopt technology at various stages [18]. Other models of technology adoption identify demographic and individual characteristics, such as gender, age, technology advancement, technology readiness, technology experience, and self-efficacy, as mediating factors that affect technology adoption [5,19-21]. Demographic, socioeconomic, and individual factors can affect deterministic factors that affect technology adoption [5]. Perceived ease of use and perceived usefulness of technology are seen to affect attitudes and behaviors that influence the adoption and use of technology [5,20,21]. Social and program factors are also deemed important, but these are not the focus of our study [19,22,23].

CHWs are at various stages of technology adoption in our setting. Some CHWs have fully adopted the technology, others have not adopted the technology at all, while the remaining CHWs have only adopted certain features of the app.

Individual Factors and Their Effects on Community Health Workers' Utilization of the Technology and Quality and Experience of Care

We identified age, education, literacy, and previous mobile experience as individual factors that can affect the way CHWs leverage the technology. Literacy and education can also have a direct impact on the quality and experience of care provided by the CHWs, regardless of their influence on the CHWs' ability to leverage the app effectively.

The Ministry of Health (MOH) had imposed a minimum education criterion of 8 years to be eligible to work as an ASHA. However, not all ASHAs are educated to this level. Utilizing technology like CommCare effectively places additional cognitive demands on the CHWs' attention and abilities [24], and it is possible that levels of literacy and education are important factors affecting quality and experience of care provided by the CHWs [25].

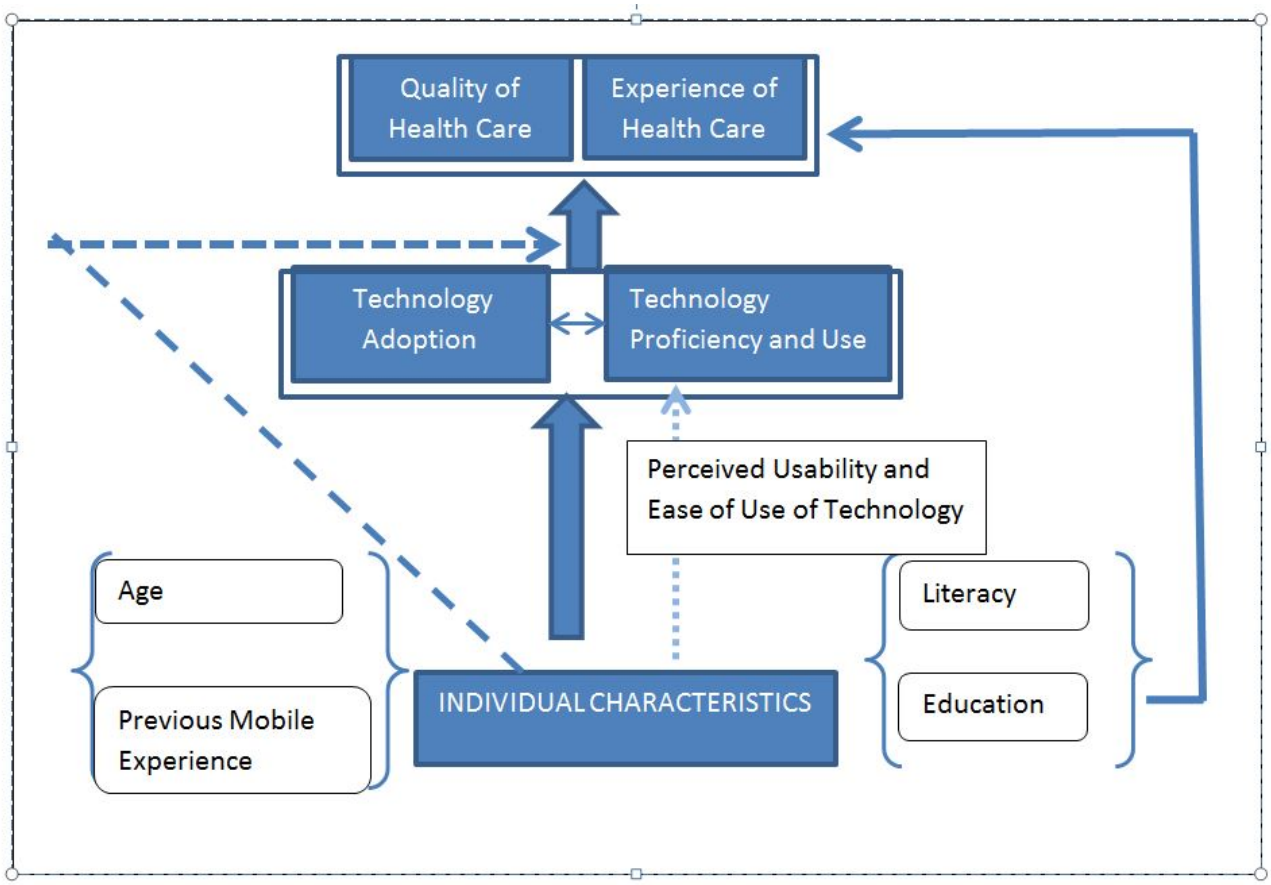
Partners implementing CommCare often cite low literacy and education among CHWs as a major challenge. In order to register patients and understand counseling messages in CommCare, CHWs require some level of comfort with reading and entering text into the app. If the CHWs are not literate, it is anticipated that they will be slow, require additional supervision and support, or be unable to adopt CommCare to facilitate their work. The quality of care they provide may also be lower since both CommCare and ASHA content training require that CHWs be able to read and write. CHWs with more years of education are able to better grasp the health information provided during the training.

CommCare includes audio prompts, images, and video to facilitate usage by lower-literacy users. An ASHA who is unable to read is still able to play the audio messages in order to understand the questions and to input appropriate responses. As such, she may be able to provide higher quality and experience of care compared to those not using or utilizing the technology effectively.

While previous mobile experience and age are not deemed to impact the quality or experience of care directly, as in the case of literacy or education, they can influence self-efficacy and perceived relevance and usability of the technology, mediating technology adoption and usage [26-28]. The impact on quality or experience of care is via the CHW's ability to use and leverage the app. To facilitate technology adoption, all ASHAs are trained in the use of mobile phones, as well as in the app and its content by CARE.

Figure 2 describes the analytical framework used in our study. This framework is adequate to study whether different levels of technology adoption and usage lead to differential effects of mHealth technology on quality and experience of care provided by the CHW. It is important to note that low adoption and usage of the technology, and low literacy and education, do not necessarily translate into low quality or experience of care. As long as the CHWs are able to leverage and adapt the technology to suit their skill set, they can provide high quality and experience of care even if they have low literacy and education levels, or are limited adopters and users of the technology. Social, cultural, and program factors are also important for mHealth technology adoption. Although these are not the focus of our study, our proposed model can benefit from including social, cultural, and program-wide factors, which also influence technology adoption and usage, and quality and experience of care.

Figure 2. Analytical framework. The flowchart shows the relationship between mHealth technology adoption and usage, and quality and experience of care. The level of technology adoption can affect the quality and experience of care provided by CHWs because of the design and content of the app. Individual factors, including literacy, education, age, and previous mobile experience, are seen as mediating factors for mHealth technology adoption and usage. They influence the quality and experience of care by affecting the way CHWs leverage the technology to do their jobs. Literacy and education can also directly influence the quality and experience of care delivered by the CHW.



Methods

Overview

We assessed the quality and experience of home visits of 13 ASHAs who are at different stages of CommCare adoption and use. We also assessed six home visits of 3 ASHAs who are not using CommCare, in order to better understand the effects of the mHealth platform on quality and experience of care. The ASHAs not using CommCare were selected based on availability and their proximity to CARE's offices in Saharsa. We collected field observations from these visits to further inform our findings.

Tools to (1) assess home visit quality and experience for the client, and (2) measure CommCare proficiency were developed as part of the study. An additional survey was administered to all the ASHAs to collect information on their background characteristics, including literacy, education, age, and previous mobile experience. The assessment tools underwent iterations after two rounds of field testing. We then shadowed each ASHA on a home visit to observe and score her visit for quality and experience. We also shadowed 3 ASHAs who were not using CommCare, in order to collect observations on how home visits are conducted without the tool, and to note any differences in quality and experience of these visits.

Data and Indicators

A total of 15 ASHAs were sampled and were classified as low (n=5), middle (n=5), and high (n=5) users of CommCare based on the number of forms submitted to CommCare's central database in the past 90 days. Home visits of 14 ASHAs were observed to assess whether ASHA literacy, CommCare proficiency, and CommCare user type—low, middle, high—affected the quality and experience of care provided. There were 2 ASHAs, both low users of CommCare, who refused to participate in the home visits, although we were able to collect their background information. We had a total of 13 ASHAs in our sample, and we shadowed 1 ASHA, a low CommCare user, on two home visits, which means we had a total of 14 home visits for our analysis.

The specific indicators are discussed in Table 2. Continuous variables, including CommCare proficiency score, visit quality score, visit experience score, age, and previous mobile experience were classified as low, middle, or high in order to test for measures of association with the categorical variables literacy, education, observed visit quality, and CommCare user type. The number of form submissions in the last 30, 60, and 90 days was used to classify the type of CommCare user, which was used as a measure for technology adoption.

CommCare proficiency was measured directly during the home visit assessment as a composite score based on whether the

ASHA could perform certain tasks in CommCare, and whether or not she used certain features of CommCare. We identified the following as features that measure CommCare proficiency and use: (1) navigation within the phone and within the app, (2) ability to select the client from a list of registered patients, (3) use of all the forms listed for the visit, (4) entering accurate answers, (5) entering text, dates, and numbers, and (6) the ability to play audio and video.

The quality of the home visit was measured through a quality score generated by scoring the visit based on (1) whether all the forms listed were used during the visit, (2) whether the visit included counseling, and (3) the number of counseling topics. For each topic, scoring was also based on (1) whether complete and accurate information was provided, (2) whether the client asked questions, and (3) whether the ASHA verified that the messages were received. Experience of home visit was measured directly during the home visit assessment. It is a composite score, generated by adding scores regarding (1) audio usage frequency, (2) video usage frequency, (3) frequency of showing images, (4) whether all of the people present were addressed, (5) whether the ASHA spoke clearly and loudly with confidence, and (6) the duration of the visit, where a visit under 10 minutes scored 1, 10 to 20 minutes scored 2, and over 20 minutes scored 3. The rationale for measuring visit experience in this way is based on the finding that the use of multimedia increases the quality of the visit [17].

Previous mobile experience was scored by asking whether AHSAs were able to perform a set of tasks on the mobile phone prior to using CommCare. A composite score was based on whether the ASHA (1) used a mobile phone, (2) owned her own phone, and (3) whether she could answer the phone, place calls, use the contact list, send and receive short message service (SMS) messages, play music, take photographs, change the phone's date and time, check her balance, and charge the phone's battery prior to using CommCare.

Table 2. List of indicators and their descriptions.

Indicators	Descriptions
CommCare user type	
Low	Users who have not submitted any forms using CommCare in the last 90 days. We selected the 5 ASHAs who submitted the least number of forms in the last 90 days for this sample.
Middle	Users who fell into the 50th percentile in terms of forms submitted in the last 30, 60, and 90 days. Those who fell between the 50th-55th percentiles for form submissions in the last 30, 60, and 90 days were the preferred middle users. In our sample, 3 users fell into this range for all three time periods and 2 fell into this range in the last 30 and 90 days.
High	Users with the highest number of form submissions in the last 30, 60, and 90 days. In our sample, 3 users had the highest number of form submissions in all three time periods, and 2 ASHAs had the highest number in the last 30 and 60 days.
CommCare proficiency [a]	
Low	ASHAs in the lowest 25th percentile of the CommCare proficiency score were categorized as low.
Middle	ASHAs within the 25th to 75th percentile of the CommCare proficiency score were categorized as middle.
High	ASHAs above the 75th percentile of the CommCare proficiency score were categorized as high.
Quality of home visit [b]	
Low	ASHAs in the lowest 25th percentile of the visit quality score were categorized as low.
Middle	ASHAs within the 25th to 75th percentile of the visit quality score were categorized as middle.
High	ASHAs above the 75th percentile of the visit quality score were categorized as high.
Observed visit quality	A second measure for visit quality based on the researcher's perception of the home visit was included. This was a subjective measure of the visit quality, classified again as low, middle, or high, based on the researcher's perception.
Experience of home visit [c]	
Low	ASHAs in the lowest 25th percentile of the visit experience score were categorized as low.
Middle	ASHAs within the 25th to 75th percentile of the visit experience score were categorized as middle.
High	ASHAs above the 75th percentile of the visit experience score were categorized as high.
Literacy level [d]	
Illiterate	The ASHA cannot read at all.
Low Literacy	The ASHA can read with difficulty, or can read some of the sentence.
Literate	The ASHA can read easily.
Education level [e]	
Low	The ASHA was educated up to 8th standard.
Middle	The ASHA was educated up to 10th standard.
High	The ASHA was educated up to, or higher than, 12th standard.
Previous mobile experience[f]	Previous mobile experience was classified as low, middle, or high based on the percentile, where those under the 25th percentile score were low, 25th-75th percentile were middle, and above the 75th percentile were high.
Age[g]	Age was classified as low, middle, or high based on the percentile, where 25th percentile and below were low, 25th-75th percentile were middle, and above 75th percentile were high.

[a]ASHAs could earn a maximum of 22 points for their CommCare proficiency score.

[b]ASHAs could receive a maximum of 22 points for their quality of home visit score.

[c]ASHAs could receive a maximum of 16 points for their visit experience score.

[d]A literacy test was administered as part of the background interviews to assess the literacy levels, where ASHAs were asked to read a sentence in Hindi out loud.

[e]Education was self-reported by the ASHA during the interview.

[f]ASHAs could score a maximum of 18 points for previous mobile experience.

[g]Age was self-reported by the ASHA during the interview.

Limitations of the Indicators

The tool to measure the visit quality and experience had not been rigorously tested prior to the formative work, and our measure of the visit quality and experience has room for improvement. Currently, the quality score is mostly dependent on the number of counseling topics, where if an ASHA counsels on three topics she will have a larger possible score even if the counseling is not necessarily of high quality, than if she counsels on one topic. Although it is likely to be more comprehensive, it is not necessarily the case that a larger number of counseling topics equals higher-quality visits. The scoring system for quality and experience of visit can also be changed so that most important aspects are weighted accordingly. Another possibility is to assess the components of quality and experience independently, without creating a composite quality or experience score. Along this line, it may also be beneficial to create a single score for both quality and experience combining the different elements into one single indicator. Despite the shortcomings, we are confident that the tool captures the quality and experience of the home visit without bias.

Empirical Strategy

The empirical strategy adopted below will be supplemented by field observations to understand four specific elements of our analytical model.

1. Does mHealth Technology Adoption Affect Quality and Experience of Care?

We will start by using both of the indicators of adoption, CommCare proficiency and usage, as well as CommCare user type based on form submissions to understand the effects of technology adoption on quality and experience of care. We will use the quality and experience scores as our dependent variables since the observed visit quality measure is more subjective.

We first estimate how CommCare adoption measured by CommCare user type affects the quality and experience of care provided by the ASHA using a simple linear regression model.

$$Quality_{ij} = \beta0+\beta1COMMCAREuser\text{-}type1_{ij} +\beta2COMMCAREuser\text{-}type3_{ij}+ _{ij} \quad (1.1)$$

$$Experience_{ij} = \beta0+\beta1COMMCAREuser\text{-}type1_{ij} +\beta2COMMCAREuser\text{-}type\ 3_{ij}+ _{ij} \quad (1.2)$$

We estimate two different specifications for equations 1.1 and 1.2 and include two out of the three different categories of CommCare user type in each specification, which allows us to compare all three groups of users against each other. The coefficient for β1 and β2 estimate the difference in quality and experience score for that CommCare user type against the CommCare user type excluded from the specification.

We then estimate specifications 1.3 and 1.4 using CommCare proficiency as the indicator for CommCare adoption. We again use a linear regression model to estimate specifications 1.3 and 1.4. β1 estimates how a 1-point increase in CommCare proficiency affects quality/experience of care.

$$Quality_{ij} = \beta0+\beta1COMMCAREProficiency_{ij} + _{ij} \quad (1.3)$$

$$Experience_{ij} = \beta0+\beta1COMMCAREProficiency_{ij} + _{ij} \quad (1.4)$$

2. Do Accredited Social Health Activist Characteristics Affect Adoption of CommCare?

We aim to see the effects of ASHA characteristics, namely literacy levels, education level, age, and previous mobile experience, on CommCare adoption and usage by estimating the following two models. Model 2.1 estimates how the ASHA characteristics affect their categorization of CommCare user type using an ordered probit model, while model 2.2 estimates how these characteristics affect their CommCare proficiency using a multivariate regression model. For education and literacy, we create binary variables. Since our sample for illiterate is small, we estimate specification 2.1 with illiterate and low literacy users combined into one group, and those with low and middle education levels combined into one group, and compare results against literate and highly educated users. We also follow this method for specification 2.2.

$$CommCare\ user\text{-}type_{ij} = \beta0+\beta1Lit_{ij}+\beta2Edu_{ij}+\beta3Age_{ij} +\beta4PrevMobileExp_{ij}+ _{ij} \quad (2.1)$$

$$CommCareproficiency_{ij} = \beta0+\beta1Lit_{ij}+\beta2Edu_{ij}+\beta3Age_{ij} +\beta4PrevMobileExp_{ij}+ _{ij} \quad (2.2)$$

We also estimated the relationship for all three categories of education, without combining the low- and middle-educated users into one category. However, this did not change our results.

3. Do Accredited Social Health Activist Characteristics, Namely Literacy and Education, Affect Quality and Experience of Care Directly?

We focus on literacy and education since age and previous mobile experience likely do not have any direct association with quality and experience of care. We estimate specifications 3.1 and 3.2 using a linear regression model. We estimate 3.1 and 3.2 with the illiterate and literate users combined into one group, and those with low and middle education levels combined into one group, and compare results against literate and highly educated users due to the small sample of illiterate users.

$$Quality_{ij}=\beta0+\beta1Lit_{ij} +\beta3Edu_{ij} + _{ij} \quad (3.1)$$

$$Experience_{ij}=\beta0+\beta1Lit_{ij}+\beta2Edu_{ij} + _{ij} \quad (3.2)$$

We also estimated the relationship for all three categories of education, without combining the low- and middle-educated users into one category. However, this did not change our results and they are not presented in this paper.

4. Do Accredited Social Health Activist Characteristics Matter for the Relationship Between CommCare Adoption and Quality and Experience of Care Identified in the First Model (1.1-1.4)?

ASHA characteristics can affect their ability to leverage the technology. We estimate how ASHA characteristics affect the relationship between CommCare adoption and quality and experience of care by using the predicted values for quality score and experience score from equations 1.1 and 1.2 as our dependent variables. These predicted values estimate the relationship between quality/experience of care and CommCare user type, a measure for adoption in our study. Using these

predicted values as our dependent variables will allow us to gauge whether the relationship between quality/experience of care and technology adoption is affected by individual characteristics.

Using predicted values from equations 1.1 and 1.2, respectively, we estimate models 4.1 and 4.2 treating these as a system of equations. The dependent variables are predicted values from models 1.1 and 1.2, which estimate the relationship between quality/experience of care and CommCare user type.

$Quality'_{ij} = \beta0+\beta1Lit_{ij}+\beta2Edu_{ij}+\beta3Age_{ij}+\beta4PrevMobileExp_{ij}+ _{ij}$ (4.1)

$Experience'_{ij} = \beta0+\beta1Lit_{ij}+\beta2Edu_{j}+\beta3Age_{ij}+\beta4PrevMobileExp_{ij}+ _{ij}$ (4.2)

Using predicted values from equations 1.3 and 1.4, respectively, we estimate equations 4.3 and 4.4. The dependent variables are predicted values from equations 1.3 and 1.4, which estimate the relationship between quality/experience of care and CommCare proficiency, a second measure for technology adoption in our study.

$Quality'_{ij} = \beta0+\beta1Lit_{ij}+\beta2Edu_{ij}+\beta3Age_{ij}+\beta4PrevMobileExp_{ij}+ _{ij}$ (4.3)

$Experience'_{ij} = \beta0+\beta1Lit_{ij}+\beta2Edu_{ij}+\beta3Age_{ij}+\beta4PrevMobileExp_{ij}+ _{ij}$ (4.4)

Results

Descriptive Statistics

Table 3 presents the descriptive statistics of the individual ASHA characteristics informing the study, and Table 4 presents the descriptive statistics for the ASHAs' home visit quality and experience scores for different levels of CommCare adoption. CommCare user type was positively and significantly correlated with CommCare proficiency, with a correlation coefficient of .771 (P=.001), significant at the 99% confidence level. There was also significant positive correlation between our two measures of quality with a correlation coefficient of .787 (P<.001), also significant at the 99% confidence level.

Table 3. Descriptive statistics of ASHA characteristics.

ASHA characteristics		Mean (SD) or n (%)
Age in years (n=15), mean (SD)		31.60 (5.86)
Previous mobile experience (n=15), mean (SD)		8.25 (4.23)
CommCare proficiency and use (n=14), mean (SD)		8.78 (4.84)
Education (n=15), n (%)		
	Low, n (%)	5 (33)
	Middle, n (%)	4 (27)
	High, n (%)	6 (40)
Literacy (n=15), n (%)		
	Illiterate, n (%)	1 (7)
	Low Literacy, n (%)	6 (40)
	Literate, n (%)	8 (53)
CommCare user type (n=15), n (%)		
	Low, n (%)	5 (33)
	Middle, n (%)	5 (33)
	High, n (%)	5 (33)
CommCare proficiency (n=14), n (%)		
	Low, n (%)	3 (21)
	Middle, n (%)	7 (50)
	High, n (%)	4 (29)

Table 4. Descriptive statistics for quality/experience of home visits for different levels of CommCare adoption.

ASHA CommCare adoption (n=14)	n (%)	Quality score, mean (SD)	Experience score, mean (SD)	Perception of visit quality			
				Mean (SD)	Low, n (%)	Middle, n (%)	High, n (%)
CommCare visits (n=14)	14 (100)	7.92 (5.24)	5.57 (2.59)	2.00 (0.88)	5 (36)	4 (29)	5 (36)
Non-CommCare visits (n=6)	6 (100)	9.17 (4.75)	4.33 (1.37)	1.50 (0.84)			
CommCare user type							
Low	4 (29)	4.25 (0.50)	3.25 (0.96)	1.00 (0)	4 (29)		
Middle	5 (36)	7.20 (6.38)	5.80 (2.39)	2.00 (0.71)	1 (7)	3 (21)	1 (7)
High	5 (36)	11.60 (4.16)	7.20 (2.59)	2.80 (0.45)		1 (7)	4 (29)
CommCare proficiency							
Low	3 (21)	4.33 (0.58)	3.67 (0.58)	1.00 (0)	3 (21)		
Middle	7 (50)	8.00 (6.22)	5.71 (3.14)	2.00 (0.82)	2 (14)	3 (21)	2 (14)
High	4 (29)	10.50 (4.43)	6.75 (1.89)	2.75 (0.50)		1 (7)	3 (21)

The mean age of the ASHAs was 31.60 years (SD 5.86) and most self-reported having 8 years of education. Out of the 15 ASHAs, only 1 (7%) was classified as illiterate in the sample, while 6 (40%) were low literacy, and 8 (53%) were literate. The mean visit quality score for ASHAs using CommCare was 7.92 (SD 5.24), while the mean experience score was 5.57 (SD 2.59). We only had 3 ASHAs that were not using CommCare included in the study, and we observed two home visits for each of these 3 ASHAs for a total of six home visits. Their visit quality and experience scores were modified versions of the CommCare visit quality and experience scores, excluding the mobile technology components. As such, the quality and experience scores are not entirely comparable across the two groups of ASHAs. However, we can look at the score for perception of visit quality as an unbiased indicator capturing the quality of the visit for both CommCare and non-CommCare ASHAs. Based on this indicator, the perception of visit quality was higher for ASHAs using CommCare compared to those not using CommCare. Similarly, mean quality and experience scores for ASHAs with higher levels of CommCare adoption were higher compared to those with lower levels of CommCare adoption using both measures of adoption.

ASHAs who were literate had a higher mean for the experience score compared to those that had low literacy or were illiterate, as seen in Table 5. However, means for quality and experience did not seem to increase along with increases in the levels of education, literacy, and previous mobile experience. Older ASHAs had lower visit quality and experience means compared to their younger compatriots.

Table 5. Quality and experience scores by ASHA individual characteristics.

ASHA characteristics (n=14)	n (%)	Quality score, mean (SD)	Experience score, mean (SD)	Perception of visit quality			
				Mean (SD)	Low, n (%)	Middle, n (%)	High, n (%)
Literacy							
Illiterate	2 (14)	4.00 (0)	3.50 (0.71)	1.00 (0)	2 (14)		
Low literacy	6 (43)	8.67 (6.53)	5.67 (3.44)	2.00 (0.89)	2 (14)	2 (14)	2 (14)
Literate	6 (43)	8.50 (4.64)	6.17 (1.83)	2.33 (0.82)	1 (7)	2 (14)	3 (21)
Education in years							
8	6 (43)	7.33 (6.25)	5.50 (2.51)	1.83 (0.98)	3 (21)	1 (7)	2 (14)
10	3 (21)	5.67 (2.89)	3.67 (2.08)	1.67 (0.58)	1 (7)	2 (14)	
≥12	5 (36)	10.00 (5.15)	6.80 (2.68)	2.40 (0.89)	1 (7)	1 (7)	3 (21)
Previous mobile experience							
Low	5 (36)	7.00 (5.61)	5.40 (2.30)	2.00 (0.71)	1 (7)	3 (21)	1 (7)
Middle	6 (43)	9.17 (5.95)	6.33 (3.01)	2.17 (0.98)	2 (14)	1 (7)	3 (21)
High	3 (21)	7.00 (4.36)	4.33 (2.51)	1.67 (1.15)	2 (14)		1 (7)
Age							
Low	3 (21)	6.00 (5.20)	5.67 (2.31)	2.00 (1.00)	1 (7)	1 (7)	1 (7)
Middle	7 (50)	10.43 (5.80)	6.86 (2.61)	2.43 (0.79)	1 (7)	2 (14)	4 (29)
High	4 (29)	5.00 (2.00)	3.25 (0.96)	1.25 (0.50)	3 (21)	1 (7)	

Correlation Coefficients and Regression Results

1. Does the Level of mHealth Technology Adoption Affect Quality and Experience of Care Provided by Accredited Social Health Activists?

Table 6 presents the correlation coefficients between CommCare adoption and quality and experience of health care. While CommCare user type is significantly associated only with perception of the visit quality, CommCare proficiency and use is significantly correlated with all three response variables: quality score, experience score, and perception of visit quality.

Table 6. Correlations between CommCare adoption and quality and experience of care.

Variables[a]	χ^2 or r[b]	*P*
Quality score and CommCare user type, χ^2_2	4.6	.33
Experience score and CommCare user type, χ^2_2	2.5	.65
Perception of visit quality and CommCare user type, χ^2_4	14.3	.006
Quality score and CommCare proficiency, *r*	.50	.07
Experience score and CommCare proficiency, *r*	.57	.03
Perception of visit quality and CommCare proficiency, χ^2_2	9.3	.06

[a]We transformed quality and experience scores into categorical variables in order to test for association with CommCare user type, which is also a categorical variable.

[b]We performed chi-square tests to look for measures of association between categorical variables, and pairwise correlations (*r*) for continuous variables.

We found that the level of technology adoption is important for both quality and experience of care. High users of CommCare, as identified by CommCare user type, provided significantly higher scores for quality and experience of care than low users of CommCare for both measures of quality. The quality score for high users of CommCare was higher by 7.35 (*P*=.04), on average, compared to low users of CommCare, which is a difference of 33.4%, significant at the 95% confidence level. Those who scored higher on CommCare proficiency also provided significantly higher quality and experience of care scores, where an additional point in the CommCare proficiency score increased the quality score by around half a point (0.54,

P=.07), and experience score by around a third of a point (0.31, *P*=.03). This amounts to a 2.5% increase in quality score, and a 1.9% increase in experience score for each additional point in CommCare proficiency, both significant at the 95% confidence level (see Table 7).

Table 7. Relationship between CommCare adoption and quality and experience of care.

ASHA characteristics (n=14)	Quality score, mean (t_{13})		Experience score, mean (t_{13})		Quality score[a], mean (t_{13})	Experience score[b], mean (t_{13})
	1[c]	2	3	4	5	6
CommCare user type						
Low	-2.950 (-0.96)		-2.550 (-1.74)			
Middle		2.950 (0.96)		2.550 (1.74)		
High	4.400 (1.51)	7.350 (2.38)	1.400 (1.01)	3.950 (2.70)		
CommCare proficiency					0.541 (2.00)	0.308 (2.43)
Constant[d]	7.200 (3.50)	4.250 (1.85)	5.800 (5.94)	3.250 (2.98)	3.172 (1.18)	2.866 (2.28)

[a]The increase in quality score as a result of a 1-point increase in proficiency.

[b]The increase in experience score as a result of a 1-point increase in proficiency.

[c]The numbers 1 to 6 in this row represent the specifications that were run for the model.

[d]The constant is the value for β0 in our model, or when all the variables are estimated at 0.

2. Do Individual Characteristics Matter for mHealth Technology Adoption and Usage?

Age is the only factor that was correlated with CommCare proficiency and usage as seen in Table 8. When we combined illiterate and low literacy users and compared with literate users, we did not find any significant differences in CommCare adoption using both measures.

Age affected CommCare user type negatively, with an increase in age increasing the likelihood of belonging to a lower category of CommCare user type (-0.105, *P*=.08). Age also affected the CommCare proficiency score negatively, with each additional year decreasing the CommCare proficiency score by 0.4 points (*P*=.09), but only when low and middle levels of literacy and education were combined into one variable. Based on these results, we can hypothesize that although illiteracy could influence adoption, in general, lower literacy, education, and previous mobile experience do not affect CommCare adoption, while age can be an influencing factor for adoption.

Further analysis, not presented here, compared CommCare proficiency and usage scores for low-literacy ASHAs and the illiterate ASHA. The illiterate ASHA had lower CommCare proficiency and usage scores, with illiteracy decreasing the CommCare proficiency score by 41% compared to lower literacy. Compared to literate ASHAs, the illiterate ASHA had a CommCare proficiency score that was 51% lower. Both these results were significant at the 95% confidence level. However, as we only had one illiterate ASHA in our sample, who we had observed during multiple home visits, we cannot treat this single ASHA as a category and these results cannot be the basis to conclude that illiteracy affects CommCare proficiency and usage. Hence, these results are not presented in this paper. Table 9 shows the relationship between CommCare adoption and ASHA characteristics.

Table 8. Correlations between CommCare adoption and ASHA characteristics.

Variables[a]	χ^2 or r[b]	*P*
CommCare user type and age, χ^2_2	4.8	.31
CommCare user type and literacy[c], *F*		.54
CommCare user type and education, χ^2_2	2.5	.65
CommCare user type and previous mobile experience, χ^2_2	6.2	.19
CommCare proficiency and age, *r*	-.5137	.06
CommCare proficiency and previous mobile experience, *r*	-.2700	.35
CommCare proficiency and literacy, *F*		.001
CommCare proficiency and education, χ^2_2	6.3	.18

[a]We transformed continuous variables, age and previous mobile experience, into categorical variables in order to test for association with CommCare user type, which is also a categorical variable.

[b]We performed chi-square tests to look for measures of association between categorical variables, and pairwise correlations (*r*) for continuous variables.

[c]We used Fisher's exact test (*F*) for literacy since we only have one observation for illiteracy.

Table 9. Relationship between CommCare adoption and ASHA characteristics.

ASHA characteristics	Specification 1		Specification 2	
	CommCare user type[a] (n=16), mean (t_{15})	*P*	CommCare proficiency[b] (n=14), mean (t_{13})	*P*
Illiterate plus low literacy	0.127 (0.17)	.87	-1.963 (-0.67)	.52
Education (low plus middle)	-0.969 (-1.21)	.23	-2.616 (-0.85)	.42
Previous mobile experience	-0.0968 (-1.24)	.21	-0.139 (-0.47)	.65
Age	-0.105 (-1.77)	.08	-0.402 (-1.90)	.09
_cut1[c]	-5.093 (-2.32)	N/A[d]		
_cut2[c]	-4.055 (-1.92)	N/A		
Constant[e]			25.86 (3.71)	.005

[a]The ordered probit model was applied for this analysis.

[b]Ordinary least-squares (OLS) regression was used for generalized linear modelling.

[c]_cut1 and _cut2 are ancillary parameters and do not have associated *P* values. The coefficients show the estimates for the cutoff points chosen by the model for our categorical dependent variable.

[d]Not applicable (N/A).

[e]The constant is the value for β0 in our model, or when all the variables are estimated at 0.

3. Do Levels of Literacy and Education Affect Quality and Experience of Care Directly?

We did not find any association between literacy levels and quality and experience of care, or education levels and quality and experience of care provided by the ASHAs. Table 10 shows the correlations between the quality and experience of care and ASHA characteristics. The results from Table 11 estimating effects of literacy and education on quality and experience of care also did not show any significant effects of literacy or education on quality and experience of care.

Table 10. Correlations between quality and experience of care and ASHA characteristics.

Correlated variables	χ^{2a}	*P*
Quality and literacy[b], *F*		.32
Experience and literacy, *F*		.14
Observed quality and literacy, *F*		.48
Quality and education, χ^2_2	7.22	.13
Experience and education, χ^2_2	5.33	.26
Observed quality and education, χ^2_2	4.55	.34

[a]Chi-square tests were used to look for measures of association between categorical variables.

[b]Fisher's exact test (*F*) was used for literacy since we only have one observation for illiteracy.

Table 11. The effect of literacy and education levels on quality and experience of care.

ASHA characteristics	Specification 1	Specification 2
	Quality score[a] (n=14), mean (t_{13})	Experience score[a] (n=14), mean (t_{13})
Illiterate plus low literacy	1.085 (0.31)	-0.010 (-0.01)
Education (low plus middle)	-3.849 (-1.06)	-1.906 (-1.08)
Constant[b]	9.783 (3.89)	6.802 (5.57)

[a]Ordinary least-squares (OLS) regression was used for generalized linear modelling.

[b]The constant is the value for β0 in our model, or when all the variables are estimated at 0.

4. Do Accredited Social Health Activist Characteristics Matter for the Relationship Between CommCare Adoption and Quality and Experience of Care?

Most ASHA characteristics also did not seem to affect the relationship between CommCare adoption and quality and experience of care. Literacy level did not seem to affect the quality of the visit by leveraging the way that ASHAs are able to use the technology, and had no direct impact on quality or experience of care. Combining illiterate and low literacy users together, we did not find any significant effect of literacy levels on the relationship between quality/experience and CommCare proficiency and CommCare user type (see Table 12).

However, other analysis (not presented here) showed that predicted values estimating the relationship between CommCare proficiency and quality of care were negatively affected for the illiterate user compared to lower literacy users (-4.95, *P*=.09). Similarly, predicted values estimating the relationship between CommCare proficiency and experience of care were also negatively affected for the illiterate user compared to lower literacy users (-2.815, *P*=.09). These findings indicate that illiteracy does seem to affect the relationship between CommCare proficiency and quality score, as well as the relationship between CommCare proficiency and experience of care. These results are not presented here due to the small sample size of illiterate ASHAs, and should be seen as an informed hypothesis for further study. Table 13 shows the effect of ASHA characteristics on the relationship between CommCare user type and the quality and experience of care.

Age also negatively affected CommCare proficiency and usage. Predicted values estimating the relationship between CommCare proficiency and quality/experience of care were both negatively affected, while there was no such effect for CommCare user type.

Table 12. The effect of ASHA characteristics on the relationship between CommCare proficiency and quality/experience of care.

ASHA characteristics	Specification 1		Specification 2	
	Quality'[a] (n=14), mean (t_{13})	*P*	Experience'[a] (n=14), mean (t_{13})	*P*
Illiterate plus low literacy	-1.063 (-0.67)	.52	-0.605 (-0.67)	.52
Education (low plus middle)	-1.417 (-0.85)	.42	-0.806 (-0.85)	.42
Age	-0.218 (-1.90)	.09	-0.124 (-1.90)	.09
Previous mobile experience	-0.0751 (-0.47)	.65	-0.0427 (-0.47)	.65
Constant[b]	17.17 (4.55)	.001	10.83 (5.05)	.001

[a]The dependent variable is the predicted value from the first model estimating the relationship between CommCare proficiency and quality and experience of care.

[b]The constant is the value for β0 in our model, or when all the variables are estimated at 0.

Table 13. The effect of ASHA characteristics on the relationship between CommCare user type and quality/experience of care.

ASHA characteristics	Specification 1		Specification 2	
	Quality'[a] (n=14), mean (t_{13})	*P*	Experience'[a] (n=14), mean (t_{13})	*P*
Illiterate plus low literacy	0.831 (0.42)	.69	0.227 (0.22)	.83
Education (low plus middle)	-2.270 (-1.12)	.29	-0.793 (-0.74)	.47
Age	-0.230 (-1.59)	.14	-0.113 (-1.50)	.16
Previous mobile experience	-0.143 (-0.70)	.50	-0.147 (-1.38)	.20
Constant[b]	17.02 (3.62)	.004	10.52 (4.25)	.001

[a]The dependent variable is the predicted value from the first model estimating the relationship between commcare proficiency and quality and experience of care.

[b]The constant is the value for β0 in our model, or when all the variables are estimated at 0.

Field Observations

Observations of Non-CommCare Users

We observed six home visits of ASHAs who were not using CommCare. Most visits without CommCare were short and incomplete. The visits focused on the immediate state of the baby or mother, rather than assessing their health since the last visit, and the information provided was targeted only to the current situation of the mother or child. For example, during a visit with a mother and her newborn who had a cough/cold, an ASHA only counseled the mother to take the child to the doctor and did not address other aspects of newborn care. In another instance, we visited a day-old newborn. The ASHA passed on inaccurate information, wrapping up the child in a blanket and advising the mother to hold him against herself, enclosed in a blanket, for skin-to-skin contact. From these visits, we can suggest that CommCare increases the comprehensiveness of home visits, and decreases instances of inaccurate counseling.

Observations of CommCare Users: Stages of Adoption

Low and middle CommCare users tended to use CommCare as a reporting tool by filling out the checklists in the app without providing counseling or elaborating on the messages. For these users, the accuracy of their reporting can be brought into question as they can simply press "yes" for all questions without asking the client for a response. Reporting "yes" for everything would bring down an ASHA's CommCare proficiency and visit quality scores, because certain questions, counseling prompts, and videos are only displayed based on the answers input for previous questions.

High users of CommCare seemed to understand the design and purpose of CommCare, and were able to move beyond using CommCare as a reporting tool. They tended to elaborate on the messages and provide more thorough counseling to the client. This may be because users first concentrated their efforts on learning how to use the app, and then moved on to grasping the content and design. ASHAs first focused on using CommCare as a reporting tool before using it as a job aid to support them during home visits, to plan their schedules, and provide targeted information to the clients. Using CommCare appropriately increased the quality of home visits, the accuracy of information passed on to the clients, and the experience of home visits for the client.

For the middle users, we believe that CommCare can play an interesting role in increasing the quality and experience of care. CommCare in Bihar had been deployed for around 6 months at the time of this study. The middle users were mostly focused on using CommCare as a reporting tool. At this stage, the multimedia in CommCare plays a key role in increasing the quality of the visit. Playing the audio and video and showing images means that the client is getting some information about

a comprehensive range of topics, even if the ASHA is not providing any counseling.

If the low and middle users, initially providing low-quality care, do not use the media in CommCare, the quality of the home visit will remain poor, as the ASHA is entirely focused on filling out the checklist, sometimes without asking the client any questions. In this case, the quality of the home visit is much higher without CommCare, as in that case the ASHA would at least pass on some messages to the client, however incomplete.

The reasons we uncovered for low CommCare use are idiosyncratic, depending on attitude, training, literacy, and age/health. Of these, attitude and training seemed to be the most influential, and it is most likely that high users were providing a high level of care without CommCare, and the low users were providing a low quality of care without CommCare. Our observations showed us that the high users were more dynamic, confident, and capable in their roles as ASHAs, while low users were less dynamic and comfortable in their roles as ASHAs. However, their attitudes and dynamism did not seem to be significantly correlated with education or literacy. Although our research tools did not include any way to measure attitude, it is possible to discern the variety of attitudes present from the ASHAs' behavior. For example, two ASHAs who were low users of CommCare refused to perform a home visit for the study, though they were literate and had higher levels of education.

Discussion

Limitations of the Study

Our formative study had some important limitations. First, our sample size was too small to establish conclusive evidence to describe the relationships we were testing. Second, we only studied differences between using mHealth tools and not using mHealth tools for quality and experience of care qualitatively via field observations. Third, social and program factors that can affect technology adoption and quality and experience of care directly and via the technology were not analyzed and were outside the scope of the study. Lastly, simply observing the ASHAs' home visits could have had an effect on performance leading to biased indicators. Observing the ASHAs can potentially induce better performance due to the perception of a supervisor or outsider being present, or can induce worse performance due to a feeling of pressure or nervousness.

Conclusions

Based on our findings, we identified two levels of CommCare adoption. The first level of adoption was where the ASHAs were still learning the design and content of the technology, and used it as a tool for reporting. The second level of adoption was where they were more proficient in using CommCare, understood how the tool is designed, and used it appropriately as a job aid for reporting, as well as for counseling during home visits. ASHAs in the first stage of adoption had lower quality and experience of home visits, compared to those in the second stage of adoption. Though the causality from proficiency to adoption is not clear, and adoption affects proficiency and vice versa, it was demonstrated that higher proficiency leads to higher adoption, and to quality and experience of care. Individual characteristics, other than illiteracy, did not seem to affect proficiency nor adoption, and further research is required to reach concrete conclusions about the effects of illiteracy on proficiency and adoption of mHealth tools.

A higher level of CommCare adoption was significantly associated with higher quality and experience of care, although it is possible that these users were already providing higher quality and experience of care. While individual characteristics, including education and previous mobile experience, studied here did not affect the stages of adoption nor the quality or experience of home visits, illiteracy can affect the quality and experience of care by influencing CommCare adoption, as can the way ASHAs leverage the technology to provide care. Using multimedia effectively was more prominent in those that displayed higher levels of CommCare adoption. Low literacy users were still able to use mHealth technology to provide higher quality and experience of care, however, illiterate users do need more support and training to understand the design and workflow of mHealth apps, and to accrue the benefits of the technology.

The small sample size in our study means that our results should be taken as informed hypotheses for further study. The relationship between levels of mHealth technology adoption and quality and experience of care can be established with a larger sample size using the methods presented in this paper. Any further research should include a reliable sample of literate, lower-literacy, and illiterate users to test the model for mHealth technology adoption and quality and experience of care presented in this paper.

Conflicts of Interest

None declared.

References

1. Qiang CZ, Yamamichi M, Hausman V, Altman D. Mobile Applications for the Health Sector. Washington, DC: The World Bank, ICT Sector Unit; 2011 Dec. URL: http://siteresources.worldbank.org/INFORMATIONANDCOMMUNICATIONANDTECHNOLOGIES/Resources/mHealth_report.pdf [accessed 2015-05-19] [WebCite Cache ID 6YePtUX0o]
2. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]

3. Braveman P, Gruskin S. Defining equity in health. J Epidemiol Community Health 2003 Apr;57(4):254-258 [FREE Full text] [Medline: 12646539]
4. Witmer A, Seifer SD, Finocchio L, Leslie J, O'Neil EH. Community health workers: integral members of the health care work force. Am J Public Health 1995 Aug;85(8 Pt 1):1055-1058. [Medline: 7625495]
5. van Biljon J, Kotzé P. Modelling the factors that influence mobile phone adoption. In: Proceedings of the 2007 Annual Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries. New York, NY: ACM; 2007 Oct 02 Presented at: The 2007 Annual Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries; October 2-3, 2007; Port Elizabeth, South Africa p. 152-161. [doi: 10.1145/1292491.1292509]
6. Kumar M, Raj A. Development in Bihar: predicaments and prospects of health indices. J Health Manag 2013 Sep 16;15(3):415-430. [doi: 10.1177/0972063413492037]
7. Census of India. 2011. Office of the Registrar General and Census Commissioner, Ministry of Home Affairs, Government of India URL: http://www.censusindia.gov.in/ [accessed 2015-05-19] [WebCite Cache ID 6YeQGZR8w]
8. Annual Health Survey Bulletin 2012-2013, Bihar.: Census of India; 2013. URL: http://www.censusindia.gov.in/vital_statistics/AHSBulletins/AHS_Bulletin_2012_13/Bihar/Bulletin_Bihar.pdf [accessed 2015-05-20] [WebCite Cache ID 6YfYOPpjR]
9. International Institute for Population Sciences, Government of India. Maharashtra, India: International Institute for Population Sciences; 2006. National Family Health Survey-3 URL: http://www.rchiips.org/nfhs/nfhs3.shtml [accessed 2015-05-20] [WebCite Cache ID 6YfYnmz4P]
10. Raju PK, Kamal M, Kumari K, Kumar A. District Health Society Saharsa, District Health Action Plan 2012-2013. Patna, India: State Health Society, Bihar; 2014. URL: http://www.statehealthsocietybihar.org/pip2012-13/districthealthactionplan/Saharsa.pdf [accessed 2015-05-20] [WebCite Cache ID 6YfZAU57n]
11. Evaluation Report on Integrated Child Development Services (ICDS). New Delhi, India: Programme Evaluation Organisation, Planning Commission, Government of India; 2011 Mar. URL: http://www.planningcommission.nic.in/reports/peoreport/peoevalu/peo_icds_v1.pdf [accessed 2015-05-20] [WebCite Cache ID 6YfZYCCto]
12. Bihar: Roadmap for Development of Health Sector. A Report of the Special Task Force on Bihar. New Delhi, India: Planning Commission, Government of India; 2007 Aug. URL: http://www.planningcommission.gov.in/aboutus/taskforce/tsk_bhs.pdf [accessed 2015-05-21] [WebCite Cache ID 6YhMi0r8X]
13. Malik G. Role of auxiliary nurse midwives in National Rural Health Mission. Nurs J India 2009 Apr;100(4):88-90. [Medline: 20527493]
14. Gosavi SV, Raut AV, Deshmukh PR, Mehendale AM, Garg BS. Journal of Mahatma Gandhi Institute of Medical Sciences. 2011. ASHAS' awareness and perceptions about their roles and responsibilities: a study from rural Wardha URL: http://medind.nic.in/jaw/t11/i1/jawt11i1p33.pdf [accessed 2015-05-19] [WebCite Cache ID 6YeQhn7Zk]
15. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752. [Medline: 16702590]
16. Anonymous. Cell phones and CHWs: a transformational marriage? Glob Health Sci Pract 2014 Feb;2(1):1-2 [FREE Full text] [doi: 10.9745/GHSP-D-14-00007] [Medline: 25276556]
17. Treatman D, Lesh N. Mobile for Development. 2012. Strengthening community health worker performance through localized multimedia URL: http://hingx.org/Share/Attachment/1040?fileName=Experience - m4d2012-treatman.pdf [accessed 2015-05-21] [WebCite Cache ID 6YhQAfkoz]
18. Rogers EM. A prospective and retrospective look at the diffusion model. J Health Commun 2004;9 Suppl 1:13-19. [doi: 10.1080/10810730490271449] [Medline: 14960401]
19. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478.
20. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319-340.
21. Phan K, Daim T. Exploring technology acceptance for mobile services. Journal of Industrial Engineering and Management 2011 Jul 14;4(2). [doi: 10.3926/jiem.2011.v4n2.p339-360]
22. Zhou T, Lu Y, Wang B. Integrating TTF and UTAUT to explain mobile banking user adoption. Comput Human Behav 2010 Jul;26(4):760-767. [doi: 10.1016/j.chb.2010.01.013]
23. Ward R. The application of technology acceptance and diffusion of innovation models in healthcare informatics. Health Policy and Technology 2013 Dec;2(4):222-228.
24. Hyyppä K, Tamminen S, Hautala I, Repokari L. The effect of mental model guiding users' action in mobile phone answering situations. In: Electronical Proceedings of the 1st Nordic Conference on Computer Human Interaction. 2000 Presented at: 1st Nordic Conference on Computer Human Interaction; October 23-25, 2000; Stockholm, Sweden.
25. Lehmann U, Sanders D. Community Health Workers: What do we Know About Them?. Geneva, Switzerland: World Health Organization; 2007 Jan. URL: http://www.who.int/hrh/documents/community_health_workers.pdf [accessed 2015-05-19] [WebCite Cache ID 6YeR6ohOx]

26. Gelderblom H, van Dyk T, van Biljon J. Mobile phone adoption: do existing models capture the actual usage of older adults? In: Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists. New York, NY: ACM; 2010 Presented at: 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists; October 11-13, 2010; Bela Bela, South Africa p. 67-74. [doi: 10.1145/1899503.1899511]
27. Renaud K, van Biljon J. Predicting technology acceptance and adoption by the elderly: a qualitative study. In: Proceedings of the 2008 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries: Riding the Wave of Technology. 2008 Presented at: 2008 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries: Riding the Wave of Technology; October 6-8, 2008; Wilderness, South Africa p. 210-219. [doi: 10.1145/1456659.1456684]
28. van Biljon J, Renaud K. A qualitative study of the applicability of technology acceptance models to senior mobile phone users. In: Song I, Piattini M, Chen YP, Hartmann S, Grandi F, Trujillo J, et al, editors. Advances in Conceptual Modeling – Challenges and Opportunities: ER 2008 Workshops CMLSA, ECDM, FP-UML, M2AS, RIGiM, SeCoGIS, WISM, Barcelona Spain, October 20-23, 2008. Berlin, Germany: Springer Berlin Heidelberg; 2008:228-237.

Abbreviations

ANM: Auxiliary Nurse Midwife
APHC: additional primary health center
ASHA: Accredited Social Health Activist
AWW: Aganwadi Worker
CHC: community health center
CHW: community health worker
HSC: health sub-center
ICDS: Integrated Child Development Services
IRB: Internal Review Board
MOH: Ministry of Health
N/A: not applicable
NRHM: National Rural Health Mission
OLS: ordinary least squares
PHC: primary health center
RCT: randomized controlled trial
SMS: short message service

Edited by G Eysenbach; submitted 15.11.14; peer-reviewed by C Bond; comments to author 21.02.15; revised version received 12.03.15; accepted 14.03.15; published 28.05.15.

Please cite as:
Kaphle S, Chaturvedi S, Chaudhuri I, Krishnan R, Lesh N
Adoption and Usage of mHealth Technology on Quality and Experience of Care Provided by Frontline Workers: Observations From Rural India
JMIR mHealth uHealth 2015;3(2):e61
URL: http://mhealth.jmir.org/2015/2/e61/
doi:10.2196/mhealth.4047
PMID:26023001

Original Paper

Mobile Access to ClinicalConnect: A User Feedback Survey on Usability, Productivity, and Quality

Bell Raj Eapen[1*], MD, MSc; Barbara Chapman[1*], BSc(N), MSc

HNHB eHealth Office, Hamilton, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Bell Raj Eapen, MD, MSc
HNHB eHealth Office
Suite 108
293 Wellington Street North
Hamilton, ON, L8L8E7
Canada
Phone: 1 9058700577
Fax: 1 9055778260
Email: eapenb@hhsc.ca

Abstract

Background: ClinicalConnect, a federated clinical viewer for South West Ontario, Canada, launched a mobile interface in June 2012.

Objective: The aim of the study was to assess usability of the mobile interface and the perceived impact on productivity of health care providers and quality of healthcare delivery.

Methods: A survey was conducted using the System Usability Scale (SUS) and questionnaires designed to measure productivity and quality based on Canada Health Infoway's Benefits Evaluation framework.

Results: The mean SUS score was 67 based on 77 responses. The mean scores for productivity and quality were 3.37 (N=74) and 3.62 (N=71), respectively, on a 5-point Likert scale where 3 was neutral.

Conclusions: Users perceived the mobile interface of ClinicalConnect as useful but were neutral about the ease of use.

(JMIR mHealth uHealth 2015;3(2):e35) doi:10.2196/mhealth.4011

KEYWORDS

mHealth; health information exchange; ClinicalConnect

Introduction

Mobile devices are rapidly becoming a part of everyday life as both communication and information tools. Recent studies indicate that this is true in the case of health care providers as well. Health care providers often use their mobile devices to access educational material, pharmaceutical compendiums, practice guidelines, clinical pathways, and electronic medical records [1].

Electronic health information exchange (HIE), in which patients' clinical data is efficiently shared between care delivery settings, is expected to produce a number of quality improvements and cost savings [2]. ClinicalConnect is a federated clinical viewer that provides health information exchange between health services and care providers of South West Ontario [3]. It is considered to be one of the largest federated HIE models in North America. A mobile version of ClinicalConnect was developed and deployed in June 2012.

Health care providers often fail to realize the intended effects of their eHealth systems due to inadequate usability [4]. Despite the widely held belief that the computerization of health information systems contributes to improved quality of patient care and care management, studies have found medical staff attitudes toward computerization to be negative [5]. In order to discern whether the ClinicalConnect mobile version was meeting user requirements, an evaluation survey was conducted in November 2013. A range of survey tools was used for the assessment with a focus on usability, and the perceived impact of the mobile version on the quality of patient care and productivity of health care providers.

Methods

Tools

The System Usability Scale (SUS), a nonproprietary validated survey tool [6], was used to assess the usability of the ClinicalConnect mobile interface. SUS includes 10 statements presented on a 5-point Likert scale, which results in an overall score from 0 to 100 that indicates the perceived usability of the interface.

Canada Health Infoway published a Benefits Evaluation Indicators Technical Report in 2006 that was subsequently updated in 2012 to provide guidance for benefits evaluation planning related to information and communications technology (ICT) in health care [7]. The framework proposes the indicators that could have an impact on productivity and quality of health care. Based on these indicators, productivity and quality questionnaires using a 5-point Likert scale [8] were designed. Productivity and quality questionnaires had 8 (Table 1) and 10 questions (Table 2), respectively.

The SUS, productivity, and quality survey tools were hosted and administered using SurveyMonkey [9].

Participants

Five hundred and four health care providers who expressed interest in participating in periodic evaluations while signing up for the HIE service were sent an e-mail with a link to the survey. A second reminder e-mail was sent if no response was received from a provider. One hundred and ten providers responded to the survey, though some surveys were incomplete. Seventy-seven responses were obtained for SUS. On productivity and quality scores, we received 74 and 71 responses, respectively.

Analysis

Results of the SUS questionnaire were recoded and normalized, and the mean SUS score and the standard deviation were then recorded. The Likert-type questions for productivity and quality were independently tabulated. Since all questions within each category measure a single concept, the values were combined into a composite score by calculating the mean and standard deviation. Scores for negatively worded questions (fifth question in the productivity matrix and second and third questions in the quality matrix) were normalized prior to the calculation of the composite score.

Results

Eighty-five (77%) of the respondents were physicians. The mean SUS score was 67 (SD 14.4) with a percentile score of 46.9 [10]. This means that the mobile version of ClinicalConnect can be considered more usable than 46.9% of all products evaluated with the SUS instrument.

Table 1 depicts the responses to the productivity-related questions and Table 2 summarizes the quality questions. The mean productivity score was 3.37 (SD 1.06) and the mean quality score was 3.62 (SD 0.99), with 3 being neutral on the 5-point Likert scale.

Table 1. Productivity questions and summary of responses (N=74).

Answer options	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
	n (%)	n (%)	n (%)	n (%)	n (%)
Accessing patient/client test results and information on my mobile device has decreased time spent tracking down or waiting for these reports	15(20)	27(36)	17(23)	13(18)	2(3)
Using a mobile device allows me to spend more face-to-face time with my patient/client	5(7)	19(26)	31(42)	18(24)	1(1)
I have more portability as I can now access ClinicalConnect with a mobile device	19(26)	47(64)	4(5)	3(4)	1(1)
The “New Results” feature draws my attention to results quickly	10(14)	34(46)	22(30)	6(8)	2(3)
It is difficult to view patient information on my mobile device screen	6(8)	21(28)	22(30)	21(28)	4(5)
I can access information faster on my mobile device than on a desktop computer	3(4)	18(24)	20(27)	17(23)	16(22)
Treatment decisions are made faster now that I can access patient/client information on my mobile device	8(11)	20(27)	32(43)	14(19)	0(0)
Access to patient information on my mobile device allows for better communication between health care providers	14(19)	31(42)	22(30)	7(9)	0(0)

Table 2. Quality questions and summary of responses (N=71).

Answer options	Strongly agree	agree	Neither agree nor disagree	Disagree	Strongly disagree
	n (%)	n (%)	n (%)	n (%)	n (%)
Education activities are enhanced when the patient/client can view their results on my mobile device with me	7(10)	16(23)	33(46)	14(20)	1(1)
There is high risk that a mobile device can be the source of a nosocomial infection	3(4)	11(15)	24(34)	24(34)	9(13)
Current results are not easy to access on a mobile device	5(7)	16(23)	16(23)	25(35)	9(13)
A mobile device allows faster access to vital patient information facilitating quicker consultation, diagnostic tests, and interventions	14(20)	29(41)	17(24)	11(15)	0(0)
Access to patient/client information anywhere, anytime enhances consultations, referrals, and handoffs	17(24)	38(54)	10(14)	5(7)	1(1)
I am more confident that my mobile device is cleaner than the desktop computer that is used by multiple people	11(15)	24(34)	23(32)	8(11)	5(7)
I can better prioritize my actions to follow-up on test results with the “New Results” feature	6(8)	29(41)	31(44)	4(6)	1(1)
I feel more confident in care decisions because I have the information I need at my fingertips	15(21)	30(42)	22(31)	3(4)	1(1)
I am less likely to order a duplicate test because I have easier access to current results	15(21)	36(51)	16(23)	3(4)	1(1)
HCPs should have access to their patient’s/client’s health information no matter where they are	33(46)	29(41)	7(10)	2(3)	0(0)

Discussion

HIE is the process of sharing electronic health information between different providers and organizations. As in most other health information systems, mobile devices are increasingly becoming popular as an HIE platform [11].

The SUS consists of 10 alternating positively and negatively worded statements scored on a 5-point Likert scale. The mean SUS score for 2324 surveys about usability conducted over a 10-year period was 70.14 with a median score of 75 [6].

Inadequate usability is a major cause of failure of eHealth systems [4], especially of mobile platforms. Our data show that health service providers perceive the mobile interface of ClinicalConnect as useful, but are neutral about the ease of use. This pattern has been noticed before in other mHealth interventions [12]. The ease of use can be affected by factors beyond mobile user interface such as ergonomic and social aspects [13].

HIE is vital for improving efficiency and quality of health care. It has been demonstrated that perceived usefulness is a stronger predictor of the use of an eHealth technology than the perceived ease of use [14]. Most of our respondents agreed that ClinicalConnect on mobile devices had a positive impact on their productivity and quality. Anywhere, anytime access to patient information on mobile devices was perceived as an important factor in faster consultations, referrals, and handoffs, and in improved communication between health care providers. However, information access on mobile devices was not considered faster than the desktop counterpart, and there was no consensus on whether mobile devices allow more face-to-face time with patients. This corroborates previous studies that showed that physicians consider improvements in the quality domain as the overarching benefit of HIE [15].

Though HIE programs have demonstrated clinical value in some situations such as emergency departments [16], a consistent empirical proof of value is lacking [17]. In a study, emergency physicians reported workflow disruptions from HIE use [18]. Our study suggests that mobile access to integrated health records is perceived as beneficial, especially to the quality of patient care. However, our sample size may be insufficient and not representative enough of the health care roles for us to statistically draw a precise conclusion.

It is important for HIE system implementations to be integrated into the care practice improvement process [4]. The study shows that bringing data from disparate health care systems to the point of care via mobile HIE systems has a perceived potential for improvements in patient care.

Acknowledgments

Funding for this ClinicalConnect mobile project and Benefits Evaluation as well as funding for the publication costs was provided by Canada Health Infoway Inc, a non-profit corporation funded by the Government of Canada. The funders had no role in study design, data collection and analysis, or decision to publish the manuscript. In-kind services support was provided by the Hamilton

Niagara Haldimand Brant (HNHB) eHealth Office and associated health care providers. The authors would like to acknowledge the support of Dale Anderson, John Haywood, Debra Hutchinson, Rob Lloyd, and Kate Black.

ClinicalConnect is a trademark of the public hospital Hamilton Health Sciences. ClinicalConnect is jointly funded by public hospitals and agencies as well as eHealth Ontario.

Conflicts of Interest

The study was conducted by the HNHB eHealth Office/Hamilton Health Sciences, the regional solution provider deploying ClinicalConnect regionally, and the service delivery partner for the HNHB local health integration network (LHIN).

References

1. Sclafani J, Tirrell TF, Franko OI. Mobile tablet use among academic physicians and trainees. J Med Syst 2013 Feb;37(1):9903 [FREE Full text] [doi: 10.1007/s10916-012-9903-6] [Medline: 23321961]
2. Lammers EJ, Adler-Milstein J, Kocher KE. Does health information exchange reduce redundant imaging? Evidence from emergency departments. Med Care 2014 Mar;52(3):227-234. [doi: 10.1097/MLR.0000000000000067] [Medline: 24374414]
3. ClinicalConnect Connecting South West Ontario. URL: http://info.clinicalconnect.ca/CC/ [accessed 2014-10-01] [WebCite Cache ID 6T0eG4pBi]
4. Scandurra I, Hägglund M, Persson A, Ahlfeldt RM. Disturbing or facilitating?--on the Usability of Swedish eHealth Systems 2013. Stud Health Technol Inform 2014;205:221-225. [Medline: 25160178]
5. Kagan I, Fish M, Farkash-Fink N, Barnoy S. Computerization and its contribution to care quality improvement: the nurses' perspective. Int J Med Inform 2014 Dec;83(12):881-888. [doi: 10.1016/j.ijmedinf.2014.08.001] [Medline: 25176353]
6. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
7. Lau F. Extending the infoway benefits evaluation framework for health information systems. Stud Health Technol Inform 2009;143:406-413. [Medline: 19380969]
8. Komorita SS. Attitude content, Intensity, and the neutral point on a Likert Scale. J Soc Psychol 1963 Dec;61:327-334. [doi: 10.1080/00224545.1963.9919489] [Medline: 14084811]
9. SurveyMonkey: Free online survey software & questionnaire tool. URL: https://www.surveymonkey.com/ [accessed 2014-10-01] [WebCite Cache ID 6T0eTf3Tm]
10. Kalz M, Lenssen N, Felzen M, Rossaint R, Tabuenca B, Specht M, et al. Smartphone apps for cardiopulmonary resuscitation training and real incident support: a mixed-methods evaluation study. J Med Internet Res 2014;16(3):e89 [FREE Full text] [doi: 10.2196/jmir.2951] [Medline: 24647361]
11. Bouri N, Ravi S. Going mobile: how mobile personal health records can improve health care during emergencies. JMIR Mhealth Uhealth 2014;2(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3017] [Medline: 25098942]
12. Vélez O, Okyere PB, Kanter AS, Bakken S. A usability study of a mobile health application for rural Ghanaian midwives. J Midwifery Womens Health 2014;59(2):184-191 [FREE Full text] [doi: 10.1111/jmwh.12071] [Medline: 24400748]
13. Svanaes D, Alsos OA, Dahl Y. Usability testing of mobile ICT for clinical settings: methodological and practical challenges. Int J Med Inform 2010 Apr;79(4):e24-e34. [doi: 10.1016/j.ijmedinf.2008.06.014] [Medline: 18789753]
14. Davis FD. Perceived Usefulness, Perceived Ease of Use,User Acceptance of Information Technology. MIS Q. Society for Information Management and The Management Information Systems Research Center 1989;13(3):319. [doi: 10.2307/249008]
15. Khurshid A, Diana ML, Luce SD. Health information exchange: metrics to address quality of care and return on investment. Perspect Health Inf Manag 2012;9:1e [FREE Full text] [Medline: 22783153]
16. Tzeel A, Lawnicki V, Pemble KR. "Hidden" value: how indirect benefits of health information exchange further promote sustainability. Am Health Drug Benefits 2012 Sep;5(6):333-341 [FREE Full text] [Medline: 24991331]
17. Overhage JM, Evans L, Marchibroda J. Communities' readiness for health information exchange: the National Landscape in 2004. J Am Med Inform Assoc 2005;12(2):107-112 [FREE Full text] [doi: 10.1197/jamia.M1680] [Medline: 15561785]
18. Thorn SA, Carter MA, Bailey JE. Emergency physicians' perspectives on their use of health information exchange. Ann Emerg Med 2014 Mar;63(3):329-337. [doi: 10.1016/j.annemergmed.2013.09.024] [Medline: 24161840]

Abbreviations

HIE: health information exchange
HNHB: Hamilton Niagara Haldimand Brant
ICT: information and communications technology
LHIN: local health integration network
SUS: System Usability Scale

Edited by G Eysenbach; submitted 06.11.14; peer-reviewed by C Dias, M Kalz; comments to author 20.02.15; revised version received 27.02.15; accepted 23.03.15; published 14.04.15.

Please cite as:
Eapen BR, Chapman B
Mobile Access to ClinicalConnect: A User Feedback Survey on Usability, Productivity, and Quality
JMIR mHealth uHealth 2015;3(2):e35
URL: http://mhealth.jmir.org/2015/2/e35/
doi:10.2196/mhealth.4011
PMID:25877226

Original Paper

Selection and Pilot Implementation of a Mobile Image Viewer: A Case Study

Christine Marie Zwart[1], PhD; Miao He[2], PhD; Teresa Wu[2], PhD; Bart M Demaerschalk[3], MD; Joseph Ross Mitchell[1], PhD; Amy K Hara[1], MD

[1]Medical Imaging Informatics, Department of Radiology, Mayo Clinic, Scottsdale, AZ, United States

[2]Department of Industrial Engineering, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, United States

[3]Neurocritical Care, Department of Neurology, Mayo Clinic, Phoenix, AZ, United States

Corresponding Author:
Christine Marie Zwart, PhD
Medical Imaging Informatics
Department of Radiology
Mayo Clinic
13400 E Shea Boulevard
Scottsdale, AZ, 85259
United States
Phone: 1 (480) 301 6988
Fax: 1 (480) 301 4303
Email: zwart.christine@mayo.edu

Abstract

Background: For health care providers, mobile image viewing increases image accessibility, which could lead to faster interpretation/consultations and improved patient outcomes.

Objective: We explored the technical requirements and challenges associated with implementing a commercial mobile image viewer and conducted a small study testing the hypothesis that the mobile image viewer would provide faster image access.

Methods: A total of 19 clinicians (9 radiologists, 3 surgeons, 4 neurologists, and 3 physician assistants) evaluated (1) a desktop commercial picture archiving and communication system (PACS) viewer, (2) a desktop viewer developed internally over 20 years and deployed throughout the enterprise (ENTERPRISE viewer) and (3) a commercial Food and Drug Administration class II-cleared mobile viewer compatible with Web browsers, tablets, and mobile phones. Data were collected during two separate 7-day periods, before and after mobile image viewer deployment. Data included image viewer chosen, time to view first image, technical issues, diagnostic confidence, and ease of use.

Results: For 565 image-viewing events, ease of use was identical for PACS and mobile viewers (mean 3.6 for all scores of a possible 4.0), and significantly worse for the enterprise viewer (mean 2.9, *P*=.001). Technical issues were highest with the enterprise viewer (26%, 56/215) compared with the mobile (7%,19/259, *P*=.001) and PACS (8%, 7/91, *P*=.003) viewers. Mean time to first image for the mobile viewer (2.4 minutes) was significantly faster than PACS (12.5 minutes, *P*=.001) and the enterprise viewer (4.5 minutes, *P*=.001). Diagnostic confidence was similar for PACS and mobile viewers and worst for enterprise viewer. Mobile image viewing increased by sixfold, from 14% (37/269, before the deployment) to 88.9% (263/296, after the deployment).

Conclusions: A mobile viewer provided faster time to first image, improved technical performance, ease of use, and diagnostic confidence, compared with desktop image viewers.

(JMIR mHealth uHealth 2015;3(2):e45) doi:10.2196/mhealth.4271

KEYWORDS

mHealth; Pilot projects; Radiology; Telemedicine; Teleradiology

Introduction

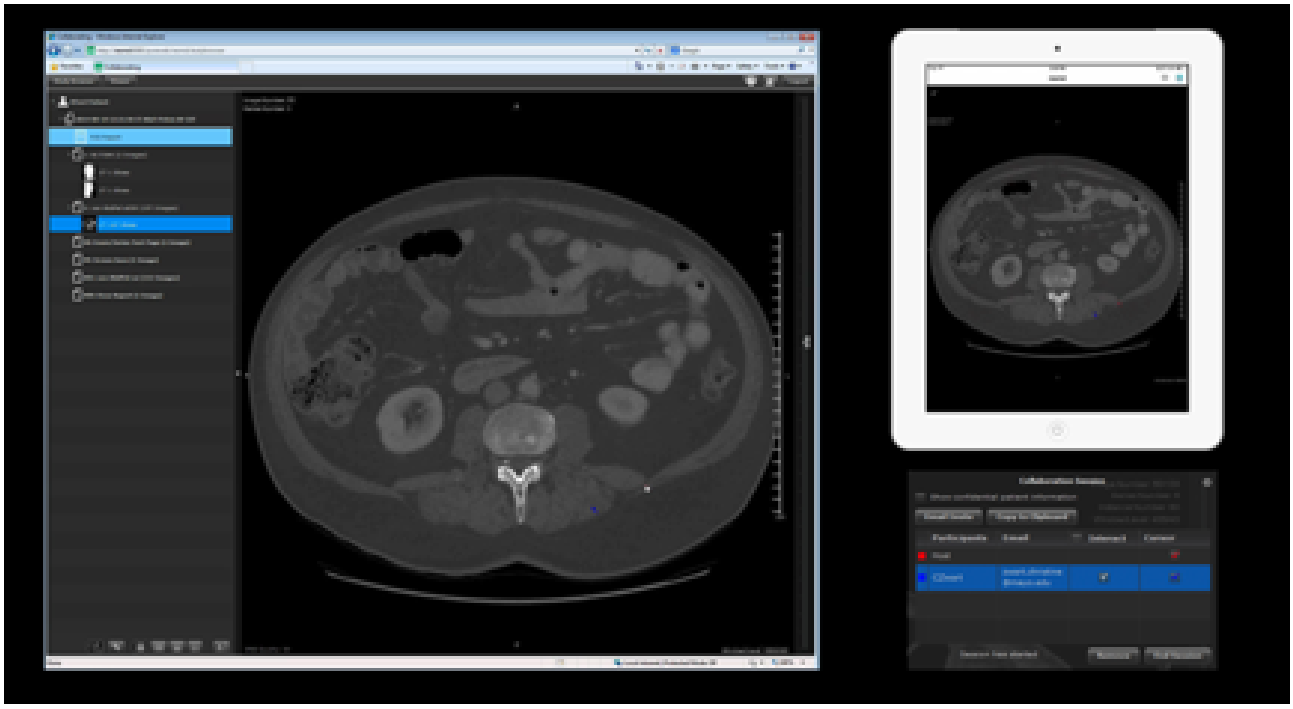
From educating to ordering tests, reporting, consulting, rounding, and sharing with patients, innovations in mobile technology are enhancing the way medicine is practiced in the 21st century. Emphasis has been placed on the availability of apps with the potential to benefit radiology residents [1,2] and staff [3,4]. The suggested toolbox of apps can contain eBooks, medical journals, note-taking apps, cloud data services, and audience polling tools, for example. The selection of a mobile image viewing solution, however, is much more complicated, requiring integration with picture archiving and communication systems (PACSs) and radiology information management systems, establishing secure logins, etc. As a result, incorporating mobile image viewers into the clinical context has been somewhat slow. Implementation has generally occurred for specific urgent care settings such as stroke or emergency medicine, rather than for general radiology use [5,6]. In this paper, we explored the technical requirements and challenges associated with implementing a commercial mobile image viewer and conducted a small study to test the hypothesis that the mobile image viewer would provide faster image access.

Methods

Selection

Several apps for diagnostic reading are currently available. Székely et al [3] identified 11 including ResolutionMD (ResMD, Calgary Scientific, Calgary, Canada) and Centricity Radiology Mobile Access and Siemens syngo.via WebViewer, now "ResolutionMD Mobile" (white-labeled versions of ResMD) [3]. The selection of ResMD for our pilot project was based on several factors, including the following: (1) Food and Drug Administration (FDA) class II clearance for diagnostic reads of all digital imaging and communication in medicine (DICOM) 3.0 imaging modalities (except mammography) on desktop Web browsers, iOS, and specific Android devices. FDA class II clearance indicates that the platform may be marketed for diagnostic use when traditional PACS workstations are not readily available. Each specific software, mobile device, and modality combination requires explicit class II clearance. (2) Vendor agnostic PACS connections—multiple PACS can be configured to connect simultaneously. (3) A client-server architecture with a "zero-footprint" client-side implementation. (4) Health Information Portability and Accountability Act compliance. In brief, this indicates that unencrypted protected health information is not stored on the mobile device after a viewing session terminates. (5) The ability to open specific series and examinations using a systematically built URL. The ability to systematically build URLs that simultaneously open the viewer and navigate to a specific series or examination has several utilities in research, conferences, and education. (6) Built-in features for advanced virtual collaboration. Advanced virtual collaboration utilities enable users to send an invitation link through email and have their viewing session streamed to both their device and that of the collaborators. All collaborators can scroll through images, window, and level, and point out features with visible cursors (Figure 1). (7) License costs were covered with a research agreement with Calgary Scientific, Inc (CSI). Additional information on the system architecture is provided in a previous study [6].

Figure 1. Overview of the ResMD interface and collaboration feature. A single ResMD session running simultaneously: in a desktop web browser (left) and as an application on an Apple iPad (top right). Interactions performed in either view are displayed interactively on both. Collaboration sessions are controlled via a collaboration window (bottom right). The collaboration window allows the "controller" to send invitations (via email), monitor who has joined the session, control the degree of interaction allowed by each user, and limit the display of patient information.



Rationale

Those desiring a mobile image viewer included radiologists who are frequently on call, especially neuroradiologists and interventional radiologists, and wanted a faster and more reliable method to review cases for imaging consulting if not at the hospital or at home. In addition to radiologists, the stroke team, which included neurologists and neurosurgeons, desired to have rapid access to view acute head computed tomography scan images. The primary perceived benefit of a dedicated mobile image viewer was more rapid image access, which allows for faster communication of imaging findings, more rapid formation of treatment plans, leading to better outcomes and lower patient care costs.

The goal of our pilot implementation of a mobile image viewer was to collect data to determine whether long-term employment of such a mobile technology was warranted. We focused on quantifying the potential speed advantages of a mobile image viewing option and collecting clinician feedback on viewer performance and preferences.

Institutional Review

We obtained Institutional Review Board approval to evaluate the chosen mobile image viewer (MOBILE) in comparison to our GE PACS workstations and a desktop viewer developed internally over 20 years and deployed throughout the enterprise (ENTERPRISE) [7].

Institutional policy mandating radiologists' use of PACS during work hours (7 am to 6 pm weekdays, excluding holidays) was not modified for this study; radiologists recorded image access data after work, when any viewer could be used. All cases were officially interpreted and dictated by radiologists using the PACS workstation and digital dictation system. If preliminary reports were given by a radiologist using a mobile device, these were subsequently reviewed by the same radiologist on PACS for final interpretation. Any discrepancies noted by the dictating radiologist between the initial non-PACS and final interpretations were recorded.

Evaluation

Our initial evaluation team of 19 clinicians included 9 radiologists (5 interventional radiologists and 4 neuroradiologists), 4 neurologists (all vascular neurologists and neurohospitalists), 3 physician assistants (2 orthopedic and 1 radiology), and 3 surgeons (2 neurosurgeons and 1 orthopedic surgeon). All participants had over 2 years' experience with institutional desktop image viewers. The 4 neurologists collectively had viewed fewer than 50 cases using the mobile image viewer (as part of an independent telestroke pilot study). None of the other 15 clinicians had prior experience with the mobile image viewer. Before the implementation of the mobile image viewer, mobile access to images was accomplished using screen-sharing or remote-desktop apps.

Users manually recorded their radiology image viewer activity during two separate periods for 7 days each (before and after mobile viewer implementation) on a standardized datasheet. Preimplementation baseline data collection occurred when users could choose only between the two desktop viewers. Postimplementation data collection occurred 3 months after the implementation of the mobile viewer, when users could choose among the PACS desktop, ENTERPRISE, or a mobile viewer. Data collection focused on which of the 3 viewers was selected most often. Because institutional policy mandated the use of PACS during work hours (7 am to 6 pm weekdays, excluding holidays), radiologists recorded image access data after work, when any viewer could be used. Nonradiologist clinicians typically do not have access to PACS workstations, and thus, recorded all image access events both during and after work.

Data recorded for each image access event included date, time, location (inside or outside the hospital), device used (mobile or desktop), system used (PACS, ENTERPRISE, or mobile), time to first image, purpose of image access, and technical issues. For the self-selected viewer system, diagnostic confidence and ease of use were graded on a Likert-type scale (1, poor; 2, fair; 3, good; and 4, excellent). For each image event, data were recorded only for the chosen viewer. The same examination was not evaluated using other viewers.

The time to first image was recorded because it does not vary by examination type or complexity. It was defined as the time from when the clinician received a request to review images (ie, verbally or via a text page) to when the first image appeared on the screen. For desktop viewers, it included the time required to get to a workstation (including drive time, if necessary), to login to the workstation, and to display the first image. For the mobile viewer, it included time to login to the virtual private network, to launch the app, and to display the first image. Participants self-recorded time to first image using the stopwatch function on their telephone or wristwatch. An electronic survey was also distributed to participants at the end of the study period to collect data regarding their user experience.

XLSTAT (Addinsoft Inc, Brooklyn, NY, USA), a statistical analysis application for Microsoft Excel (Microsoft Inc, Redmond, WA, USA), was used to conduct the statistical testing. For all statistical tests, $P<.05$ was considered statistically significant. For continuous sample data (eg, access time), one-way analysis of variance was conducted to measure whether mean values differed significantly among viewers. When significant differences were found (ie, $P<.05$), the Tukey (honestly significant difference) method was used to conduct pairwise comparisons among the three systems. For ordinal sample data (eg, diagnostic confidence and ease of use, rated as 1, poor; 2, fair; 3, good; and 4, excellent), a Kruskal-Wallis test was conducted to determine whether measurements from the three viewers came from a single distribution. When significant differences ($P<.05$) were found, the Dunn method was used to conduct pairwise comparisons among the three systems. For nominal sample data (eg, technical issues, rated as 0, none; 1, difficulty logging in; 2, slow scrolling speed; 3, could not load all images; and 4, other), the chi-square test was conducted.

Technical Details

Supporting the mobile image viewing frontend was a dedicated computer server running Red Hat Enterprise Linux 6.5 (RHEL6.5, Raleigh, NC, USA). It had two Intel Xeon E5-2643 processors (3.3 GHz, 4 physical, 8 virtual cores), 64-GB random

access memory, 2-TB raided disk space, and two NVIDIA Quadro 6000 graphics cards.

Installation of the server requires a static Internet protocol address and an assigned hostname in the domain name system lookup table. Our server uses RHEL6.5 installed using a boot disk provided by CSI. Deploying the server is facilitated by RPM package manager and YellowdogUpdater, Modified.

For use within our hospital environment, the server was configured to connect to two systems, namely, a lightweight directory access protocol (LDAP) server to provide login authorization and a PACS. The LDAP and PACS systems were configured to accept this connection. These connections and configuration steps required participation and assistance from hospital information technology and radiology informatics staff.

We used a dedicated LDAP pool fed by the enterprise system to determine which hospital personnel had access to the mobile image viewing server without having to maintain usernames and passwords.

The DICOM standards facilitate configuring the server to communicate with PACS. The server functions as a DICOM network node and any PACS can be configured to allow query and move operations to it. Depending on institution procedures and preferences, the server can be configured to search the PACS directly for images based on patient name, patient ID, modality, scan date, and/or accession number. A c-move operation is used to pull images from PACS directly to the server random access memory. The server then performs rendering operations in response to user interactions on one or more client systems. The resulting two-dimensional images are then streamed interactively to one or more client devices (eg, tablets). In cases where the radiological reports are included in the PACS as a structured report object, the report will come through and be displayed as well. Alternatively, the server can be configured to pull reports from a Mitra reports broker (Mitra, Waterloo, ON, Canada) or using a plug-in to the Softek Illuminate reports interface (Softek, Prairie Village, KS, USA). It is also possible to run the viewer (ie, view images only) without a reports connection; we did this out of necessity for the first 2years of our pilot before implementing the Softek solution. Advanced users can use the Web or mobile interface to perform three-dimensional reconstructions, measurements, image markup, and screen captures, which can be pushed back to the PACS if your institution allows it. Our facility has chosen for the flow of images to be one way (from PACS).

Results

Clinical Experience

Data before and after mobile viewer implementation were collected from all 19 clinicians, for a total of 565 data points (269 preimplementation and 296 postimplementation): 259 using MOBILE, 215 using ENTERPRISE, and 91 using PACS viewers. Because radiologists collected data only when on call, most of their data points were collected outside the hospital (87.7%, 142/162). Most data from other clinicians were collected within the hospital (76.7%, 309/403) and during work hours (78.6%, 308/392). Mobile devices used were iPad2, iPad3, and iPhones, and Wi-Fi or 3G was used to connect to the Internet; desktop devices included laptops and clinical desktops (personal computers) connected via Ethernet or Wi-Fi. The relative device usage is shown in Figure 2.

Table 1 summarizes the scores for diagnostic confidence, ease of use, and overall technical issues by user group. Diagnostic confidence was rated good to excellent for all three viewing techniques but slightly higher scores were provided for PACS (mean 3.8), compared with mobile (mean 3.7) or ENTERPRISE (mean 3.4) viewers. The difference in diagnostic confidence between PACS and mobile viewers was not statistically significant (P=.08). Diagnostic confidence with ENTERPRISE was significantly lower (P=.001) than with the other two systems. No discrepancies were reported by radiologists between preliminary interpretations using the mobile viewer (n=71) and final interpretations on PACS. Preliminary interpretations were not rendered for other cases.

Table 1. Mobile radiology image viewers compared with conventional desktop viewers: qualitative results by user group.

		Data points, n	Diagnostic confidence	Ease of use	Fraction of cases with technical issues, n (%)
All users					
	PACS	91	3.8	3.6	7/91 (7.7)
	ENTERPRISE	215	3.4	2.9	56/215 (26.0)
	MOBILE	259	3.7	3.6	19/259 (7.3)
Radiologists (n=9)					
	PACS	89	3.8	3.6	7/89 (7.9)
	ENTERPRISE	2	3.0	3.0	0/2 (0.0)
	MOBILE	71	3.4	3.5	8/71 (11.3)
Neurologists (n=4)					
	PACS	0	N/A	N/A	N/A
	ENTERPRISE	29	3.6	3.4	6/29 (20.7)
	MOBILE	41	3.9	3.8	3/41 (7.3)
Surgeons (n=3)					
	PACS	0	N/A	N/A	N/A
	ENTERPRISE	43	3.0	2.6	27/43 (62.8)
	MOBILE	27	3.6	3.8	0/27 (0.0)
Physicians Assistants (n=3)					
	PACS	2	3.0	2.0	0/2 (0.0)
	ENTERPRISE	91	3.2	2.5	14/91 (15.4)
	MOBILE	120	3.6	3.5	8/120 (6.7)

The PACS and mobile viewers had identical ease-of-use ratings (mean 3.6), which were significantly superior to the ENTERPRISE rating (mean 2.9, *P*=.001). Technical issues were reported more frequently with CUSTOM (26%, 56/214) than with PACS (8%, 7/91, *P*=.003) or mobile viewers (7%, 19/259, *P*=.001). The mobile viewer also had significantly less frequent technical issues than PACS (*P*=.007). The most common technical complaints were slow scrolling speed through images (ENTERPRISE, 28/56), inability to load images (mobile, 11/19), and log-in problems (PACS, 5/7). All of these technical issues impeded the ability of the user to evaluate the study efficiently. Although slow scrolling was inefficient, it did allow examination review unlike the other issues.

The average time to first image was fastest with mobile viewers at 2.4 minutes (ENTERPRISE, 4.5 minutes and PACS, 12.5 minutes; Table 2). The average time to first image was significantly faster with mobile viewers, compared with PACS (*P*=.001) and ENTERPRISE (*P*=.001). ENTERPRISE, however, was significantly faster than PACS (*P*=.001). For the mobile viewer, the average time to first image remained less than 3 minutes, regardless of time of day or location. Time to first image for both ENTERPRISE and PACS was influenced by delays in getting to a usable workstation.

Table 2. Time to first image in minutes by type of image viewer.

Time/location	PACS	ENTERPRISE	MOBILE
All data points	12.5	4.5	2.4
Inside hospital	11.4	3.9[a]	2.7[b]
Outside hospital	12.7	7.9[a]	2.0[b]
During work hours (weekdays 7 am to 6 pm)	9.6	4.3[a]	2.2[b]
After work hours	12.9	5.3[a]	2.5[b]

[a]*P*=.01 vs PACS

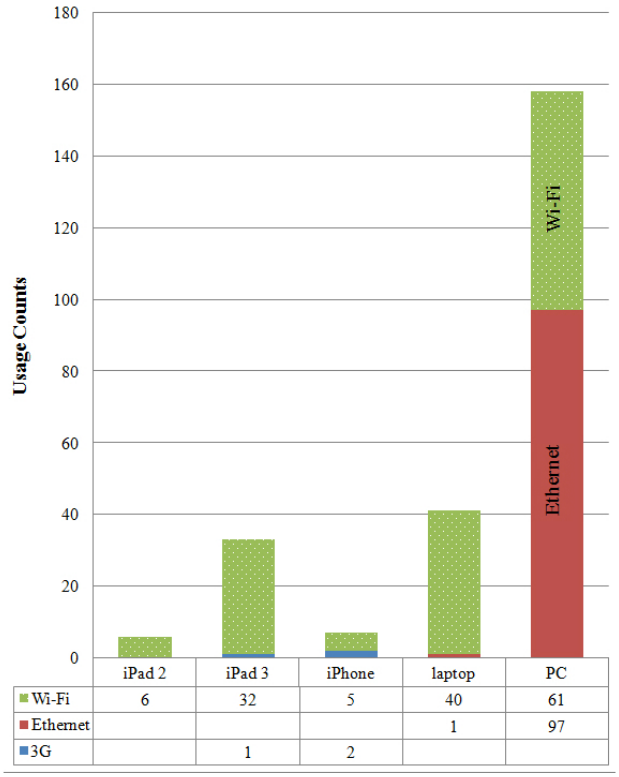
[b]*P*<.01 vs PACS and ENTERPRISE custom viewer

The predominant purpose of image access before and after the implementation of mobile viewer was for decision making (67.6%, 200/296). Once the mobile viewer became available, image use for patient education increased from 18% (48/268) to 29% (86/296). Following its implementation, most patient education episodes were conducted with the mobile viewer (99%, 85/86).

At baseline, the most commonly used image viewer by nonradiologists was ENTERPRISE (180/182), whereas by radiologists, it was PACS (85/87). However, following its implementation, the mobile viewer became the most commonly used viewer by both nonradiologists (85.1%, 188/221) and radiologists (95%, 71/75). The use of mobile devices for image viewing increased more than sixfold from baseline to postimplementation (from 14%, 37/269, to 88.9%, 263/296, respectively).

Of the 19 users, 18 completed the poststudy survey (8 radiologists and 10 nonradiologists). Most users reported that they used the mobile viewer a few days each week (10/18, 56%). The remainder of users reported daily use (3/18, 17%), use only when on call (3/18, 17%), or rare or infrequent use (2/18, 11%). Overall, the mobile viewer was the preferred program for image viewing outside the hospital (11/18, 61%), preferred by more nonradiologists (7/10) than radiologists (4/8). Inside the hospital, nonradiologists preferred ENTERPRISE (5/10), whereas radiologists preferred PACS (7/8). Overall, the desire to permanently implement the mobile viewer was rated as moderate (n=7) or high (n=8) by most users (15/18, 83%), with the remaining users rating it as mild (n=2) or neutral (n=1). None of the respondents reported a preference to not implement the mobile image viewer (n=0).

Figure 2. Usage counts are shown broken down by device and Internet connection type.



Discussion

Clinical Experience

One of the major benefits we found was the two to six times faster time to first image (ie, 2-10 minutes faster) using the mobile image viewer, compared with either of the desktop programs. Time to first image was defined as the time from a request to review images to when the image first appeared on the screen. We focused on time to first image as a metric that would be useful to compare different systems and be consistent regardless of the examination size or complexity. The longer times with desktop viewers were likely due to the following two main factors: (1) desktop viewer access often required travel time to the hospital or home especially after work hours and (2) both desktop viewers, unlike the mobile viewer, simultaneously launched other programs (eg, dictation system), which consume time. The perceived benefits of more rapid image access included faster communication of imaging findings, more rapid formation of treatment plans, and improved outcomes leading to lower patient care costs.

Although the ability to quickly access images is important, image viewers must also provide high-quality images. Overall, the study demonstrated comparable diagnostic confidence between mobile and PACS viewers. In addition, although this study was not designed to address diagnostic accuracy, radiologists found no discrepancies in 71 studies interpreted with both mobile and PACS viewers. Further studies designed to directly assess the diagnostic quality of mobile viewing options in specific clinical contexts (eg, [8]) would be necessary before modifying diagnostic read protocols to include mobile options.

Since our initial trial period we have provided access to 277 users, comprising hospital staff and physicians including all radiologists and fellows. We make use of custom software to parse the system-generated log files to evaluate usage statistics by user, time of day, and day of week. We also continue to survey our user base both formally and informally.

Technical Experience

Our experience with installing and integrating the back-end infrastructure necessary for a mobile image-viewing platform has been largely positive. We used hardware that closely resembles the high-end CSI-recommended servers and supports our current user base of nearly 500 (we generally have fewer than 5 simultaneous users). Servers can vary in price significantly (from US $5,000 to over US $15,000) depending on the number of simultaneous users accessing the system and whether or not advanced (three-dimensional) visualization capabilities will be enabled. The specific hardware needs of an organization would be based on the volume and intensity of the expected user base and may necessitate multiple servers.

The specific software installation and configuration process was relatively straightforward for a system administrator with basic Linux experience and greatly simplified by utilizing vendor suggestions for server configuration and operating system. Connectivity with enterprise and radiology informatics and computing services (LDAP and PACS) is essential.

Our PACS and RIS store images and reports separately, and as a result, our initial implementation did not include the radiology reports. Users consistently identified this as the largest shortcoming of the pilot. Our recent introduction of reports via the Softek interface has been an important step in increasing nonradiologist, nonemergent use of the app. By contrast, virtual collaboration is routinely touted as the largest benefit of the product (beyond rapid and mobile access) and is fully facilitated and enabled by the software. Maximizing the potential of this feature still requires active effort to integrate its use into the clinical routine.

Challenges

Although the mobile image viewer is now available to all clinicians, it is still used by only a minority of the staff. We attribute this to several factors. One is the current lack of integration with more commonly used mobile apps such as the mobile electronic medical record (EMR). Having the mobile image viewer embedded into the mobile EMR would make it more easily accessible to users and not require them to have multiple apps open when evaluating the patient. In addition, the interface is different than the custom desktop interface used currently, requiring the user to learn a new method. Finally, even infrequent experiences of technical difficulties as significant as failing to load images (the most common issue seen with the mobile image viewer) are sufficient to sour users on use of the mobile image viewer in their clinical practice.

Based on our clinical demonstrations, younger users including residents and fellows seem more interested and less intimidated by this technology and we believe that focusing on trainees for more widespread use could be beneficial. In this way, the knowledge could travel “up” to more senior clinicians. Finally, user support is currently limited to a few people in our department. We are currently involved with efforts to have this technology accepted and supported by institutional resources, which could provide round-the-clock support.

Future

We believe that mobile viewing technology with virtual collaboration technology has the potential to improve the speed and quality of care we deliver. Our future efforts are focused on integrating this system with the EMR and obtaining institutional support for more widespread implementation.

Conclusions

The technical implementation and upkeep of the system are manageable but a significant and successful pilot or a roll out of this type of platform, or both, requires a dedicated team to train the user base and support workflow integration. Our initial clinical experiences suggest that user perceptions and quantifiable speed benefits afforded by a mobile image viewing option support the long-term adoption of such a platform.

Acknowledgments

The ResMD software license has been provided at no cost through a research agreement with CSI. The manuscript has not been reviewed or approved by CSI in any way. We are grateful for the participation of the following persons (nonauthors) in the pilot study: Maria I Aguilar, MD; Lloyd G Barker, PAC; Barry D Birch, MD; Dan J Capampangan, MD; Dwight D Channer; Brian W Chong, MD; Richard J Claridge, MD; William G Eversman, MD; Joseph M Hoxworth, MD; Leland S Hu, MD; Eric A Huettl, MD; Joyce K Lee Iannotti, MD; J Scott Kriegshauser, MD; Sailendra G Naidu, MD; Kent D Nelson, MD; Ameet C Patel, MD; Naresh P Patel, MD; Kyra R Pokorny, PAC; Helen J Ross, MD; and Lisa A Rotellini Coltvet, PAC.

Conflicts of Interest

J Ross Mitchell is a co-founder and the founding scientist of CSI. Intellectual property developed in his laboratory has been transferred to CSI and included in some of their products. The value of CSI shares may increase if these products become commercially successful. Dr Mitchell and his immediate family own approximately 3% of CSI. In accordance with Mayo Clinic Conflict of Interest policy: (1) research results from Dr Mitchell's laboratory involving CSI products must be corroborated by a nonconflicted Mayo Clinic Voting Staff member; and (2) an Oversight Committee has been established to oversee Dr Mitchell's research activities related to CSI products and intellectual property.

References

1. Bedi HS, Yucel EK. "I Just bought my residents iPads… now what?" The integration of mobile devices into radiology resident education. AJR Am J Roentgenol 2013 Oct;201(4):704-709. [doi: 10.2214/AJR.13.10674] [Medline: 24059358]
2. Sharpe EE, Kendrick M, Strickland C, Dodd GD. The radiology resident iPad toolbox: An educational and clinical tool for radiology residents. J Am Coll Radiol 2013 Jul;10(7):527-532. [doi: 10.1016/j.jacr.2013.02.007] [Medline: 23647869]
3. Székely A, Talanow R, Bágyi P. Smartphones, tablets and mobile applications for radiology. Eur J Radiol 2013 May;82(5):829-836. [doi: 10.1016/j.ejrad.2012.11.034] [Medline: 23312700]
4. Richardson ML. Software aids for radiologists: Part 2, essential apps for handheld devices. AJR Am J Roentgenol 2012 Dec;199(6):W714-W722. [doi: 10.2214/AJR.12.9500] [Medline: 23169744]
5. Demaerschalk BM, Vargas JE, Channer DD, Noble BN, Kiernan TEJ, Gleason EA, et al. Smartphone teleradiology application is successfully incorporated into a telestroke network environment. Stroke 2012 Nov;43(11):3098-3101 [FREE Full text] [doi: 10.1161/STROKEAHA.112.669325] [Medline: 22968466]
6. Mitchell JR, Sharma P, Modi J, Simpson M, Thomas M, Hill MD, et al. A smartphone client-server teleradiology system for primary diagnosis of acute stroke. J Med Internet Res 2011;13(2):e31 [FREE Full text] [doi: 10.2196/jmir.1732] [Medline: 21550961]
7. Erickson BJ, Ryan WJ, Gehring DG, Beebe C. Image display for clinicians on medical record workstations. J Digit Imaging 1997 Aug;10(3 Suppl 1):38-40 [FREE Full text] [Medline: 9268835]
8. John S, Poh AC, Lim TC, Chan EH, Chong LR. The iPad tablet computer for mobile on-call radiology diagnosis? Auditing discrepancy in CT and MRI reporting. J Digit Imaging 2012 Oct;25(5):628-634 [FREE Full text] [doi: 10.1007/s10278-012-9485-3] [Medline: 22562174]

Abbreviations

CSI: Calgary Scientific, Inc
DICOM: digital imaging and communication in medicine
EMR: electronic medical record
FDA: Food and Drug Administration
LDAP: lightweight directory access protocol
PACS: picture archiving and communication system
ResMD: ResolutionMD

Edited by G Eysenbach; submitted 22.01.15; peer-reviewed by BS Kang, R Toomey; comments to author 23.03.15; revised version received 03.04.15; accepted 03.04.15; published 27.05.15.

Please cite as:
Zwart CM, He M, Wu T, Demaerschalk BM, Mitchell JR, Hara AK
Selection and Pilot Implementation of a Mobile Image Viewer: A Case Study
JMIR mHealth uHealth 2015;3(2):e45
URL: http://mhealth.jmir.org/2015/2/e45/
doi:10.2196/mhealth.4271
PMID:26018587

Original Paper

Identifying Quality Indicators Used by Patients to Choose Secondary Health Care Providers: A Mixed Methods Approach

Dominic King[1*], MB ChB, MRCS; Sameer Zaman[2*], BSc, MBBS; Saman Sara Zaman[2], BSc, MBBS; Gurnaaz Kaur Kahlon[2], BSc, MBBS; Aditi Naik[2], BSc, MBBS; Amar Singh Jessel[2], BSc, MBBS; Niraj Nanavati[2], BSc, MBBS; Akash Shah[2], BSc, MBBS; Benita Cox[2], PhD; Ara Darzi[1], MD

[1]Imperial College London, London, United Kingdom

[2]Imperial College London, South Kensington Campus, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Dominic King, MB ChB, MRCS
Imperial College London
10th Floor QEQM
St. Mary's Hospital
London,
United Kingdom
Phone: 44 7971500964
Fax: 44 7971500964
Email: dominic.king@imperial.ac.uk

Abstract

Background: Patients in health systems across the world can now choose between different health care providers. Patients are increasingly using websites and apps to compare the quality of health care services available in order to make a choice of provider. In keeping with many patient-facing platforms, most services currently providing comparative information on different providers do not take account of end-user requirements or the available evidence base.

Objective: To investigate what factors were considered most important when choosing nonemergency secondary health care providers in the United Kingdom with the purpose of translating these insights into a ratings platform delivered through a consumer mHealth app.

Methods: A mixed methods approach was used to identify key indicators incorporating a literature review to identify and categorize existing quality indicators, a questionnaire survey to formulate a ranked list of performance indicators, and focus groups to explore rationales behind the rankings. Findings from qualitative and quantitative methodologies were mapped onto each other under the four categories identified by the literature review.

Results: Quality indicators were divided into four categories. Hospital access was the least important category. The mean differences between the other three categories hospital statistics, hospital staff, and hospital facilities, were not statistically significant. Staff competence was the most important indicator in the hospital staff category; cleanliness and up-to-date facilities were equally important in hospital facilities; ease of travel to the hospital was found to be most important in hospital access. All quality indicators within the hospital statistics category were equally important. Focus groups elaborated that users find it difficult to judge staff competence despite its importance.

Conclusions: A mixed methods approach is presented, which supported a patient-centered development and evaluation of a hospital ratings mobile app. Where possible, mHealth developers should use systematic research methods in order to more closely meet the needs of the end user and add credibility to their platform.

(JMIR mHealth uHealth 2015;3(2):e65) doi:10.2196/mhealth.3808

KEYWORDS

mHealth; patient choice; mobile phone; hospital ratings

Introduction

Background

Patient choice has come to prominence in the United Kingdom with the advent of the National Health Service (NHS) Choose and Book and representation in key health policies such as Choice Matters [1] and High Quality Care for All: NHS Next Stage Review [2]. Patients in many countries (including the United Kingdom, Netherlands, and the United States) can now vote with their feet and choose health providers that fit best with their preferences and needs [3]. Providing choice is thought to be important in promoting competition between providers, with the goal of improving both the quality and efficiency of care [4]. While the actual evidence supporting the contribution of patient choice to cost control and quality of care is mixed [5], improving patient choice and shared decision making remains a stated objective of different health systems worldwide.

Service users trying to select between different health providers can use information from a variety of sources. Increasingly, patients are using websites that provide information about the comparative quality of health care from different providers [6]. For example, patients can compare hospitals using a wide range of quality and performance indicators such as waiting times, staffing, and patient safety on the website of the Care Quality Commission, the independent regulator of health care in England [7]. Service users are also turning to information based on the experiences of other patients when making a choice of provider. While patients have long used the experiences of friends and family in choosing hospitals, service users can now more systematically access collected information on patient experience (eg, Consumer Assessment of Health care Providers and Systems reports in the United States [8] and NHS Choices in the United Kingdom [9]). There is evidence that information based on patient experience is considered at least as important by service users choosing between different providers as different performance indicators provided by hospitals or reporting bodies [10].

An interesting recent development is the advent of patient rating websites such as PatientOpinion [11] and IWantGreatCare [12] (United Kingdom) and Rate MDs [13] (United States and Canada), where patients express views about the care they have received in much the same way as they might rate a hotel on a travel website. Users rating and commenting on the health care they receive is only set to increase with growing access to the Internet (particularly through increasingly ubiquitous mobile phone and tablet devices) [14]. Our research team has previously shown that online discretionary patient ratings can be useful in providing reliable information about health care quality [15,16].

This paper describes the process that was undertaken in the development of a hospital ratings platform for a consumer health care app. The aim to incorporate the best available evidence lies in sharp contrast to the majority of health related apps [17]. By working with patients and members of the public, we also sought to meet end-user needs often overlooked [18]. To develop the hospital ratings service, research was undertaken to determine which factors were considered important to individuals when choosing a nonemergency health care provider. This paper describes how established research methods can be used in the development of new mHealth apps.

Objectives

The aim of this study was to generate a list of quality indicators from the general public that were deemed important when choosing nonemergency secondary health care services along with the rationales for these choices, with the intention of using the findings in a new mHealth hospital ratings platform. Further, we aimed to illustrate the importance of rigorous research methodologies to underpin the development of mHealth technologies. The study was considered as part of a service evaluation, and ethics approval was not required. The study was conducted in London between November 2011 and June 2013.

Methods

Literature Review

As the first part of a mixed methods approach to identify which factors were considered most important to people when choosing nonemergency secondary health care providers in the United Kingdom, a review of existing literature (both academic and grey literature) was conducted. Publications were included only if they described patient choice in the *United Kingdom* health system so as to avoid confounding factors in the context of other health systems. For example, although there are many relevant articles from the United States, the differences between the largely privately funded United States and publicly funded United Kingdom systems may influence what users consider important when choosing a hospital. Patient choice did not feature prominently in national health policy until more recently [1,2]. A pilot literature review revealed a dearth of high quality research formally investigating patient choice in secondary health care prior to 2005. The patient choice agenda was first investigated when Lewis began studying patients' attitudes towards choice of hospital in the context of waiting times for cardiothoracic surgery [19]. Studies were therefore excluded if they had been carried out prior to 2005.

Survey

The aim of the questionnaire was to formulate a ranked list of quality indicators. The survey was completed by participants with a member of the research team at hand to explain any terms or answer questions. Care was taken to ensure that facilitators did not directly ask questions to avoid leading or influencing participants choices. A power calculation [20] for a study comparing the attitudes regarding choosing secondary health care between two groups (general public and out-patients attending clinic appointments at the hospital) determined a target total sample size of 400 (population of London, 2001: 7,172,091 [21]; standard of error 0.05), thus two groups of approximately 200 participants. Members of the general public were recruited (n=201), and the data are presented here. Data collected from the second group (patients at the hospital) are beyond the scope of this manuscript. An initial pilot questionnaire was conducted on 20 individuals prior to its wider use to identify and correct any unforeseen problems. Quota sampling was used to estimate size of target groups to ensure accurate representation of ages and genders [22]. Four age categories were formed by

combining existing categories from 2001 United Kingdom national census data [21]; target proportions for each age category were based on urban population proportions from the same census. Due to time and resource constraints, convenience sampling was then used to collect data, with checks to ensure collected data approximated the census age proportions. A higher proportion of people aged 18 to 35 years were included compared to people 60 years and older to account for lower usage of mHealth apps in the latter age group as described by previous investigators [23].

Inclusion criteria were English-speaking adults (18 years and older) UK residents. In order to ensure data collection was feasible under time and resource constraints, convenience sampling was then employed to recruit participants in six separate locations in central and greater London to provide greater geographical spread and wider generalizability of the data. Questionnaire collection ceased once the number of participants in each demographic group approached the estimated targets.

Informed consent was obtained for each participant completing the questionnaire. Participants were asked to provide demographic information and rank a predetermined list of quality indicators in order of importance. An ordinal scale was used, where respondents were asked to rank the factors in order of importance, first within their categories and then the categories themselves. This allowed us to assess the *relative* importance of the factors, as opposed to *absolute* importance [24]. Data were collated using Excel (Microsoft Corporation); SPSS (IBM Corporation) was used to undertake statistical analysis. Differences in mean ranks within categories were determined using the Friedman test, with $P<.05$ considered to be statistically significant.

Focus Groups

Focus groups were used to discuss the rationales behind the quality indicators considered to be important. Convenience sampling was used to recruit participants due to time and cost constraints, and referrals from initial recruits were used for further recruiting. We sought to recruit an equal representation of genders and ages in order to increase the generalizability of the results. Due to resource and time constraints it was not possible to match the age stratification of focus group participants with questionnaire respondents. The median age of participants at pilot focus groups was 40 years; therefore, participants were stratified by age and gender using this as a marker of division (males under 40 years, females under 40 years, males over 40 years, females over 40 years) to enable timely data collection and efficient analyses. Four focus groups were conducted, each comprising 6 individuals, with a gradual shift from broad open questions to narrow, focused questions [25]. Written consent was gained from each participant in advance. A final script of questions for the focus groups was confirmed following a restructuring of a preliminary script that had been piloted. Each focus group lasted for approximately 90 minutes and was led by one researcher acting as an impartial facilitator and one as an assistant moderator. All recordings were transcribed verbatim, and thematic analysis [26] was used to identify common themes.

Results

Characteristics of Quality Indicators

Literature Review

The full findings of the literature review are beyond the scope of this manuscript but we include key details pertinent to subsequent survey and focus group development. Searches of the grey literature were included due to a paucity of peer-reviewed publications. Five publications were identified for critical review [24,27-30]. Regular surveys commissioned by the UK Department of Health regarding the subject of choice in health care were also examined. A summary of included literature is presented in Textbox 1. From these, a list of choice factors important to patients selecting a health provider was devised. The factors identified in the literature review were separated into four categories of quality indicators: hospital statistics, hospital staff, hospital facilities, and hospital access (Textbox 2). These categories formed the questionnaire and informed the discussion topics for focus groups.

Textbox 1. Summary of the literature review: key factors guiding patient choice.

Understanding Patients' Choices at the Point of Referral [28]

Areas of investigation

- Factors influencing patients when choosing hospitals
- Developing an algorithm to predict demand for particular services

Key findings

- Views provider quality as extremely important: 80%
- Values low mortality rates, infection rates, and readmission rates: 90%
- Views waiting times as important: 55%
- Views primary care provider influence as important: 60% (most important factor: 2%)
- Views travel as important: 30% (most important factor: 15%)
- Preference for lower travel costs was observed

Patient Choice: How Patients Choose and Providers Respond [24]

Areas of investigation

- Patient considerations when choosing health care
- Primary care provider response to the notion of patient choice and subsequent support of patient choice

Key findings

- Considers personal experience: 41%
- Judges primary care provider advice as important: 36%
- Factors identified in order of importance (graded out of 3):
  - Cleanliness (2.6)
  - Quality of care (2.5)
  - Standard of facilities (2.1)
  - Friendliness (2.1)
  - Waiting time (2.1)
  - Experience (2.0)
  - Proximity (2.0)
  - Waiting room (1.8)
  - Convenience of appointment time (1.8)
  - Consultant of choice (1.7)
  - Fixtures and fittings (1.5)
  - Accessibility (1.2)
  - Food (1.2)
  - Travel Cost (1.0)

Report on the National Patient Choice Survey [30]

Area of investigation

- The single most important factor patients consider when choosing a secondary health care provider

Key findings

- Rates proximity to home/work as single most important factor: 38%
- Factors reported as being most important:
  - Previous experience of the hospital: 12%

- Waiting times: 10%
- Previous good experience: 6%
- Quality of care: 5%
- Accessibility: 5%

Choosing a High Quality Hospital: The Role of Nudges, Scorecard Design, and Information [29]

Areas of investigation

- Information important to patients when choosing a hospital
- How presentation of information affects decisions

Key findings

- Values information relevant to the patient (eg, their consultant, condition)
- Format of information plays a role in its interpretation (ie, only patients with high levels of numeracy can interpret mortality ratios)
- Factors deemed important: waiting times, MRSA rates, quality of service, doctors' expertise, cleanliness, distance, being treated with respect

London Patient Choice Project Evaluation: A Model of Patients' Choices of Hospital from Stated and Revealed Preference Choice Data. [27]

Areas of investigation

- Factors used by patients when deciding to accept alternative treatment
- Weighing the relevant factors
- Trade-offs patients make when considering different factors

Key finding

- Less likely to take up offer of quicker treatment elsewhere if the alternative hospital has a worse reputation or the appointment involves increased travel time, results in patient paying for transport or requires nonlocal follow-up care.

Textbox 2. Quality indicators identified by the literature review.

Hospital statistics [24,28-30]

- MRSA infection rates
- Readmission rates
- Mortality rates
- Wound infection rates
- Waiting times

Hospital staff [24,29]

- Friendliness
- Respectfulness
- Competence

Hospital facilities [24,29]

- Cleanliness
- Hygiene
- Availability of single-sex wards
- Quality of food
- Standard of facilities

Hospital access [24,27-30]

- Distance from home
- Cost of travel
- Time to travel
- Car parking availability

Questionnaire

Members of the general public completed the questionnaire (93 male, 108 female, n=201). The age spread of the sample compared to 2001 population proportions can be seen in Figure 1. Respondents ranked quality indicators in order of importance within their specified categories (ie, 1 through 5 with 1 being the most important). Based on mean rankings, the quality indicators were arranged in order of preference within respective categories (Table 1). Similarly respondents were asked to rank the overall categories (statistics, staff, facilities, and access) (Table 2). The Friedman test was used to determine the statistical significance of the differences between the mean ranks obtained for the quality indicators within and between categories to determine the true order (Multimedia appendix 1). Final ranked order of quality indicators and categories was determined after statistical analyses (Table 3).

While three of the categories (statistics, staff, facilities) were deemed equally important, quality indicators under the category of access were considered to be of less importance. Within each group some indicators were seen as being more important than others. Regarding staff, competence was seen as being significantly more important than friendliness and respectfulness. In terms of facilities, up-to-date facilities and the cleanliness of the premises were seen as equally important but more so than the other factors. In the category of statistics, infection rates, mortality rates, complication rates, and waiting times were of equal importance; statistics regarding readmission rates were seen as less important. Regarding access, ease of travel was more important that the cost and availability of car parking.

Table 1. Mean rankings of quality indicators within each category.

Categories	Quality indicators	Mean rankings
Hospital statistics	Infection rates	2.2
	Mortality rates	2.8
	Waiting times	3.0
	Complication rates	3.2
	Readmission rates	3.8
Hospital staff	Competence	1.3
	Friendliness	2.3
	Respectfulness	2.3
Hospital facilities	Clean premises	1.8
	Up-to-date equipment	2.0
	Good food	4.1
	Disabled facilities	4.2
	Single sex wards	4.4
	Appealing appearance	4.5
Hospital access	Ease of travel	1.8
	Cost/availability of car parking	2.1

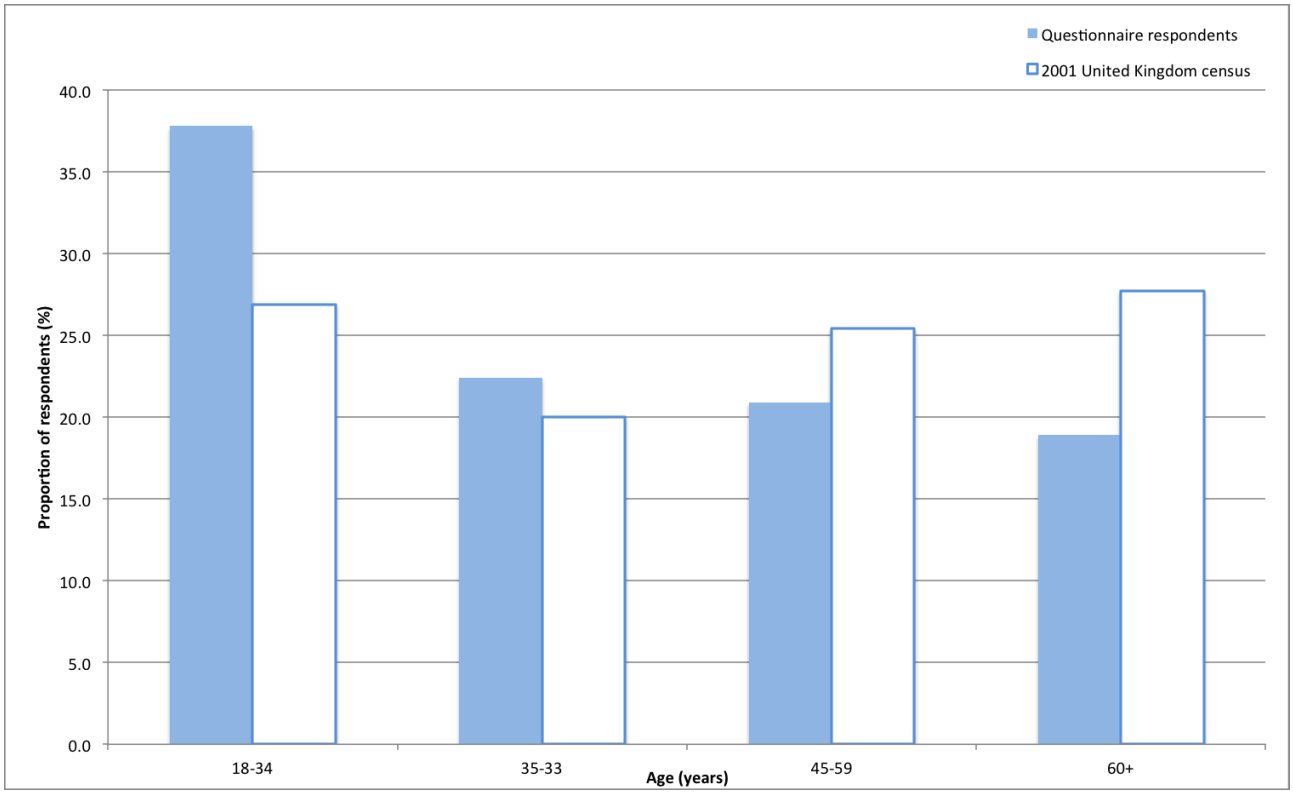
Table 2. Mean rankings between categories.

Overall groups	Mean rankings
Hospital facilities	2.2
Hospital staff	2.3
Hospital statistics	2.5
Hospital access	3.1

Table 3. Overall rankings of quality indicators within and between categories.

Categories		Ranking of quality indicators
Hospital statistics		
	More important	Infection rates, mortality rates, complication rates, waiting times
	Less important	Readmission rates
Hospital staff		
	More important	Competence
	Less important	Friendliness, respectfulness
Hospital facilities		
	More important	Clean premises, up-to-date equipment
	Less important	Good food, disabled facilities, single sex wards, appealing appearance
Hospital access		
	More important	Ease of travel
	Less important	Cost/availability of car parking
Overall categories		
	More important	Hospital statistics, staff, facilities
	Less important	Hospital access

Figure 1. Age spread of questionnaire respondents compared with 2001 United Kingdom census respondents.



Focus Groups

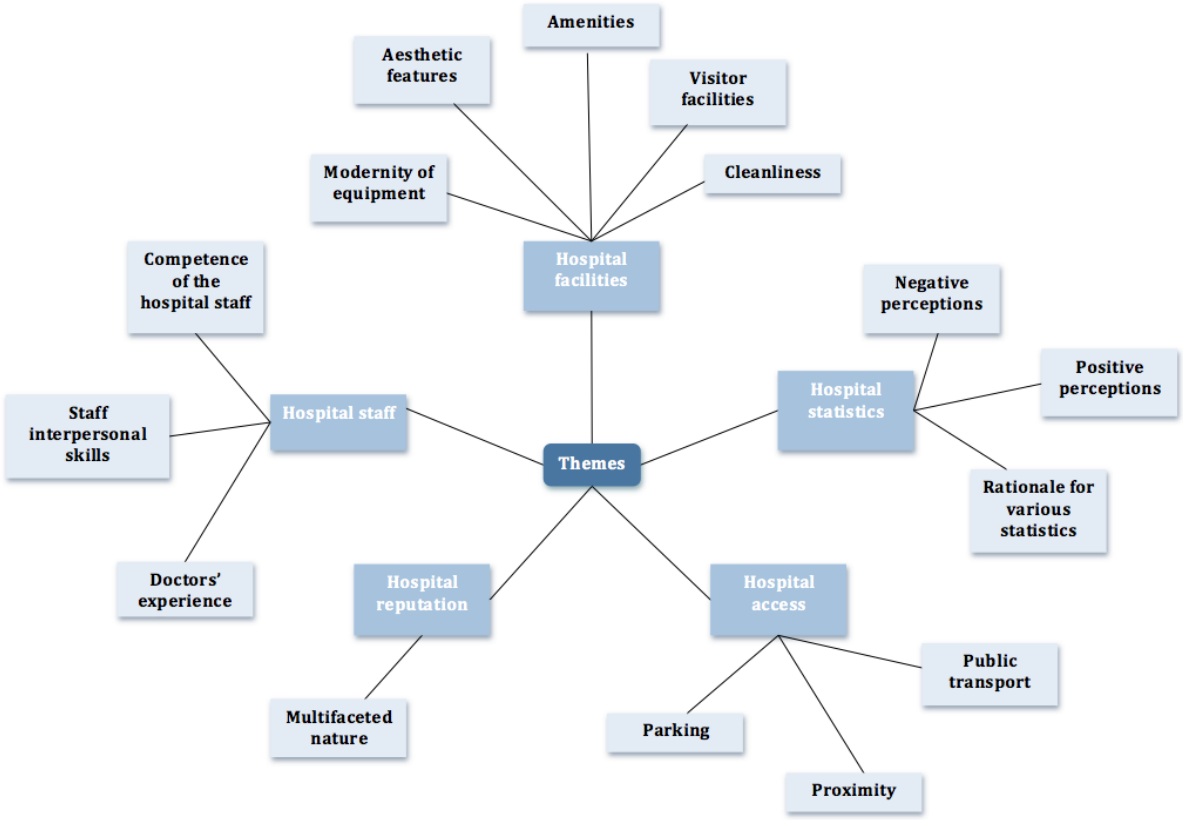
Four focus groups were used to explore the rationales behind rankings formulated from the questionnaire. Thematic analysis was conducted by performing manual coding [26], from which a collective list of codes was assembled. Overarching subthemes, and subsequently themes, were identified and reviewed. The themes, subthemes, and codes for different preferences established during the analysis are presented in Table 4 and are visually represented in Figure 2. The findings from both the quantitative and qualitative methodologies were mapped to each other under the four categories identified by the systematic literature review.

Table 4. Themes, subthemes, and codes from focus groups.

Theme	Subtheme	Codes
Hospital reputation	Multifaceted nature of reputation	Important because it encompasses everything
		Important because it reflects the facilities at the hospital
		Important because it reflects the competence of the hospital staff
Hospital statistics	Rationale for choosing various statistics	Infection rates are important because they are frequently reported to the media
		Waiting times are important because being treated quickly is my main concern
		Waiting times are important because they reflect the hospital's efficiency
		Departmental statistics are more relevant because they are specific to the situation
		Waiting times are important because I do not want to spend too much time at the hospital
		MRSA rates are important because of the risks faced by visitors
		Mortality rates may not be the best indicators because better hospitals may undertake more challenging cases
	Negative perceptions of statistical descriptors	Not relevant in the context of routine procedures
		Not important because they can be manipulated
		Not important because they are negatively exaggerated in the media
	Positive perceptions of statistical descriptors	Important because they are true facts about hospital quality
Hospital staff	Competence of hospital staff	Seeing specialists is important because they are more skilled
		Competence is the most important because my main aim is being treated properly
		Important because I want to be treated correctly, regardless of friendliness
		Important because it reflects staff experience
		Most important because I would travel further to ensure it
		Important because it encompasses interpersonal skills too
	Staff interpersonal skills	Important because I expect to be treated fairly
		Important for nurses because they are responsible for making you comfortable
		Important because I feel more reassured with doctors and nurses that I know
		Important because they have an impact on recovery rates
	Doctors' experience	Younger doctors are not good because they are inexperienced
		Qualifications are important because they reflect competence
Hospital facilities	Modernity of equipment and cleanliness of hospital	Important for outpatients because there is only a limited time to experience it
		Not important because it is assumed to be equally up-to-date at all hospitals
		Cleanliness of the hospital is the most important factor for outpatients because they are only there for a short time
	Aesthetic features and amenities of patient comfort	Important because it reflects the comfort of the hospital
		Not important because it is assumed that all hospitals are equally clean
		Important because poor aesthetics can lead to depression
		Not important because they do not affect health care

Theme	Subtheme	Codes
		Not important so long as staff is competent
		Important because I would like to see the hospital before choosing to be treated there
		TV facilities are important because one might be staying at the hospital for an extended time
		Facilities are not important because they are subject to individual experience
	Facilities for visitors	Food and drink facilities are important to ensure comfort for visitors
		Visiting times at the hospital should be flexible because the convenience of visitors is important
		Overnight facilities for visitors are important so they can spend longer time with the patient
Hospital access	Parking at and around the hospital	Parking charges are important because the may affect my visitors
		Not important because I do not have a car
		Availability is important because I drive
	Proximity of the hospital	Not important because I am willing to travel further if other factors are better satisfied
		More important so that visitors can visit me easily
		Important because I do not have a car
	Public transport	Important because parking at/around the hospital is too expensive

Figure 2. Thematic map of qualitative data from focus groups.



Interpretation of Data

Hospital Statistics

Qualitative analyses of the focus groups showed that people consider a wide range of statistical descriptors. Participants reasoned that information regarding infection rates was an important consideration, mainly due to the extensive media coverage: this may account for why infection rates were ranked joint highest in this category during the quantitative analyses. Elsewhere, the qualitative data disagreed with the survey findings, asserting that readmission rates were also considered to be important by the focus group participants because people felt the rates reflected the success of a specific treatment or condition. Furthermore, unlike the survey respondents, the focus group participants did not feel that mortality rates were a valuable quality indicator because it was reasoned that a hospital may undertake more challenging cases, which could inflate mortality rates despite the hospital faring well on other indicators such as staff competence. Participants also speculated that in many cases departmental statistics may be more meaningful than those describing the hospital overall due to the variation between departments within a hospital, leading to potential misrepresentation of the overall hospital statistics. This interesting finding was not captured by the quantitative analyses and highlights a potential avenue for future research.

Hospital Staff

Staff competence was identified to be the most important factor by the quantitative analyses, followed by friendliness and respectfulness equally. Qualitative analyses were in agreement, with participants confirming the importance of receiving treatment by competent staff, with some even stating that they would travel further than their nearest hospital in order to secure treatment by a competent doctor. However, the focus groups identified that there is no single measure by which the public could judge competence. Rather competence was defined as a compound of experience, qualifications, place of education, or even the possession of excellent interpersonal skills.

Hospital Facilities

The quantitative analysis identifies cleanliness and modernity of equipment as the two highest ranked indicators. Rationales elicited from the focus groups shed light on why this may be the case. Members of the focus groups felt cleanliness to be very important in hospitals but did not necessarily seek out data about it when making a choice of hospital. It was suggested that this was a consequence of the assumption that cleanliness is the same in all hospitals. There was less consensus regarding the importance of modern equipment, although some certainly felt access to the latest technologies to be important.

Hospital Access

The quantitative analyses showed that this category was less important than the other three. Within this category, ease of traveling to the hospital was significantly more important than parking and cost of travel. Focus groups revealed that the proximity of the hospital to home or work was an important consideration. However, this is very much dependent on the severity of illness and the availability of treatment, with participants expressing that they may be willing to travel beyond their most proximal hospital in order to benefit from a higher quality of care. Therefore, it appears that the importance of this indicator may depend on the context of the decision.

Discussion

Key Findings and Recommendations

Service users across many health systems are now offered a wider choice of health care providers and increasingly have at their disposal a wide variety of factors to consider when making these decisions. The rapid adoption of mobile phones and tablet devices has enhanced access to information about different hospitals by making it possible for patients to view and share this information at any time and while on the move [31].

This study collates the existing literature regarding which factors are considered important for consumers in this context, contributes a categorized and ranked list of quality indicators, and reconciles the rationales underpinning these decisions. Furthermore, this study demonstrates how this information can be harnessed in the context of developing a robust user-generated ratings platform for use on mobile communication technologies.

Although mobile technologies are frequently put forward as a solution to challenges in health informatics, there is often a lack of rigorous research underpinning their development and evaluation. This project illustrates the importance of sound research methodology when developing these strategies by employing a mixed methods approach to reconfigure the ratings service based on factors that the public held to be important in choosing nonemergency health providers.

Findings included that staff competence was the most important factor within the hospital staff category, with participants asserting that they would travel further than their nearest hospital to secure treatment under a doctor they perceived to be more competent. However, the qualitative analyses revealed that there is no single measure by which competence could be judged; rather it was a compound of many factors including amount of experience, qualifications, place of education, and interpersonal skills.

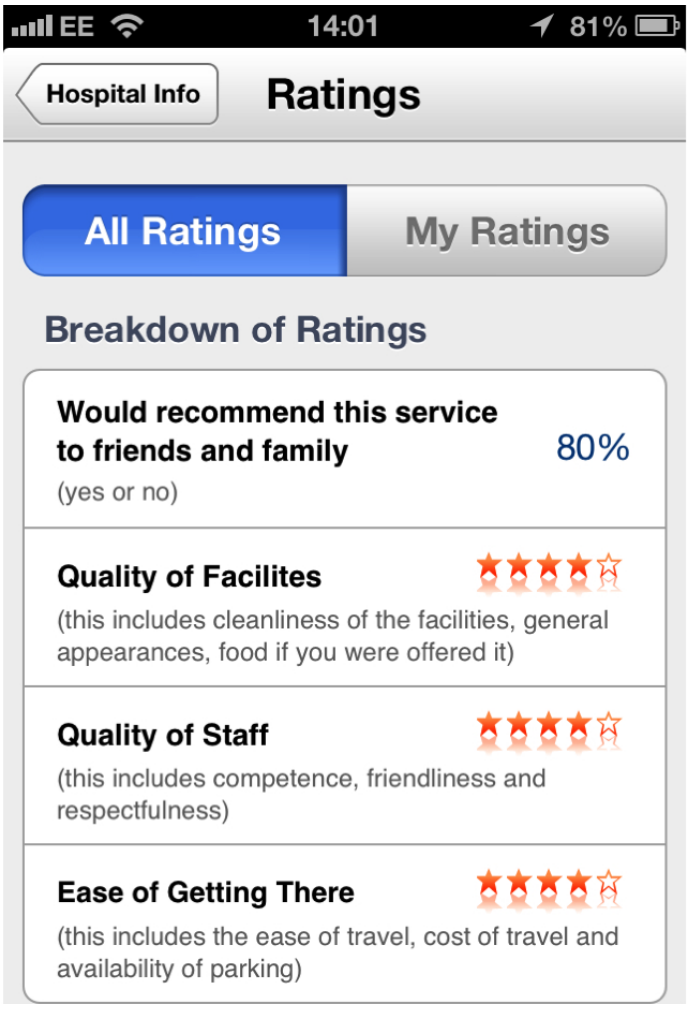
Cleanliness and modernity of equipment stood out as the two most important hospital facilities. This is concurrent with previous reports that people consider information about cleanliness when researching a hospital [32,33]. Qualitative analyses discovered that while this was a factor deemed to be highly important, it was not widely sought after. Participants suggested this might be due to a commonly held assumption that hospitals are of equal cleanliness, therefore only those hospitals with a remarkably poor reputation for cleanliness may be of note. This was also the case for modernity of equipment. Future mHealth developers should reflect on this subtlety in order to include factors that are not only important but also highly sought after to avoid information overload for users.

Participants could not differentiate level of importance between various types of hospital statistics. Hospital-wide statistics may be of limited use to users who would be more interested in department-specific statistics. Moreover, users appreciate that

overall hospital statistics may not be an accurate representation of an their department of interest due to interdepartmental variation. Conversely it may be argued that an inability to compare the importance of statistical descriptors may reflect that they are poorly understood by users. This may explain equal significance attributed to individual statistics within this category and highlights the need for the careful inclusion of statistics that are relevant to the user's individual health encounter in mHealth platforms (See Figure 3). Care must be taken by developers to ensure that presentation of statistics, including color coding or a glossary of terms, aids user interpretation. These subtleties may not be appreciated without formal research methods informing these strategies.

The fact that the categories of hospital staff, hospital facilities and hospital statistics were deemed equally important illustrates that users' demands for information about hospitals are extensive and varied. mHealth developers should aim to provide information about these categories equally in order to reflect and satisfy these demands. Adequate provision of these varied factors requires an equally varied presentation of information. For example, participants asserted that graphs and percentages provided objective evidence of statistical measures, whereas past users' reviews were more useful in capturing complex domains such as staff competence. Therefore we recommend that mHealth developers include a range of formats as this study illustrates that each caters to different, and equally important, categories of quality indicators.

Figure 3. Screenshot of Wellnote ratings platform.



Limitations

An important limitation of the study is that the questions were asked outside the context of mobile phones and mHealth. This was a purposeful decision as it was felt that doing so may lead to reduced applicability of this research. Further research is required specifically investigating whether the information consumers want in the context of an mHealth app is any different from the factors that are important when choosing secondary health care in general. The original study was adequately powered for a comparison between patient groups and the general public; the analyses included here may therefore by underpowered due to resource constraints. This study was unable to match age stratification between quantitative and qualitative stages. We recommend that future investigators attempt to do so to allow closer mapping of the two datasets.

Conclusion

The huge interest in developing apps for mobile phone and tablet platforms to enhance health outcomes and service delivery—widely termed mHealth—has led to an "enthusiastic proliferation of untested methods" [17]. An evidence base needs to be developed to make this field credible and address the needs

of the end-user. More attention needs to be paid to structuring app development in theory or best practice [34].

This study used a mixed methods approach to find that information about hospital staff, hospital facilities and hospital statistics are equally important to people when choosing a hospital. Information about getting to the hospital is least important. Staff competence is most important regarding hospital staff, which is a multifactorial domain best captured by past users' reviews; cleanliness and modernity of equipment are most important regarding hospital facilities but are not actively sought after. People find it difficult to compare relative importance between various hospital statistics. Barriers to understanding statistics may be removed by use of graphs and percentages.

Users of health care demand a wide and varied range of information about hospitals. mHealth developers must determine which information is most relevant to their users' needs and provide this in an accessible format. Less important information must be identified and removed to avoid information overload. A sophisticated appreciation of the complex needs of mHealth users is possible when these strategies are underpinned by rigorous research methods. This study demonstrates how a mixed methods approach can enhance mHealth solutions.

Conflicts of Interest

Dominic King and Ara Darzi are part of the team responsible for designing and releasing the Wellnote by Dr. Darzi iPhone app, which includes an applet for the rating and reviewing of health services by users. The data presented in this study were originally collected for the purpose of updating the Wellnote app. The Wellnote app was not mentioned to participants at any point during data collection stages. The analyses, discussion, and conclusion presented in this article are strictly independent of the Wellnote app and are for broad application to the context of mHealth hospital rating services.

Multimedia Appendix 1

Tabulated *P* values of statistical comparison of mean rankings of quality indicators within and between categories.

[PNG File, 82KB - mhealth_v3i2e65_app1.png]

References

1. Choice matters 2007-2008: putting patients in control. London: Department of Health (UK) URL: http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_076330.pdf [accessed 2015-05-12] [WebCite Cache ID 6YTsstinY]
2. Darzi A. High quality care for all: NHS next stage review final report. London: Department of Health (UK) URL: http://webarchive.nationalarchives.gov.uk/20140609052416/https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228836/7432.pdf [accessed 2015-05-12] [WebCite Cache ID 6YTtDdgEm]
3. Dixon A, Robertson R, Bal R. The experience of implementing choice at point of referral: a comparison of the Netherlands and England. Health Econ Policy Law 2010 Jul;5(3):295-317. [doi: 10.1017/S1744133110000058] [Medline: 20462469]
4. Vrangbaek K, Robertson R, Winblad U, Van de Bovenkamp H, Dixon A. Choice policies in Northern European health systems. Health Econ Policy Law 2012 Jan;7(1):47-71. [doi: 10.1017/S1744133111000302] [Medline: 22221928]
5. Buntin MB, Damberg C, Haviland A, Kapur K, Lurie N, McDevitt R, et al. Consumer-directed health care: early evidence about effects on cost and quality. Health Aff (Millwood) 2006 Dec;25(6):w516-w530 [FREE Full text] [doi: 10.1377/hlthaff.25.w516] [Medline: 17062591]
6. Fasolo B, Reutskaja E, Dixon A, Boyce T. Helping patients choose: how to improve the design of comparative scorecards of hospital quality. Patient Educ Couns 2010 Mar;78(3):344-349. [doi: 10.1016/j.pec.2010.01.009] [Medline: 20171822]
7. Care Quality Commission. URL: http://www.cqc.org.uk [accessed 2015-06-01] [WebCite Cache ID 6Yyel8Sh2]
8. Consumer Assessment of Healthcare Providers and Systems.: Agency for Healthcare Research and Quality URL: https://cahps.ahrq.gov [accessed 2015-06-01] [WebCite Cache ID 6YyetLLmA]
9. NHS Choices.: National Health Service URL: http://www.nhs.uk [accessed 2015-06-01] [WebCite Cache ID 6YyexxgdM]
10. de Groot IB, Otten W, Dijs-Elsinga J, Smeets HJ, Kievit J, Marang-van de Mheen PJ, CHOICE-2 Study Group. Choosing between hospitals: the influence of the experiences of other patients. Med Decis Making 2012 Nov;32(6):764-778. [doi: 10.1177/0272989X12443416] [Medline: 22546750]
11. PatientOpinion. URL: https://www.patientopinion.org.uk/ [accessed 2015-06-01] [WebCite Cache ID 6Yyf397gp]
12. IWantGreatCare. URL: https://www.iwantgreatcare.org [accessed 2015-06-01] [WebCite Cache ID 6YyfGAhiJ]
13. RateMDs. URL: https://www.ratemds.com [accessed 2015-06-01] [WebCite Cache ID 6Yyg9tqs9]
14. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
15. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Web-based patient ratings and objective measures of hospital quality. Arch Intern Med 2012 Mar 12;172(5):435-436. [doi: 10.1001/archinternmed.2011.1675] [Medline: 22331980]

16. Greaves F, Pape UJ, Lee H, Smith DM, Darzi A, Majeed A, et al. Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. J Med Internet Res 2012;14(5):e146 [FREE Full text] [doi: 10.2196/jmir.2280] [Medline: 23076301]
17. van HA, Tomlinson M, Swartz L. Point of care in your pocket: a research agenda for the field of m-health. Bull World Health Organ 2012 May 1;90(5):393-394 [FREE Full text] [doi: 10.2471/BLT.11.099788] [Medline: 22589575]
18. Whittaker R. Issues in mHealth: findings from key informant interviews. J Med Internet Res 2012;14(5):e129 [FREE Full text] [doi: 10.2196/jmir.1989] [Medline: 23032424]
19. Lewis R. More patient choice in England's National Health Service. Int J Health Serv 2005;35(3):479-483. [Medline: 16119571]
20. Padua RN, Santos RG. Fundamentals of Educational Research and Data Analysis. Quezon City Philippines: Katha Publishing Co. Inc; 1998.
21. UK Data Service: Census Support.: Office of National Statistics; 2001. 2001 Census: aggregate data (England and Wales) URL: http://infuse.mimas.ac.uk [accessed 2015-06-01] [WebCite Cache ID 6YyaslGZp]
22. Pizzo E. Session 7: developing research tools (lecture). Imperial College London: Health Services Research; 2010.
23. Aitken M, Gauntlett C. Patient apps for improved healthcare: from novelty to mainstream. 2013 Oct. IMS Institute for Healthcare Informatics URL: http://www.imshealth.com/deployedfiles/imshealth/Global/Content/Corporate/IMS%20Health%20Institute/Reports/Patient_Apps/IIHI_Patient_Apps_Report.pdf [accessed 2015-06-01] [WebCite Cache ID 6YyDtz5k0]
24. Dixon A, Robertson R, Appleby J, Burge P, Devlin N, Magee H. The Kings Fund. 2010 Feb 01. Patient Choice: How patients choose and providers respond URL: http://www.kingsfund.org.uk/sites/files/kf/Patient-choice-final-report-Kings-Fund-Anna%20Dixon-Ruth-Robertson-John-Appleby-Peter-Purge-Nancy-Devlin-Helen-Magee-June-2010.pdf [accessed 2015-06-01] [WebCite Cache ID 6YyCsdDZ4]
25. Morgan DL. Focus groups. Ann Rev Sociol 1996;22:129-152.
26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3:77-101.
27. Burge P, Devlin N, Appleby J, Rohr C, Grant J. London Patient Choice project evaluation: a model of patient's choices of hospital from stated and revealed preference choice data. RAND Corporation 2005 [FREE Full text]
28. Burge P, Devlin N, Appleby J, Gallo F, Nason E, Ling T. Understanding patient's choices at the point of referral. Kings Fund and the RAND Corporation 2006 [FREE Full text]
29. Dixon A, Boyce T, Fasolo B, Reutskaja E. Choosing a high quality hospital: the role of nudges, scorecard design and information. The Kings Fund 2010 [FREE Full text]
30. Dixon S. Report on the National Patient Choice Survey: February 2010 England.: Department of Health; 2010. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216073/dh_117096.pdf [WebCite Cache ID 6YycT0d9j]
31. Estrin D, Sim I. Health care delivery. Open mHealth architecture: an engine for health care innovation. Science 2010 Nov 5;330(6005):759-760. [doi: 10.1126/science.1196187] [Medline: 21051617]
32. Sofaer S, Crofton C, Goldstein E, Hoy E, Crabb J. What do consumers want to know about the quality of care in hospitals? Health Serv Res 2005 Dec;40(6 Pt 2):2018-2036 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00473.x] [Medline: 16316436]
33. Vonberg RP, Sander C, Gastmeier P. Consumer attitudes about health care acquired infections: a German survey on factors considered important in the choice of a hospital. Am J Med Qual 2008;23(1):56-59. [doi: 10.1177/1062860607310915] [Medline: 18187591]
34. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]

Edited by K Jethwani; submitted 26.08.14; peer-reviewed by J Bloor, S Agboola; comments to author 17.09.14; revised version received 17.11.14; accepted 28.12.14; published 05.06.15.

Please cite as:
King D, Zaman S, Zaman SS, Kahlon GK, Naik A, Jessel AS, Nanavati N, Shah A, Cox B, Darzi A
Identifying Quality Indicators Used by Patients to Choose Secondary Health Care Providers: A Mixed Methods Approach
JMIR mHealth uHealth 2015;3(2):e65
URL: http://mhealth.jmir.org/2015/2/e65/
doi:10.2196/mhealth.3808
PMID:26048441

Original Paper

Doctors and the Etiquette of Mobile Device Use in Trauma and Orthopedics

Oliver Blocker[1], MRCS, MSc; Lydia Hayden[1], MB BS; Alison Bullock[1], PhD

The Cardiff Unit for Research and Evaluation in Medical and Dental Education (CUREMeDE), Cardiff University, Cardiff, United Kingdom

Corresponding Author:
Oliver Blocker, MRCS, MSc
The Cardiff Unit for Research and Evaluation in Medical and Dental Education (CUREMeDE)
Cardiff University
Cardiff University School of Social Sciences, Glamorgan Building
King Edward VII Avenue
Cardiff, CF10 3WT
United Kingdom
Phone: 44 2920870780
Fax: 44 2920874000
Email: drblocker@gmail.com

Abstract

Background: The etiquette surrounding the use of mobile devices, so-called "mobiquette," has been previously identified as a barrier to use in an educational context.

Objective: To investigate the influence of mobile device use on patient and staff opinions in the trauma and orthopedics department at a teaching hospital in Wales.

Methods: A survey of patients at the bedside and staff in their work environment was undertaken. Data included age, frequency of observed use, suspected main reason for use, and whether doctors' use of a mobile device positively or negatively influenced participants' opinions of them as a professional and as a person.

Results: A total of 59 patients and 35 staff responded. The modal age range was 40 to 54 years old. Most patients (78%) never see doctors using mobile devices in the workplace, compared with 3% of staff. The main reason for use was thought to be "communicating with colleagues" (48%) followed by "Internet use/applications for work reasons" (40%). Approximately 40% of patients' opinions of doctors were positively influenced by device use, compared with 82% of staff. This difference between patient and staff opinions was statistically significant for both professional (P<.001) and personal (P=.002) opinions.

Conclusions: Patients are likely to have a negative opinion of doctors using mobile devices in the workplace. This can be balanced by the more positive opinions of colleagues. We advise doctors to remember "mobiquette" around patients.

(JMIR mHealth uHealth 2015;3(2):e71) doi:10.2196/mhealth.4122

KEYWORDS

education, medical; cell phones; patient-physician relationship

Introduction

Mobile technology is being used with the intention of enhancing the learning of medical students and doctors in the workplace and the evidence of its value is growing [1,2]. What remains under-researched is the opinions of patients and colleagues regarding doctors' use of mobile devices for learning on the ward. The term "mobiquette" was coined by Ellaway and Masters in 2008 [3] to describe the etiquette of mobile device use and appropriate mobiquette has been identified as a challenge to device use in the workplace [1,2]. Even if there was consensus about what is considered "appropriate," there is concern that learners' (eg, students and trainees) interactions with devices will be misinterpreted. Without looking over a user's shoulder, we cannot know if the mobile technology is being used for professional, educational purposes or personal reasons (ie, texting, social media, or Internet browsing). Beyond the etiquette issue, another concern is the fear of superficial learning [4] and the erosion of the traditional practice of internalizing knowledge, replacing it with an ability to locate information in "the cloud." If the concept of "I may not know the answer, but I know where to find it" becomes more prevalent in medicine, it may require a change in both doctors' and

patients' perceptions of practice. Given these negative associations, it would be reasonable to have measures in place to ensure that mobiquette is observed by those using mobile technology in a clinical environment. The iDoc project [1], for example, advises its participants to inform their colleagues that they will be using mobile devices to retrieve clinical information and to reassure them that their use is work related. When using a mobile device in front of patients, trainees are advised to inform them what they are doing and to potentially involve patients in the process by sharing a view of the information on the screen. However, importantly, users need to exercise judgement around patients [5]. Professionalism, which includes establishing and maintaining partnerships with patients and colleagues [6], extends to the various situations faced by a doctor, including judgements about whether and how to share information from a mobile device.

Too readily assumptions are made about mobile technology, its users, and people's opinions about device use. Doctors now in training are mainly in their 20s and early 30s and can be included in the demographic group known as "millennials" [7] or the "net generation" [5], but it should not be assumed that they have the technical skills or attitudes to use mobile technology appropriately around other people. Age may also be unrelated to the opinions people form about mobile device use. Part of the purpose of the research we report here was to explore these assumptions by examining the relationship between patient and staff ages and their opinions of mobile device use.

Our primary research question was, "Does mobile device use influence patient or staff opinions of doctors?" Our null hypothesis was that across our respondents as a whole, mobile device use would not influence group opinions either positively or negatively and that there would be no difference between patients' and staff members' views. Our secondary research question was, "Is there a relationship between what patients and staff believe devices are being used for and age?"

Methods

Overview

We used a survey of patients at the bedside and staff in their work environment. A hard copy questionnaire (Multimedia Appendix 1) was issued face-to-face by OB (who surveyed staff) and LH (who surveyed patients) on 1 day in September 2013. The respondents completed the questionnaire in front of the doctor-researcher and were given an opportunity to clarify any questions.

Setting and Sample

The setting was a trauma and orthopedics department at a teaching hospital in Wales where the authors (OB and LH) were doctors in training. Participants were a convenience sample of inpatients on 2 adult orthopedic wards and the staff working in various environments in the orthopedics department of the same hospital. These environments included the operating theater, fracture and elective outpatient clinics, inpatient wards, and secretarial and management departments. There were practical reasons for the discrepancy in the environments for the patient and staff populations: patients on the ward are easily sampled in reasonably large numbers and usually have time to talk to doctor-researchers. In the theater environment the majority of patients are anaesthetized and undergoing an operation, and it was judged inappropriate to survey patients in the orthopedics clinic. There was a high patient-to-staff ratio on the ward and ward staff comprised 2 main groups (nursing and therapy staff). Therefore, to ensure adequate staff numbers on the study date and to expand the variation of staff groups, the surveyed staff population was extended to junior and senior grades of all health care professionals in the multidisciplinary team found throughout the orthopedics surgery department. The exception to this was doctors who had completed their training, as the survey was related to opinions of doctors currently in training.

The Survey Instrument

Data on age, frequency of observed use, and main reason for device use were collected. Age was presented in 15-year ranges from "less than 25 years" to an upper age range of "85 plus years."

Frequency of observed use was classified as regularly, occasionally, and never. Suspected main reason for device use was a single tick box from a selection of 6 options (see Textbox 1).

Textbox 1. Suspected main reason for mobile device use given as options in the questionnaire.

Suspected main reason for device use:

1. Communicating with friends
2. Social media/Facebook
3. Gaming
4. Internet use for personal reasons
5. Communicating with colleagues
6. Internet/electronic textbooks/medical apps for work reasons

The first 4 options cover common uses for connected, mobile devices for nonwork reasons (the hospital in this survey had open-access Wi-Fi for both patients and staff). We recognize that options 1-3 are a subset of option 4 ("Internet use for personal reasons"). We purposefully sequenced these items so that option 4 would pick up other forms of personal Internet

use, such as shopping. Device use for professional education includes using electronic textbooks stored on a device, mobile applications (eg, a medical calculator), or accessing Internet-based medical information (eg, UpToDate). Communication with colleagues using a connected mobile device (ie, one that's connected to the Internet via Wi-Fi or a mobile telephone network) could include use of messaging systems.

Opinion Questions

To determine a respondent's opinion regarding a doctor using a mobile device in their presence, 2 questions were asked:

1. How does a doctor using a phone at work/the bedside influence your opinion of them as a professional?
2. How does a doctor using a phone at work/the bedside affect your personal opinion of them?

These were closed questions with the response options of "positively" or "negatively." The 2 questions therefore addressed potential differences in interactions—between the professional (views on the trainee as a doctor) and the personal (views on the trainee as a person). A "don't know" option was not included. While we appreciated that this may have forced respondents to express an opinion that they did not hold, we assumed that the sample population (ie, patients over the age of 16 and staff members over the age of 18) would certainly have encountered mobile devices and would have some opinion about their use around other people. The literature on research methodology would suggest that data quality is not enhanced by the inclusion of a "don't know" option [8].

Analysis

All data were analyzed in SPSS (IBM SPSS Statistics for Macintosh, Version 20.0). Statistical tests of significance (Chi-square and Fisher's exact) were used to test difference between patient and staff responses. Correlation was tested using Pearson's 2-tailed test of significance.

Ethical Considerations

Using the National Health Service's (NHS) Heath Research Authority online decision tool [9], we determined that this project was not classified as research so approval to conduct the survey was sought from and granted by the trauma and orthopedics department. The data were collected from a volunteer sample. All participants were assured of confidentiality and anonymity. Verbal consent to participate was obtained. Privacy was enhanced by using a paper questionnaire on a clipboard that could be closed over, concealing their responses from others in the room.

Results

There were 94 respondents in total; 59 patients and 35 staff members. Five patients and 9 staff members required explanation of the opinion questions.

The modal age range for all respondents was 40 to 54 years old. For patients, the modal range was 70 to 84 years old (27% of patient respondents). There were 4 patients over age 85, 2 who were 90 years old, and 2 who were 91 years old. The modal range for staff was 40 to 54 years old (60% of staff respondents). There were no staff members over 69 years old.

The results for the frequency of observed use by patients at the bedside and staff in the work environment are shown in Table 1.

Table 1. Results for frequency of observed use of devices in the workplace/bedside by patients and staff.

Frequency of observed use	Patient % (n)	Staff % (n)	Total % (n)
Regularly	2% (1)	46% (16)	20% (17)
Occasionally	20% (10)	51% (18)	32% (28)
Never	78% (40)	3% (1)	47% (41)
Missing data	8	0	8

The results for observed frequency of use were notably different for the 2 groups: 78% of patients compared with 3% of staff never saw doctors use mobile devices, and 2% of patients compared to 46% of staff regularly saw doctors using mobile devices in the workplace. These results were significantly different (Chi-square, $P<.001$).

For all respondents, the top suspected main reasons for use was thought to be "communicating with colleagues" (48%) followed by "Internet use/applications for work reasons" (40%) (Table 2). There were no significant differences between patient and staff groups for the suspected main reason for use (Chi-square, $P=.335$). Neither patients nor staff suspected that doctors were using social media or games on their devices.

Table 2. Patients' and staff members' suspected main reason for mobile device use in the workplace.

Perceived main reason for use	Patients% (n)	Staff% (n)	All respondents% (n)
Communicating with colleagues	47% (27)	52% (17)	48% (44)
Internet/electronic textbooks/ medical apps for work reasons	43% (25)	33% (11)	40% (36)
Internet use for personal reasons	9% (5)	6% (2)	8% (7)
Communicating with friends	2% (1)	9% (3)	4% (4)
Social media/Facebook	0	0	0
Gaming	0	0	0
Missing data	1	2	3

Overall, 42% of patients' opinions of doctors as a professional were positively influenced by device use, compared with 82% of staff (Table 3). Of total respondents, their professional (57%, n=53) and personal (56%, n=52) opinions of doctors were overall positively influenced by mobile device use. There was strong correlation between the results for the 2 opinion questions (Pearson's correlation 2-tailed, $P<.001$); only 1 respondent gave differing answers.

Table 3. Influence of opinion of doctor results, by group.

Group	Influence opinion of doctor as a professional		Influence opinion of doctor as a person		Total No. (missing data)
	Positively	Negatively	Positively	Negatively	
Patient	42% (25)	57% (34)	44% (26)	56% (33)	59 (0)
Staff	82% (28)	18% (6)	76% (26)	24% (8)	34 (1)
Totals	57% (53)	43% (40)	56% (52)	44% (41)	93 (1)

The opinions of staff and patients differed greatly, with significantly more patients than staff being negatively influenced by mobile device use. This was true for both patient and staff opinions of doctors as professionals (57% vs 18%, Fisher's exact test 1-sided, $P<.001$) and their personal opinions of doctors (56% vs 24%, Fisher's exact test 1-sided, $P=.002$). Age and opinions were investigated, but due to the difference in the modal age ranges for patients and staff, these ranges were collapsed into 2 groups; under 55 years old and 55 years old and older. The patients' split between these 2 groups was relatively balanced (44%, n=26, in the under age 55 group). In contrast the majority of staff were in this group (89%, n=31). No significant differences were found in the combined patient and staff group for professional (Fisher's exact test 1-sided, $P=.063$) nor personal opinions (Fisher's Exact test 1-sided, $P=.087$). We repeated the analysis for the patient-only group, and no significant relationships were found (Fisher's exact test 1-sided, for professional opinion $P=.398$; for personal opinion $P=.291$). We did not repeat this analysis for staff due to low numbers in the over age 55 group.

Discussion

Principal Findings

It would appear from this study that most patients in the orthopedics wards are not observing doctors in training using mobile devices at the bedside, whereas staff members on the wards and in other work environments are seeing them used regularly. This would fit with the concerns that doctors have about using mobile devices in front of patients [1] and also the workplace environment, which means that staff members, including doctors, share the same nonclinical work spaces where mobile devices are commonly used. When they do see mobile devices being used, the majority of both patients and staff believe that their use is for work-related communication (48%) or educational reasons (40%). It was encouraging that no respondents thought mobile devices were being used for gaming or social media, but it is important to remember that the uses were suspected or perceived by the respondents and not actual, observed use. Unless mobile device use is being directly observed in close proximity (such as looking over a user's shoulder) it is almost impossible to tell if a Web browser is showing Facebook or an online textbook. This result would suggest that when both patients and staff members do see doctors using mobile devices, they assume that their use is work-related. The authors suspect that in real life, this may not always be the case.

While for approximately 40% of patients their opinion of the doctor was positively influenced by mobile device use, this result was half that of the staff group (82%). In this study, the majority of patients' opinions were negatively influenced by device use. Compared to staff, patients were significantly more likely to have their opinions of doctors both as professionals and as people negatively influenced by mobile device use. The patient group was more heterogeneous than the staff group, who as health care professionals were more likely to be using mobile devices in the workplace for the same reasons as doctors. They may have greater insight into how mobile devices can support doctors' work and, therefore, would be less likely to form negative opinions of their use.

The relationship between the influence of device use on the respondents' opinions of doctors as professionals and as people would suggest that these opinions are similar. The negative influence of device use on patients' opinions is countered by

the more positive influence it has on the opinions of colleagues. The lack of significant relationships regarding age and opinion is noteworthy; it debunks an assumption that opinions regarding mobile device use are age related.

Strengths

One of the strengths of this study is that it is rare for patients' opinions to be sought on these matters. Much of the current research is focused on doctors' use of mobile technology [1,2,10,11]. This is 1 of few, if any, known studies to investigate patients' (and other health care professionals') opinions. The study is suitably powered; it is generally accepted that Fisher's exact test requires samples of 30 in the groups being compared. A retrospective power calculation was performed on the approximate combined differences between patient and staff responses for the opinion questions, using the formula devised by Lehr [12]. If the power of the proposed hypothesis test is fixed at 80% and the level of significance of the 2-tailed test set at 5%, the number needed in each group is 25.

Limitations

The limits of this study are the narrow population group; 1 department in 1 hospital in Wales and a convenience sample of both groups. It would be inappropriate to extrapolate the results to all patients and all doctors in the NHS in the United Kingdom. There is also an element of researcher bias, as a result of the face-to-face distribution of the survey. The researcher (OB) was known to all staff participating in the study. Bias was minimized by emphasizing that there were no right or wrong answers and ensuring confidentiality and anonymity. Five patients and 9 staff members required explanation of the opinion questions. The most common comment on the survey instrument was about the lack of a "don't know/no opinion" option. While there were no refusals to participate among patients, not all staff members in each environment visited on the study day were available to participate, and 1 staff member refused to answer the opinion questions.

Conclusions

Observing doctors using mobile devices was viewed negatively by a majority of patients, but positively by most staff members. Yet the majority of respondents thought that the main reasons for mobile device use was for work-related information retrieval and communication. No respondent thought doctors were using devices inappropriately for gaming or social media. Given the perception of appropriate mobile device use, it is interesting that this did not positively influence the opinions of patients. These opinions about device use showed no relationship to the age of the respondent. This is important. It reinforces the danger of making age-specific assumptions.

Doctors are using mobile devices at work but not in front of patients. Does the discrepancy between perceived appropriate use of devices and negative influence on patient opinions mean that patients need a better understanding of mobile technology in the workplace? Can doctors play a role in this? We recommend that doctors continue to be advised to be mindful of the etiquette regarding mobile device use in front of patients and colleagues.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey instrument.

[PDF File (Adobe PDF File), 39KB - mhealth_v3i2e71_app1.pdf]

Multimedia Appendix 2

Presentation from Medicine 2.0 2014, Malaga.

[PPTX File, 6MB - mhealth_v3i2e71_app2.pptx]

References

1. Hardyman W, Bullock A, Brown A, Carter-Ingram S, Stacey M. Mobile technology supporting trainee doctors' workplace learning and patient care: an evaluation. BMC Med Educ 2013;13:6 [FREE Full text] [doi: 10.1186/1472-6920-13-6] [Medline: 23336964]
2. Davies BS, Rafique J, Vincent TR, Fairclough J, Packer MH, Vincent R, et al. Mobile Medical Education (MoMEd) - how mobile information resources contribute to learning for undergraduate clinical students - a mixed methods study. BMC Med Educ 2012;12:1 [FREE Full text] [doi: 10.1186/1472-6920-12-1] [Medline: 22240206]
3. Ellaway R, Masters K. AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment. Med Teach 2008 Jun;30(5):455-473. [doi: 10.1080/01421590802108331] [Medline: 18576185]
4. Wallace S, Clark M, White J. 'It's on my iPhone': attitudes to the use of mobile computing devices in medical education, a mixed-methods study. BMJ Open 2012;2(4) [FREE Full text] [doi: 10.1136/bmjopen-2012-001099] [Medline: 22923627]
5. Bullock A, de Jong JP. Technology-enhanced learning. In: Swanwick T, editor. Understanding medical education: Evidence, theory and practice. Oxford: Wiley-Blackwell; 2014:149-158.

6. General Medical Council. Good Medical Practice. In: GMC Publications: London. Manchester: General Medical Council; 2013.
7. Raines C. Managing millennials. In: Connecting Generations: The Sourcebook for a New Workplace. Menlo Park, CA: Crisp/Course Technology; 2003:172.
8. Bryman A. Social research methods. Oxford: Oxford University Press; 2008:244.
9. National Health Service. Heath Research Authority decision tool Web site URL: http://hra-decisiontools.org.uk [accessed 2015-06-17] [WebCite Cache ID 6ZNZZBBrH]
10. Bullock A, Fox F, Barnes R, Doran N, Hardyman W, Moss D, et al. Transitions in medicine: trainee doctor stress and support mechanisms. Journal of Workplace Learning 2013 Aug 02;25(6):368-382. [doi: 10.1108/JWL-Jul-2012-0052]
11. Bullock A. Does technology help doctors to access, use and share knowledge? Med Educ 2014 Jan;48(1):28-33. [doi: 10.1111/medu.12378] [Medline: 24330114]
12. Petrie A. Statistics in orthopaedic papers. J Bone Joint Surg Br 2006 Sep;88(9):1121-1136 [FREE Full text] [doi: 10.1302/0301-620X.88B9.17896] [Medline: 16943459]

Edited by G Eysenbach; submitted 09.12.14; peer-reviewed by S Murthy; comments to author 03.03.15; revised version received 07.04.15; accepted 27.04.15; published 26.06.15.

Please cite as:
Blocker O, Hayden L, Bullock A
Doctors and the Etiquette of Mobile Device Use in Trauma and Orthopedics
JMIR mHealth uHealth 2015;3(2):e71
URL: http://mhealth.jmir.org/2015/2/e71/
doi:10.2196/mhealth.4122
PMID:26116061

Original Paper

Testing the Feasibility and Psychometric Properties of a Mobile Diary (myWHI) in Adolescents and Young Adults With Headaches

Anna Huguet[1,2*], PhD; Patrick J McGrath[1,3,4*], PhD; Michael Wheaton[1*], Binf; Sean P Mackinnon[5*], PhD; Sharlene Rozario[1*], BSC; Michelle E Tougas[1,5*], MSc; Jennifer N Stinson[6,7*], PhD; Cathy MacLean[8*], MD

[1]IWK Health Centre, Centre for Research in Family Health, Halifax, NS, Canada

[2]Dalhousie University, Department of Community Health and Epidemiology, Halifax, NS, Canada

[3]Dalhousie University, Departments of Community Health and Epidemiology, Pediatrics and Psychiatry, Halifax, NS, Canada

[4]Nova Scotia Health Authority, Halifax, NS, Canada

[5]Dalhousie University, Department of Psychology and Neuroscience, Halifax, NS, Canada

[6]The Hospital for Sick Children, Child Health Evaluative Sciences, Toronto, ON, Canada

[7]University of Toronto, Lawrence S Bloomberg Faculty of Nursing, Toronto, ON, Canada

[8]Memorial University of Newfoundland, Faculty of Medicine, Health Sciences Centre, St John's, NL, Canada

[*]all authors contributed equally

Corresponding Author:
Anna Huguet, PhD
IWK Health Centre
Centre for Research in Family Health
5850/5980 University Avenue, 8th Floor Children's Building, Room K8527, PO Box 9
Halifax, NS, B3K 6R8
Canada
Phone: 1 902 470 7521
Fax: 1 902 470 6534
Email: anna.huguet@iwk.nshealth.ca

Abstract

Background: Headaches are prevalent among teens and young adults. Self-monitoring is essential for managing headaches and can be accomplished with the help of electronic headache diaries. An increasing number of electronic headache diaries exist, yet the absence of quality standards compromises their use for research and clinical purposes.

Objective: Our goal was to develop and test the usability, feasibility, and psychometric properties of an electronic diary iPhone application for self-monitoring by adolescents and young adults with headaches.

Methods: We used an iterative participatory design to develop and test our electronic headache diary. Participants aged 14-28 years old with recurrent headaches were recruited internationally. Screening and consent were conducted online. Following completion of an online pre-questionnaire, participants downloaded the diary to use in their natural environment for 14 days. An online post-questionnaire was completed following testing. The diary's usability and feasibility were tested first and determined to be complete when improvements to the diary did not result in a statistically significant impact on indicators of feasibility and adherence. Interviews were conducted with participants of usability and feasibility testing. The psychometric properties of the diary were then tested, and a case study analysis of one participant was completed.

Results: Three cycles to test the usability and feasibility were conducted. Each cycle included 11-19 unique participants ranging in age from 16 to 28 years. Following the testing period for each cycle, 15% to 25% of participants took part in the post-cycle interview. Participants perceived the final version of the diary as useful, easy to learn, and efficient to use. Psychometric properties were then tested with a sample of 65 participants (6 aged 14-17 years old; 59 aged 18-28 years old). All items in the diary had substantial between- and within-subjects variability (percent of variance for the two participant groups ranged from 20.64 to 75.60 and 23.74 to 79.21, respectively). Moreover, the Migraine Disability Assessment (MIDAS) included in the diary had adequate between-subjects reliability (R1F=0.66, RKF=0.98), but low within-subjects reliability (RC=0.51). Critical elements of the diary demonstrated adequate convergent and concurrent validity, particularly in the older age group (18-28 years). The validity of some critical elements of the diary could not be explored in the younger age group due to the small subgroup size. The case study provides an example of the potential utility of the diary.

Conclusions: Our electronic headache diary was shown to be a usable and feasible self-monitoring tool when used by adolescents and young adults with headaches for 14 days. This study provides preliminary support of its psychometric properties. Our diary has the potential for helping users to better understand their headaches and, consequently, to change behaviors to improve self-management of their headaches. Its effectiveness as a component of an intervention will be the focus of future research.

(JMIR mHealth uHealth 2015;3(2):e39) doi:10.2196/mhealth.3879

KEYWORDS

headache; diary; smartphone; feasibility; psychometric properties

Introduction

Tension-type headache (HA) and migraine are ranked as the second and third most common diseases in the world, affecting 20.8% and 14.7% of the world population, respectively [1]. Migraine is ranked as the eighth leading cause for disability [1]. Due to the burden of these conditions, effort toward improving care is warranted.

Consistent with the International Headache Society guidelines [2], health care professionals often advise diary use to self-monitor headaches [3,4]. Self-monitoring enables recognition of temporal behavior patterns, allows individuals to become informed and actively self-manage headaches, facilitates treatment decision-making and treatment tailoring, and offers a measure of treatment efficacy [5]. Self-monitoring is particularly useful for people with recurrent headaches, whose episodes usually occur in response to unrecognized triggers [6]. Diaries can help individuals understand headache patterns and identify triggers [7], which is a basic treatment strategy for headaches [8,9]. Behavioral management of these triggers can result in fewer headaches [10]. Findings from meta-analyses indicate that behavioral interventions, usually including self-monitoring, are also effective at reducing headaches [11].

In addition to the clinical advantages of self-monitoring with headache management, diaries also have research benefits. Diaries allow researchers to test hypotheses of within-subject relations over time as an extension of prior cross-sectional research (eg, finding whether headache episodes more likely to occur in the context of a putative trigger) [12,13].

While paper diaries have long been used for self-monitoring headaches [14], advances in technology have afforded widespread use of electronic diaries (e-diaries) [10,15-18]. E-diaries offer several advantages, including increased adherence, accuracy, acceptability, and efficiency [14,18-21]. Two recent systematic reviews identified 5 e-diaries used in research [22] and 38 in Canadian mobile app stores for iOS and Android platforms [23]. The quality of these self-monitoring tools is questionable in the absence of any existing standards. Current headache e-diaries have several limitations, including not using a participatory design [24], recording insufficient data to provide understanding of headache patterns, lacking evidence that demonstrates feasibility and psychometric properties, and lacking research on the impact of these diaries on health outcomes [11,22].

Our overall goal was to create a usable, feasible, and psychometrically sound electronic headache diary for people aged between 14 and 28 years old who have recurrent headaches that also overcame some of the identified weaknesses of existing diaries. The specific objectives of this study were to: (1) test and improve the usability and feasibility of a new iPhone-based diary in terms of adherence, learnability, acceptability, efficiency, and accuracy through the use of iterative cycles; (2) test the psychometric properties (eg, reliability/convergent and concurrent validity) of this diary when used for assessment purposes; and (3) illustrate its potential utility with a case study analysis.

Methods

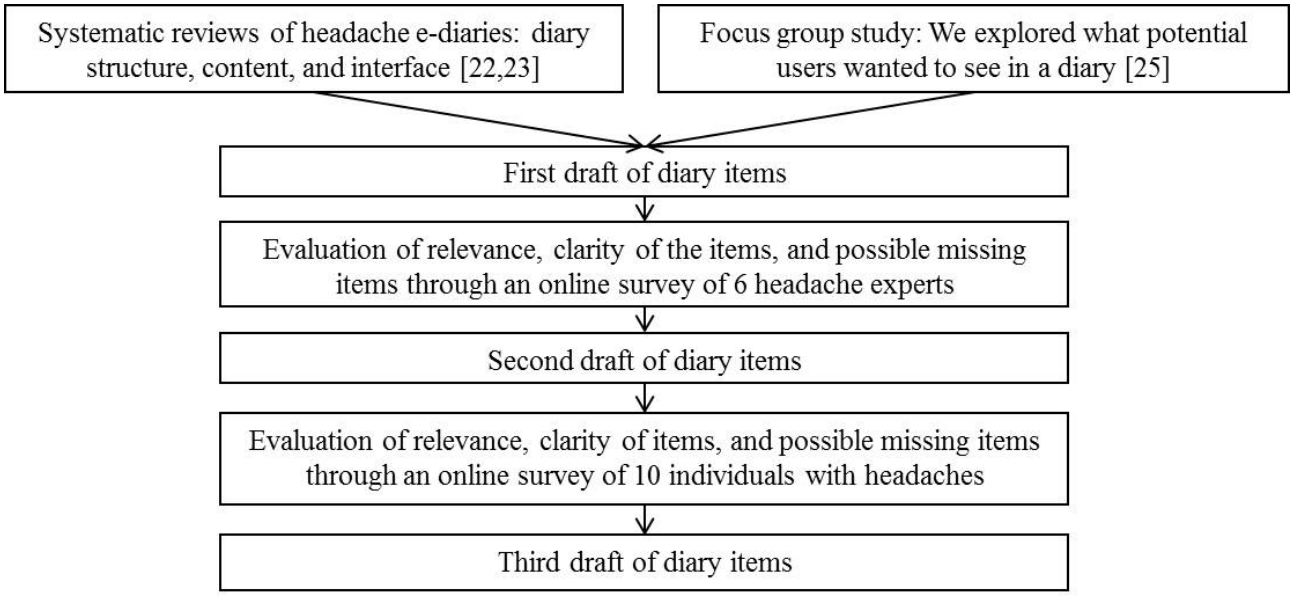
Overview

Participants were eligible if they self-identified as: (1) being aged 14-28; (2) having proficiency with English speaking, reading, and writing; (3) having a headache frequency of 2 or more episodes each month for the last 3 months; and (4) owning an iPhone with a data plan or wireless Internet access. Participants were excluded if they self-identified as having: (1) cognitive and/or developmental delays; (2) not visited a physician to exclude an organic disorder or traumatic injury as the cause for their headaches; and/or (3) significant visual impairment or blindness. Participants who consented but did not complete the pre-assessment questionnaire were also excluded.

Concept and Development Process of the Electronic Headache Diary

Several steps were taken to create our electronic headache diary application (see Figure 1). Once diary items and features were well-defined, an iterative process that involved 3 cycles of designing, testing, reviewing, and refining the e-diary app was followed. Initially, low-fidelity paper diary prototypes, followed by high-fidelity software-based diary prototypes, were tested with volunteers in the lab. Potential end users were then involved in 3 iterative cycles of testing high-fidelity, software-based diary prototypes in participants' natural environments. The final high-fidelity prototype was used to test the psychometric properties of the diary. The evaluation of the diary with potential end users during the iterative cycle phase is the focus of this manuscript.

Figure 1. Defining the concept of myWHI diary.



Evaluation of the Usability, Feasibility, and Psychometric Properties of the Diary

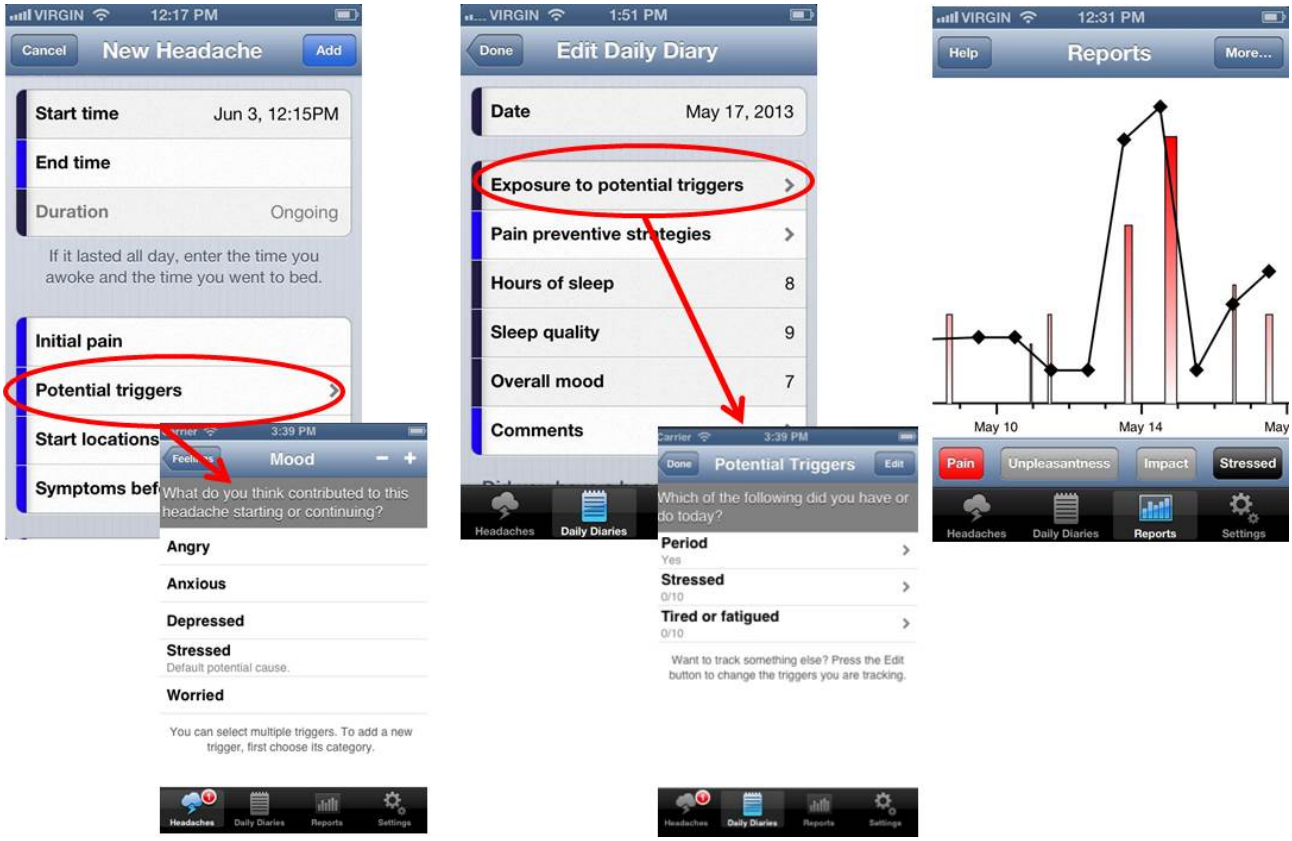
Participants involved in usability/feasibility testing and psychometric properties testing followed the same procedure unless indicated. Participants were recruited internationally through online advertisements on social networks, classified ads, and mailing lists. We used an automated Web-based system for efficient online screening and consent that closely mimicked the electronic signup process that will be followed when the application is released to the public. Potential participants were asked screening questions by the automated system. Those identified as eligible were automatically asked to provide online consent (assent and parent authorization if between 14-16 years old). If consent was provided, an online pre-questionnaire was completed. The researcher then emailed each new participant with instructions for downloading the iPhone application along with the request to use the diary for 14 days. After this testing period, an online post-questionnaire was completed and 15% to 25% of participants involved in the usability/feasibility testing were randomly selected for a 45-minute, end-of-study interview via phone or Skype. Participants involved in psychometric testing were not interviewed. Participants were reimbursed for their time. This study was approved by the IWK Health Centre's Research Ethics Board.

Measures

Electronic Headache Diary

Our electronic headache diary was called the myWireless Headache Intervention diary (myWHI diary). It was designed to help users become more aware of headache symptoms and patterns. It tracks temporal, sensory, and affective aspects of headaches, and headaches' impact on daily life, potential triggers, and coping behaviors. The diary incorporates ad-hoc and validated paper measures (ie, MIDAS/PedMIDAS and NRS-11). It includes outcome measures recommended by IMMPACT, PedIMMPACT [25-27], and guidelines for both pharmacological and behavioral clinical trials of headache [28-30] to facilitate use of the diary in a scientific trial context. The diary encourages users to complete a headache entry for every headache. First, users report occurrence by specifying start time, initial intensity, starting location, potential trigger(s), and prior symptoms (Figure 2.1). At the end of the headache, users are encouraged to report ending time, headache quality, highest pain intensity, level of unpleasantness, associated symptoms, change in headache location, and medication taken or strategies utilized to cope with the headache. At the end of each day, regardless of whether they had a headache, users are asked to enter additional information into the daily diary (eg, overall mood, hours and quality of sleep; see Figure 2.2). In addition, participants record headache impact on daily activities if a headache occurred that day. As all items are optional, users can keep track of information most relevant to them. The diary provides visual graphs of headaches over time in terms of occurrence, intensity, duration, and level of headache-related interference in daily functioning. To highlight potential causal relationships, users can also see how these parameters are related with the tracked potential triggers (Figure 2.3).

Figure 2. Screenshots of the myWHI diary interface. 2.1 Headache entry 2.2 Daily diary entry 2.3 Reports.



Pre-questionnaire

A closed-ended questionnaire gathered demographics, iPhone usage, and headache characteristics. Based on headache characteristics and considering the International Headache Classification (IHC) criteria [31], participants were classified as having migraine-like headaches, tension-like headaches, mixed headaches (meeting criteria for both migraine-like and tension-like headaches), or unclassified headaches.

Post-questionnaire

A closed-ended questionnaire was administered to participants of usability/feasibility testing and psychometric testing. The questions posed to the 2 testing groups were different, as explained below.

Post-Questionnaire for Usability/feasibility Testing

This questionnaire evaluated the diary in terms of helpfulness, usefulness, efficiency, visual appeal, and ability to be understood through ad-hoc questions.

Post-Questionnaire for Psychometric Testing

This questionnaire included the following standardized tools:

1. The Numerical Rating Scale (NRS-11): An 11-point self-report pain intensity scale commonly used. It has been psychometrically tested in children aged 8 years and above, adolescents [32], and adults [33,34].
2. The Mental Health Inventory-5 (MHI-5): A 5-item self-report measure to assess psychological distress during the past month [35]. A higher score indicates better mental health. MHI-5 has repeatedly been shown to be valid and reliable in adolescents and adults [36,37].
3. The Patient Reported Outcomes Information Measurement System (PROMIS) Short Form—Sleep Disturbance (adult version): A measure developed to assess the qualitative aspects of sleep. Its psychometric properties have been shown to be valid for adults [38]. This scale was administered to participants aged 18 years and over.
4. The PROMIS Short Form—Pain Interference (adult version): An 8-item self-report measure to assess pain-related interference with: (1) physical functioning, (2) psychological functioning, and (3) social functioning. It has been tested in adults [39]. This scale was administered to participants aged 18 years and over.
5. The Sleep/Wake Behavior Problems Scale of the Sleep Habits Survey, a scale that evaluates erratic sleep/wake behaviors [40], and the PROMIS Pediatric Short Form—Pain Interference (child version), an 8-item self-report measure to assess pain-related interference with functioning [41], were administered to participants between age 14 and 17 years old. However, this information was not used in our analyses due to the reduced number of participants between the ages of 14 and 17 years old (see results).

End-of-study interview

A semi-structured interview administered to 15% to 25% of participants involved in each cycle of usability/feasibility testing was facilitated by one researcher guided by a script. The questions focused on the participants' experiences using the

diary, including barriers, usefulness, burdens, goals, and suggestions for improvement.

Data Analytic Strategy

Usability and Feasibility Testing of the Diary

Analysis was exploratory in nature. Using SPSS 20 predictive analytics software, descriptive statistics (median and range) were calculated for pre- and post-assessment data. Usage of the diary was automatically collected from the system. Mann-Whitney U and Chi-square tests were used to evaluate differences between consecutive testing cycles for continuous and categorical variables used as indicators of feasibility and adherence. We used non-parametric methods and report the median (mdn) instead of the mean because of our small sample size. Qualitative data collected through end-of-study interviews were transcribed, coded, and analyzed using an inductive thematic analysis [42].

Psychometric Properties Testing of the Diary

We hypothesized that the diary would be reliable and the most essential components of the diary would be valid. We hypothesized that convergent construct validity would be supported by high correlations between the data derived from the items of the diary that assess headache occurrence, intensity, unpleasantness, mood, and sleep and headache impact, with data obtained through questions asking for the same information retrospectively mostly using well-established single-point measures. We also hypothesized that if the most essential parts of diary had concurrent validity then a good number of variables collected through the diary would be correlated with other single-point measures that assess related constructs. Specifically, we hypothesized that the total number of headache episodes, total headache time, highest headache intensity, and unpleasantness of headaches would be moderately to strongly related to levels of pain-related interference, sleep-related impairment, and emotional functioning as assessed retrospectively through well-established single-point measures.

Reliability of most variables measured by the diary was assessed using generalizability theory analysis. Following Cranford et al [43] we used generalizability theory to decompose the variance of daily variables and to calculate reliability estimates using VARCOMP procedures in SPSS 20. For single-item variables, variance was decomposed into person, day, and person-by-day components. The person component represents the between-subjects portion of the variance that remains stable across all 14 days, and the person-by-day component represents the within-subjects component that varies by day. Thus, a variable with a high proportion of person variance is highly stable over time, which is analogous to having high test-retest reliability. Because MIDAS/PedMIDAS, which measures the impact of headache on the subject's daily functioning, is a multi-item measure, the variance was decomposed into person, day, item, person-by-day, person-by-item, day-by-item, and error variability. This information is used to calculate 3 estimates of between- and within-subjects reliability [43]. R_{1F} is equivalent to calculating Cronbach's alpha for each of the 14 days separately, and taking the average of all 14 alphas. R_{KF} values are equivalent to averaging all 14 days of data into a single composite index, then calculating Cronbach's alpha once. R_C values represent within-subjects reliability, which represents the precision of the measure for measuring systematic variance from day-to-day.

The convergent construct validity and criterion validity of the diary were assessed by correlating several items of the diary with well-established 1-point measures. Because diagnostics (ie, kurtosis, skewness, histograms, P-P plots) suggested many variables were non-normally distributed, we used Spearman's rank-order correlation coefficient (r_s). Cohen's criterion (small = 0.10; medium = 0.30; large = 0.50; Cohen et al [44]) was used to interpret effect sizes. Variables from the diary were combined across all 14 days into a single value per participant using various functions (eg, sums, averages, maximum value) to facilitate comparison with variables measured only once. A full list of all convergent and criterion validity tests can be found in Table 7. The minimum required sample size to achieve 80% power assuming to find at least medium-to-large effect size (r_s=0.40) and an alpha of 0.05 was 46 people.

Case Study Analysis

A representative example of a participant of psychometric properties testing was selected.

Results

Table 1 provides the descriptive statistics of the participants involved in evaluation of the diary. In the usability/feasibility testing cycles, the majority of participants were young adult females with median ages of 24 to 25 years old (range, 16-28 years). The most frequent type of headache was migraine-like headache. In psychometric testing, the median age was similar, 22 years old (range, 14-28); only 6 of the 65 participants were 14 to 17 years old, and 59 were 18 to 28 years old. As in usability/feasibility testing, the majority of participants were female and the most common type of headache was migraine-like headache.

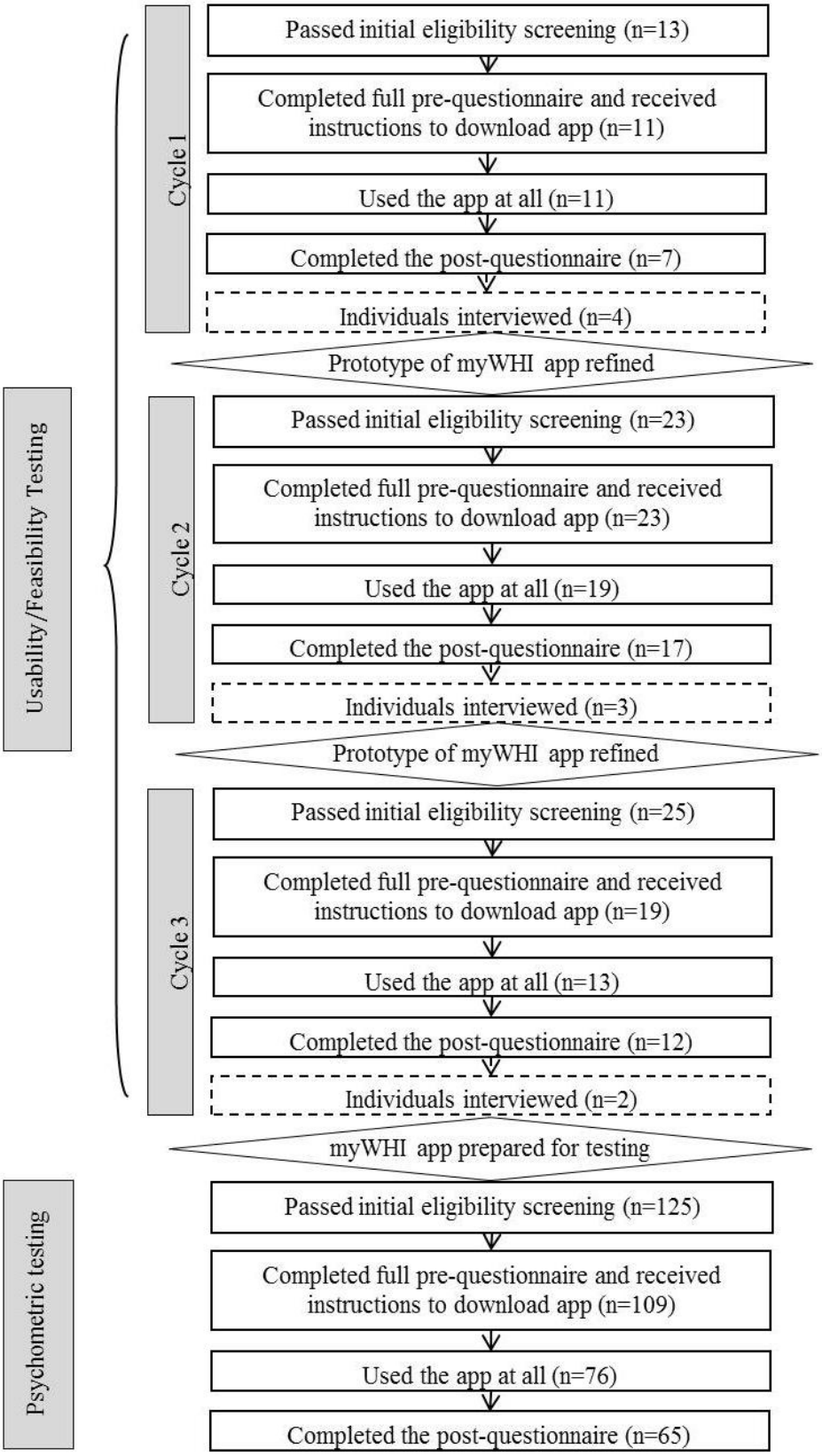
Table 1. Descriptive statistics of participants involved in the evaluation of the myWHI diary.

		Usability/Feasibility testing (n=43)			Psychometric testing (n=65)
		Cycle 1 (n=11)	Cycle 2 (n=19)	Cycle 3 (n=13)	
Age					
	Median years (range)	25 (16-28)	24 (16-28)	25 (20-28)	22 (14-28)
Gender					
	Males	1	4	1	9
	Females	10	15	12	56
Headache diagnosis[a]					
	Migraine-like headache	9	14	11	46
	Tension-like headache	0	0	0	3
	Mixed headache	0	0	0	0
	Unclassified headache	1	1	0	1
	>1 headache diagnosis	1	4	2	8

[a]*Migraine-like headache*: Headache with at least 2 of the following: (a) unilateral location, (b) pulsating quality, (c) moderate or severe pain intensity, (d) aggravation by or causing avoidance of routine physical activity. And during the headache, experiencing one of the following: (a) nausea and/or vomiting, (b) photophobia and phonophobia. *Tension-like headache*: Headache with at least 2 of the following: (a) bilateral location, (b) pressing or tightening (non-pulsating) quality, (c) mild or moderate intensity, (d) not aggravated by routine physical activity such as walking or climbing stairs. And during the headache experiencing both: (a) no nausea or vomiting, (b) no more than one of photophobia or phonophobia. *Mixed headache*: Headache meeting all criteria to be classified as both migraine-like and tension-like headache. *Unclassified headache*: Headache that does not meet all clinical categories outlined in the IHC criteria and defined above to classify headaches.

Figure 3 shows the flowchart of participation in the evaluation of the diary. Results of usability/feasibility testing are based on participants who used the diary application at least once during Cycle 1 (n=11), Cycle 2 (n=19), and Cycle 3 (n=13). Results of psychometric testing are based on participants who completed the post-questionnaire (n=65).

Figure 3. Flowchart of participation in the evaluation of the myWHI diary.



Usability and Feasibility Testing of the Diary

Adherence

Table 2 shows indicators of adherence. In Cycle 1, headache entries tended to be entered a long time after reported start time (mdn=13.59 h; range, 1.73-101.28 hours). Headaches were often not recorded on the actual day that the episode was reported to have occurred. However, most important headache items were answered with each entry. A minority of participants completed all 14 daily diary entries (18%, n=2) or 75% of the 14 daily

diary entries (18%, n=2). The items encouraged to be answered every day were usually entered the day that the daily entry was made. Following Cycle 1, changes were made to the diary to make the application clearer, which we expected would increase adherence. The most significant changes are shown in Table 4.

In Cycle 2, statistically significant improvements in adherence with headache entries were found following refinements to the first prototype. Participants in Cycle 2 completed their headache entries closer to the time pain began than did participants in Cycle 1 (Cycle 2 mdn=3.83 h; range, 0.09-19.92 hours, vs Cycle 1 mdn=13.59 h; U=32.00, z=-3.01, *P*=.003). Adherence with the daily diary entries also improved, but did not reach statistical significance (26% of participants, n=5, completed all 14 daily entries in Cycle 2 vs 18% of participants, n=2, in Cycle 1; $\chi^2_{(1)}$=0.26, *P*=.69; 53% of participants, n=10, completed 75% of the 14 daily diary entries in Cycle 2 vs 18% of participants, n=2, in Cycle 1; $\chi^2_{(1)}$=3.44, *P*=.12). As observed in Cycle 1, participants in Cycle 2 also tended to answer all of the items when completing a headache or daily entry (see Table 2). Following Cycle 2, minor changes were made to the diary primarily to increase adherence (see Table 4).

In Cycle 3, the level of participant adherence utilizing the diary for headache entry remained acceptable with no statistically significant differences found between Cycle 2 and Cycle 3 (see Table 2). As observed in Cycle 2, the majority of participants' headache entries during Cycle 3 were made on the same day that the episode occurred. Once participants created the headache entry, they tended to report initial information about their headache right away. The level of adherence of participants utilizing the diary for entering the daily diary entries was not statistically different from Cycle 2. Participants completed the majority of daily entries in real-time with only a minority of daily entries entered retrospectively. Because significant improvements in feasibility indicators of the diary were not observed in Cycle 3, we decided not to make further changes. This was the final version of the diary used to test the psychometric properties.

Table 2. . Median level of adherence with the myWHI diary for Cycles 1-3[a].

			Cycle 1 (n=11)	Cycle 2 (n=19)	Cycle 3 (n=13)
Headache (HA) entries					
	HA entries made the same day that HA was reported to happen, per participant		57% (0%-100%)	100%[c] (23%-100%)[b]	82%[c] (31%-100%)[c]
	Times that the following items were answered when entering an HA episode, per participant				
		Initial intensity	100% (56%-100%)	100% (0%-100%)[c]	100% (50%-100%)[c]
		Start location	100% (56%-100%)	100% (25%-100%)[c]	100% (0%-100%)[c]
		Potential triggers	99% (50%-100%)	100% (0%-100%)[c]	85% (0%-100%)[c]
		Highest intensity	100% (56%-100%)	100% (50%-100%)[c]	100% (50%-100%)[c]
Daily diary entries					
	Daily diary entries made		42.8% (0%-100%)	78.57% (21.43%-100%)[c]	85.71% (0%-100%)[c]
	Daily entries made retrospectively for another day, per participant		22% (0%-50%)	0% (0%-50%)[c]	14% (0%-73%)[c]
	Times that the following items were answered when entering a daily diary, per participant				
		Hours of sleep	100% (20%-100%)	100% (86%-100%)[c]	100% (29%-100%)[c]
		Sleep quality	100% (80%-100%)	100% (79%-100%)[c]	100% (29%-100%)[c]
		Overall mood	100% (100%-100%)	100% (77%-100%)[c]	100% (29%-100%)[c]
		Pain interference	61% (0%-100%)	75% (0%-100%)[c]	100% (50%-100%)[c]
		Pain unpleasantness	100% (56%-100%)	100% (50%-100%)[c]	100% (40%-100%)[c]
		Pain qualities	100% (56%-100%)	100% (50%-100%)[c]	100% (50%-100%)[c]

[a]Reported values are the median percentage values across Cycles. The ranges of all reported values are presented in parentheses.

[b]Difference between values for consecutive Cycles (Cycle 1 vs Cycle 2; Cycle 2 vs Cycle 3) is statistically significant (P<.05).

[c]Difference between values for consecutive Cycles (Cycle 1 vs Cycle 2; Cycle 2 vs Cycle 3) is not statistically significant.

Usage of Diary Features

Table 3 summarizes how participants used the features of the diary.

Table 3. Usage of diary features for Cycles 1-3.

		Cycle 1 (n=11)	Cycle 2 (n=19)	Cycle 3 (n=13)
Web reports				
	% (#) visited reports	N/A	0% (0)	0% (0)
Reminders				
	% (#) disabled reminders	9% (1)	5% (1)	7% (1)
	% (#) modified reminder time	36% (4)	47% (9)	31% (4)
	Range of reminder time	5:00pm-10:30pm	9:00pm-11:00pm	7:00pm-10:00pm
Customizable list of potential triggers to record daily				
	% (#) edited default trigger list	73% (8)	58% (11)	54% (7)
	Mdn # (range) of potential triggers concurrently recorded, per participant	2 (0-4.2)	3.62 (2-21.5)	3.67 (2-9.2)
	Top 5 most common selected triggers per Cycle: % (#) of participants who tracked each trigger	Stressed: 91% (10)	Stressed: 100% (19)	Stressed: 85% (11) [a]
		Tired/fatigued: 83% (9)	Tired/fatigued: 100% (19)	Tired/fatigued: 77% (10)
		Period: 37% (4)	Period: 74% (14)	Period: 77% (10)
		Lack of sleep: 27% (3)	Computer use: 21% (4)	
		Caffeine, less than usual: 27% (3)	Missing meal/ hunger: 21% (4)	
Comments section				
	% (#) of participants who used the headache comments	45% (5)	50% (9)	46% (6)
	Mdn % (range) times the headache comments section used	22.22% (14.29%-50%)	40% (16.67-100)	50% (9.09-60)
	% (#) of participants who used daily diary comments	45% (5)	58% (11)	64% (7)
	Mdn % (range) times daily diary comments section used	33.33% (14.29%-100%)	27.27% (7.69%-80%)	14.28% (7.14%-50%)

[a]For Cycle 3, there were only 3 top triggers.

Learnability, Acceptability, and Efficiency

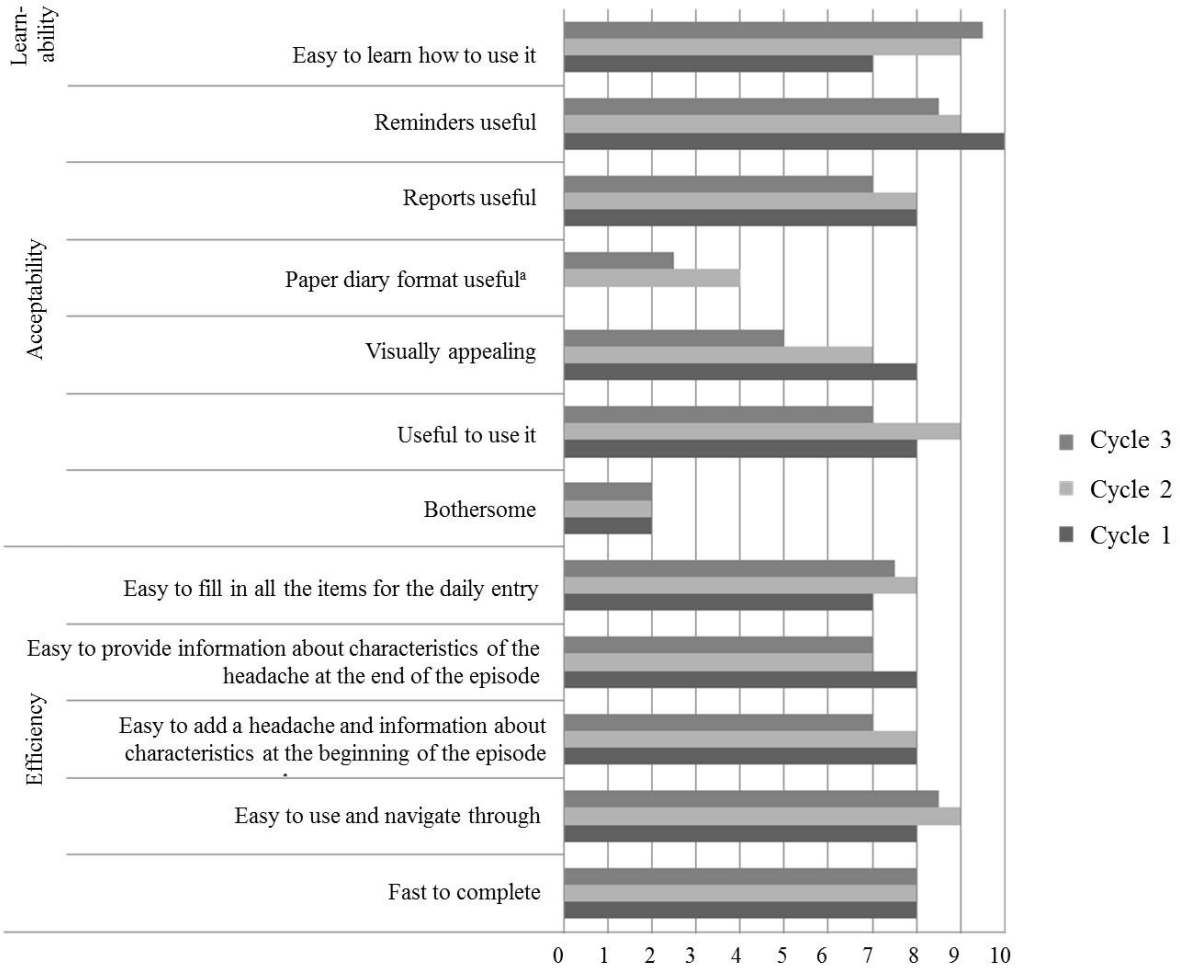
Information Collected Through the Online Post-Questionnaires

Figure 4 shows participants' opinions on attributes of the diary. In Cycle 1, 7 of 11 participants completed the post-questionnaire. In terms of acceptability, all participants reported that they would recommend this diary to others. Five participants (71%) expressed interest in continuing to use the diary (eg, for 6 more weeks). Six participants (55%) continued using the diary for 2 weeks following their 14-day trial, with 3 (27%) continuing to use it beyond that. This behavior occurred in the absence of any incentives or encouragement. In terms of efficiency, the headache diary item that was repeatedly reported to be the most difficult to complete was "potential triggers" (n=4, 57%).

In Cycle 2, 17 of 19 participants completed the post-questionnaire. Participants' opinions on the feasibility of the diary remained positive and improved when contrasted with opinions in Cycle 1 (Figure 4). The levels of acceptability and efficiency between the first 2 cycles were not statistically different, with the exception of fewer participants reporting trouble recording "potential triggers" for headache entries during Cycle 2, the item with the most problems for participants in Cycle 1 (Cycle 1: 57%, 4 of 7 participants vs Cycle 2: 0%, 0 of 19 participants; $\chi^2_{(1)}=11.66$, $P<.001$). For learnability, participants in Cycle 2 reported the diary to be easier to learn how to use than did participants in Cycle 1 (z=-2.22, $P=.03$).

In Cycle 3, 12 of 13 participants completed the post-questionnaire. The learnability, acceptability, and efficiency of the diary remained as positive as users' experience during Cycle 2, and no statistically significant improvements were found (Figure 4).

Figure 4. Participants' feedback on attributes of the myWHI diary (mdn values). aNo participant used the paper-format diary in Cycle 1.



Information Collected Through the End-of-Study Interviews

In Cycle 1, 4 participants were interviewed. Thematic analysis revealed 3 distinct themes:

1. Poor understanding: Many participants neglected to use features of the diary because they either did not know how the features worked, or were unaware that they existed (eg, participants were unaware that they could customize the triggers tracked on a daily basis). This lack of understanding was the prevailing message of the first cycle. As a potential solution, participants suggested the addition of tutorials.

2. General endorsements, likes, and dislikes: Participants felt satisfied with the diary and perceived it as useful because it taught them what to pay attention to; helped increase awareness of headaches, identification of triggers, and effectiveness of medications; helped guide self-care behaviors; and helped in reporting headaches to physicians and family. Participants felt that the diary was easy to use overall, but identified some difficult and confusing situations, such as: setting the start and end time of headaches as wake and sleep times, respectively; including not applicable items (eg, symptoms before and after the headache) for those with constant headache; having to develop the habit of using the diary; identifying potential headache triggers within the diary's hierarchical presentation; and the slowing of the application due to network connections.

3. Suggestions for improvement: To improve ease of use, participants suggested: setting default answers to daily items, allowing participants to remove unused items, adding more reminders to complete either ongoing headache entries or daily entries, and providing the most frequently entered options at the top of lists to expedite data entry. They also suggested providing additional reports and adding the ability to export diary data.

In Cycle 2, 3 participants were interviewed with 3 distinct themes identified:

1. Good understanding: Participants demonstrated a good understanding of the functionality of the application. They found the tutorials that were added after Cycle 1 to be clear and useful.

2. General endorsements, likes, and dislikes: Participants found the application easy to use, relevant, and useful. They reported that using the application gave them a more accurate idea of their headaches (ie, type of pain, intensity, and potential triggers). However, they did not speak highly of some sections or functionalities, such as: no interviewed participant reported using the Web-based reports that were added after Cycle 1, instead emphasizing the convenience of reviewing reports on the phone and indicating that they were likely to consult the

reports more often after using the diary for a longer period of time. They did consult the iPhone reports, but not frequently. Participants also disliked being unable to enter daily diary entries retrospectively, being unable to enter daily entries that extended past midnight without starting a new day, or having to enter day-long headaches for constant headache.

3. Suggestions for improvement: For ease-of-use improvements, participants suggested: adding the ability to be reminded about incomplete items in headache entries and the ability to see the day of the week when entering dates.

In Cycle 3, 2 participants were interviewed and the following theme was identified:

1. General endorsements, likes, and dislikes: Both participants liked the application, reporting that it was clear, easy to use, and useful. Despite their satisfaction with the overall application, a few features were not always perceived as useful. Whereas 1 participant used the comment section throughout headache episodes to track perceived changes, the other reported that the section was not useful. While 1 participant found the reports appealing and easy to interpret, the other found them confusing. Other aspects that were raised as a source of dissatisfaction or not used were: the loading time, the difficulty deciding what daily triggers to track due to the large number of available options, and the Web-based reports. They expressed again that access to the reports from within the mobile application would be more beneficial than accessing them through the Web.

Summary of Diary Changes

As part of the iterative design process, the diary was refined following Cycles 1 and 2 based on participant feedback. Table 4 summarized the most important changes made.

Table 4. Most important changes made to the diary during testing.

Reason for change	Change made
Cycle 1	
To improve learnability	• Added automatic help/instruction slides to explain how the application works when the user first launches the application. • Added help buttons to the headaches, daily diary, and reports tabs to view the help/instructions slides again at any time. • Moved the feature that allows users to customize the list of potential triggers from the settings area to the daily diary entries to make it clearer that the list can be modified at any time. • Added detail text (ie, “This is a default”) to the triggers setup as a default for tracking daily exposure, so that users can easily recognize the suggested default triggers. • Added a “More” button to the reports tab to explain accessing the Web-based reports. MyWHI 2.0 is the first version of the application with Web-based reports.
To improve efficiency	• Added a “Frequently used” category containing the 5 most-reported triggers. • Updated the order of medications to include the most recently used at the top, followed by those never taken ordered alphabetically. • Modified several lists of answer options to be sorted alphabetically.
To improve accuracy	• Modified the date selection to disallow selection of future dates and times.
To improve acceptability	• Added more PC-accessible-only reports to provide the user with more information from their entered data. • Added question text for “Potential triggers” to the top of the screen when searching for the appropriate response option within the hierarchical tree structure to help users to keep the question of what might trigger a headache in mind.
Cycle 2	
To improve adherence	• Added a badge to the headaches tab to indicate ongoing headaches. • Added an alert message to remind participants to create subsequent daily diary or headache entries sooner, if a diary entry is created for a previous day or a headache is entered more than 3 hours after the reported start time. • Added an alert to the “Duration” item, which is automatically reported by the system once the user has entered “Start time” and “End time.” The item turns red if the headache duration is a negative value. • Added a reminder message that shows up if headache duration exceeds 24 hours. The message reminds the participant that headaches lasting more than 1 day should be recorded as a separate headache for each day.
To improve learnability	• Added a reminder for the user to look at the reports. This shows up after the 5th, 10th, 15th, etc. daily diary entry is added until the user views reports.
To improve acceptability	• Removed the restriction on creating a diary entry if one already exists. This was preventing users from adding diary entries for previous days. • Revised the date format in headaches and daily diary entries to “Monday, June 6” instead of “June 6.”
To improve efficiency	• Changed medication search to prioritize medications that contain the search string at the beginning (eg, searching “ty” brings up “Tylenol” rather than “Atyopanex”).
To improve accuracy	• Fixed bugs (eg, fixed a bug that caused an error when saving empty MIDAS responses).

Psychometric Properties Testing of the Diary

The 65 participants completed a mean of 10.32 daily diary entries (SD=4.94) during the 14-day period. A total of 33 participants (65%) completed all 14 daily diary entries and 44 (67%) completed more than 75% of the daily diary entries.

Generalizability theory results for single-item measures recorded using the NRS-11 scale are located in Table 5. The largest portion of variance for headache start intensity, maximum intensity, and duration was explained by person variability (62.79% to 75.60%) suggesting that these variables are highly stable across individuals across 14 days. In contrast, unpleasantness, amount of sleep, quality of sleep, and mood was primarily characterized by person-by-day variability (60.27% to 79.21%) suggesting that these variables tended to vary substantially from day to day. However, all variables had substantial person and person-by-day variability suggesting that these variables capture both trait-like individual differences and state-like daily fluctuations. Reliability for PedMIDAS was not calculated due to the small sample size of users between the ages of 14 and 17 years old. Generalizability theory results for MIDAS (Table 6) shows that MIDAS has substantial person and person-by-day variance, but also has a substantial amount of measurement error. The various estimates of reliability were excellent (R_{KF}=0.98), adequate (R_{1F}=0.66), and somewhat low (R_C=0.51), depending on the measure used. These results suggest that MIDAS has good reliability if data from all 14 days are aggregated to create a single estimate for a person, but may not be suitable for measuring day-to-day variations in headache impact.

Table 5. Variance components for single-item measures of the diary.

Source of variance		Person[a]	Day[b]	Person-by-day[c]	Total[d]
Start intensity					
	0-10 NRS	2.495	0.077	1.234	3.806
	% overall variance	65.55%	2.03%	32.42%	100.00%
Highest intensity					
	0-10 NRS	3.076	0.128	1.694	4.899
	% overall variance	62.79%	2.62%	34.59%	100.00%
Duration					
	Difference between starting and ending time (in minutes)	102,960.884	891.587	32,333.475	136,185.946
	% overall variance	75.60%	0.65%	23.74%	100.00%
Unpleasantness					
	0-10 NRS	2.387	0.005	3.627	6.019
	% overall variance	39.654%	0.075%	60.27%	100.00%
Amount of sleep					
	Hours of sleep	0.969	0.007	3.332	4.207
	% overall variance	20.64%	0.16%	79.21%	100%
Quality of sleep					
	0-10 NRS	1.276	0.000	3.486	4.762
	% overall variance	26.80%	0%	73.20%	100%
Mood					
	0-10 NRS	0.955	0.002	3.023	3.980
	% overall variance	23.99%	0.06%	75.95	100%

[a]Person = variance due to between-person differences across all days.

[b]Day = variance due to differences between days across all persons.

[c]Person-by-day = variance due to between-person differences at different days.

[d]Total = sum of all variances.

Table 6. Variance components for multi-item measures of the diary.

Source of variance	Headache impactMIDAS[a]	% overall variance
Person[b]	0.060	27.91%
Day[c]	0.002	0.93%
Item[d]	0.017	7.91%
Person-by-day[e]	0.034	15.81%
Person-by-item[f]	0.005	2.33%
Day-by-item[g]	0.000	0.00%
Error[h]	0.097	45.12%
Total	0.22	100.00%
Between Subjects Reliability (R_{1F})[i]	0.66	
Between Subjects Reliability (R_{KF})[i]	0.98	
Within Subjects Reliability (R_C)[i]	0.51	

[a]PedMIDAS measure was not explored due to the small sample size that completed this measure.

[b]Person = variance due to between-person differences across all days and items.

[c]Day = variance due to differences between days across all persons and items.

[d]Item = variance due to responses to scale items across all persons and days.

[e]Person-by-day = variance due to between-person differences at different days across all items.

[f]Person-by-item = variance due to between-persons differences in responses to scale items across all days.

[g]Day-by-item = variance due to differences between days in responses to scale items across all persons.

[h]Error = Person x Day x Item interaction plus random error (unknown sources of variance).

[i]$R_{1F,}$ $R_{KF,}$ and R_C are forms of internal consistency calculated using formulas from Cranford et al (2006) [43].

Table 7 summarizes the correlations used to validate several sections of the diary. Although we administered The Sleep/Wake Behavior Problems Scale of the Sleep Habits Survey [40] and the child version of the PROMIS Pediatric Short Form—Pain Interference [41] this information was not used in our analyses due to the reduced number of participants between the ages of 14 and 17 years old (n=6). All convergent construct validity correlations were statistically significant, and 4 of 5 had large effect sizes (r_s>|0.50|). These correlations ranged from -0.26 (average sleep quality recorded with the PROMIS Sleep Disturbance scale) to 0.81 (highest headache pain recorded on the diary with retrospective recall of highest intensity). A total of 8 of the 12 concurrent validity correlations were statistically significant with medium to large effect sizes, ranging from -0.22 (average headache unpleasantness and MHI-5 emotional functioning) to 0.55 (highest headache pain intensity on the diary and the adult PROMIS Pain Interference scale). Due to a very small sample of adolescent participants (n=6), 10 planned correlations using adolescent measures (for participants between 14 and 17 years old) were omitted.

Table 7. Convergent and concurrent validity Spearman rank-order correlations for sections of the diary.

myWHI diary variables	Convergent construct validation		Concurrent validation		
How it was calculated	Post- questionnaire measures	r_s[a]	myWHI diary criteria	Post- questionnaire measures	r_s[a]
Headache entries					
Occurrence					
			Pain interference[a]	PROMIS Pain Interference	0.293[c] (n=59)
			Emotional functioning	MHI-5	-0.196 (n=57)
Total number of recorded HA episodes	Recall of number of HA episodes in past 14 days	0.630[e] (n=63)	Sleep-related impairment[b]	PROMIS Sleep Disturbance	0.165 (n=59)
Duration					
			Pain interference[b]	PROMIS Pain Interference	0.422[d] (n=51)
			Emotional functioning	MHI-5	-0.360[d] (n=56)
Sum of all headache durations in minutes	--	--	Sleep-related impairment[b]	PROMIS Sleep Disturbance	0.192 (n=51)
Highest pain					
			Pain interference[b]	PROMIS Pain Interference	0.549[d] (n=52)
			Emotional functioning	MHI-5	-0.341[d] (n=52)
Recorded worst pain intensity ratings	Recall of worst pain intensity on a NRS-11	0.809[e] (n=57)	Sleep-related impairment[b]	PROMIS Sleep Disturbance	0.369[d] (n=52)
Unpleasantness					
			Pain interference[b]	PROMIS Pain Interference	0.519[d] (n=50)
			Emotional functioning	MHI-5	-0.221 (n=55)
Average of the unpleasantness ratings	--	--	Sleep-related impairment[b]	PROMIS Sleep Disturbance	0.430[d] (n=50)
Daily diary entries					
Sleep quality[b]					
Average of sleep quality ratings	PROMIS Sleep Disturbance	-0.264[c] (n=56)	-- --	-- --	--
Overall mood					
Average of overall mood ratings	MHI-5	0.652[e] (n=62)	--	--	--
Headache impact[b]					
Average of MIDAS scores	PROMIS Pain Interference	0.693[e] (n=45)	-- --	-- --	-- --

[a]Coefficients represent Spearman rank-order correlations (r_s).

[b]Only participants aged between 18 and 28 years old were considered for these analyses.

[c]$P<.05$

[d]$P<.01$

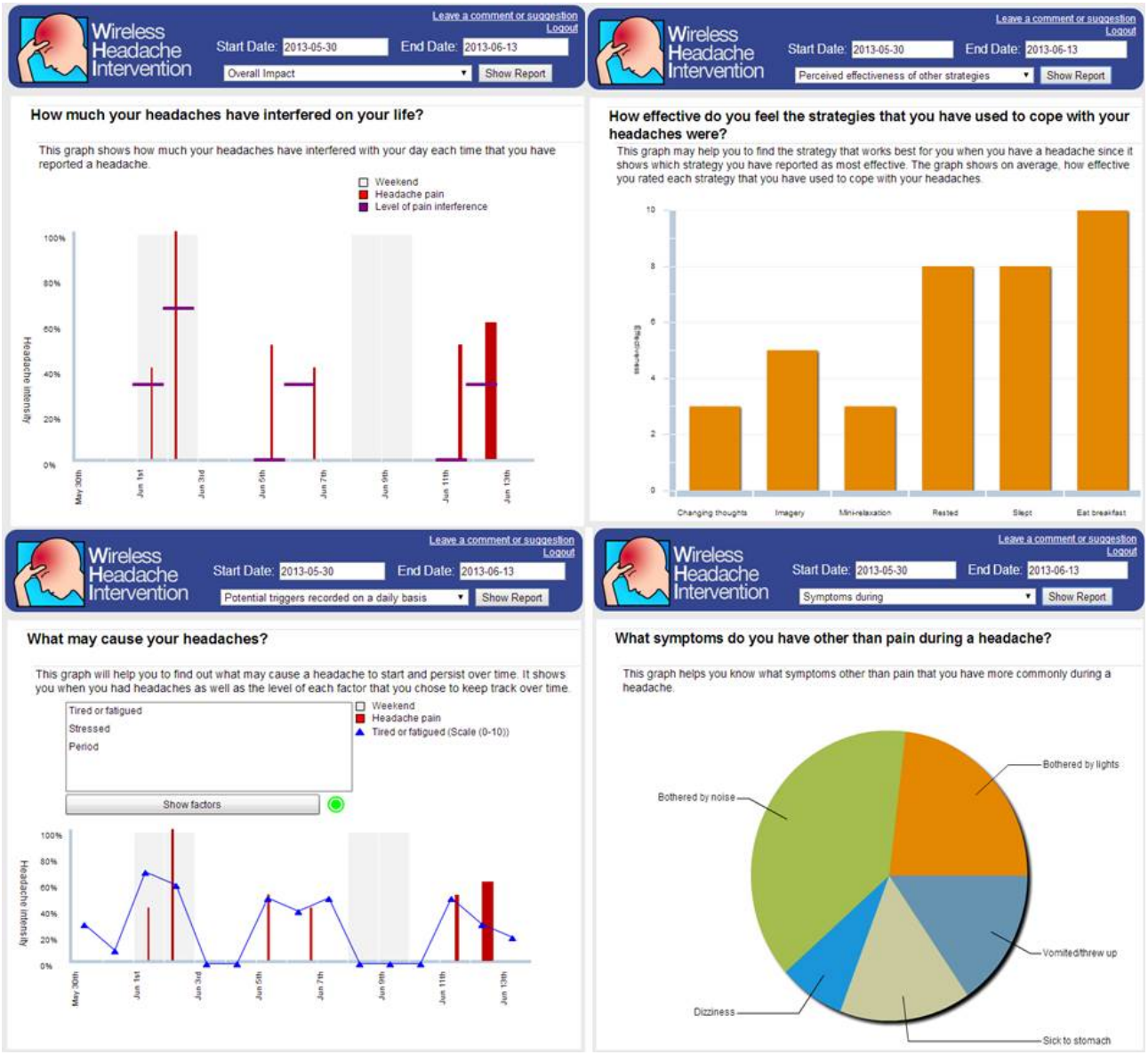
[e]$P<.001$

Case Study Analysis

A case study analysis of 1 participant of psychometric properties testing was conducted. The results are presented to illustrate how it is possible to improve the quality of self-reported data, consequently helping users and health care professionals to better understand headaches and improve health care decisions. Figure 5 shows screenshots of the Web-based reports of a

21-year-old female participant. This participant experienced migraine-like headaches, and recorded 6 episodes mainly of moderate headache pain over the 14 days. Most episodes lasted less than 1 hour, 2 episodes lasted around 2 and 8 hours, respectively, and caused mild to moderate interference in her life. The participant reported symptoms such as dizziness or hearing changes before the onset of half of the episodes, and often reported accompanying symptoms such as vomiting and light or noise sensitivity. The participant daily kept track of potential triggers, including: menstrual period, stress, and tiredness. There may be relationships between her stress and the occurrence of headache episodes, but it is too premature to extract conclusions from these limited observations. Besides taking ibuprofen on 2 occasions, which did not seem to help ease the pain in a timely manner, the participant reported different strategies to cope with her headache episodes, and perceived eating breakfast, rest, and sleep as the most effective. However, because of the limited duration of using the headache diary, again no conclusions about the effectiveness of these strategies can be drawn, however, this may be clinically helpful.

Figure 5. Screenshots of the Web-based reports of one of the participants.



Discussion

Principal Results

After 3 iterations of designing, testing, reviewing, and refining, we created a feasible electronic headache diary. At the end of this iterative process, young adults with headaches perceived the application as useful, easy to learn, and efficient to use. Although adherence with the final version evaluated in Cycle 3 was not perfect, most participants made active use of the application throughout the testing period. The option for participants to omit items when completing the diary could have had drawbacks to this research by threatening maintenance of internal consistency while diminishing the usefulness of the diary in clinical settings. However, participants typically provided an answer to each item included in the diary when either completing a daily entry or recording a headache episode during the testing period. Moreover, most participants completed a majority of the daily entries and recorded most headache episodes on the day they occurred. Substantial differences have

been observed in levels of adherence with electronic headache diaries, with rates of 75%-98% [22]. There is no consensus on the acceptable minimum adherence level in self-monitoring systems. The ultimate goal when developing a diary such as this is to provide individuals with headaches (and their health care providers) with the greatest insight into their condition while imposing minimum time and effort on maintaining a diary. This delicate balance poses a challenge. It is possible that the comprehensiveness of our application, although recognized as a strength by most participants, may have reduced adherence due to the time and effort required to complete all of the items. However, it may also take time to develop a habit of using the diary. We found that 35% of participants were still creating diary entries 2 weeks after the conclusion of their 14-day participation, and 18% created more diary entries after the study period than they had within it. This suggests that level of adherence with the diary may be maintained or even improved over time. This is most apparent with the Web-based reports, which can provide enhanced value after regular use of the application over time. It is challenging to motivate users to regularly enter data when the benefits are uncertain and meaningful reports may be weeks or even months away. Therefore, despite the possible issues with the time and effort required due to the comprehensiveness of our application, we do not plan to simplify this application, mainly because simplicity does not automatically imply that this will help patients improve adherence, and most importantly, we do not know the minimum level of adherence required to achieve better outcomes. In addition to the observed levels of participant adherence when filling in the diary, we also observed that the features incorporated in the diary (ie, the reminders, the comment sections, and the customizable list of potential triggers for the users to track their exposure on a daily basis and explore their potential relevance) were commonly used by the participants, with the exception of the Web-based report generation system that was external to the application.

We assessed reliability of measures included in the final version of the diary using generalizability theory analyses. For single-item measures, these analyses revealed that headache intensity and duration measures were primarily trait-like variables that were highly stable over time. In contrast, headache unpleasantness, amount/quality of sleep, and mood measures were more state-like, and varied considerably from day-to-day. The MIDAS had excellent internal consistency when all 14 days are combined together into a single composite variable; however, it does not appear to be a good measure for reliably measuring variability in headache-related interference from day to day. This said, poor within-subjects reliability is typical of many measures that otherwise have excellent psychometric properties [43]. Consequently, we recommend that patients and health care providers who use the diary not to examine day-to-day fluctuations in MIDAS scores. Instead, they should calculate an overall index of headache-related interference by averaging MIDAS scores across many measurement occasions and look for changes over longer periods of time using small-N, A-B-A designs. Finally, we also explored convergent and criterion validity for some critical elements of the myWHI diary, especially when used by the oldest subgroup.

Strengths and Weaknesses of the Mywhi Application

The myWHI diary overcomes several drawbacks identified in the electronic diaries described and used in the scientific literature [22] or available on the market [23]. It allows users to track information for both headache and headache-free days, and has undergone formal usability, feasibility, and psychometric testing. Unlike the myWHI diary, the usability, feasibility, and reliability of current headache diaries are often unknown. Available diaries have been fundamentally designed to only log headaches and the main variables associated with them (ie, intensity, duration, and timing). Diaries often do not collect information in the absence of headaches, as they are event-contingent, making it difficult for the user (or health care professionals) to understand what is precipitating the headaches, and what strategies may prevent new headaches; the digital headache diary (DHD) is one exception [45]. The myWHI diary meets the criteria, which our team defined a priori, for an “ideal” diary application; it is intended to help individuals with headache to better understand and manage their headaches, while providing relevant data to health professionals [23]. The myWHI diary was created with the help of headache experts, it measures clinically relevant headache variables, it allows customization to make the application clinically relevant to the individual user, and it generates reports displaying relationships between variables.

Despite the high levels of satisfaction with the application, we also identified weaknesses. There were weaknesses with our Web-based report generation system that is external to the application, which was not used by the participants. Incorporating a robust set of reports directly in the mobile application is something to consider. The relatively slow responsiveness of the application influences user experience; we stored data on a remote server for research purposes, which introduced delays users are not accustomed to when using mobile apps. This could be resolved in a new version of the application that stores data locally, with the potential for a data synchronization system to maintain availability of usage data to researchers. We acknowledge the need to adapt the diary for other mobile phone platforms or Web-based application. Finally, taking into account that the MIDAS scores seem inappropriate for day-to-day reporting, it would be desirable to replace this measure with a new one that could examine daily changes. However, more psychometric work is required to develop a new measure.

Limitations of the Study

This study has some limitations. First, participants across a wide range of ages were included. We encountered recruitment difficulties for participants between 14 and 17 years old, which limits generalizability of our psychometric findings to this age group. Second, introducing a routine such as using this diary every day is neither easy nor fast. When testing usability and feasibility of the diary, we offered participants a prototype version of the diary to use for 14 days. With this time period, which is shorter than the 28-day period that is often recommended for assessment in headache studies [28,29,46], the use of the diary was not effortlessly adapted as daily routine. This could have negatively affected use of the diary, quality of

the suggestions, and even bias modifications taken as result of the collected data. A longer trial duration could have minimized these problems and provide a more accurate picture of feasibility data when this diary is used as consistently recommended in the literature. Third, findings from interviews conducted in each cycle of usability/feasibility were not representative of the participating sample in each cycle as only 15%-25% of participants were involved in the interviews. Despite this inherent characteristic of qualitative research, the data helped to provide richer insight of some participant experiences using the diary. Similarly, the case study is not generalizable; however, it provides an example of how this diary may be helpful. Finally, allowing participants to freely omit some items on the questionnaire may result in greater levels of missing data. In a clinical context, practitioners should work with patients closely to determine which questionnaire items are most important to track on a case-by-case basis.

Conclusions and Next Steps

This research represents the necessary first steps toward creating a feasible and psychometrically sound electronic diary for young adults with headaches. Users should be continuously involved during the design of applications [47] and we involved users from the onset [48]. The application was evaluated using formal testing cycles with participants once we had a high-fidelity and fully functional prototype of the application to provide an accurate and realistic setting for evaluation, following Prinz's recommendations [49]. This diary may be useful not only for individuals with headaches but also for medical doctors who want to collect accurate and thorough information. Without a diary, medical doctors may be forced to make care decisions on the basis of limited retrospective information collected during brief and sporadic encounters. They will now have an accurate picture of their patients' headaches in order to tailor headache treatments to each patient, which ultimately may improve treatment outcomes. The diary records of frequency, severity, duration, location, qualities of headaches, the level of disability, and its associated symptoms can help to consolidate headache diagnosis. The exposure to potential triggers and perceived precipitating factors can help identify headache triggers, and its records of frequency of analgesic use and other coping strategies can help to determine the best methods for managing pain.

Since there is still no evidence indicating the multiple potential benefits of electronic diaries in the field of headache [50] our plan for the future is to evaluate the impact of this diary at multiple levels (eg, does the use of myWHI save clinician time? does it facilitate diagnosis? does it improve patient outcomes?). The authors plan to make this diary publically available soon at the myWHI website.

Acknowledgments

The authors thank Carl von Bayer, PhD, Mark Connelly, PhD, Bo Larsson, MD, Mattias Linde, MD, Tonya Palermo, PhD, Laurie Pereles, MD, Allan Purdy, MD, and 10 anonymous individuals with headache for their participation in the review of an initial draft of the diary items. The authors would also like to thank Zack Miller for his technical assistance in helping us to extract data from the diary system for data analysis. Dr. McGrath's research is supported by a Canada Research Chair. This project has been funded by a Canadian Institutes of Health Research (CIHR) operating grant (grant #97981).

Conflicts of Interest

The authors have no conflicts of interest to declare.

References

1. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, de Vaccaro Karen Courville, Des Jarlais Don C, Fowkes Francis Gerry R, Narayan K M Venkat, De Leòn Felipe Rodriguez, Stapelberg Nicolas J C, van der Werf Marieke J, Williams Sean R M, Zaidi Anita K M, Murray Christopher J L, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2163-2196. [doi: 10.1016/S0140-6736(12)61729-2] [Medline: 23245607]
2. Headache Classification Committee of the International Headache Society (IHS). The International Classification of Headache Disorders, 3rd edition (beta version). Cephalalgia 2013 Jul;33(9):629-808. [doi: 10.1177/0333102413485658] [Medline: 23771276]
3. Dowson AJ, Sender J, Lipscombe S, Cady RK, Tepper SJ, Smith R, van Duijn N P, et al. Establishing principles for migraine management in primary care. Int J Clin Pract 2003;57(6):493-507. [Medline: 12918889]
4. Peng KP, Wang SJ. Migraine diagnosis: screening items, instruments, and scales. Acta Anaesthesiol Taiwan 2012 Jun;50(2):69-73. [doi: 10.1016/j.aat.2012.05.002] [Medline: 22769861]
5. Winkler R, Underwood P, Fatovich B, James R, Gray D. A clinical trial of a self-care approach to the management of chronic headache in general practice. Soc Sci Med 1989;29(2):213-219. [Medline: 2665109]
6. Baldacci F, Vedovello M, Ulivi M, Vergallo A, Poletti M, Borelli P, et al. How aware are migraineurs of their triggers? Headache 2013 May;53(5):834-837. [doi: 10.1111/head.12083] [Medline: 23534912]
7. Moloney MF, Aycock DM, Cotsonis GA, Myerburg S, Farino C, Lentz M. An Internet-based migraine headache diary: issues in Internet-based research. Headache 2009 May;49(5):673-686. [Medline: 19472446]

8. Nicholson RA, Buse DC, Andrasik F, Lipton RB. Nonpharmacologic treatments for migraine and tension-type headache: how to choose and when to use. Curr Treat Options Neurol 2011 Feb;13(1):28-40. [doi: 10.1007/s11940-010-0102-9] [Medline: 21080124]
9. WHO. Neurological Disorders: Public Health Challenges. Geneva, Switzerland: World Health Organization; 2006.
10. Martin PR. Behavioral management of migraine headache triggers: learning to cope with triggers. Curr Pain Headache Rep 2010 Jun;14(3):221-227. [doi: 10.1007/s11916-010-0112-z] [Medline: 20425190]
11. Huguet A, McGrath PJ, Stinson J, Tougas ME, Doucette S. Efficacy of psychological treatment for headaches: an overview of systematic reviews and analysis of potential modifiers of treatment efficacy. Clin J Pain 2014 Apr;30(4):353-369. [doi: 10.1097/AJP.0b013e318298dd8b] [Medline: 23823250]
12. Connelly M, Bickel J. An electronic daily diary process study of stress and health behavior triggers of primary headaches in children. J Pediatr Psychol 2011 Sep;36(8):852-862 [FREE Full text] [doi: 10.1093/jpepsy/jsr017] [Medline: 21441572]
13. van den Brink M, Bandell-Hoekstra EN, Abu-Saad HH. The occurrence of recall bias in pediatric headache: a comparison of questionnaire and diary data. Headache 2001 Jan;41(1):11-20. [Medline: 11168599]
14. Sorbi MJ, Mak SB, Houtveen JH, Kleiboer AM, van Doornen Lorenz J P. Mobile Web-based monitoring and coaching: feasibility in chronic migraine. J Med Internet Res 2007;9(5):e38 [FREE Full text] [doi: 10.2196/jmir.9.5.e38] [Medline: 18166526]
15. 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]
16. 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 Feb;107(3):213-219. [Medline: 14736583]
17. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [Medline: 12173681]
18. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003 Apr;24(2):182-199. [Medline: 12689739]
19. Affleck G, Zautra A, Tennen H, Armeli S. Multilevel daily process designs for consulting and clinical psychology: a preface for the perplexed. J Consult Clin Psychol 1999 Oct;67(5):746-754. [Medline: 10535241]
20. Burton C, Weller D, Sharpe M. Are electronic diaries useful for symptoms research? A systematic review. J Psychosom Res 2007 May;62(5):553-561. [doi: 10.1016/j.jpsychores.2006.12.022] [Medline: 17467410]
21. Lewandowski AS, Palermo TM, Kirchner HL, Drotar D. Comparing diary and retrospective reports of pain and activity restriction in children and adolescents with chronic pain conditions. Clin J Pain 2009 May;25(4):299-306 [FREE Full text] [doi: 10.1097/AJP.0b013e3181965578] [Medline: 19590478]
22. Stinson JN, Huguet A, McGrath P, Rosenbloom B, Soobiah C, White M, et al. A qualitative review of the psychometric properties and feasibility of electronic headache diaries for children and adults: where we are and where we need to go. Pain Res Manag 2013;18(3):142-152 [FREE Full text] [Medline: 23748255]
23. Hundert AS, Huguet A, McGrath PJ, Stinson JN, Wheaton M. Commercially available mobile phone headache diary apps: a systematic review. JMIR Mhealth Uhealth 2014;2(3):e36 [FREE Full text] [doi: 10.2196/mhealth.3452] [Medline: 25138438]
24. Skeels MM, Pratt W. Participatory design with health consumers. AMIA Annu Symp Proc 2008:1136. [Medline: 18999037]
25. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
26. McGrath PJ, Walco GA, Turk DC, Dworkin RH, Brown MT, Davidson K, von Baeyer Carl L, PedIMMPACT. Core outcome domains and measures for pediatric acute and chronic/recurrent pain clinical trials: PedIMMPACT recommendations. J Pain 2008 Sep;9(9):771-783. [doi: 10.1016/j.jpain.2008.04.007] [Medline: 18562251]
27. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003 Dec;106(3):337-345. [Medline: 14659516]
28. Bendtsen L, Bigal ME, Cerbo R, Diener HC, Holroyd K, Lampl C, International Headache Society Clinical Trials Subcommittee. Guidelines for controlled trials of drugs in tension-type headache: second edition. Cephalalgia 2010 Jan;30(1):1-16. [doi: 10.1111/j.1468-2982.2009.01948.x] [Medline: 19614696]
29. Penzien DB, Andrasik F, Freidenberg BM, Houle TT, Lake AE, Lipchik GL, American Headache Society Behavioral Clinical Trials Workgroup. Guidelines for trials of behavioral treatments for recurrent headache, first edition: American Headache Society Behavioral Clinical Trials Workgroup. Headache 2005 May;45 Suppl 2:S110-S132. [doi: 10.1111/j.1526-4610.2005.4502004.x] [Medline: 15921503]
30. Tfelt-Hansen P, Block G, Dahlöf C, Diener HC, Ferrari MD, Goadsby PJ, International Headache Society Clinical Trials Subcommittee. Guidelines for controlled trials of drugs in migraine: second edition. Cephalalgia 2000 Nov;20(9):765-786. [Medline: 11167908]

31. von Baeyer Carl L, Spagrud LJ, McCormick JC, Choo E, Neville K, Connelly MA. Three new datasets supporting use of the Numerical Rating Scale (NRS-11) for children's self-reports of pain intensity. Pain 2009 Jun;143(3):223-227. [doi: 10.1016/j.pain.2009.03.002] [Medline: 19359097]
32. Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. Pain 1999 Nov;83(2):157-162. [Medline: 10534586]
33. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain 1986 Oct;27(1):117-126. [Medline: 3785962]
34. Berwick DM, Murphy JM, Goldman PA, Ware JE, Barsky AJ, Weinstein MC. Performance of a five-item mental health screening test. Med Care 1991 Feb;29(2):169-176. [Medline: 1994148]
35. Rumpf HJ, Meyer C, Hapke U, John U. Screening for mental health: validity of the MHI-5 using DSM-IV Axis I psychiatric disorders as gold standard. Psychiatry Res 2001 Dec 31;105(3):243-253. [Medline: 11814543]
36. Strand BH, Dalgard OS, Tambs K, Rognerud M. Measuring the mental health status of the Norwegian population: a comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). Nord J Psychiatry 2003;57(2):113-118. [doi: 10.1080/08039480310000932] [Medline: 12745773]
37. Buysse DJ, Yu L, Moul DE, Germain A, Stover A, Dodds NE, et al. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. Sleep 2010 Jun;33(6):781-792 [FREE Full text] [Medline: 20550019]
38. Amtmann D, Cook KF, Jensen MP, Chen WH, Choi S, Revicki D, et al. Development of a PROMIS item bank to measure pain interference. Pain 2010 Jul;150(1):173-182 [FREE Full text] [doi: 10.1016/j.pain.2010.04.025] [Medline: 20554116]
39. Wolfson AR, Carskadon MA. Sleep schedules and daytime functioning in adolescents. Child Dev 1998 Aug;69(4):875-887. [Medline: 9768476]
40. Headache Classification Subcommittee of the International Headache Society. The International Classification of Headache Disorders: 2nd edition. Cephalalgia 2004;24 Suppl 1:9-160. [Medline: 14979299]
41. Varni JW, Stucky BD, Thissen D, Dewitt EM, Irwin DE, Lai JS, et al. PROMIS Pediatric Pain Interference Scale: an item response theory analysis of the pediatric pain item bank. J Pain 2010 Nov;11(11):1109-1119 [FREE Full text] [doi: 10.1016/j.jpain.2010.02.005] [Medline: 20627819]
42. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
43. Cranford JA, Shrout PE, Iida M, Rafaeli E, Yip T, Bolger N. A procedure for evaluating sensitivity to within-person change: can mood measures in diary studies detect change reliably? Pers Soc Psychol Bull 2006 Jul;32(7):917-929 [FREE Full text] [doi: 10.1177/0146167206287721] [Medline: 16738025]
44. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: L Erlbaum Associates; 1988.
45. Allena M, Cuzzoni MG, Tassorelli C, Nappi G, Antonaci F. An electronic diary on a palm device for headache monitoring: a preliminary experience. J Headache Pain 2012 Oct;13(7):537-541 [FREE Full text] [doi: 10.1007/s10194-012-0473-2] [Medline: 22842873]
46. Silberstein S, Tfelt-Hansen P, Dodick DW, Limmroth V, Lipton RB, Pascual J, Task Force of the International Headache Society Clinical Trials Subcommittee. Guidelines for controlled trials of prophylactic treatment of chronic migraine in adults. Cephalalgia 2008 May;28(5):484-495. [doi: 10.1111/j.1468-2982.2008.01555.x] [Medline: 18294250]
47. Lynch T, Gregor S. User participation in decision support systems development: Influencing system outcomes. Eur J Inf Syst 2004 Nov 02;13(4):286-301. [doi: 10.1057/palgrave.ejis.3000512]
48. Huguet A, Stinson J, Mackay B, Watters C, Tougas M, White M, et al. Bringing psychosocial support to headache sufferers using information and communication technology: lessons learned from asking potential users what they want. Pain Res Manag 2014;19(1):e1-e8 [FREE Full text] [Medline: 24511572]
49. Prinz W, Mark G, Pankoke-Babatz U. Designing groupware for congruency in use. In: CSCW '98: computer supported cooperative work: ACM 1998 Conference on Computer Supported Cooperative Work: proceedings, Seattle, Washington, November 14-18, 1998. New York, NY: The Association for Computing Machinery; 1998:373-382.
50. Marceau LD, Link C, Jamison RN, Carolan S. Electronic diaries as a tool to improve pain management: is there any evidence? Pain Med 2007 Oct;8 Suppl 3:S101-S109. [doi: 10.1111/j.1526-4637.2007.00374.x] [Medline: 17877520]

Abbreviations

DHD: daily headache diary
HA: Headache
MIDAS: Migraine Disability Assessment Scale
Mdn: Median
myWHI diary: myWireless Headache Intervention diary
PROMIS: Patient Reported Outcomes Measurement Information System

Edited by G Eysenbach; submitted 20.09.14; peer-reviewed by N Maleki, L Simons; comments to author 24.12.14; revised version received 14.02.15; accepted 02.03.15; published 08.05.15.

Please cite as:
Huguet A, McGrath PJ, Wheaton M, Mackinnon SP, Rozario S, Tougas ME, Stinson JN, MacLean C
Testing the Feasibility and Psychometric Properties of a Mobile Diary (myWHI) in Adolescents and Young Adults With Headaches
JMIR mHealth uHealth 2015;3(2):e39
URL: http://mhealth.jmir.org/2015/2/e39/
doi:10.2196/mhealth.3879
PMID:25956377

Original Paper

Development and Pilot Testing of the Eating4two Mobile Phone App to Monitor Gestational Weight Gain

Catherine Knight-Agarwal[1], MNutDiet; Deborah Lee Davis[2]; Lauren Williams[3]; Rachel Davey[1]; Robert Cox[1]; Adam Clarke[1]

[1]The University of Canberra, Canberra, Australia

[2]University of Canberra, Canberra, Australia

[3]Griffith University, Gold Coast, Australia

Corresponding Author:
Catherine Knight-Agarwal, MNutDiet
The University of Canberra
University drive
Bruce
Canberra, 2617
Australia
Phone: 61 0407806663
Fax: 61 62015111
Email: Cathy.Knight-Agarwal@canberra.edu.au

Abstract

Background: The number of pregnant women with a body mass index (BMI) of 30kg/m^2 or more is increasing, which has important implications for antenatal care. Various resource-intensive interventions have attempted to assist women in managing their weight gain during pregnancy with limited success. A mobile phone app has been proposed as a convenient and cost-effective alternative to face-to-face interventions.

Objective: This paper describes the process of developing and pilot testing the Eating4Two app, which aims to provide women with a simple gestational weight gain (GWG) calculator, general dietary information, and the motivation to achieve a healthy weight gain during pregnancy.

Methods: The project involved the development of app components, including a graphing function that allows the user to record their weight throughout the pregnancy and to receive real-time feedback on weight gain progress and general information on antenatal nutrition. Stakeholder consultation was used to inform development. The app was pilot tested with 10 pregnant women using a mixed method approach via an online survey, 2 focus groups, and 1 individual interview.

Results: The Eating4Two app took 7 months to develop and evaluate. It involved several disciplines--including nutrition and dietetics, midwifery, public health, and information technology--at the University of Canberra. Participants found the Eating4Two app to be a motivational tool but would have liked scales or other markers on the graph that demonstrated exact weight gain. They also liked the nutrition information; however, many felt it should be formatted in a more user friendly way.

Conclusions: The Eating4Two app was viewed by participants in our study as an innovative support system to help motivate healthy behaviors during pregnancy and as a credible resource for accessing nutrition-focused information. The feedback provided by participants will assist with refining the current prototype for use in a clinical intervention trial.

(JMIR mHealth uHealth 2015;3(2):e44) doi:10.2196/mhealth.4071

KEYWORDS

pregnancy; mobile phone; antenatal care; maternal obesity; intervention

Introduction

The proportion of overweight and obese pregnant women is increasing, which has important implications for antenatal care [1-3]. Of particular concern is that, compared to women who have given birth previously, young first-time mothers have a greater likelihood of gaining excess weight during pregnancy and many are entering pregnancy with a high BMI [4,5]. One recent Australian study reported that 43% of pregnant women were overweight or obese [6]. Maternal obesity and excessive

GWG have well-documented associations with adverse outcomes for the woman, such as increased risk of gestational diabetes, and for the baby, such as higher risks of hypoglycemia and macrosomia [3]. High GWG has been independently associated with an increased risk of childhood obesity, suggesting that influences occurring in the fetal environment are contributing to obesity onset [7].

Various resource-intensive interventions have attempted to assist women to manage their weight gain during pregnancy with varied success. One systematic review and meta-analysis of randomized and nonrandomized controlled trials (RCTs) suggested that dietary interventions during pregnancy may be effective in decreasing total GWG. A trend toward reduced prevalence of gestational diabetes in overweight and obese women was also reported [8]. Another systematic review of RCTs concluded that the effect of providing antenatal dietary interventions for overweight or obese pregnant women on maternal and infant health outcomes remains unclear [9]. Qualitative studies show health professionals struggle to support women to manage their weight gain in pregnancy due to a paucity of resources, concern over raising the sensitive issue of weight, the fact that weighing women in pregnancy is not routine, and a lack of high level guidance [10].

More recently, a systematic review examined new communication technologies and their potential to support lifestyle interventions during pregnancy [11]. The authors concluded that there is a paucity of data and RCTs examining the effectiveness of communication technology with pregnant women, particularly among those who are overweight and obese. A mobile phone app has been proposed as one such convenient and cost-effective approach to maternal obesity interventions. Mobile phones are ubiquitous in Australia, with an estimated 8.67 million users. Mobile app downloads increased by 85% over a 12-month period from 2.42 million users downloading a mobile app in June 2011 to 4.45 million in June 2012 [12].

A wide range of apps are available that focus on general weight management. Hebden et al developed and qualitatively evaluated a series of apps aimed at modifying lifestyle behaviors associated with excess weight gain in young adulthood [13]. Lee et al developed a mobile phone-based diet game for weight control [14] and Bexelius et al formulated an app for monitoring physical activity [15]. Hearn et al developed a mobile phone app and accompanying website for perinatal women to help track their weight, diet, physical activity, emotional well-being and sleep patterns [16]. After careful review of both Android and iPhone apps, the researchers identified very few that focused specifically on weight management in pregnancy and none, to our knowledge, that contained a visual graphing function for women to input and follow their GWG. The aims of this study were to: (1) develop an app providing real-time feedback on gestational weight gain progress and credible, up-to-date nutrition advice for pregnancy; and (2) identify any usability problems with the app and determine participant satisfaction with the product.

Methods

Development of the Eating4Two App

The overall content and usability for the Eating4Two app was shaped by qualitative investigation and supported by an evidence-based approach. The study team first met in August 2013 to discuss the project plan and an IT specialist was employed shortly after to write the software program for the app. It was decided that an Android 2.1 and higher platform be used as this was the most cost-effective choice [17]. The study team met on a weekly basis in order to resolve issues and review the progress of the project. A simple, visual graphing function that allows the user to configure their start BMI, estimate due date, and weigh in throughout the pregnancy and that provided real-time feedback was developed as part of the app. The opening screen of the app requires users to agree to the statement that the app does not replace the care and advice provided by a health professional. In order to progress to the app functions, the user must indicate that they agree with the statement. Alternatively, they are given the option to exit. Once the “I agree” tab has been pressed, users enter the configuration screen shown in Figure 1. Gestation (in weeks) is calculated by setting a date (usually provided by ultrasound), from first day of last menstrual period, or by current gestational week. The next step requires input of weight and height in either metric or imperial units. To appeal to a broad range of potential consumers, BMI is automatically calculated and appears on the screen (BMI is the division of weight in kilograms by height in meters squared). Once the “Save” tab has been pressed, the recommended GWG is calculated to align with Institute of Medicine guidelines based on BMI at conception. This function allows the user to check, at any time during their pregnancy, where they are positioned on the graph in terms of weight gain (see Figure 2). Numbers on the graph were purposely omitted to reduce weight-related stress as appropriate trends within each participant’s recommended weight gain range were felt to be more important than numbers aiming for an exact weight.

Another function, the “Library” tab, contains information about nutrients, foods, menus, behaviors, and symptoms especially relevant to pregnancy (see Figure 2). This information was drawn from the Commonwealth Government of Australia’s National Health and Medical Research Council [18,19], the Dietitian’s Association of Australia [20], and Food Standards Australia and New Zealand [21]. Photographs demonstrating recommended food portions provide a visual guide of what a “standard serving” looks like; for example, a half a cup of green peas [22] (see Figure 3). All 5 food groups are represented [18,19].

Table 1. Dietary information contained in the Eating4Two app library.

Topic	Information under this topic
General dietary intake	Healthy eating advice (eg, dieting while pregnant discouraged). However, there is no need to eat double than usual.
Alcohol	The potential risks of consuming alcohol during pregnancy are outlined. There is no known completely safe level of alcohol consumption during pregnancy.
Energy	During the first trimester, energy (kJ) intake should remain about the same as it was prior to pregnancy. Second and third trimester energy (kJ) should increase by about 600 kJ/day. Practical examples of how to achieve this were provided.
Iron	During pregnancy, requirements for this mineral are elevated. It is recommended that women consume approximately 27 mg/day. Practical dietary examples of how to achieve this were provided.
Folate	During pregnancy, requirements for this vitamin are elevated. It is recommended that women consume approximately 600 mcg/day (plus a 400 mcg supplement/day). Practical dietary examples of how to achieve this were provided.
Calcium	During pregnancy, requirements for this mineral are elevated. It is recommended that women consume approximately 1,000 mg/day. Practical dietary examples of how to achieve this were provided.
Iodine	During pregnancy, requirements for this mineral are elevated. It is recommended that women consume approximately 200 mcg/day. Practical dietary examples of how to achieve this were provided.
Zinc	This mineral is widely available from a variety of foods, making it possible for pregnant women to achieve their requirements through a balanced diet alone.
Nausea	Commonly experienced in the first trimester. Always discuss symptoms with health care provider. Nutrition-related information to relieve nausea (eg, eat small amounts of food often) was provided.
Heartburn	Commonly experienced in the third trimester. Always discuss symptoms with health care provider. Nutrition-related information to relieve heartburn (eg, stay upright after eating) was provided.
Constipation	May occur at any stage during pregnancy. Always discuss symptoms with health care provider. Nutrition-related information to relieve constipation (eg, drink plenty of fluids) was provided.
Tiredness	May occur at any stage during pregnancy and for some women may be ongoing. Nutrition-related information to assist women (eg, suggestion to make a batch of meals all at once and have a ready supply in the freezer) was provided.
Listeria	A bacterium that can contaminate food and cause listeriosis. It's a rare infection, but it's very serious if contracted during pregnancy. Dietary information was included on what to avoid.
Mercury	Fish is an important part of a healthy diet. Some types of fish may contain mercury. Practical examples of what and how much fish is ideal to eat during pregnancy were provided.
Cravings	May be driven by hormonal changes. Unlikely to be the body's way of indicating a nutrient deficiency exists. Discuss unusual or inappropriate cravings with health care provider. Healthy tips on dealing with cravings (eg, cravings for ice cream may be satisfied by eating frozen berries) were provided.
Serving sizes	Photographs of recommended serving sizes (eg, nuts, lean meat, bread) were included.

Figure 1. Configuration screen.

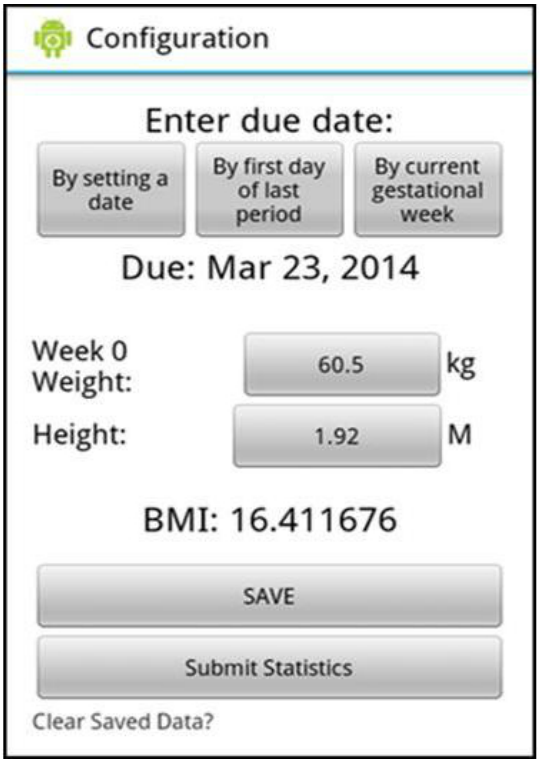


Figure 2. Gestational weight gain calculator screen.

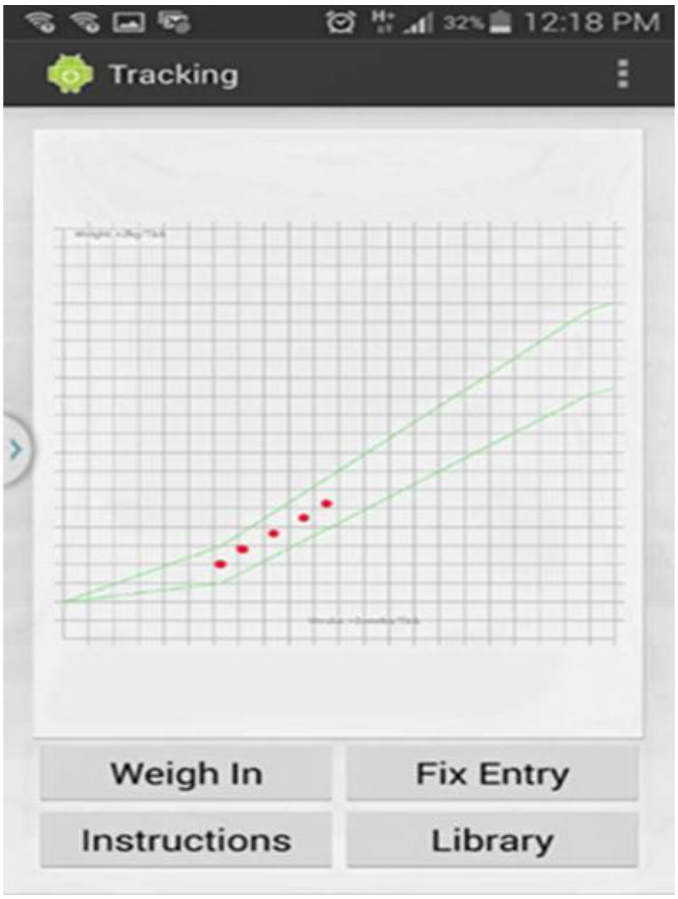


Figure 3. Example of nutrition library screen.

Stakeholder Review

A stakeholder group was established at the end of August 2013 to further inform development. This group included the researchers, an obstetrician, and 4 women of child-bearing age. In mid-September 2013, the stakeholder group met and reviewed existing dietary and antenatal care apps plus a preliminary draft of the Eating4Two app. Stakeholders provided invaluable feedback on the app's content and functionality. The first version of the app was finalized in November 2013.

Pilot Testing the Eating4Two App

Ethical approval for the evaluation of the app was sought and obtained from the University of Canberra Human Research Ethics Committee (ref: 13–173). An Eating4Two study information flyer was circulated via the university intranet and on a local mother's group Facebook page. Women were eligible to participate if they were at no more than 30 weeks gestation, had a singleton uncomplicated pregnancy, were 18 years or older, and were able to attend meetings at the university. Ten pregnant women were recruited and invited to attend an Eating4Two introductory session at the University of Canberra. The first part of the session included both verbal and written information explaining the background and objectives of the study. Women were then invited to provide signed consent. The second part of the session involved downloading the app to each woman's mobile phone and providing a practical demonstration on how to use it. All women owned and regularly used mobile phones. However, 50% of the study participants had access to an iPhone. The University of Canberra supplied these women with a prepaid Huawei Ascend Y300 (powered by Android) for the duration of the 6-week evaluation period. Participants were asked to weigh themselves at the same time every day and on the same day each week (ie, Friday morning at 7) using the same bathroom scale for each weigh-in and to insert the result immediately into the app. No more than 1 weigh-in per week was recommended since researchers wanted to prevent angst and weight-related stress. Once weight was measured, women were encouraged to insert this immediately into the app. They were encouraged to view the nutrition information as desired. Participants were provided with contact technical support in the event they required assistance. Women began using the app from late November 2013 to early January 2014. We provided each participant with a AUD $100 gift voucher to compensate them for their time. Women were not informed of the gift vouchers until after they had agreed to participate in the study.

Evaluation process

An online survey that focused on usability was emailed to all participants midway through the evaluation period. The first 2 questions were:

1. Have you found it useful to graph and monitor your weight change with the app?

2. What do you think about the nutrition information contained in the app?

Participants were asked to provide responses on a 5-point Likert scale [23] where 1 means very unhelpful and 5 means very helpful.

Questions 3, 4, and, 5 were open-ended and asked:

1. What do you like best about the app?
2. What do you like least about the app?
3. Have you any additional comments regarding the mobile phone app?

At the end of the evaluation period, 2 focus groups (FG) [24] and 1 individual interview were conducted. FG1 (n=6), FG2 (n=3), and individual interview (n=1) were held in January, February, and March 2014, respectively. A facilitatory style was employed, which included the use of verbal and nonverbal cues and sought not to influence answers. Each focus group and the individual interview gave participants the opportunity to add any information relating to the topic that may have been missed. Sessions were audio recorded and supported by written notes. The focus group discussions and the individual interview were transcribed verbatim and entered into a word processing document.

The open-ended comments from the survey and transcripts were analyzed using inductive thematic analysis, which allows findings to emerge from those that dominate in the raw data. The primary aim of this type of analysis is to establish clear links between the research objectives and the research findings and to ensure these links are both transparent and justifiable [25,26].

Results

Participants

Ten women living in or close to Canberra, Australia, participated in the pilot. Participants were 18 years of age or older with a singleton, uncomplicated pregnancy. All participants provided signed consent and were given unique numbers to ensure anonymity (see Table 2).

Table 2. Demographic information of participants at the beginning of the 6-week pilot study.

Participant identification No.	Gestation, No. of weeks	Parity	Prepregnancy BMI
1	17	0	Not provided
2	19	0	Not provided
3	19	1	20.5
4	30	0	23.8
5	13	0	24.5
6	15	0	Not provided
7	17	0	23.9
8	22	1	23.3
9	30	1	22.1
10	19	1	26.1

Survey

For the first question, 70% (n=7) of participants reported finding the app a helpful tool to graph and monitor their weight change, 20% (n=2) found it neither helpful nor unhelpful, and 10% (n=1) found it unhelpful (mean=3.6; median=4). For the second question, 10% (n=1) of participants reported finding the nutrition information contained in the app very helpful, 50% (n=5) found it helpful, 30% (n=3) neither helpful nor unhelpful, and 10% (n=1) found it unhelpful (mean=3.6; median=4). Women acknowledged that the Eating4Two app is a motivational tool that is generally simple to use. Women reported that they would have liked scales or other markers on the graph to demonstrate exact weight gain and that the nutrition information should be formatted in a more user friendly way.

Focus Groups and Individual Interview Findings

Overview

Participants provided very similar feedback to the online survey. Women found the Eating4Two app to be a motivational tool but would have liked more explicit feedback regarding their GWG. They liked the nutrition information, which included food photos demonstrating portion sizes. However, participants felt that the food photos should be clearer with labels attached and that the nutrition library should be made easier to navigate. Overall, 3 themes relating to user experience and acceptability were identified and are presented below.

Theme 1: Functionality Is Important When Navigating an App.

Most participants described the Eating4Two app as "very user friendly" [FG1, P6] and viewed its function as being an efficient and convenient way to monitor GWG. As 1 woman stated:

> *It takes 3 seconds to input the data every couple of days. . . . I suppose I am looking at the effort verses reward, which is actually quite good* [FG1, P1]

Whereas other women expected the app to require less user input, saying:

> *It takes 2 presses on the screen to bring the numbers up when you type your weight in . . . once to click in the box and another to bring up the numbers. It would have helped to have the numbers come up automatically.* [FG2, P9]

Women expressed a desire to know exactly how their weight was tracking. One participant commented:

> *I would have at least liked scales on the graph—that would have been good.* [FG2, P7]

This was supported by another woman, who said:

> *A little feedback from what I was putting in would have been helpful . . . so you know what the number range is in the projection that it gives you because you have no concept of whether you are 1 kilo below the line or 2 or 5 or 10.* [FG1, P3]

The written and visual information regarding quantity and type of foods to eat during pregnancy was seen to be an important inclusion. One woman commented that "the serving sizes . . . having that information really helped" [FG1, P6] and another woman described this function as a unique feature, saying:

> *I haven't seen another app or even just literature that describes usually what a standard serving size is supposed to look like . . . to see it, I thought was really good.* [FG2, P8]

However, other participants questioned the clarity of photos, with 1 saying:

> *I didn't actually find them that easy to gauge the serving sizes or sometimes what food I was looking at.* [FG1, P3]

Participants suggested that the nutrition component could be improved by inserting a label under each photo stating quantity and type of food and by presenting the information "in categories and on different pages" [FG1, P1] instead of in 1 scroll-down sheet. Participants suggested additional functions that may improve the usability of the Eating4Two app; for example, a routine pop-up message "as a reminder of the day to weigh in" [FG2, P7] or a prompt to bring up conversations with a health care provider about any weight- or nutrition-related issues.

Theme 2: the Eating4two App Helps Motivate Women to Be Healthy During Pregnancy.

Women expressed motivation to stay healthy during pregnancy. TheEating4Two app was viewed as a type of support system to help promote motivating behaviors, as 1 participant stated:

> *I think it was really good as it encouraged me to keep on track.* [FG1, P3]

The plethora of dietary information available was seen as confusing, but the app simplified this process. For 1 woman this was especially true, she said:

> *I think I most struggle with things I can't eat. People present you with food like ahhh, can I eat that? So I found that [app function] quite handy* [FG1, P5]

This woman emphasized that the app not only motivated her to make healthy food choices but also provided her with reassurance that the choices she did make were appropriate.

Despite maternal obesity being a growing problem, health care practice has moved way from weighing pregnant women in recent years [10].The participants confirmed this, stating that they were not weighed as part of routine antenatal care. Using the app motivated them to self-monitor their weight progress and, for some, the app encouraged them to initiate conversations with their health care providers regarding their gestational weight gain, with 1 woman saying:

> *When I noticed there wasn't any gain for ages . . . I had occasion to raise it* [FG2, P8]

One woman commented on the variability of weight over a day and its influence on when she weighs in, saying:

> *The time of day I weigh myself really affects whether my weight is in the recommended range. So I find myself cheating, like weighing myself first thing in the morning more often to feel better about my weight so it is in the normal range. . . . My weight will easily fluctuate 2-3kg a day based on fluid retention* [Individual interview, P10]

This participant offered the suggestion that weight gain results may be more helpful if they were color-coded according to morning, noon, or night measures.

Theme 3: How Far Can an App Go?

Pregnancy is a time when many women develop a strong desire to learn as much as they can about keeping healthy for the sake of themselves and their unborn baby [27]. Nevertheless, there is a huge amount of pregnancy-related information available, particularly via the World Wide Web, which can sometimes be confusing and contradictory. As 1 participant commented:

> *I have put myself on a self-imposed, strictly no googling policy. . . . I have a sore finger, I have a brain tumor . . . that is the sort of thing Google will tell you. All my information about diet and things has come from a 1-page printout that my GP gave me . . . and from what's in the [Eating4Two] app.* [FG1, P1]

This comment raises the issue of credibility. Some women felt more comfortable obtaining information from sources deemed reputable—such as from their doctor, midwife, or university—rather than from "other apps" or websites where little or no information is provided regarding authorship or institution of origin, etc.

Other participants were happy to access a wide variety of different resources. One woman admitted to trying out a number of pregnancy apps in addition to the Eating4Two app. Apart from ours, the apps she continued to use had certain features that she found engaging, such as:

> *One app doubles the quiz for you each week . . . questions are dependent on what week [of pregnancy] you are at so it will be like, oh, this week is really important that you do this. . . . It will say, "Hey you*

can improve your score this week if you do this much extra exercise." [FG1, P4]

Some women felt that the information contained in the Eating4Two app was not comprehensive enough and needed to include more direct reference to the peer review literature and commentaries from experts in the field, saying:

Personally, I need a lot more information in there to be able to recommend it to someone else. . . . [I]f there is sort of references . . . to professionals in the area, a lot more detail then yes, you know, I would recommend it. [FG1, P2]

Nevertheless, most women expressed a desire to continue using the Eating4Two app post-evaluation. Participants enjoyed using the Eating4Two app and felt it was a valuable adjunct to routine antenatal care, but that it required some refinements in order to improve usability and make the information it contains more informative. One woman stated:

The information provided could have related a bit more to where you were on the [weight] scale. So, if you were a bit under, then there could be a section on evaluat[ing] your eating or something that could link the information with the graph. [FG2, P7]

Participants acknowledged that the Eating4Two app should be used in conjunction with traditional modes of antenatal care. However, there were differing views regarding the type and quantity of information that a maternity app should contain and what functions it should or shouldn't claim to deliver. One participant felt that such apps had the potential not only to provide clinical input but should be personally reassuring, she commented:

When someone is pregnant they want reassurance . . . the fact that someone may be a bit or way over the line for me means nothing, but it would mean a lot more if there is backup. [FG1, P2]

Others recognized that every woman is different and that each pregnancy has its own unique challenges. However, they were very clear in their beliefs that front-line advice should be obtained from a doctor or midwife. One woman stated:

I think you still have to appreciate that it is an app and it can't possibly replace health care advice [from a health professional]. [FG1, P1]

Discussion

Overview

Evaluation of the usability of mobile phone apps is crucial for success, so developers can adapt and improve them in the rapidly changing world of mobile technology. However, very few studies have been published which focus on the development and use of mobile phone apps for individual dietary change or weight gain monitoring. None to our knowledge have reported on pregnancy apps though a number of studies testing the efficacy of mobile phone apps are currently in progress [27,28]. Hebden et al developed and tested 4 mobile phone apps aimed at improving nutrition and physical activity lifestyle behaviors during young adulthood. Qualitative feedback provided little suggestion for content change in these apps, with the major concerns being slow running speed and requirement to log in to the apps [13]. In another qualitative investigation, Dennison et al reported that healthy young adults displayed some interest in using health care apps that support behavior change. Legitimacy, effort required, and immediate effects on mood emerged as important influences relating to app usage [29].

Our study reinforced these results plus found that women were interested in not only learning new information but in recording, and having as a reference, the exact details of their gestational weight status. Wennberg et al described pregnant women as having an eagerness to learn everything about pregnancy and seeking that information from multiple sources including health care providers, friends, magazines, and the Internet [30]. Some participants in the Eating4Two pilot test accessed information from only a few sources, whereas others admitted to accessing many. Interestingly, women in the current study expressed a desire for the Eating4Two app to contain more prompts as a reminder to undertake certain activities such as weighing in and pop-up messages that provide support but include action-oriented dialogue when it is required. In contrast, Dennison et al and Hebden et al both reported in that users of mobile phone apps often expressed irritation at receiving alerts and some became demotivated by viewing records that showed they were not meeting a goal [29,30].

Tripp et al emphasized that mobile phone apps have the potential to influence health behaviors in expectant women by providing almost-anywhere, at-any-time interactive and often personalized information at the push of a button. Due to the increasing popularity of apps, particularly among women of child-bearing age, they acknowledge that traditional antenatal services need to accommodate the use of such technology [31]. The participants in our study reported that they were not weighed as part of routine antenatal care, so the Eating4Two app provided them with the opportunity to self-monitor GWG. The fact that this led some women to initiate conversations with their health care providers regarding appropriate GWG and pregnancy-related nutrition, suggests that the app has the potential to improve nutrition related quality of care.

Dennison et al reported that users of behavior-change apps may lose motivation over time and engage in only transient use. Whether this view is applicable to pregnant users remains to be seen [29]. There is evidence that pregnant women are motivated to keep healthy not only for themselves but their unborn baby. Verbeke and De Bourdeaudhuj found that pregnant women had higher intakes of fruit and dairy products and reduced consumption of foodstuffs deemed to have elevated safety risks than nonpregnant women [32]. A study by Szwajcer et al concluded that pregnancy could indeed be a life-changing event leading to increased nutrition awareness and could influence future dietary-related behaviors [33]. Conversely, Shub et al found that women in their study had poor knowledge of GWG recommendations and the complications associated with excess weight gains [34]. The sensitive nature, and sometimes avoidance, of weight-related discussions between a client and their maternity caregiver has been suggested as 1 reason for this lack of knowledge [10]. As a result, opportunities for healthy

lifestyle education, at what is an opportunistic time for many women, are missed.

Bridging gaps in knowledge is an important step toward improving outcomes for both women and offspring. Encouraging the use of apps like Eating4Two is 1 way this may be achieved. In saying this, Tripp et al has acknowledged that the proliferation of health apps may reduce pregnant women's reliance on health care professionals for advice [31]. There are obviously potential risks with this scenario. It is at this point that we should provide a rationale for our choice of data collection methods. Both focus groups and individual interviews are techniques relying on qualitative methodology. Focus group discussions allow multiple perspectives to be heard and participants may feel encouraged to share their experiences in a homogenous group [35]. While we would have preferred only to use focus groups, 1 participant was unable to attend either focus group discussion. Rather than miss out on her feedback, we used the identical protocol to interview her individually. During the analysis stage, we were careful to consider any potential differences between responses provided by members of the focus group discussions and the individual participant interviewed. No new themes arose from the data collected from the individual participant, thus 1 method corroborated the other. Even though it was not initially planned this way, the use of both methods (and resulting corroboration) does serve as a form of triangulation [24,25,35].

Limitations

There are some limitations that must be acknowledged. This study contained only a small sample size of women, thus limiting the validity of results. Their views may not reflect those of pregnant women elsewhere. The pre-pregnancy BMIs we obtained were self-reported and are only an estimate of weight status. We did not collect participant's age or educational status, which is a potential limitation of the study. However, the aim of our research was to test usability of the Eating4Two app in women of childbearing age. Therefore, regular use of a mobile phone and being pregnant were deemed to be the most useful demographics to collect. It is also important to point out that the Eating4Two app has been designed as an adjunct to usual maternity care and in no way should replace the individual advice provided by a health professional.

Conclusion

TheEating4Two app was viewed by the majority of participants in our study as a support system to help motivate healthy behaviors during pregnancy and as a credible resource for accessing information, with the GWG graphing function perceived as a unique and useful feature. However, women wanted the Eating4Two app to display absolute figures rather than ranges, to include more comprehensive pregnancy-related resources, and be formatted in a more user friendly fashion. This feedback will assist with the refinement of the current prototype and will be used as part of a pilot RCT of pregnant women.

Acknowledgments

The development and evaluation of the Eating4Two app was funded, in part, by a 2013 grant from the faculty of health at the University of Canberra (Award Number: 2013/05). The Eating4Two team would also like to thank Dr David Knight of RANZCOG for his comments on the manuscript.

Conflicts of Interest

None declared.

References

1. Rasmussen KM, Catalano PM, Yaktine AL. Curr Opin Obstet Gynecol. USA: PMC; 2009 Dec 21. New guidelines for weight gain during pregnancy: what obstetrician/gynecologists should know URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2847829/ [accessed 2015-05-12] [WebCite Cache ID 6UI0C3Pp2]
2. Cedergren M. Effects of gestational weight gain and body mass index on obstetric outcome in Sweden. Int J Gynaecol Obstet 2006 Jun;93(3):269-274. [doi: 10.1016/j.ijgo.2006.03.002] [Medline: 16626716]
3. McIntyre HD, Gibbons KS, Flenady V JCallaway L J. Overweight and obesity in Australian mothers: epidemic or endemic? Med J Aust 2012;196(3):184-188. [doi: 10.5694/mja11.11120] [Medline: 22339524]
4. Chu SY, Callaghan WM, Bish CL, D'Angelo D. Gestational weight gain by body mass index among US women delivering live births, 2004-2005: fueling future obesity. Am J Obstet Gynecol 2009 Mar;200(3):271.e1-271.e7. [doi: 10.1016/j.ajog.2008.09.879] [Medline: 19136091]
5. Siega-Riz AM, Viswanathan M, Moos MK, Deierlein A, Mumford S, Knaack J, et al. A systematic review of outcomes of maternal weight gain according to the Institute of Medicine recommendations: birthweight, fetal growth, and postpartum weight retention. Am J Obstet Gynecol 2009 Oct;201(4):339.e1-339.14. [doi: 10.1016/j.ajog.2009.07.002] [Medline: 19788965]
6. Athukorala C, Rumbold AR, Willson KJ, Crowther CA. The risk of adverse pregnancy outcomes in women who are overweight or obese. BMC Pregnancy Childbirth 2010;10:56 [FREE Full text] [doi: 10.1186/1471-2393-10-56] [Medline: 20849609]

7. Herring SJ, Rose MZ, Skouteris H, Oken E. Optimizing weight gain in pregnancy to prevent obesity in women and children. Diabetes Obes Metab 2012 Mar;14(3):195-203 [FREE Full text] [doi: 10.1111/j.1463-1326.2011.01489.x] [Medline: 21851516]
8. Oteng-Ntim E, Varma R, Croker H, Poston L, Doyle P. Lifestyle interventions for overweight and obese pregnant women to improve pregnancy outcome: systematic review and meta-analysis. BMC Med 2012;10:47 [FREE Full text] [doi: 10.1186/1741-7015-10-47] [Medline: 22574949]
9. Dodd JM, Grivell RM, Crowther CA, Robinson JS. Antenatal interventions for overweight or obese pregnant women: a systematic review of randomised trials. BJOG 2010 Oct;117(11):1316-1326. [doi: 10.1111/j.1471-0528.2010.02540.x] [Medline: 20353459]
10. Knight-Agarwal CR, Kaur M, Williams LT, Davey R, Davis D. The views and attitudes of health professionals providing antenatal care to women with a high BMI: a qualitative research study. Women Birth 2014 Jun;27(2):138-144. [doi: 10.1016/j.wombi.2013.11.002] [Medline: 24295597]
11. O'Brien OA, McCarthy M, Gibney ER, McAuliffe FM. Technology-supported dietary and lifestyle interventions in healthy pregnant women: a systematic review. Eur J Clin Nutr 2014 Jul;68(7):760-766. [doi: 10.1038/ejcn.2014.59] [Medline: 24781682]
12. Australian Government Australian Communications and Media Authority (ACMA). Available http://www.acma.gov.au/webwr/_assets/main/lib310665/report-3-smartphones_tablets-summary.pdf. Canberra: Australian Government; 2014. Communications report 2011-12 series: Report 3 – Smartphonestablets take upuse in Australia summary report URL: http://www.acma.gov.au/webwr/_assets/main/lib310665/report-3-smartphones_tablets-summary.pdf [accessed 2015-05-12] [WebCite Cache ID 6YSv1utlw]
13. Hebden L, Cook A, van der Ploeg Hidde P, Allman-Farinelli M. Development of smartphone applications for nutrition and physical activity behavior change. JMIR Res Protoc 2012;1(2):e9 [FREE Full text] [doi: 10.2196/resprot.2205] [Medline: 23611892]
14. Lee W, Chae YM, Kim S, Ho SH, Choi I. Evaluation of a mobile phone-based diet game for weight control. J Telemed Telecare 2010;16(5):270-275. [doi: 10.1258/jtt.2010.090913] [Medline: 20558620]
15. Bexelius C, Löf M, Sandin S, Trolle LY, Forsum E, Litton JE. Measures of physical activity using cell phones: validation using criterion methods. J Med Internet Res 2010;12(1):e2 [FREE Full text] [doi: 10.2196/jmir.1298] [Medline: 20118036]
16. Hearn L, Miller M, Lester L. Reaching perinatal women online: the Healthy You, Healthy Baby website and app. J Obes 2014 [FREE Full text] [doi: 10.1155/2014/573928] [Medline: 24872891]
17. Charland A, Leroux B. Mobile application development. Commun. ACM 2011 May 01;54(5):49. [doi: 10.1145/1941487.1941504]
18. Commonwealth Government of Australia. National Health and Medical Research Council. 2012. Nutrient Reference Values for Australia and New Zealand URL: https://www.nhmrc.gov.au/guidelines-publications/n35-n36-n37 [accessed 2015-05-12] [WebCite Cache ID 6YSvSqkgt]
19. Commonwealth Government of Australia. National Health and Medical Research Council. Healthy eating during your pregnancy – Eat for health URL: https://www.eatforhealth.gov.au/guidelines [accessed 2015-05-12] [WebCite Cache ID 6YSvlUown]
20. Dietitians Association of Australia. Dietitians Association of Australia website. 2013. Smart eating for your pregnancy URL: http://daa.asn.au/for-the-public/smart-eating-for-you/nutrition-a-z/pregnancy/ [accessed 2015-05-13] [WebCite Cache ID 6YU4dYP5G]
21. Food Standards Australia New Zealand. Food Standards Australia New Zealand website. 2013. Listeria and Food URL: http://www.foodstandards.gov.au/publications/Pages/listeriabrochuretext.aspx [accessed 2015-05-13] [WebCite Cache ID 6YU4GukXs]
22. Knight-Agarwal C. Food Photographs. University of Canberra 2013.
23. Jamieson S. Likert scales: how to (ab)use them. Med Educ 2004 Dec;38(12):1217-1218. [doi: 10.1111/j.1365-2929.2004.02012.x] [Medline: 15566531]
24. Massey OT. A proposed model for the analysis and interpretation of focus groups in evaluation research. Eval Program Plann 2011 Feb;34(1):21-28. [doi: 10.1016/j.evalprogplan.2010.06.003] [Medline: 20655593]
25. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
26. Thomas DR. A General Inductive Approach for Analyzing Qualitative Evaluation Data. American Journal of Evaluation 2006 Jun 01;27(2):237-246. [doi: 10.1177/1098214005283748]
27. Leanne Redman. clinicaltrials.gov. 2012 May 31. Personalised management of body weight during pregnancy URL: http://clinicaltrials.gov/show/NCT01610752 [accessed 2014-11-23] [WebCite Cache ID 6UHuJmsCo]
28. JiWon Choi. clinicaltrials.gov.: University of California, San Francisco The MOTHER (Mobile Technologies to help Enhancing Regular Physical Activity) trial URL: http://clinicaltrials.gov/show/NCT01461707 [accessed 2014-11-23] [WebCite Cache ID 6UHxHyGod]

29. 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]
30. Wennberg AL, Lundqvist A, Högberg U, Sandström H, Hamberg K. Women's experiences of dietary advice and dietary changes during pregnancy. Midwifery 2013 Sep;29(9):1027-1034. [doi: 10.1016/j.midw.2012.09.005] [Medline: 23427852]
31. Tripp N, Hainey K, Liu A, Poulton A, Peek M, Kim J, et al. An emerging model of maternity care: smartphone, midwife, doctor? Women Birth 2014 Mar;27(1):64-67. [doi: 10.1016/j.wombi.2013.11.001] [Medline: 24295598]
32. Verbeke W, De BI. Dietary behaviour of pregnant versus non-pregnant women. Appetite 2007 Jan;48(1):78-86. [doi: 10.1016/j.appet.2006.07.078] [Medline: 17005297]
33. Szwajcer EM, Hiddink GJ, Koelen MA, van Woerkum Cees M J. Nutrition awareness and pregnancy: implications for the life course perspective. Eur J Obstet Gynecol Reprod Biol 2007 Nov;135(1):58-64. [doi: 10.1016/j.ejogrb.2006.11.012] [Medline: 17178186]
34. Shub A, Huning EYS, Campbell KJ, McCarthy EA. Pregnant women's knowledge of weight, weight gain, complications of obesity and weight management strategies in pregnancy. BMC Res Notes 2013;6:278 [FREE Full text] [doi: 10.1186/1756-0500-6-278] [Medline: 23866845]
35. Smith JC, editor. Qualitative Psychology: A Practical Guide to Research Methods. In: Qualitative Psychology: A Practical Guide to Research Methods. London, UK: Sage Publications Ltd; 2008.

Abbreviations

BMI: body mass index
FG: focus group
GWG: gestational weight gain
P: participant
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 23.11.14; peer-reviewed by AL Wennberg; comments to author 18.02.15; revised version received 10.03.15; accepted 14.03.15; published 05.06.15.

Please cite as:
Knight-Agarwal C, Davis DL, Williams L, Davey R, Cox R, Clarke A
Development and Pilot Testing of the Eating4two Mobile Phone App to Monitor Gestational Weight Gain
JMIR mHealth uHealth 2015;3(2):e44
URL: http://mhealth.jmir.org/2015/2/e44/
doi:10.2196/mhealth.4071
PMID:26048313

Original Paper

A Novel System for Supporting Autism Diagnosis Using Home Videos: Iterative Development and Evaluation of System Design

Nazneen Nazneen[1], PhD; Agata Rozga[1], PhD; Christopher J Smith[2], PhD; Ron Oberleitner[3], MBA; Gregory D Abowd[1], PhD; Rosa I Arriaga[1], PhD

[1]Georgia Institute of Technology, Atlanta, GA, United States

[2]Southwest Autism Research & Resource Center, Phoenix, AZ, United States

[3]Behavior Imaging Solutions, Boise, ID, United States

Corresponding Author:
Nazneen Nazneen, PhD
Georgia Institute of Technology
Technology Square Research Building 85 Fifth Street NW
Atlanta, GA, 30332
United States
Phone: 1 4049033916
Fax: 1 (404)894 7452
Email: nazneen@gatech.edu

Abstract

Background: Observing behavior in the natural environment is valuable to obtain an accurate and comprehensive assessment of a child's behavior, but in practice it is limited to in-clinic observation. Research shows significant time lag between when parents first become concerned and when the child is finally diagnosed with autism. This lag can delay early interventions that have been shown to improve developmental outcomes.

Objective: To develop and evaluate the design of an asynchronous system that allows parents to easily collect clinically valid in-home videos of their child's behavior and supports diagnosticians in completing diagnostic assessment of autism.

Methods: First, interviews were conducted with 11 clinicians and 6 families to solicit feedback from stakeholders about the system concept. Next, the system was iteratively designed, informed by experiences of families using it in a controlled home-like experimental setting and a participatory design process involving domain experts. Finally, in-field evaluation of the system design was conducted with 5 families of children (4 with previous autism diagnosis and 1 child typically developing) and 3 diagnosticians. For each family, 2 diagnosticians, blind to the child's previous diagnostic status, independently completed an autism diagnosis via our system. We compared the outcome of the assessment between the 2 diagnosticians, and between each diagnostician and the child's previous diagnostic status.

Results: The system that resulted through the iterative design process includes (1) NODA smartCapture, a mobile phone-based application for parents to record prescribed video evidence at home; and (2) NODA Connect, a Web portal for diagnosticians to direct in-home video collection, access developmental history, and conduct an assessment by linking evidence of behaviors tagged in the videos to the Diagnostic and Statistical Manual of Mental Disorders criteria. Applying clinical judgment, the diagnostician concludes a diagnostic outcome. During field evaluation, without prior training, parents easily (average rating of 4 on a 5-point scale) used the system to record video evidence. Across all in-home video evidence recorded during field evaluation, 96% (26/27) were judged as clinically useful, for performing an autism diagnosis. For 4 children (3 with autism and 1 typically developing), both diagnosticians independently arrived at the correct diagnostic status (autism versus typical). Overall, in 91% of assessments (10/11) via NODA Connect, diagnosticians confidently (average rating 4.5 on a 5-point scale) concluded a diagnostic outcome that matched with the child's previous diagnostic status.

Conclusions: The in-field evaluation demonstrated that the system's design enabled parents to easily record clinically valid evidence of their child's behavior, and diagnosticians to complete a diagnostic assessment. These results shed light on the potential for appropriately designed telehealth technology to support clinical assessments using in-home video captured by families. This assessment model can be readily generalized to other conditions where direct observation of behavior plays a central role in the assessment process.

(JMIR mHealth uHealth 2015;3(2):e68) doi:10.2196/mhealth.4393

KEYWORDS

asynchronous telemedicine system; in-home behavior recording; naturalistic observation diagnostic assessment; NODA Connect; NODA smartCapture; remote autism diagnosis

Introduction

Background and Motivation

According to the Centers for Disease Control and Prevention, the prevalence of autism in the United States has been increasing dramatically, from 1 in 150 to 1 in 68 children between 2000 and 2010 [1,2]. Over the same period, the median age of diagnosis remained relatively stable, around 53 months [1,2].

One key challenge with respect to diagnosing autism is the significant time lag (20-60 months) between the age at which parents first become concerned about their child's development and the age at which the child receives a diagnosis from a qualified professional [3-5]. Moreover, many ethnic minorities, low-income families, and rural communities lack access to health care professionals with autism-specific expertise, resulting in delays in diagnosis [6-9]. Even in urban communities where services are more widely available, timely access to diagnostic services is often hampered by long waiting lists. Delays in diagnosis can lead to delays in early intervention services that have been shown to improve future learning capabilities and developmental outcomes [10-13].

Another key challenge with respect to diagnosing autism is that although clinical professionals acknowledge that observing behavior in the natural environment (eg, the home) is preferred for a comprehensive assessment, in practice behavioral observations are limited to a single in-clinic observation [10,11,14]. There are various barriers to more widespread use of home-based observation [14-18]. It is time consuming and resource intensive for clinicians to travel to each family's home to conduct an observation, and impractical to do so for remotely located families. Even when home visits are feasible, the presence of an unfamiliar observer may cause children to alter their behavior due to their awareness of being observed. Such reactivity poses a threat to the validity of any data that are collected. In addition, child behaviors of interest may not occur during the short span of a clinician's home visit.

In this paper, we consider the opportunity of designing a telehealth solution to support remote diagnosis of autism through parent-recorded behavioral evidence of child suspected to have autism in the home. Telehealth technology can connect families with clinicians and accelerate the diagnostic process. Indeed, such technologies have recently been investigated as a means of supporting the delivery of treatment for individuals with autism spectrum disorder, including remote coaching of parent-implemented early intervention programs [19-21], behavioral assessments [22,23], and professional development [24]. Few attempts have been made at exploring the potential for such technologies to support diagnostic assessments [25,26].

Most current telehealth technologies support a real-time interaction between a remotely located clinician and a caregiver or patient. By contrast, "store-and-forward" telehealth systems support video recordings of live events, which are subsequently shared with a clinical expert for review and assessment. The latter asynchronous approach, which we have adopted, offers several key advantages particularly relevant to the case of remote diagnosis of autism. It enables families to record videos in their home, in the course of their day-to-day activities, which ensures the capture of natural expressions of child behavior that are widely acknowledged as crucial to making an accurate and comprehensive diagnostic assessment [10,11,14]. Moreover, because home recordings can be carried out over the course of several days, they may mitigate some of the consequences of a single clinic-based or live telehealth assessment. These include the child's reactivity, child's current mood or level of fatigue, or the likelihood that low-frequency behaviors may not be observed. From a practical standpoint, it minimizes the need to coordinate schedules with a clinician, and reduces the need for remotely located families to travel long distances to a clinic.

Research Questions and Contribution

Our research addresses two key challenges in designing a system for remote autism diagnosis using home videos. Perhaps the most important challenge is how to enable parents to easily record clinically relevant video evidence of their child's behavior. Diagnostic assessments are typically designed to enable the diagnostician to observe the child under more or less structured periods and different situations for a rich sampling of the child's behavior. Parents can record and share their concerns about their child, but may not know the specific types of situations and behaviors the diagnostician needs to observe. Thus, the first research challenge is how the parent-recorded video can be turned into meaningful clinical evidence through the use of the right kind of technology and the design of the right user experience with that technology. This paper summarizes our work on identifying and evaluating specific design features that contribute to ease of use of the in-home recording system and clinical validity of parent-recorded video evidence. The second challenge involves supporting diagnosticians in completing the remote autism diagnosis. This involves enabling diagnosticians to review the videos in a systematic and structured way so that they can map the situations and behaviors that they observe in the video to the Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnostic criteria [27,28]. In other words, the parent-recorded video of child behavior must produce observations that become evidence to support clinical judgment. This paper summaries our work on identification and evaluation of specific design features that support diagnosticians in completing a remote diagnostic assessment.

The system that resulted from this work includes two components: NODA smartCapture and NODA Connect. NODA smartCapture is a mobile phone-based application that enables parents to easily record clinically relevant prescribed video evidence of their child's behavior. It supports recording and

uploading 4, up to 10-minute long naturalistic observation diagnostic assessment (NODA) scenarios, that were chosen based on pilot research on video-based diagnosis of autism [29]. These scenarios include (1) the child playing alone, (2) the child playing with a sibling or peer, (3) a family mealtime, and (4) any behavior that is of concern to the parent. The first 3 scenarios provide opportunities for typical social communication and play-based behaviors, whereas the last scenario allows parents to share evidence of a behavior that is of particular concern to them. The NODA Connect is a Web portal for diagnosticians to direct in-home video collection, access the child's developmental history, and conduct a remote diagnostic assessment by linking evidence of behaviors tagged in the videos to DSM criteria. Relying on clinical judgment, the diagnostician renders an opinion about the child's diagnostic outcome.

Paper Outline

The rest of this paper is structured as follows. We first describe interviews with relevant stakeholders about our system concept, that is, "remote assessment based on in-home video evidence." Next, we outline the iterative design of NODA smartCapture and NODA Connect. We then describe an in-field evaluation, in which 5 families used NODA smartCapture in their homes to collect prescribed behavioral evidence and 2 diagnosticians reviewed videos from each family and independently performed an autism diagnosis using NODA Connect. We report results on ease of use of NODA smartCapture, clinical validity of recorded evidence, and NODA Connect's support for diagnosticians in completion of a remote diagnostic assessment. We conclude with a discussion on the utility and limitations of the system, potential design enhancements, our vision for adoption of the system within current autism diagnostic practices, and how such a prescription, collection, and assessment model can be generalized to other clinical applications.

Methods

Overview

First, interviews were conducted with parents of children with autism and clinicians to seek input from these key stakeholders about the overall concept of the system. Next, NODA smartCapture was iteratively designed, informed by experiences of families using it in a controlled experimental home-like setting. The NODA Connect was designed using a participatory design process involving a collaborating diagnostician with 20 years of experience in autism diagnosis and a domain expert in autism. Finally, an in-field evaluation was conducted with families and diagnosticians.

Stage 1: Insight from Stakeholders

We conducted a series of structured one-on-one interviews, each lasting 2 hours, with parents (n=7) of children with autism and clinicians (n=11) who work with this population. These interviews allowed us to gather input from these key stakeholders about the overall system concept of "remote assessment based on in-home video evidence," and about its feasibility and potential utility. During the interview, a mock-up design based on previous studies [29,30] was presented to elicit feedback from stakeholders. The mock-up design included a mobile phone-based recording application to record the NODA scenarios and a Web-based assessment portal to review the videos and tag behaviors for assessment.

Stage 2: Iterative Development

smartCapture Iterative development

The NODA smartCapture application resulting from stage 1 was iteratively improved, informed by the experiences of families using it in a home-like experimental setup. Our goal was to identify specific design features that would enable parents to easily record clinically valid video evidence by analyzing the usage pattern of the recording application in a controlled setting. Families (n=8) and their child with autism as well as any siblings (n=18) visited the Georgia Tech Aware Home [31] for 2 hours, 1 family at a time. The Aware Home has the look of a single-family home (fully equipped kitchen; living room, bedrooms, and bathroom; furniture; TV, etc) except there are a number of cameras installed throughout, enabling both recording and viewing a live feed of interactions happening in different parts of the home (Figure 1).

Parents were asked to use the NODA smartCapture to record the 4 NODA scenarios, for up to 10 minutes each. Before the study, the Aware Home was set up with toys and items for the snack (to simulate the family mealtime). Ceiling-mounted cameras allowed us to observe the family from another room live as they were using NODA smartCapture, and to record the whole session for subsequent review.

After the video recording, each parent completed an interview and was asked to rate the ease of use of the system on a scale ranging from 1 (not easy to use) to 5 (very easy to use). In addition to this feedback from parents and the video evidence recorded by them using NODA smartCapture, video recordings from the fixed ceiling cameras were reviewed to gain further insight into how parents used the system.

The collaborating diagnostician was asked to rate each parent-collected video, for its clinical validity, on a scale of 0-2 and give a justification for the assigned rating. A rating of "0" means that the video is not clinically valid for conducting assessment whereas a rating of "2" indicates that the video is clinically valid. A rating of "1" indicates that the video is clinically valid but an additional video might be required to fully assess the associated scenario. By analyzing the collaborating diagnostician's reasons for the assigned rating, we identified specific issues that lowered the clinical value of a video evidence.

After the first 4 of the total 8 families completed their participation, the design of NODA smartCapture was revised based on initial findings about ease of use and issues lowering clinical utility. The revised system was tested with the remaining 4 families. Once all families completed participation, NODA smartCapture was subsequently improved based on findings from the experience of the last 4 families.

Figure 1. Aware Home setup for families to experience NODA smartCapture.

NODA Connect Iterative Development

The NODA Connect Web portal was designed, through an iterative design process, for diagnosticians to direct in-home video-collection process and conduct remote diagnostic assessment. Our goal was to identify specific features that would support diagnosticians in completing the diagnostic assessment for autism based on parent-collected videos, developmental history information, and their clinical judgment. The initial design of NODA Connect was informed by previous pilot research [29] and feedback from stakeholders solicited during the Stage 1 interviews about the system concept. However, major design contributions came from a participatory design process involving a collaborating diagnostician who had 20 years of experience in conducting autism diagnosis and a domain expert in autism. Participatory design is a common method in the technology design community whereby the designer works closely with the target user to collaboratively iterate on the design of a technology [32]. Before the in-field evaluation in the final stage of the research, the design of the NODA Connect platform was further improved based on findings from a pilot assessment conducted via NODA Connect by the collaborating diagnostician.

Stage 3: In-Field Evaluation

The iterative design process described in Stage 2 resulted in a final design of the remote autism diagnostic assessment system that was then evaluated in the field. During this evaluation, the parents used NODA smartCapture in their homes to record behavior evidence and the diagnosticians used NODA Connect to review and tag the videos, and to complete a diagnostic assessment. We recruited 4 families with at least 1 child with a previously confirmed diagnosis on the autism spectrum and 1 family with a typically developing child. Children were between 2 and 6 years of age (average 4 years). Parents were not given any prior training on NODA smartCapture. They were hand-delivered a kit that included NODA smartCapture preinstalled on an iPod touch and a tripod for mounting the iPod. During an in-home deployment that lasted an average of 2 weeks, each family was asked to complete a brief child developmental history online and use the NODA smartCapture application to record and upload the 4 10-minute NODA scenarios. The collaborating diagnostician remotely guided the in-home evidence-collection process by reviewing the videos as they were uploaded and sending alerts to the family as needed to request that they rerecord a particular scenario.

We recruited 3 diagnosticians experienced in autism diagnosis and unfamiliar with our system to complete the independent diagnostic assessments via NODA Connect. Each family's videos were reviewed by at least 2 diagnosticians, who were blind to the diagnostic status of the child. After completing each diagnostic assessment via NODA Connect, the diagnosticians concluded whether the child had autism or was typically developing. They were prompted to assign confidence ratings to the diagnostic outcome: "How confident are you that the child has autism?" and "How confident are you that the child is typically developing?" on a scale from 1 (not confident) to 5 (extremely confident). Including both of these ratings allowed diagnosticians to indicate diagnostic uncertainty in cases where they were confident that the child does not have autism but also did not think the child was typically developing. Other than the child's age, no other information was disclosed to the diagnosticians about the child's developmental history until they completed the diagnostic assessment via NODA Connect and reached a decision about the child's diagnostic outcome. At the end of this process, a follow-up interview was conducted for the diagnosticians to reflect on their experience of remote diagnostic assessment. The child's previous diagnosis and developmental history were revealed during the interview.

Data analysis of the in-field evaluation of NODA smartCapture consisted of assessing its ease of use based on parent ratings, the quality of the videos recorded by parents, and the system log about parents' reliance on help menu and navigation patterns though NODA smartCapture. The collaborating diagnostician rated the clinical utility of parent-recorded videos using the same scale as described earlier for Stage 2. Data analysis of the in-field evaluation of NODA Connect consisted of analysis of how diagnosticians completed diagnostic assessment by tagging videos and completing DSM checklist through NODA Connect. To conduct this analysis, videos of screen capture when diagnosticians were conducting assessment through NODA Connect were examined. In addition, for each child, we compared the outcome of the assessment between the 2 diagnosticians, and between each diagnostician and the child's previous diagnostic status.

Results

Ease of Use of NODA smartCapture

Based on experiments in the controlled home-like setting in Stage 2, three main features were added to NODA Connect to facilitate ease of use. First, icons on the home screen clearly depict each of the 4 NODA scenarios parents are being asked to record (Figure 2). Second, clear and redundant cues for the recording status were added so that a parent would know whether a video is being recorded or not, and how many minutes of recording have elapsed (Figure 2). Third, in addition to the "Stop Recording" button, an autostop feature that automatically stops the recording after 10 minutes was included to enable one-click recording. Once video recording is completed, parents can upload it directly or save it on the device to review the video first before uploading it.

These features were implemented based on results from the first 4 participating families of the controlled experiment. Before these features were implemented in NODA smartCapture, the first 4 participating parents gave an average ease-of-use rating of 3 on a 5-point scale, and the number (4 recordings, 1/scenario) and length of recorded videos (maximum 10 minutes) were not consistent with the instructions that were given. Once these features were implemented, the next set of 4 parents gave an average ease-of-use rating of 4, which was also maintained in the field evaluation. The second set of 4 parents who participated in the controlled experiment and the 5 parents who participated in the in-field evaluation all collected the right number of videos of appropriate length according to the given instructions. In addition, the log analysis confirmed that parents were able to use NODA smartCapture easily during the in-field evaluation. Parents did not rely much on the help menu even without any prior training for using NODA smartCapture. Log analysis showed that only 2 families of the 5 accessed the help menu, 1 and 2 times, respectively. In addition, on average 73% (22/30) of the time, parents took the shortest path from selecting a recording scenario to starting a recording. Because there could be reasons other than complexity of the NODA smartCapture that can contribute to stopping a recording and not completing it, analysis of the workflow focused only on scenario selection to starting a recording.

Figure 2. NODA smartCapture. (1) Home screen showing 4 NODA scenarios, as well as status of ones recorded. (2) Each scenario has recording instructions for parents as prescription. Pressing "Ready" proceeds to recording interface. (3) Recording mode with clear time-elapsed status and a green boundary to reinforce recording mode.

Clinical Utility of Video Evidence

Based on data analysis from the controlled experiment in a home-like setting in Stage 2, we identified two key features that increase the clinical utility of the recorded videos. These include an embedded prescription feature and a notification feature.

While rating the clinical utility of videos recorded by parents in the controlled experiment, the collaborating diagnostician identified two sets of issues that negatively influenced utility. The first set related to the set up of the recordings. The most common set-up-related issue was incorrect field of view. For example, parents often captured videos where the face of the child with autism, the relevant toys, or the person that the child was interacting with were not clearly visible on camera, either because of the way the camera was set up or mounted, or because it was too zoomed in or zoomed out. In some cases, parents followed the child around while holding the camera, which was both distracting to the child and prevented the parent from actively playing with the child during the recording. Other times, parents would not set up the camera in advance of recording and would start recording while they are still setting up the camera on the mounting device. The second set of issues related to the frequency and quality of interaction between the child and the parent. Some parents interacted with the child excessively, preventing the clinician from observing what the child does naturally when left alone. Other times there was insufficient interaction between the parent and the child, and the diagnostician wished to observe how the child might react

to the parent's attempts to interact with him/her or whether the child would direct his/her attention to something. Thus, both excessive and insufficient interaction between the parent and the child can make it difficult for a diagnostician to reliably assess the child's level of functioning.

In response to these two sets of issues, and in consultation with the collaborating diagnostician, we embedded explicit instructions within the NODA smartCapture interface. This clinical prescription (Figure 2) included specific instructions for the parent about how to set up and frame each recording (staging), and how to interact with the child during the recording (social presses). These instructions were intended to maximize the likelihood that the parent records the right kind of video evidence of their child's behavior from the diagnostician's perspective. For each of the 4 recording scenarios, we established a set of directions to improve the staging of the recording and a set of social presses that the parent was asked to present to the child. Staging instructions covered the set up of the camera and the environment, such as (1) making sure the child's face and relevant objects and social partners are in the field of view of the camera; (2) suggestions for appropriate play items, such as toys and books; and (3) across all scenarios, parents were asked to set up the camera ahead of time, and to use a mounting device (tripod). Instructions for social presses included specific actions the parent needed to take during the recording, such as calling the child's name, pointing to an object to see whether the child will look toward it. These actions represented the types of social presses that a diagnostician might use while assessing the child in person.

In addition to the explicit directions embedded within the NODA smartCapture interface, we realized (through our discussions with the diagnosticians) that diagnosticians may want to guide parents during the in-home recording process. For example, the diagnostician may wish to ask the parent to rerecord a scenario because the lighting conditions were poor, or because they want the parent to try a social press again. Therefore, a notification system was included in the system (added before the in-field evaluation) whereby the diagnosticians could send notifications to NODA smartCapture from the NODA Connect Web portal. This feature was not intended to support real-time messaging; rather, it was intended to enable the diagnosticians to review the videos uploaded by a parent for appropriateness in advance of the video being used for the diagnostic assessment, and ask for additional recording as needed.

After incorporating these prescription and notification features, the clinical validity ratings of the parent-collected videos increased from 81% (13/16) in the experimental controlled setting to 96% (26/27) in the in-field evaluation. In total, 10 notifications were sent to families during the field study. Six of these notifications were about instructions for parents to include particular social presses and 4 messages were about confirming the status of recording. Utilization of the notification system reflects its usefulness. Furthermore, the participating diagnosticians in the field evaluation were also asked to rate the usefulness of the videos after they completed remote diagnostic assessment on a 5-point scale (1 indicates "not useful" to 5 indicates "very useful"). Diagnosticians' rating and qualitative feedback during the follow-up interview confirmed that the videos collected by parents during the in-field evaluation were clinically useful (average rating of 4) for conducting remote diagnostic assessment.

Completion of Diagnostic Assessment via NODA Connect

The iterative design process of NODA Connect in Stage 2 helped finalize features that support diagnosticians in completing a diagnostic assessment based on the videos recorded by parents. These include the following: (1) a set of predefined tags that the diagnostician can use to flag specific child behaviors in the videos; (2) an integrated DSM checklist where each tag assigned by a clinician is mapped to the relevant DSM subcriterion; and (3) access to the child's developmental history entered by the parent into the system.

Once diagnosticians receive appropriate video recordings, they can review them and begin tagging them with behaviors relevant to diagnosing autism (Figure 3). The NODA Connect has a built-in set of tags representing specific behavioral markers such as "no eye contact" or "repetitive play," which were created based on the diagnostic criteria for autism within the DSM. The list of tags was compiled by the collaborating diagnostician and the autism domain expert, and vetted through conversations with several other clinical experts. In total, there were 66 tags included in NODA Connect. These tags included both atypical (n=57) behavior tags (representing atypical development) and typical (n=9) behavior tags (representing typical development). The goal of the tagging step is to have the diagnostician watch the videos for any evidence of atypical or typical behavior, and flag moments in time when that behavior occurs, without yet considering specific DSM criteria.

Once all the videos for a child are viewed and tagged, the diagnostician can review the DSM diagnostic checklist (Figure 4). At the time this research study was conducted, the DSM-IV was still in use, and thus, formed the basis of the diagnostic checklist in the NODA Connect assessment portal. Subsequent to the release of the DSM-V, tags and the diagnostic checklist were updated to reflect the new framework.

The DSM checklist contains categories of symptoms, with specific subcriteria for each category. The DSM-IV included the following diagnostic categories: (1) qualitative impairments in social interaction; (2) qualitative impairments in communication; and (3) restricted, repetitive and stereotyped patterns of behavior, interests, and activities. Each of these 3 categories included 4 subcriteria. Within NODA Connect, each tag inserted in the videos by the diagnostician during the video review step is automatically mapped to the relevant subcriterion, and shows up as a video snippet (Figure 4).

Within the DSM checklist, the diagnosticians can review the tags and then check a Yes/No box to indicate whether, based on the tags and their clinical judgment, the child meets that specific criterion. Once the entire DSM checklist is filled, the diagnostician makes a determination about the child's diagnosis of autism based on the DSM criteria, developmental history, and clinical judgment. Note that within NODA Connect the diagnosticians have access to the child's developmental history that parents fill during the in-home video-collection phase,

although during our in-field evaluation we restricted access to this checklist so that diagnosticians would remain blind to the child's true diagnostic status.

Results from the in-field evaluation confirmed that NODA Connect features support diagnosticians in completing a diagnostic assessment. Overall, in 91% of assessments (10/11) via NODA Connect, diagnosticians reached a decision about diagnostic outcome that matched with the child's previous diagnostic status.

Analysis of the NODA Connect usage pattern during the in-field evaluation showed that there was not much variability in the time taken to complete tagging and filling the DSM checklist by different diagnosticians. Across all assessments, the total time taken to complete tagging all the videos of a child and filling out the DSM checklist was on average 62 minutes (SD 14.8 minutes). Diagnosticians spent the least amount of time (average 37 minutes) completing the assessment for the child who was typically developing. After removing the two assessments for this child from the analysis, the average time spent tagging videos and completing the DSM checklist was even more consistent (average 68 minutes, SD 8.3 minutes).

Figure 3. NODA Connect: Web-based assessment portal video tagging.

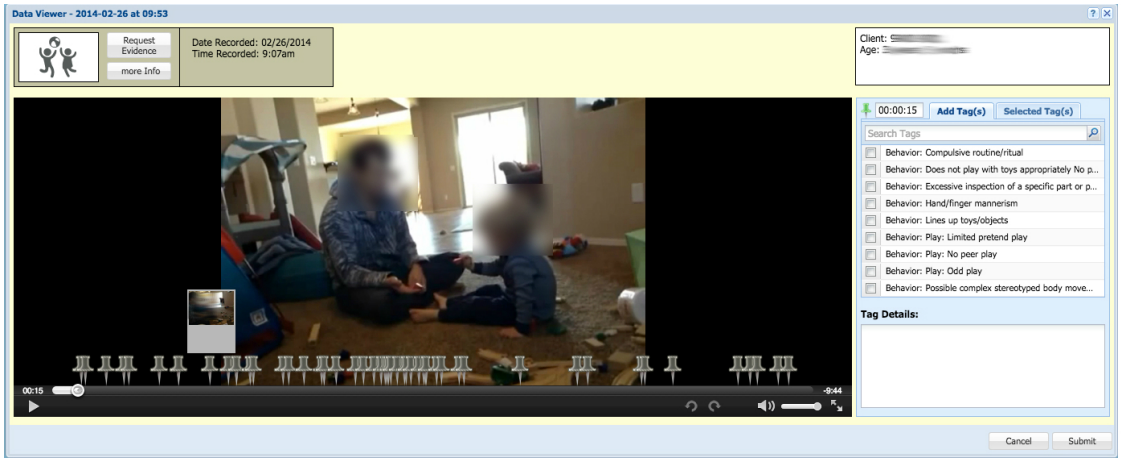
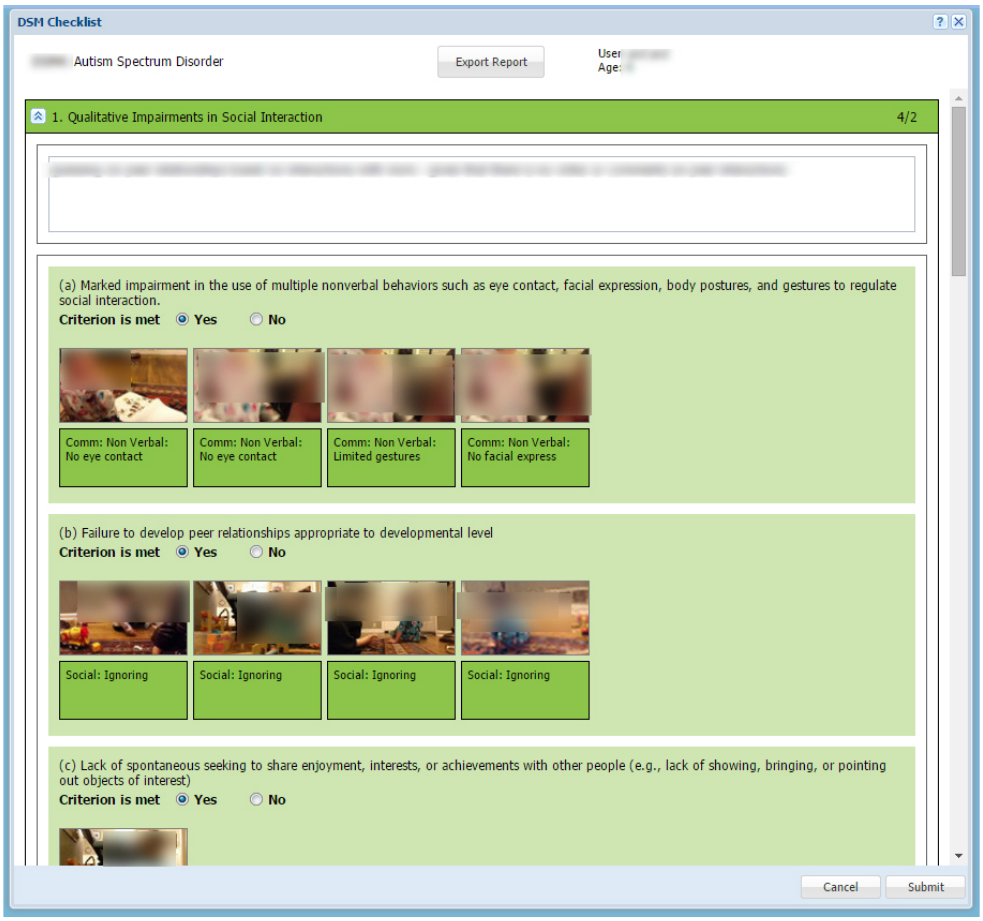


Figure 4. NODA Connect: Web-based assessment portal showing the DSM checklist screen.



Comparative Results of Diagnostic Outcomes

The main focus of the work presented here was to iteratively develop and evaluate the design of the two systems, NODA smartCapture and NODA Connect. In addition, given that during the in-field evaluation the diagnosticians assigned a diagnosis to the child upon completing the assessment, we were able to compare the diagnosis conducted through NODA Connect with the child's previous diagnosis as indicated in the child's medical record. For 4 of the 5 children (3 children with a previous diagnosis of autism and 1 typically developing child), both the remote diagnosticians independently arrived at the same diagnostic decision, and in agreement with the child's actual diagnostic status. For the fifth child (with a previous autism diagnosis), one diagnostician matched the diagnosis in the child's record but the other did not, although the latter indicated with high confidence that the child was not typically developing. A third diagnostician independently reviewed this case via NODA Connect and also confirmed the diagnosis in the child's medical record. Overall, in 91% of assessments (10/11) via NODA Connect, diagnosticians reached a decision about diagnostic outcome that matched with the child's previous diagnostic status.

Discussion

Principal Findings

The iterative design approach undertaken in this study enabled us to identify specific features of a store-and-forward telehealth platform that supports remote diagnosis of autism using videos recorded by families in their homes. The results of the in-field evaluation of NODA smartCapture and NODA Connect demonstrated that our system design allowed parents to easily capture clinically useful evidence of child behavior, and diagnosticians to complete a diagnostic assessment of autism with high confidence. See Multimedia Appendix 1 for the most recent version of NODA Capture and Connect resulted from this work.

This section discusses the perspectives of the various stakeholders on the perceived utility and limitations of the system, potential design enhancements, our vision for the large-scale adoption of the system within current autism diagnostic practices, and how our prescription, collection, and assessment model can be generalized to other clinical assessment applications.

Perceived Utility and Limitations

During initial stakeholder interviews in Stage 1, parents and clinicians considered the concept of video collection and sharing of in-home behavioral evidence potentially valuable for a variety of reasons. They reported that this approach can allow clinicians to observe otherwise inaccessible behaviors (eg, less-frequent behaviors, behavior triggers at home) in their natural context, and to view family-child interactions. Moreover, it can efficiently connect parents and clinicians for timely assessment of the child as, unlike current practice, clinicians can have immediate access to the behavior evidence. However, during the same interviews, parents and clinicians also highlighted several potential barriers to the adoption of an in-home video-recording system. The most commonly mentioned concerns were system complexity, privacy concerns, and child's reactivity. Parents suggested that having explicit data capture and sharing policies, and control over data collection and sharing would alleviate privacy concerns. They also indicated that they would be willing to sacrifice some privacy concerns to get help with a more timely diagnosis for their child. Parents and clinicians also highlighted that the recording device may cause the child to react differently than he or she would otherwise. However, clinicians reported that for them, the child's reactivity to being recorded would not necessarily invalidate the clinical utility of the video evidence, as such reactivity happens during clinic-based observations as well. Parents and clinicians appreciated that the recording application could be installed on mobile phones and tablets because these are everyday objects that children are used to seeing and the reactivity effects would thus likely be minimal.

During the in-field evaluation, diagnosticians appreciated that the system helped them conduct an autism diagnosis based on naturalistic behavioral evidence. They also highlighted that, unlike direct observation, video observation would allow them to go back in time to review and verify certain observations, if required. Among the 3 participating diagnosticians in the in-field evaluation, 2 had no previous experience with video observation. These 2 reported that before the study they were reluctant and skeptical about the value of in-home video recording for diagnostic assessment. The third diagnostician had previously participated in other research efforts that involve video observation for assessment and interventions of children with autism and had previously found these methods valuable. However, all the diagnosticians, irrespective of their initial biases, reported that using the remote diagnosis system left them feeling it was extremely valuable and effective for remote autism diagnosis. However, the diagnosticians also identified potential situations when in-home behavior evidence along with a brief developmental history may not be sufficient to complete a diagnostic assessment of autism. These situations included (1) when the child is too young (<2 years old); (2) when the child has very subtle characteristics of autism; and (3) when the child's level of functioning is very limited. According to the diagnosticians, in all these cases it may be difficult to make a judgment about the child's overall development level, which is required for comparison with the child's social profile. In such cases, supplementary evidence in addition to video evidence would be required, which, depending on the situation, could be a parent report, a standard developmental assessment, or even direct observation of the child.

Technology Enhancements

Advanced technology features can be incorporated into the existing NODA smartCapture and NODA Connect for built-in intelligence. For example, there are a number of factors associated with staging (lighting conditions, audio quality, field of view, whether the child's face is in view, etc) that a recording system could automatically detect during the recording and alert parents to rerecord without the need for the diagnostician to review the videos first. In addition, results from the in-field evaluation indicated that on average tagging videos took 84.6% (52.5/62 minutes) of the total time spent on completing 1

diagnostic assessment via NODA Connect. The amount of time spent on tagging could be significantly reduced if the Web-based assessment system were to include an automated tagging process. For example, certain detectable behaviors such as response to name call, a smile, giving or taking an object, eye contact, stereotypical behaviors are reasonable candidates to be automatically detected within collected video evidence, given the recent advances in automated video analysis [33-35]. Because these recognition techniques would not be perfect, the system could suggest potential tags in the video timeline and allow the diagnosticians to confirm or reject them. As another example, the assessment system could learn the diagnostician's tag assignment behaviors and highlight the most frequently assigned tags so the diagnostician could quickly locate them.

Diagnostic Workflow and Field Adoption

An open question for future research is to explore a feasible workflow for wide-scale adoption of our remote diagnostic system. One workflow that we envision involves a referral mechanism for remote diagnostic assessment like any other laboratory tests. Pediatricians are often the first medical professionals to identify children as potentially showing early signs of autism, and are responsible for referring families to a specialist for further assessment. In the proposed workflow, the pediatrician can refer the family for a remote diagnostic assessment. Upon connecting with the remote assessment service, the parents download NODA smartCapture directly to their mobile phone. A diagnostician at an affiliated diagnostic center can then guide the in-home evidence-collection procedure and complete the diagnostic assessment through NODA Connect. Finally, an electronic diagnostic report summarizing diagnostician's video observation, DSM checklist, and diagnostic outcome can be shared with the pediatrician, who then shares it with parents.

Overall, this workflow has two potential benefits. First, it engages pediatricians, which is beneficial because research suggests that pediatrician involvement in the referral and diagnostic process can result in more timely diagnosis [23,36-39]. A pediatrician sees children at regular intervals during the early years of development and is in the best position to note early warning signs and take appropriate timely action. Second, this workflow model may allow autism diagnostic centers to serve more families by remotely assessing children for the purposes of triage. Children whose diagnostic outcome is not clear through this remote procedure can be seen in person for a more comprehensive diagnostic assessment.

Generalizability of Our Approach

Our remote diagnostic assessment system is based on a prescription, behavior specimen collection, and assessment model. Analogous to traditional medical specimen collection and assessment process, this model involves (1) a clinician's prescription for behavior specimens (in the form of short videos) to be collected; (2) in-home collection of behavior specimens by parents; and (3) the assessment of behavior specimens by a remotely located clinician.

This is a generic model that is transferable, beyond remote autism diagnosis, to other clinical situations where an analysis of behavior by a professional is key to the clinical assessment. Any condition or situation in which observation of behavior in the natural environment is of value, and for which those behaviors can be specified to ensure relevant examples are recorded, is a candidate use case. During stakeholder interviews, the participating clinicians suggested a number of potential use cases where this model could be applicable and valuable. One such use case is to sort and prioritize families on waiting lists for clinical services to expedite the intake process. Sorting and prioritizing the waiting list is crucial, because timely access to diagnostic and intervention services is often hampered by long waiting lists at centers and clinics. In addition, a system based on such a model may be valuable for providing treatment and follow-up services to remotely located patients who do not have easy access to the clinic. Another use case is parent training, such as those involving clinicians training parents to implement an intervention at home.

Although the prescription, collection, and assessment model along with its high-level design features (embedded prescription, guided capture through notification feature, tagging, video observation, assessment based on mapped tags) are both generic, it must be customized within the context of its end-use-case scenario. For instance, the embedded prescribed instructions in the recording application can be contextualized through a new prescription writing pad feature within the Web-based assessment portal. One example of successful transfer and customization of our approach is the use case of medication management. In-home behavior specimens captured and shared through the mobile phone-based system allow physicians to monitor medication side effects and note any improvements in symptoms between office visits using the Web-based assessment portal. In a preliminary evaluation, physicians highlighted that this medication administration system assisted them in monitoring patients with autism spectrum disorder more comprehensively and accurately than using subjective reports provided by caregivers during office visits [40].

Conclusions

The in-field evaluation demonstrated that the system's design enabled parents to easily record clinically valid evidence of their child's behavior, and diagnosticians to complete a diagnostic assessment for autism. These results shed light on the potential for appropriately designed telehealth technology to support clinical assessments using in-home video captured by families. This assessment model can be readily generalized to other conditions where direct observation of behavior plays a central role in the assessment process.

The results of this paper are not a final statement on the clinical validity of diagnostic outcome; rather, this paper reports on the design of the remote autism diagnosis system that resulted from an iterative design process and has shown a promising conclusion from an evaluation in the field. The next step is to validate the diagnostic outcome through a clinical trial in which a large sample of children would be assessed via both remote autism diagnosis system and standard in-person diagnostic assessments for comparison.

Acknowledgments

We would like to thank all participants for their time and valuable feedback. All phases of this study were supported by a National Institutes of Health grant (Grant No 9 R44 MH099035-04).

Conflicts of Interest

RO and GA have conflicts of interests. Mr RO is the CEO of Behavior Imaging Solutions, the company that will commercialize NODA as part of the NIMH SBIR grant. Dr GA was a coadvisor for NN during her graduate school period, which presents a conflict of interest that is registered with and managed by Georgia Institute of Technology. The remaining authors have no conflicts of interest to disclose.

Multimedia Appendix 1

The most recent version of NODA smartCapture and Connect.

[PDF File (Adobe PDF File), 1MB - mhealth_v3i2e68_app1.pdf]

References

1. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2002 Principal Investigators, Centers for Disease Control and Prevention. Prevalence of autism spectrum disorders: Autism and developmental disabilities monitoring network, 14 sites, United States, 2002. MMWR Surveill Summ 2007 Feb 9;56(1):12-28 [FREE Full text] [Medline: 17287715]
2. Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators, Centers for Disease Control and Prevention (CDC). Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. MMWR Surveill Summ 2014 Mar 28;63(2):1-21 [FREE Full text] [Medline: 24670961]
3. Howlin P, Asgharian A. The diagnosis of autism and Asperger syndrome: Findings from a survey of 770 families. Dev Med Child Neurol 1999 Dec;41(12):834-839. [Medline: 10619282]
4. Sivberg B. Parents' detection of early signs in their children having an autistic spectrum disorder. J Pediatr Nurs 2003 Dec;18(6):433-439. [Medline: 15058541]
5. Wiggins LD, Baio J, Rice C. Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. J Dev Behav Pediatr 2006 Apr;27(2 Suppl):79-87. [Medline: 16685189]
6. Thomas KC, Ellis AR, McLaurin C, Daniels J, Morrissey JP. Access to care for autism-related services. J Autism Dev Disord 2007 Nov;37(10):1902-1912. [doi: 10.1007/s10803-006-0323-7] [Medline: 17372817]
7. Moldin SO, Rubenstein JL. Understanding Autism: From Basic Neuroscience to Treatment. Boca Raton, FL: CRC Press; Apr 25, 2006.
8. Shattuck PT, Durkin M, Maenner M, Newschaffer C, Mandell DS, Wiggins L, et al. Timing of identification among children with an autism spectrum disorder: Findings from a population-based surveillance study. J Am Acad Child Adolesc Psychiatry 2009 May;48(5):474-483 [FREE Full text] [doi: 10.1097/CHI.0b013e31819b3848] [Medline: 19318992]
9. Mandell DS, Novak MM, Zubritsky CD. Factors associated with age of diagnosis among children with autism spectrum disorders. Pediatrics 2005 Dec;116(6):1480-1486 [FREE Full text] [doi: 10.1542/peds.2005-0185] [Medline: 16322174]
10. Matson JL. Clinical Assessment and Intervention for Autism Spectrum Disorders. Amsterdam, The Netherlands: Elsevier/Academic Press; Mar 2008.
11. Siegel B. Getting the Best for Your Child with Autism: An Expert's Guide to Treatment. New York: Guilford Press; Jan 1, 2008.
12. McEachin JJ, Smith T, Lovaas OI. Long-term outcome for children with autism who received early intensive behavioral treatment. Am J Ment Retard 1993 Jan;97(4):359-372. [Medline: 8427693]
13. Rattazzi A. The importance of early detection and early intervention for children with autism spectrum conditions. Vertex 2014;25(116):290-294. [Medline: 25546644]
14. Ford T. In: McLeod BD, Jensen-Doss A, Ollendick TH, editors. Diagnostic and Behavioural Assessment in Children and Adolescents: A Clinical Guide. New York: Guilford Press; Apr 4, 2014:159.
15. Yoder P, Symons F. Observational Measurement of Behavior. New York: Springer Publishing Company; Feb 16, 2010.
16. Suen HK, Ary D. Analyzing Quantitative Behavioral Observation Data. Hillsdale, NJ: Lawrence Erlbaum; Jan 1, 1989.
17. Frick PJ, Barry CT, Kamphaus RW. Clinical Assessment of Child and Adolescent Personality and Behavior. New York: Springer; Dec 22, 2009.
18. Kazdin AE. Artifact, bias, and complexity of assessment: The ABCs of reliability. J Appl Behav Anal 1977;10(1):141-150 [FREE Full text] [Medline: 16795543]

19. Wainer AL, Ingersoll BR. Increasing access to an ASD imitation intervention via a telehealth parent training program. J Autism Dev Disord 2014 Jul 18 Epub ahead of print. [doi: 10.1007/s10803-014-2186-7] [Medline: 25035089]
20. Vismara LA, Young GS, Rogers SJ. Telehealth for expanding the reach of early autism training to parents. Autism Res Treat 2012;2012:121878 [FREE Full text] [doi: 10.1155/2012/121878] [Medline: 23227334]
21. Baharav E, Reiser C. Using telepractice in parent training in early autism. Telemed J E Health 2010;16(6):727-731. [doi: 10.1089/tmj.2010.0029] [Medline: 20583950]
22. Wacker DP, Lee JF, Padilla Dalmau YC, Kopelman TG, Lindgren SD, Kuhle J, et al. Conducting functional communication training via telehealth to reduce the problem behavior of young children with autism. J Dev Phys Disabil 2013 Feb 1;25(1):35-48 [FREE Full text] [doi: 10.1007/s10882-012-9314-0] [Medline: 23543855]
23. Goldstein F, Myers K. Telemental health: A new collaboration for pediatricians and child psychiatrists. Pediatr Ann 2014 Feb;43(2):79-84. [doi: 10.3928/00904481-20140127-12] [Medline: 24512157]
24. Vismara LA, Young GS, Stahmer AC, Griffith EM, Rogers SJ. Dissemination of evidence-based practice: Can we train therapists from a distance? J Autism Dev Disord 2009 Dec;39(12):1636-1651 [FREE Full text] [doi: 10.1007/s10803-009-0796-2] [Medline: 19582564]
25. Parmanto B, Pulantara IW, Schutte JL, Saptono A, McCue MP. An integrated telehealth system for remote administration of an adult autism assessment. Telemed J E Health 2013 Feb;19(2):88-94. [doi: 10.1089/tmj.2012.0104] [Medline: 23230821]
26. Terry M. Telemedicine and autism: Researchers and clinicians are just starting to consider telemedicine applications for the diagnosis and treatment of autism. Telemed J E Health 2009 Jun;15(5):416-419. [doi: 10.1089/tmj.2009.9965] [Medline: 19548820]
27. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR), 4th Edition, Text Revision. Washington, DC: American Psychiatric Association; Jun 2000.
28. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5. Washington, DC: American Psychiatric Association; May 2013.
29. Smith CJ, Oberleitner R, Treulich K, McIntosh R, Melmed R. International Meeting for Autism Research (IMFAR). 2009 May 8. Using a behavioral imaging platform to develop a naturalistic observational diagnostic assessment for autism URL: https://imfar.confex.com/imfar/2009/webprogram/Paper4954.html [accessed 2015-06-12] [WebCite Cache ID 6ZB4m2Oa2]
30. Nazneen N, Rozga A, Romero M, Findley AJ, Call NA, Abowd GD, et al. Supporting parents for in-home capture of problem behaviors of children with developmental disabilities. Pers Ubiquit Comput 2011 May 1;16(2):193-207. [doi: 10.1007/s00779-011-0385-1]
31. Aware Home Research Initiative (AHRI). Aware Home Research Initiative. URL: http://www.awarehome.gatech.edu/ [accessed 2015-03-03] [WebCite Cache ID 6WlKAtKgT]
32. Sears A, Jacko JA. The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications. Mahwah, NJ: Lawrence Erlbaum Associates; 2003.
33. Bidwell J, Rozga A, Essa I, Abowd GD. Measuring child visual attention using markerless head tracking from color and depth sensing cameras. 2014 Nov 12 Presented at: Proceedings of the 16th ACM International Conference on Multimodal Interaction, ICMI; 2014; Istanbul, Turkey p. 447-454. [doi: 10.1145/2663204.2663235]
34. Rehg J, Abowd GD, Rozga A, Romero M, Clements M, Sclaroff S, et al. Decoding children's social behavior. New York: IEEE; 2013 Jun 23 Presented at: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2013; Portland, OR p. 3414-3421. [doi: 10.1109/CVPR.2013.438]
35. Rao H, Kim JC, Rozga A, Clements MC. Detection of laughter in children's speech using spectral and prosodic acoustic features. 2013 Aug 23 Presented at: Proceedings of the 14th Annual Conference of the International Speech Communication Association, Interspeech; 2013; Lyon, France p. 1399-1403.
36. Committee on Children With Disabilities. American Academy of Pediatrics: The pediatrician's role in the diagnosis and management of autistic spectrum disorder in children. Pediatrics 2001 May;107(5):1221-1226. [Medline: 11331713]
37. Zuckerman KE, Mattox K, Donelan K, Batbayar O, Baghaee A, Bethell C. Pediatrician identification of Latino children at risk for autism spectrum disorder. Pediatrics 2013;132(3):445-453 [FREE Full text] [doi: 10.1542/peds.2013-0383] [Medline: 23958770]
38. Swanson AR, Warren ZE, Stone WL, Vehorn AC, Dohrmann E, Humberd Q. The diagnosis of autism in community pediatric settings: Does advanced training facilitate practice change? Autism 2013;18(5):555-561. [doi: 10.1177/1362361313481507] [Medline: 23847130]
39. Hart-Hester S, Noble SL. Recognition and treatment of autism: The role of the family physician. J Miss State Med Assoc 1999 Nov;40(11):377-383. [Medline: 10568088]
40. Reischl U, Oberleitner R. Telehealth technology enabling medication management of children with autism. In: Duffy V, editor. Advances in Human Factors and Ergonomics in Healthcare. Boca Raton, FL: CRC Press; 2012.

Abbreviations

DSM: Diagnostic and Statistical Manual of Mental Disorders
NODA: naturalistic observation diagnostic assessment

Edited by G Eysenbach; submitted 04.03.15; peer-reviewed by A Ellertson, J Suri; comments to author 26.04.15; accepted 26.04.15; published 17.06.15.

Please cite as:
Nazneen N, Rozga A, Smith CJ, Oberleitner R, Abowd GD, Arriaga RI
A Novel System for Supporting Autism Diagnosis Using Home Videos: Iterative Development and Evaluation of System Design
JMIR mHealth uHealth 2015;3(2):e68
URL: http://mhealth.jmir.org/2015/2/e68/
doi:10.2196/mhealth.4393
PMID:26085230

Original Paper

Automated Personalized Feedback for Physical Activity and Dietary Behavior Change With Mobile Phones: A Randomized Controlled Trial on Adults

Mashfiqui Rabbi[1], MS; Angela Pfammatter[2], PhD; Mi Zhang[3], PhD; Bonnie Spring[2], PhD; Tanzeem Choudhury[1], PhD

[1]Cornell University, Department of Information Science, Ithaca, NY, United States

[2]Northwestern University, Department of Preventive Medicine, Chicago, IL, United States

[3]Michigan State University, Department of Electrical and Computer Engineering, East Lansing, MI, United States

Corresponding Author:
Mashfiqui Rabbi, MS
Cornell University
Department of Information Science
219 Gates Hall
Ithaca, NY, 14850
United States
Phone: 1 6036671797
Fax: 1 6072555196
Email: ms2749@cornell.edu

Abstract

Background: A dramatic rise in health-tracking apps for mobile phones has occurred recently. Rich user interfaces make manual logging of users' behaviors easier and more pleasant, and sensors make tracking effortless. To date, however, feedback technologies have been limited to providing overall statistics, attractive visualization of tracked data, or simple tailoring based on age, gender, and overall calorie or activity information. There are a lack of systems that can perform automated translation of behavioral data into specific actionable suggestions that promote healthier lifestyle without any human involvement.

Objective: MyBehavior, a mobile phone app, was designed to process tracked physical activity and eating behavior data in order to provide personalized, actionable, low-effort suggestions that are contextualized to the user's environment and previous behavior. This study investigated the technical feasibility of implementing an automated feedback system, the impact of the suggestions on user physical activity and eating behavior, and user perceptions of the automatically generated suggestions.

Methods: MyBehavior was designed to (1) use a combination of automatic and manual logging to track physical activity (eg, walking, running, gym), user location, and food, (2) automatically analyze activity and food logs to identify frequent and nonfrequent behaviors, and (3) use a standard machine-learning, decision-making algorithm, called multi-armed bandit (MAB), to generate personalized suggestions that ask users to either continue, avoid, or make small changes to existing behaviors to help users reach behavioral goals. We enrolled 17 participants, all motivated to self-monitor and improve their fitness, in a pilot study of MyBehavior. In a randomized two-group trial, investigators randomly assigned participants to receive either MyBehavior's personalized suggestions (n=9) or nonpersonalized suggestions (n=8), created by professionals, from a mobile phone app over 3 weeks. Daily activity level and dietary intake was monitored from logged data. At the end of the study, an in-person survey was conducted that asked users to subjectively rate their intention to follow MyBehavior suggestions.

Results: In qualitative daily diary, interview, and survey data, users reported MyBehavior suggestions to be highly actionable and stated that they intended to follow the suggestions. MyBehavior users walked significantly more than the control group over the 3 weeks of the study (P=.05). Although some MyBehavior users chose lower-calorie foods, the between-group difference was not significant (P=.15). In a poststudy survey, users rated MyBehavior's personalized suggestions more positively than the nonpersonalized, generic suggestions created by professionals (P<.001).

Conclusions: MyBehavior is a simple-to-use mobile phone app with preliminary evidence of efficacy. To the best of our knowledge, MyBehavior represents the first attempt to create personalized, contextualized, actionable suggestions automatically from self-tracked information (ie, manual food logging and automatic tracking of activity). Lessons learned about the difficulty of manual logging and usability concerns, as well as future directions, are discussed.

Trial Registration: ClinicalTrials.gov NCT02359981; https://clinicaltrials.gov/ct2/show/NCT02359981 (Archived by WebCite at http://www.webcitation.org/6YCeoN8nv).

(JMIR mHealth uHealth 2015;3(2):e42) doi:10.2196/mhealth.4160

KEYWORDS

mobile health; mHealth; mobile phone sensing; smart systems; context-aware systems; physical activity; self-management; personal health care; machine learning; artificial intelligence

Introduction

In 2010, the World Health Organization (WHO) attributed 63% of deaths to noncommunicable diseases that are largely preventable [1]. The Centers for Disease Control and Prevention (CDC) estimates that in the US nearly 200,000 deaths annually could be prevented based upon modifications in diet, exercise, and obesity [2]. Obesity alone affects more than one-third of the adult population [3] and burdens the US with an estimated US $190 billion annually in health care costs [4].

A rapid rise has occurred in the development of mobile phone apps and wearable devices to address diet and physical activity. While empirical data is lacking for some commercial apps and sensor-based technologies [5,6], a number of scientific studies have explored the impact of novel technology-supported behavior change strategies on physical activity [7-9]. For example, Weegen et al [10] applied behavior change theories to design a mobile app that visualized a summary of physical activity logs and gave clinicians feedback to support their promotion of physical activity. Food logging has proved to be more difficult, burdensome, and time consuming than tracking physical activity. Recent work, however, has attempted to use image-based systems to decrease burden and enhance accuracy in food tracking with some success [11-13]. The ubiquity and ever-presence of mobile phones gives them the potential to perform assessment and intervention in the right place at the right time.

Although these methods show promise, they continue to fall short by not providing context-specific, relevant, personalized help at the moment when the individual needs it to make healthier choices. The science of how to present daily physical activity and dietary intake data back to users also has been at a suboptimal state. To date, feedback has been limited to one of three categories: (1) overall numeric summaries [7,8,14] (eg, step counts), (2) tailored suggestions that only adapt to personal characteristics (eg, age, gender) and overall behavior (eg, daily calories consumed and burned) [15], and (3) visualizations that incorporate little processing [16]. Simple goals are offered, but without actionable insights on when, where, and how to achieve them. Visualization of large amounts of minimally processed data produces a related problem—information overload without clear steps to behavior change. Providing personalized, in-the-moment, actionable guidance that prompts smaller, but more frequent, changes in existing behavior has potential for greater impact. A deeper look into physical activity and dietary intake data can reveal patterns of both healthy and unhealthy behavior that could be leveraged for personalized feedback. With current technologies, this can be achieved automatically, without human interpretation.

Given these observations, MyBehavior was created to address some shortcomings of current mHealth interventions. MyBehavior uses a machine-learning model—multi-armed bandit (MAB)—to automatically create contextualized and personalized suggestions based on the individual's physical activity and dietary intake data collected solely from a mobile phone. Moreover, MyBehavior is one of the very few mHealth apps designed on the basis of established behavioral theory. As such, the system reflects and incorporates the contemporary state of the behavioral science knowledge about how to foster healthful change. Based on effective behavior change principles, MyBehavior provides low-effort suggestions that request small changes to users' existing repeated behaviors. To the best of our knowledge, MyBehavior is the first mHealth app that encourages healthy behavior change by automatically providing low-effort suggestions based on the user's context and personal information.

The objective of this study was to evaluate a new behavior change technology—MyBehavior—using a mixed-method approach as suggested by others [17]. We focused on (1) whether the users intended to follow the automated MyBehavior suggestions, (2) early indications of behavior change empowered by automated suggestions, and (3) participant feedback that could inform user experience and guide future design of automated health feedback systems.

The outcomes from this study will be used to further refine the features and messages of MyBehavior to optimize its effect on physical activity and dietary intake.

Methods

Study Procedure

To evaluate the feasibility of MyBehavior, a small 3-week, two-group randomized control trial (RCT) was conducted. The team that supervised the trial included the builders of the MyBehavior app and authors of this paper. This team recruited participants through advertisements placed around the Cornell University campus. In the advertisement, we invited participants to test a new mobile app to help them stay on track for physical activity and food intake. Recruitment was restricted to participants who owned an Android mobile phone and had an interest in fitness. Prior to the study, the investigators arranged face-to-face meetings with the participants and acquired their informed consent. Participants also completed a brief survey to provide demographic data and information about their prior experience with mobile technologies and weight loss/fitness apps. All participants attended a training session, where they installed MyBehavior on their primary mobile phone and received basic instructions, including how to enter their gender,

height, and weight and how to set up a weekly weight goal (ie, lose weight, maintain weight, or gain weight). During the first week, users received a daily summary of their activities and food intake. This baseline week was intended to resemble many modern mobile health apps [5,6] without suggestions on what behaviors to change.

After the first week, the experimenters conducted an in-depth, semistructured interview with participants about their experience to date and then randomized participants into control and experimental groups. A random number generator was used for randomization. Assignment was single blind, as the study participants did not know their condition, while experimenters had full knowledge about the assignments.

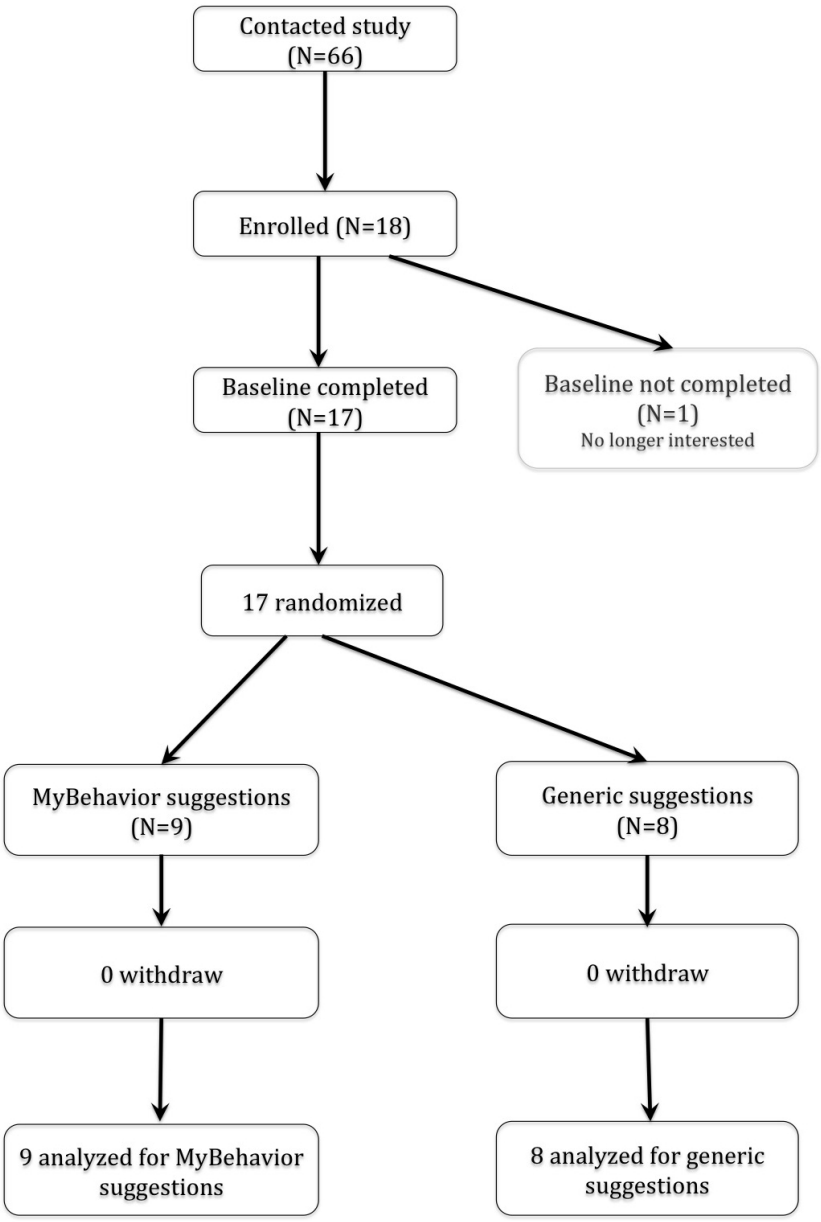
We provided MyBehavior's personalized context-sensitive suggestions to the experimental group, while the control group received generic prescriptive recommendations generated from a pool of 42 suggestions for healthy living, such as "walk for 30 minutes" and "eat fish for dinner." A certified fitness professional created these generic suggestions after following National Institutes of Health resources [18,19]. An external nutrition counselor also reviewed the suggestions to ensure that they were both healthy and achievable. The list of these 42 suggestions is included as Multimedia Appendix 1 in this paper. For the following 2 weeks, participants continued to log behaviors and receive their respective suggestions on their mobile phones. During the entire study period, we asked participants to complete Web-based daily diaries to better understand their experience in following the suggestions provided. At the conclusion of the 3-week period, all participants were asked to complete a brief survey about the suggestions provided and were interviewed again face-to-face about their experience with the app.

This study was approved by Cornell University Institutional Review Board (1302003617) and a protocol was registered retrospectively at ClinicalTrails.gov (NCT02359981).

Participants

We recruited 18 participants, 17 of whom completed the study. Of the 17 participants, there were 13 students (76%), 4 professionals (24%), 8 females (47%), and 9 males (53%), and all were between the ages of 18 and 49 (mean 28.3, SD 6.96, lower quartile [q_{25}]=22, median [q_{50}]=26.3, upper quartile [q_{75}]=36). All participants reported low-to-moderate levels of physical activity. The majority of participants were experienced mobile phone users—9 participants (53%) had previous experience using a food diary, and 6 participants (35%) had previously kept an exercise log. After the randomization, participants in the groups were similar in terms of level of active lifestyle and experience with using mobile-based self-management tools. Our sample size was determined based on earlier literature [17,20,21] that suggested that small studies ($n \geq 4$) are more suitable to test early feasibility of novel behavior change technologies like MyBehavior. See Figure 1 for the flow of participants in the trial.

Figure 1. Flow of participants in the MyBehavior trial.



MyBehavior Mobile App

Overview

MyBehavior is comprised of five key modules: (1) physical activity tracking, (2) food logging, (3) life-log generation, (4) physical activity and food clustering, and (5) suggestion generation.

Physical Activity Tracking

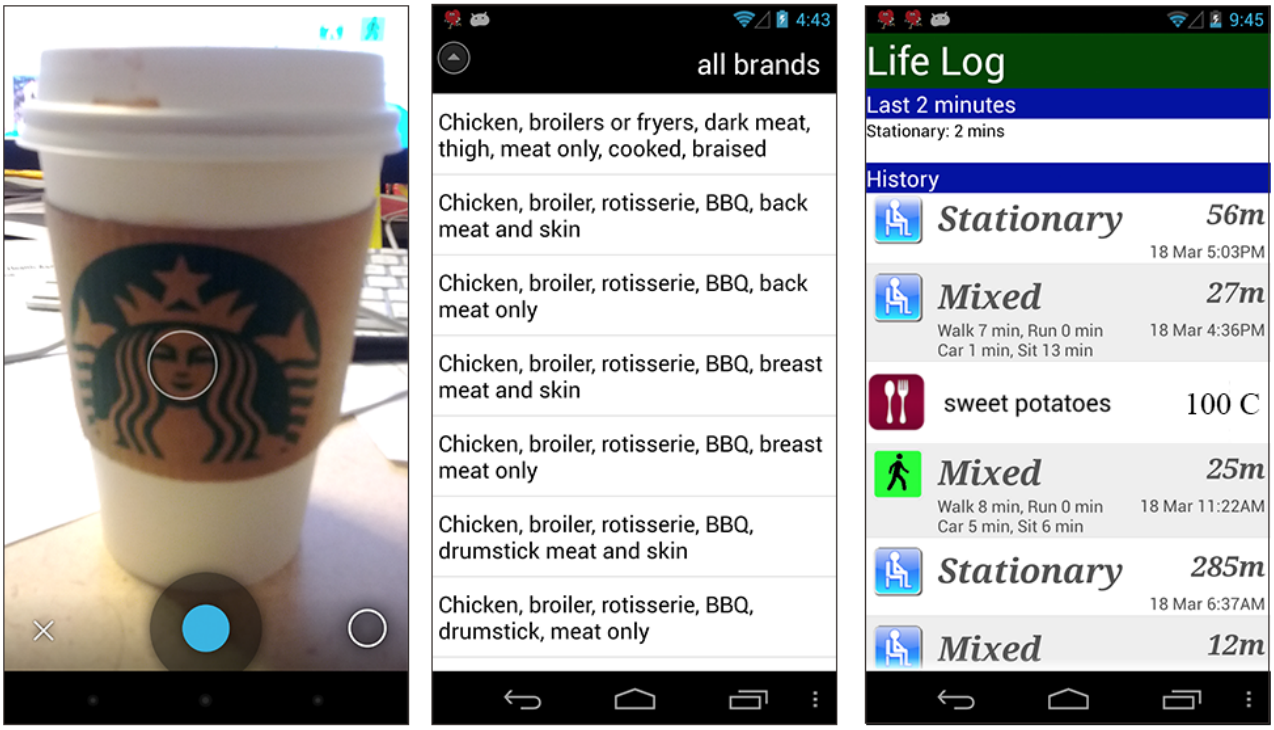
MyBehavior uses the accelerometer and the Global Positioning System (GPS) sensor inside the mobile phone to continuously keep track of an individual's physical activities. A number of statistical features (eg, mean, variance, zero-crossing rate) are extracted from the sensor data and a machine-learning model—Gaussian Mixture Model (GMM) [22]—is applied to map the extracted feature values into the four most common daily physical activities—walking, running, stationary (sitting or standing), and driving. The technical details of this mobile sensing framework can be found in our previous work [23]. For physical activities that cannot be automatically recognized, MyBehavior provides users with a list of about 800 different physical activities from the compendium of physical activities [24]. Users can manually select the specific physical activity from the list and record the start and end time for the activity. In addition to tracking physical activities, MyBehavior calculates the calories expended during these activities based on the standard Metabolic Equivalents of Task (METS) [25].

Food Logging

Users select food items from a database and enter the consumed quantity to get the corresponding calorie intake. The United States Department of Agriculture (USDA) [26] maintains this database, containing more than 8000 types of food.

MyBehavior provides several features to make the food logging experience easier for users. First, users can take photos of their food. These photos serve as a memory aid when users are prompted to input their food information at 9:30pm every night (see Figure 2). Second, to facilitate entry of frequently repeated food choices, MyBehavior allows users to input food which reuses a prior meal (eg, add yesterday's breakfast) and prioritizes food items that were selected previously. Finally, MyBehavior provides users with an option to directly input calorie information taken from the label of prepackaged foods (eg, a soft drink can or yogurt cup).

Figure 2. MyBehavior app screenshots: (a) taking photo of a food item, (b) searching for foods from the USDA database, and (c) Life Log, a chronological list of activity and food log events.



Life-Log Generation

MyBehavior generates as a "life log," a chronological list of activity and food log events, as shown in Figure 2. The log includes food, automatically sensed physical activity, and manually logged exercise entries as life events. To create concise and meaningful activity entries, MyBehavior processes the data into two stages.

In the first stage, activity predictions, which happen every 1 second, are aggregated every minute and labeled automatically. In the second stage, the contiguous activities having the same label are combined into a single entry. For example, if a user is stationary for 50 minutes, MyBehavior will generate a "stationary" activity entry into the life log with a 50-minute duration. Other common life events include a sequence of different activities that happen within a short time interval (ie, 15 minutes). An example might be the following: walk to the bus stop, wait for a few minutes, ride the bus, and walk to the office after exiting the bus. In this example, MyBehavior generates a "mixed" activity entry in the life log (eg, taking the bus from home to work) by combining the multiple activity sequences that happen within a 15-minute window.

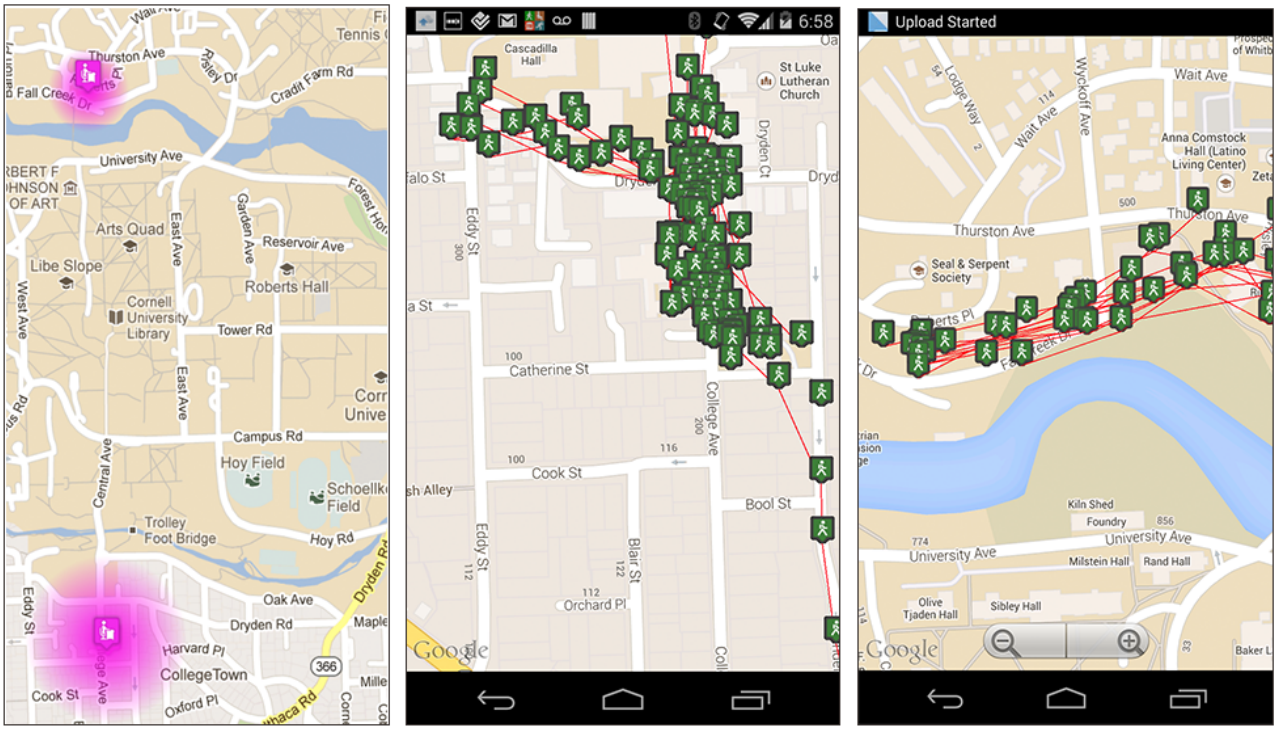
Physical Activity and Food Clustering

To enable suggestion generation delivery to the experimental group only, MyBehavior used the life logs to cluster similar physical activities and similar food items. The food similarity matching process follows a simple logic—food is clustered based on similar food ingredients. For example, MyBehavior will detect if a user is repeatedly having high-calorie burgers with similar ingredients and form the cluster "burger" that groups together the same or similar types of burgers.

Regarding clustering physical activities, manually tracked activities are clustered based on the type of activities similar to food clustering. Automatically tracked activities, tagged with location information, are clustered by places they occur. Clusters are found using unsupervised machine-learning techniques to identify similarity. As indoor localization is often accurate up to 150 meters, any stationary activities that fall within 150 meters of each other are clustered together. For example, a user's stationary activities in the office are typically in close proximity to each other and, as such, MyBehavior clusters these office locations into a single unit that represents the user's stationary behavior in the office. Walking and running activities are more difficult to cluster because MyBehavior needs to determine whether two activity trajectories look similar and happen at a similar location. To group similar walking or running events, MyBehavior uses an algorithm derived from the literature on handwriting recognition [27]. In handwriting recognition, the task is to find a canonical letter that matches the shape or trajectory of a handwritten letter. The analogous

task in MyBehavior is to find whether a new walking trajectory (eg, office to coffee shop) matches previous walking trajectories. Figure 3 shows some clusters generated by this technique. The image on the left represents a user's stationary episode in the office and home, whereas the middle and right-hand images show two walking clusters generated by two different users. The middle image represents a user's walks near the office, while the cluster in the right-hand image represents another user's daily walks from home to a bus stand.

Figure 3. Clusters generated from user activities: (a) locations where user A stayed stationary, (b) location traces for user B where he walked around his office, and (c) walking traces of user A from his house to the bus stop.



Suggestion Generation

After clustering user behaviors, MyBehavior uses an *exploit-explore strategy* to automatically generate suggestions based on users' past physical activities and food intake. This suggestions-generating strategy is grounded in contemporary behavioral science theories: (1) learning theory [28], (2) social cognitive theory [29], and (3) the Fogg Behavior Model (FBM) [30]. Behavior analysis applies learning theory first to assess whether a person has the skills needed to perform a behavior [28]. If so, the next step is to increase or decrease the target behavior's frequency by harnessing its antecedents (ie, its setting and cues) and consequences (ie, reinforcement). For example, if a health suggestion asks a user to swim but the user can't swim (ie, he never acquired the skills), the user will not follow the suggestion. On the other hand, if a person has performed a behavior before, even if rarely, the skills can be assumed present. The Fogg Behavior Model applies theoretical principles to technology design by creating tools to prompt *low-effort* actions that can be triggered even when motivation is low [30]. Thus, MyBehavior suggests (ie, cues or triggers) a frequent behavior (eg, a particular walk) that the person often does in a particular life context. This small, low-effort change simply increases the frequency of a behavior that the person already does. Sometimes instead, MyBehavior suggests an infrequent behavior (eg, bike ride) that would burn more calories and that the person has shown he/she can do, but does only rarely. Social cognitive theory [29], the most widely used behavioral theory, suggests that in order to voluntarily initiate an action, a person needs a sense of self-efficacy or confidence that he/she will be able to perform it. The more frequently the person can be triggered to ride a bike *repeatedly* in a certain *context* where bikes are accessible, the more self-efficacy increases, the less effortful the behavior becomes, and the more likely that bike riding becomes a habit.

MyBehavior *exploits* the frequency principle by suggesting activities that users perform repeatedly. In addition, the algorithm favors actions that are not only frequent, but also result in higher calorie expenditure. For example, short 1-minute walks inside the office, though very frequent, are likely to be superseded in the suggestion-generation engine by a less frequent, but higher-calorie-burning, gym class. On the other hand, if the person rarely visits the gym but walks 30 minutes to work several times a week, the recommender engine will rank the walk higher than the gym since the aggregate calorie loss—*frequency x calories* burned each instance—is higher.

For stationary activities, the recommender engine suggests small changes, such as walking 3 minutes for every hour spent stationary. The right-most image in Figure 4 shows a prioritization order of MyBehavior suggestions where simply adding 3-minute walks to the user's hour-long stationary episodes burns more calories compared to rarely occurring gym visits.

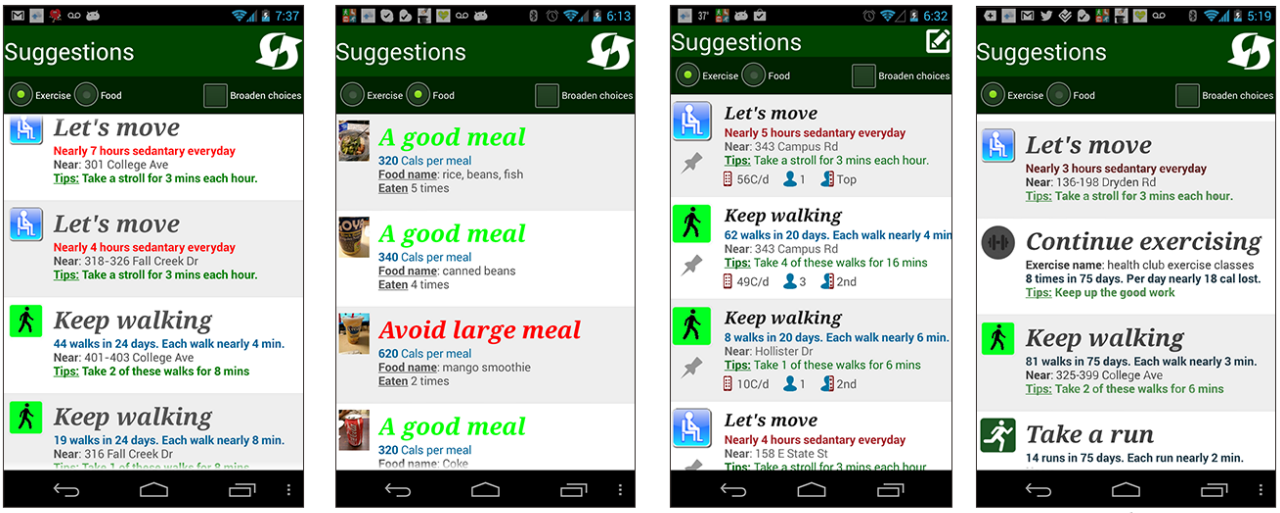
Exploit suggestions generated solely from users' frequent past behavior may not generate sufficient energy expenditure to cause weight loss. Consequently, MyBehavior periodically suggests higher-calorie-burning activities to entice the user to try out and adopt. *Explore* suggestions target infrequent, high-calorie-burning behaviors that the user can turn into a more regular activity. Future behavior is only imperfectly predicted by past behavior, and it could be the case that users will increase infrequent activities if suggested. Hence, if a user walks regularly near her office but sometimes goes to the gym or takes a long walk home, MyBehavior exploits this knowledge by suggesting walking near the office most of the time and by sometimes suggesting a gym visit or a long walk. If the new suggestion sticks and the user starts going to the gym regularly as a result, then MyBehavior learns to target the gym as an exploit suggestion rather than an explore suggestion.

When generating food suggestions, a separate set of suggestions is created based on the exploit-explore strategy. First, MyBehavior distinguishes between meals and snacks. Then it takes into account both intake frequency and calories similar to the physical activity suggestions. Thus, a user's frequent healthy low-calorie meals are exploited and are encouraged to be continued. During exploration, a random selection of infrequent low-calorie meals/snacks from the past is suggested. Here, the expectation is that users will take up some of these infrequent meals and make them frequent in the future.

At the start of every day, MyBehavior generates 10 food and 10 activity suggestions. Of these, 90% are from the users' most frequent activities (ie, exploit) and 10% are from the users' infrequent behaviors (ie, explore). This split of 90% exploit and 10% explore was heuristically chosen based on previous literature [31]. This kind of exploit-explore strategy, well grounded in artificial intelligence research, falls under a wider decision-making framework called multi-armed bandit [31]. MAB models have been well studied for modeling dynamic systems where situations can change over time. In our case, user behavior is not fixed and can change over time under MyBehavior's influence (see Figure 4, left-most and right-most images). The exploit-and-explore strategy models this dynamic nature of human behavior effectively. MyBehavior exploits the most common user behaviors that promote energy balance to produce short-term health gain. To target long-term health, it occasionally explores infrequent higher-energy-expending behaviors to discover actions that the user might repeat in the future, leading to sustained energy balance that could boost weight loss.

Figure 4 shows different generated suggestions that encourage the user to either continue positive activities (ie, low-calorie foods, walking, or exercise), make small changes in some situations (ie, stationary activities) (left-most image), or avoid negative activities (ie, frequent large meals) (second image). The first and third images in Figure 4 show suggestions for two different users and the first and last images show suggestions for the same user that change over time. A video demonstrating different features of MyBehavior is included as Multimedia Appendix 2.

Figure 4. Screenshots showing recommended suggestions for exercise and food: (a) physical activity suggestions made by MyBehavior, (b) food suggestions made by MyBehavior, (c) physical activity suggestions for a different user, and (d) physical activity suggestions for the same user as in (a), but at a different point in time.



Measures

First, we used a suggestion-rating survey to evaluate user *intentions* to follow the suggestions. Participants completed this survey after the 3-week study concluded. Participants rated the suggestions, by indicating on a 1-to-5 scale, whether they would be willing and able to do the recommended action on an average day—5 (Strongly Agrees that he/she can follow the suggestion), 1 (Strongly Disagrees). Each participant rated suggestions that she/he saw during the study in an online form. Experimental group participants rated 15 top-ranked—top 8 physical activity and top 7 food—personalized MyBehavior suggestions of their own. On the other hand, the control group participants rated 10 randomly chosen generic prescriptive suggestions. In addition, we quantitatively measured *behavior change* for all participants using logs of daily physical activity and dietary intakes.

The daily diary and the in-depth, semistructured interviews measured participant feedback regarding the suggestions. For

the daily diaries, we queried (1) whether they looked at MyBehavior's suggestions, and (2) whether they made or wanted to make any changes after seeing the suggestions. The semistructured interviews covered users' general overall experience with MyBehavior and the quality of the suggestions. Specifically, we inquired about awareness, behavior change, and of any software improvement they would like to see. In addition, in the interview, we asked clarifying questions that explained quantitative results observed from the data.

Analysis Plan

Regarding the user's intention of following MyBehavior's suggestions, we gathered ratings for suggestions on a secure website and analyzed the data using RStudio. Since the ratings were in ordinal scale, we used a nonparametric Mann-Whitney U test [32] for statistical significance and effect size.

We measured behavior changes by analyzing activity and dietary logs for statistical significance using MATLAB (MathWorks, Inc) statistics toolbox and RStudio. For each user, we computed median walking length and calories per food item. We considered medians across entire weeks over other central measures since they are less susceptible to spurious noise or outliers (eg, occasional intake of very-high-calorie food or atypical, unusually lengthy walk). We did not report changes in running and manually logged exercises in the data analysis as they often require higher effort and are tough to change within the 3 weeks of the experiment. In our analysis, we first considered the number of positive changes. A *positive change* is defined as a downward trend in median calories in meals, or an upward trend toward longer-length walks over the first week to the third week. We used the Fisher Exact Test [32] to measure the number of positive changes as an effect of MyBehavior. Because of small sample size, the Fisher Exact Test is used instead of the chi-square test for independence. We used a two-sample independent Student's *t* test to measure statistical significance for total walk lengths and total food calories consumed per day. We computed differences in walking distances instead of total number of calories burned, since a walk of a fixed distance can result in a different amount of calories burned for different individuals [25]. We calculated the effect size of walking and eating behavior changes with Cohen's d measure.

Finally, face-to-face, semistructured interviews were audio recorded and transcribed. Interview transcripts and daily diaries were then broken down into themes using thematic analysis [33].

Results

Adherence

A total of 17 participants completed the 3-week study, yielding almost 2.1 million recorded physical activity instances, amounting to more than 8000 hours of physical activity. During the same period, participants labeled nearly 850 images of food with annotations.

User Acceptance of MyBehavior Suggestions

In the suggestion-rating survey, the experimental group (mean 3.4, SD 1.2, q_{25}=2.75, q_{50}=3, q_{75}=4), with MyBehavior suggestions, intended to follow personalized suggestions more than the control group (mean 2.5, SD 1.6, q_{25}=1, q_{50}=2, q_{75}=4) intended to follow the generic suggestions. A nonparametric Mann-Whitney U test [32] found this difference to be statistically significant (P<.001, 95% CI 0-1.001, effect size = 0.99).

Physical Activity

Figure 5 shows the distribution, in the form of box plots, of walking lengths over time for the experimental (left-hand image) and the control (right-hand image) groups. For each week of the study, we computed these distributions for the different users. To ease interpretation, we joined the median per week with thick green or red lines for each user. A green line implies a positive change as discussed in the data analysis section. A red line indicates the reverse negative trend. We used a log scale for walking-length distribution since walking-length distributions have heavy tails [34].

For walking, 78% (7/9) of participants in the experimental group (Figure 5, left-hand image) showed positive trends, whereas 75% (6/8) of participants in the control group (Figure 5, right-hand image) exhibited negative trends. A Fisher Exact Test found this ratio in the number of positive changes between the experimental and control groups statistically significant (P=.05) [35]. In addition, MyBehavior users walked an average of 10 minutes more per day within the experiment phase (ie, from the first to the third week). However, we did not observe any change for the control group. A two-sample *t* test found this difference in change of walking duration to be significant (t_{15}= 2.1, P=.055, 95% CI -0.23 to 19.052, d=0.9).

Qualitative data from daily diary and face-to-face interviews largely supported this quantitative result. However, we also observed some important subtleties. First, participants in the experimental group described the activity suggestions to be actionable and relevant to their lives. Control group participants appreciated that the generic suggestions reminded them of good habits. However, they often faced problems incorporating the suggestions into their daily lives. The following quotes were taken from the daily diaries of participants.

> *Those suggestions are quite good, which reminds me not to sit too long in one place.* [Experimental group participant #1]
>
> *The exercise suggestions made me want to do some more activities and be less stationary. Seeing how long I have been stationary and the low frequency of activity made me want to make a change.* [Experimental group participant #5]
>
> *Try to get up from my desk more often...added "walk" notes to my calendar.* [Experimental group participant #2]
>
> *I did some walking where I normally walk. The app now shows I walked there 26 times. The app makes me feel that I can do it again since I have done the*

> *same walk many times.* [Experimental group participant #7]
>
> *The suggestions encourage me to do/plan exercises for the near future...It reminds me that some foods are better than others.* [Control group participant #1]
>
> *They seem like good generic suggestions. The kind you would read...as tips in a health magazine or some such...* [Control group participant #4]

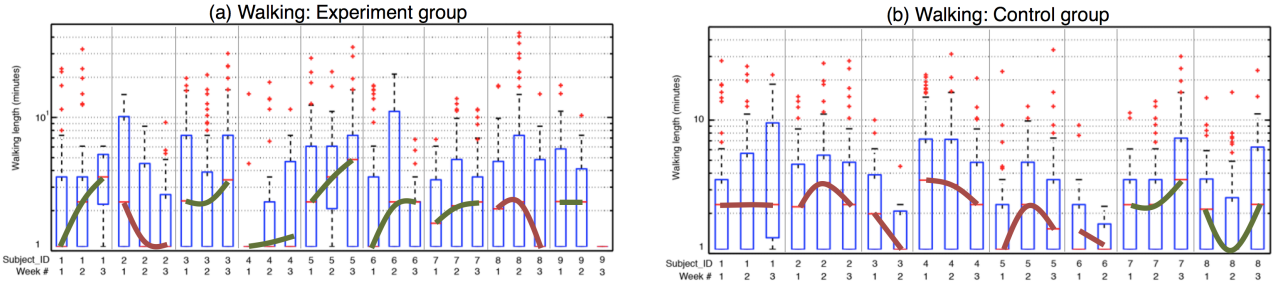
Some MyBehavior users reported that even the nonfrequent explore suggestions were actionable and expressed interest in acting on them. For instance, experimental group participant #7 said the following in his/her daily diary:

> *I saw a walk to my nearest bus stand listed. Normally, I drive my car to go to my office. But looking at the extra walking I got while going to the bus stop makes me think about doing it often and making it a habit.* [Experimental group participant #7]

Results from interviews also revealed that participants at various stages of active lifestyle reacted to suggestions differently [15]. For the experimental group, participants who were considering making changes expressed that they became more self-conscious about their behavior and they were eager to follow the suggested changes (eg, starting to walk more near home, or continuing runs on treadmills). Comparatively, users likely maintaining an active lifestyle expressed that the suggestions reflected their current healthy behavior and considered them as good reinforcements. However, participants in the maintenance phase wanted to change their stationary behavior in the office with occasional small walks. For the control group, users were frustrated because the suggestions were not always feasible and did not blend with their routines and lifestyle. Control group users maintaining an active lifestyle were unaffected by generic suggestions and continued their regular behavior across weeks. For example, control group participants #7 and #8 were maintaining participants and their behavior showed no negative trends in Figure 5 (right-hand image). Control group users who did not already have a maintaining lifestyle gradually became less active or made poorer food choices after the initial phase of the study.

Finally, on a few occasions, MyBehavior suggestions were hard to follow or did not reflect user preferences. For example, one user reported in the interview that he used to play soccer with his friends but his friends recently moved to a new location. He could no longer play soccer, which MyBehavior was suggesting. In addition, often user-preferred activities are not top MyBehavior suggestions. For instance, one user preferred to swim even though she did not do it often. Finally, experimental group participant #8 (subject 8 in Figure 5, left-hand image, with negative trends) reported an inability to follow MyBehavior suggestions because of a looming work deadline during the study.

Figure 5. Box plots showing the distribution of walking lengths for the experimental group (a) and for the control group (b) over the 3-week study. We joined the medians of distributions and showed each trend as a thick green line (increasing trend) or red line (decreasing trend) for walking length.



Dietary Behavior

Figure 6 shows the distribution, in the form of box plots, of meal calories for the experimental group (left-hand image) and the control group (right-hand image). For each week of the study, we computed these distributions for different users. Similar to walking-behavior graphs, we joined medians across weeks to show positive or negative changes for each user.

For caloric intake, 78% (7/9) of participants in the experimental group showed positive trends (green lines in Figure 6, left-hand image), and 57% (4/7) of participants in the control group showed negative trends (red lines in Figure 6, right-hand image—1 participant had insufficient data). However, a Fisher Exact Test found this to be nonsignificant (P=.15). For control group participants, we also found their average median calories per day to increase by 211 calories (mean 211.7, SD 263.07, q_{25}=-31.25, q_{50}=187.5, q_{75}=429.35) from the first week to the third week. Comparatively, the experimental group showed an average calorie per day decrease of nearly 100 calories (mean -99.3, SD 481.27, q_{25}=-527.83, q_{50}=-37.3, q_{75}=87.5) from the first week to the third week. This change was not significant in a two-sample t test (t_{12}=1.3234, P=.21, 95% CI -201 to 822.96, d=0.72).

In qualitative feedback, similar to physical activity suggestions, experimental group users found the suggestions to be more actionable and reported to make more changes compared to control group users who found the suggestions to be hard to work on. This feedback is illustrated in the following quotes from participants' daily diaries.

> *The pictures of my meals are very useful to keep track of what I've been eating in the past. People tend to forget about their habits, but pictures in this case are a nice way to bring your food history in front of your eyes.* [Experimental group participant #9]
>
> *The suggestions remind me that some foods are better than others.* [Control group participant #1]

It recommends me to eat stuff that I don't have at home. [Control group participant #4]

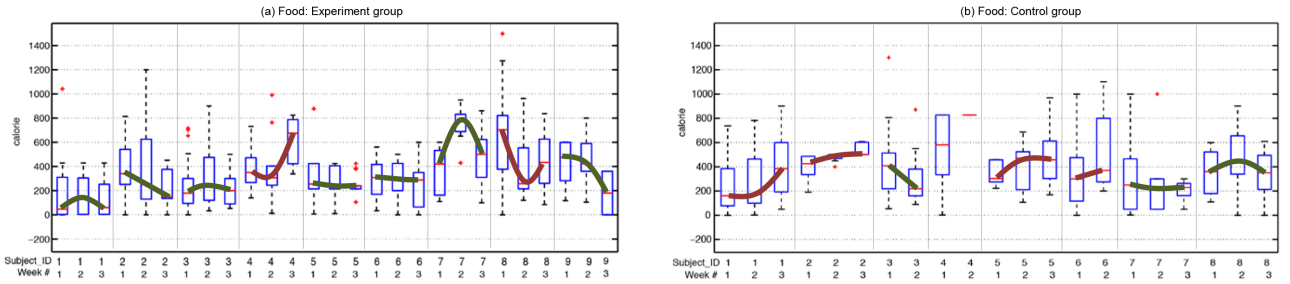
These suggestions don't take into account my dietary restrictions. [Control group participant #5]

Similar to activity explore suggestions, MyBehavior users often found the explore suggestions to be actionable.

I just wanted to see what it was...These ones [explore suggestions] seemed to pick up some "good" food habits. [Experimental group participant #4]

Finally, users reported manual food logging to be time consuming in the interview. However, they also reported that this manual process made them more aware of their foods. Consequently, control group participants reported making dietary changes without personalized suggestions.

Figure 6. Box plots showing the distribution of food calories for the experimental group (a) and for the control group (b) over the 3-week study. We joined the medians of distributions and showed each trend as a thick green line (increasing trend) or red line (decreasing trend) for median food calorie intake.



Discussion

Principal Findings

To our knowledge, MyBehavior is the first system to automatically provide personalized suggestions that relate to users' lifestyles. In the quantitative results, MyBehavior users demonstrated superior behavior changes compared to the control group. Qualitative measures from the face-to-face interviews and the daily diaries confirmed that the suggestions indeed were perceived to be personalized to their lives. This concordance of superiority in both quantitative behavior change and qualitative user perception makes MyBehavior's automated health feedback approach very promising and provides support for longitudinal studies and future investigations into automated personalization approaches.

Specifically, in our evaluation, users rated that they could follow MyBehavior personalized suggestions more than the control condition suggestions. Results also revealed a significant change in walking behaviors for MyBehavior users. In qualitative measures, users reported MyBehavior activity suggestions to be more actionable. Interestingly, although users qualitatively reported the dietary suggestions to be more actionable, dietary behavior changes were not found to be different between the groups. This finding could be due to the manual-logging nature of food intake being sufficient for behavior change alone. The manual process of food logging might produce self-awareness and reflection. Indeed, past research demonstrates that simple logging can improve one's food consumption behavior [16]. However, food logging is an arduous process and it is often hard to continue for an extended period. Thus, we need longer studies to determine if food logging along with suggestions could aid in sustained behavior change. Furthermore, we had a small sample in the study with inadequate statistical power. Thus, larger trials are necessary to further elucidate the effects of food logging and these types of suggestions on eating behavior.

Nonetheless, MyBehavior explores a unique space for health feedback. Earlier studies in this domain predominantly focused on overall behavior [7,14], tailoring [36], or self-tracking [33] without deeper data analysis and personalization. MyBehavior takes a data mining approach to automatically find contextualized suggestions from logged data. This automated approach also relieves users from the burden of self-analyzing their data. Thus, MyBehavior is a marked departure from previous self-monitoring programs found in the literature, where users themselves decide on how to make changes on their own [33]. MyBehavior suggestions relate to a user's existing behaviors, making them actionable as the user is told *where* and *when* to act on them. Furthermore, unique sets of suggestions are generated for each user based on their routine and lifestyle. The literature on N-of-1 approaches [37,38,39] argue that such personalization should yield better efficacy than one-size-fits-all or tailored-suggestion approaches [15], where similar suggestions are provided to users with similar characteristics (eg, age, gender, daily calorie intake, and loss).

Despite this promising direction, the automated data-driven personalization approach of MyBehavior brings its own challenges. Manual logging of food and exercise, in addition to automated logging, are necessary for proper functioning of MyBehavior. Qualitative interviews revealed that manual food and exercise logging were often burdensome. Future iterations could use crowdsourcing-based semiautomated approaches to decrease the burden of manual food journaling [12]. Finally, interviews also highlighted the importance of considering contextual changes in users' lives and preferences. Thus, giving users control in deciding which suggestion they want to follow is required for well-accepted personalization [7].

Limitations

An important limitation is the short-term and small-scale nature of the study, which makes it difficult to make definitive conclusions. However, the study helped us to identify the potential efficacy of MyBehavior and pinpoint design improvements for future deployments. Indeed, Klasanja et al

[17] argued that such short-term studies with similar evaluation goals as in our study are often more suitable for new and untested behavior change technologies like MyBehavior. Another limitation was that the nonpersonalized suggestions were sometimes too specific, for example, "walking with a dog." In the daily diaries, some users reported that they could not follow this suggestion since they did not own a dog. While designing generic suggestions, we tried to find suggestions that most users could follow, without being overly generic. However, there will always be exceptions where a suggestion does not fit one's lifestyle.

Despite these limitations, this pilot study demonstrates the potential of using automated personalization for actionable health feedback. As we move into an age where increasingly more people are tracking their health with mobile and other technologies, we believe MyBehavior's automated technique holds great potential to provide feedback that can be used to improve health outcomes at scale.

Conclusions

MyBehavior is the first mobile health app that can encourage healthy behavior change by automatically providing low-effort suggestions based on the context and personal information of users. The pilot user study demonstrated the feasibility and acceptability of MyBehavior. Users considered MyBehavior's personalized, contextualized suggestions to be more actionable and to require less effort to implement than generic prescriptive suggestions. Preliminary evidence of behavior change shows that a high percentage of MyBehavior users did more physical exercise, yet the potential impact on eating behaviors remains unclear. The addition of more human control over the suggestions and providing easier logging mechanisms for food and exercise were identified as key areas of improvement. Future directions should include addressing the identified shortcomings of the system and testing its effectiveness in fostering health behavior change in a larger longitudinal trial.

Acknowledgments

We acknowledge Emiliano Miluzzo, Erin Carroll, Fabian Okeke, Jean Costa, Steven Voida, and Chantelle Farmer for their early contributions to this research. This work was supported by Intel Science and Technology Center for Pervasive Computing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Generic suggestions used in the control group.

[PDF File (Adobe PDF File), 40KB - mhealth_v3i2e42_app1.pdf]

Multimedia Appendix 2

A video showing how MyBehavior works.

[MP4 File (MP4 Video), 42MB - mhealth_v3i2e42_app2.mp4]

Multimedia Appendix 3

CONSORT-EHEALTH checklist V1.6.1 [40].

[PDF File (Adobe PDF File), 1MB - mhealth_v3i2e42_app3.pdf]

References

1. Alwan A. Global Status Report on Noncommunicable Diseases 2010. Geneva, Switzerland: World Health Organization; 2011. URL: http://www.who.int/nmh/publications/ncd_report_full_en.pdf [accessed 2015-05-02] [WebCite Cache ID 6YEj9OBF0]
2. Centers for Disease Control and Prevention (CDC). Million hearts: strategies to reduce the prevalence of leading cardiovascular disease risk factors--United States, 2011. MMWR Morb Mortal Wkly Rep 2011 Sep 16;60(36):1248-1251 http://www.cdc.gov/mmwr/pdf/wk/mm6036.pdf [FREE Full text] [Medline: 21918495]
3. Kaplan R, Stone A. Bringing the laboratory and clinic to the community: mobile technologies for health promotion and disease prevention. Annu Rev Psychol 2013;64:471-498. [doi: 10.1146/annurev-psych-113011-143736] [Medline: 22994919]
4. Cawley J, Meyerhoefer C. The medical care costs of obesity: an instrumental variables approach. J Health Econ 2012 Jan;31(1):219-230. [doi: 10.1016/j.jhealeco.2011.10.003] [Medline: 22094013]
5. MyFitnessPal. URL: http://www.myfitnesspal.com/ [accessed 2015-05-07] [WebCite Cache ID 6YLirNtca]
6. Moves. URL: https://www.moves-app.com/ [accessed 2015-05-02] [WebCite Cache ID 6YEijxud8]
7. Consolvo S, McDonald DW, Toscos TT, Chen MY, Froehlich J, Harrison B, et al. Activity sensing in the wild: a field trial of ubifit garden. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY:

ACM; 2008 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 5-10, 2008; Florence, Italy p. 1797-1806. [doi: 10.1145/1357054.1357335]
8. James JL, Mamykina L, Lindtner S, Delajoux G, Strub HB. Fish'n'Steps: encouraging physical activity with an interactive computer game. In: Proceedings of the 8th International Conference on Ubiquitous Computing. Berlin, Germany: Springer-Verlag; 2006 Presented at: UbiComp'06: 8th International Conference on Ubiquitous Computing; September 17-21, 2006; Orange County, CA p. 261-278.
9. Consolvo S, Everitt K, Smith I, Landay JA. Design requirements for technologies that encourage physical activity. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2006 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 22-27, 2006; Montréal, PQ p. 457-466.
10. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]
11. Daugherty BL, Schap TE, Ettienne-Gittens R, Zhu FM, Bosch M, Delp EJ, et al. Novel technologies for assessing dietary intake: evaluating the usability of a mobile telephone food record among adults and adolescents. J Med Internet Res 2012;14(2):e58 [FREE Full text] [doi: 10.2196/jmir.1967] [Medline: 22504018]
12. Noronha J, Hysen E, Zhang H, Gajos KJ. Platemate: crowdsourcing nutritional analysis from food photographs. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. New York, NY: ACM; 2011 Presented at: The 24th Annual ACM Symposium on User Interface Software and Technology; October 16-19, 2011; Santa Barbara, CA p. 1-12.
13. Thomaz E, Parnami A, Essa I, Abowd GD. Feasibility of identifying eating moments from first-person images leveraging human computation. In: Proceedings of the 4th International SenseCam and Pervasive Imaging Conference. New York, NY: ACM; 2013 Presented at: The 4th International SenseCam and Pervasive Imaging Conference; November, 18-19, 2013; San Diego, CA p. 26-33.
14. Lane ND, Mashfiqui M, Lin M, Yang X, Lu H, Ali S, et al. Bewell: A smartphone application to monitor, model and promote wellbeing. In: Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare. 2011 Presented at: 5th International ICST Conference on Pervasive Computing Technologies for Healthcare; May 23-26, 2011; Dublin, Ireland p. 23-26. [doi: 10.4108/icst.pervasivehealth.2011.246161]
15. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [Medline: 10170434]
16. Siek KA, Connelly KH, Rogers Y, Rohwer P, Lambert D, Welch JL. When do we eat? An evaluation of food items input into an electronic food monitoring application. In: Proceedings of the Pervasive Health Conference and Workshops. New York, NY: IEEE; 2006 Presented at: Pervasive Health Conference and Workshops; November 29-December 1, 2006; Innsbruck, Austria p. 1-10. [doi: 10.1109/pcthealth.2006.361684]
17. Predrag K, Consolvo S, Pratt W. How to evaluate technologies for health behavior change in HCI research. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 2011 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC p. 3063-3072. [doi: 10.1145/1978942.1979396]
18. National Heart, Lung, and Blood Institute, National Institutes of Health. 2011 Sep 26. Getting started and staying active URL: http://www.nhlbi.nih.gov/health/health-topics/topics/phys/getstarted.html [accessed 2015-05-07] [WebCite Cache ID 6YLlZqFNZ]
19. National Heart, Lung, and Blood Institute, National Institutes of Health. 2013 Feb 13. Healthy eating plans URL: http://www.nhlbi.nih.gov/health/educational/wecan/eat-right/sample-plans.htm [accessed 2015-05-02] [WebCite Cache ID 6YEhYdWTL]
20. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
21. Dallery JJ, Cassidy RN, Raiff BR. Single-case experimental designs to evaluate novel technology-based health interventions. J Med Internet Res 2013;15(2):e22 [FREE Full text] [doi: 10.2196/jmir.2227] [Medline: 23399668]
22. Bishop CM. Pattern Recognition and Machine Learning. New York, NY: Springer-Verlag; 2006.
23. Lu H, Yang J, Liu Z, Lane ND, Choudhury T, Campbell AT. The Jigsaw continuous sensing engine for mobile phone applications. In: Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems. New York, NY: ACM; 2010 Presented at: 8th ACM Conference on Embedded Networked Sensor Systems; November 3-5, 2010; Zurich, Switzerland p. 71-84. [doi: 10.1145/1869983.1869992]
24. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. Med Sci Sports Exerc 2011 Aug;43(8):1575-1581. [doi: 10.1249/MSS.0b013e31821ece12] [Medline: 21681120]
25. Wikipedia. 2015 Feb 16. Metabolic equivalent URL: http://en.wikipedia.org/wiki/Metabolic_equivalent [accessed 2015-05-07] [WebCite Cache ID 6YLmIETNM]

26. USDA National Nutrient Database for Standard Reference, Release 21. Beltsville, MD: US Department of Agriculture, Agricultural Research Service; 2008. URL: http://www.ars.usda.gov/News/docs.htm?docid=18880 [accessed 2015-05-07] [WebCite Cache ID 6YLnRQFBb]
27. Eiter T, Mannila H. Technischen Universität Wien, Knowledge Based Systems Group. Vienna, Austria: Technischen Universität Wien; 1994 Apr 25. Computing discrete Frechet distance URL: http://www.kr.tuwien.ac.at/staff/eiter/et-archive/cdtr9464.pdf [accessed 2015-05-07] [WebCite Cache ID 6YLp2RwPg]
28. Madden GJ, editor. APA Handbook of Behavior Analysis. Washington, DC: American Psychological Association; Sep 16, 2013.
29. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
30. Fogg BJ. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology. New York, NY: ACM; 2009 Presented at: 4th International Conference on Persuasive Technology; April 26-29, 2009; Claremont, CA URL: http://bjfogg [doi: 10.1145/1541948.1541999]
31. Powell WB. Approximate Dynamic Programming: Solving the Curses of Dimensionality. 2nd edition. Hoboken, NJ: Wiley-Interscience; Sep 2011.
32. Ott RL, Longnecker MT. An Introduction to Statistical Methods and Data Analysis. 6th edition. Belmont, CA: Brooks/Cole Cengage Learning; 2010.
33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. González MC, Hidalgo CA, Barabási AL. Understanding individual human mobility patterns. Nature 2008 Jun 5;453(7196):779-782. [doi: 10.1038/nature06958] [Medline: 18528393]
35. Fisher RA. On the interpretation of X2 from contingency tables, and the calculation of P. J R Stat Soc 1922 Jan;85(1):87-94. [doi: 10.2307/2340521]
36. Kukafka R. Tailored health communication. In: Lewis D, Eysenbach G, Kukafka R, Stavri PZ, Jimison HB, editors. Consumer Health Informatics. New York, NY: Springer; 2005:22-33.
37. Wiggins JS. Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior, Minnesota Press, Minneapolis. In: Cicchetti D, Grove WM, editors. Thinking Clearly About Psychology: Essays in Honor of Paul E Meehl. Minneapolis, MN: University of Minnesota Press; 1991:89-113.
38. Tashakkori A, Teddlie C. Mixed Methodology: Combining Qualitative and Quantitative Approaches. Thousand Oaks, CA: SAGE Publications, Inc; 1998.
39. Estrin D. Small data, where n=me. Commun ACM 2014 Apr 01;57(4):32-34. [doi: 10.1145/2580944]
40. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

CDC: Centers for Disease Control and Prevention
FBM: Fogg Behavior Model
GMM: Gaussian Mixture Model
GPS: Global Positioning System
MAB: multi-armed bandit
METS: Metabolic Equivalents of Task
q25: lower quartile
q50: median
q75: upper quartile
RCT: randomized control trial
USDA: United States Department of Agriculture
WHO: World Health Organization

Edited by G Eysenbach; submitted 19.12.14; peer-reviewed by J Galante; comments to author 16.01.15; revised version received 19.02.15; accepted 17.04.15; published 14.05.15.

Please cite as:
Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T
Automated Personalized Feedback for Physical Activity and Dietary Behavior Change With Mobile Phones: A Randomized Controlled Trial on Adults
JMIR mHealth uHealth 2015;3(2):e42
URL: http://mhealth.jmir.org/2015/2/e42/
doi:10.2196/mhealth.4160
PMID:25977197

Original Paper

What Overweight Women Want From a Weight Loss App: A Qualitative Study on Arabic Women

Aroub Abdulaziz Alnasser[1], MSc; Abdulrahman Saleh Alkhalifa[1]; Arjuna Sathiaseelan[2], PhD; Debbi Marais[3], PhD

[1]King Saud University, Food Science and Nutrition, Riyadh, Saudi Arabia

[2]University of Cambridge, Cambridge, United Kingdom

[3]University of Aberdeen, Aberdeen, United Kingdom

Corresponding Author:
Aroub Abdulaziz Alnasser, MSc
King Saud University
Food Science and Nutrition
PO Box 86683
Riyadh, 11632
Saudi Arabia
Phone: 966 114556208
Fax: 966 114026995
Email: aroub@ksu.edu.sa

Abstract

Background: Overweight and obesity are international public health issues. With mobile and app use growing globally, the development of weight loss apps are increasing along with evidence that interventions using technology have been effective in the treatment of obesity. Although studies have been conducted regarding what content health professionals would recommend within weight loss apps, there are limited studies that explore users' viewpoints. There is specifically a paucity of research that takes the cultural background of the user into consideration, especially in Middle Eastern countries where the lives and weight loss intervention needs of women not only vary vastly from the West, but the obesity rate is also increasing exponentially.

Objective: The current study sought to explore the proposed features of an Arabic weight loss app by seeking the experiences and opinions of overweight and obese Saudi Arabian users in order to design a mobile phone app to fit their needs.

Methods: Focus group discussions were conducted with a purposive sample of volunteer overweight and obese Saudi women (BMI ≥ 25) who were older than 18 years and who owned a mobile phone. The most common Arabic and English weight loss mobile apps were downloaded to initiate dialogue about app usage and to get their opinions on what an ideal weight loss app would look like and the features it would include. All transcribed, translated discussions were thematically analyzed, categorized for each of the main topics of the discussion, and specific quotations were identified.

Results: Four focus groups were conducted with a total of 39 participants. Most participants owned an Android mobile phone and only a few participants were aware of the availability of health-related apps. Barriers to weight loss were identified including: motivation, support (social and professional), boring diets, customs, and lifestyle. Diverse themes emerged as suggestions for an ideal weight loss app including: Arabic language and culturally sensitive; motivational support and social networking; dietary and physical activity tools; and a tailorable, user-friendly interface.

Conclusions: This study identifies weight loss app features from the users' perspective, which should be considered in the development of a weight loss app for this population.

(JMIR mHealth uHealth 2015;3(2):e41) doi:10.2196/mhealth.4409

KEYWORDS

weight loss; focus groups; smartphone; mobile apps; Arabic; qualitative research

Introduction

Numerous influential studies have shown that overweight and obesity has increased markedly, making it a major international health issue, especially among women. According to the World Health Organization [1], the most recent evidence suggests that women are twice as likely as men to be obese in Europe, the Eastern Mediterranean, and the Americas. This growing trend

indicates the need for effective weight control interventions that are accessible to people all over the world regardless of language or location. As the prevalence of obesity around the world has increased, so has the use of mobile technology, particularly in the Middle East and Saudi Arabia [2,3]. This indicates that this technology may be useful as a weight loss intervention. A recent systematic review [4] found that technology interventions, regardless of device type, are an efficient approach for behavioral treatment of obesity in order to achieve or maintain a healthy weight. The recent growth in the development of mobile phone apps has created an opportunity to use them as an intervention tool to both treat and prevent obesity. However, there has been very little research done to discover what users want and need in a weight loss app.

Mobile technology for self-recording and other behavioral modification aspects has been found to be a useful approach to enhancing health [5]. Various apps have been developed for this purpose, but there is evidence that English as well as Arabic apps do not comply with evidence-informed practices for weight management [6,7,8]. Although it is important to ensure that development of an app for weight loss intervention takes these evidence-based weight loss practices into consideration, it is also important to include participants' opinions and preferences, especially when designing an app for a specific culture. Some research has been conducted to determine the opinion of health professionals relating to weight loss apps; however, there is a paucity of evidence regarding user preferences, which has shown to be crucial in app development for weight loss [9]. This is exemplified by a qualitative study conducted solely with physicians and dieticians in Qatar to develop an obesity management app, which lacked any user contribution [2]. Furthermore, a recent study done on English-language weight loss apps in the UK shows that participants are interested in weight loss apps that provide structure, ease of use, personalized features, and accessibility between devices. However, this study neglected to elect participants who had preexisting overweight and obesity issues, and because the needs of users with existing conditions differ from those who are not suffering from these problems, this cannot be overlooked [10].

This study, therefore, aimed to explore the proposed features for an Arabic weight loss app by seeking the experiences and opinions of overweight and obese female Saudi Arabian users in order to design a mobile phone app to fit their needs.

Methods

Participants

A qualitative cross-sectional study was conducted. Focus group discussions were conducted with a purposive sample of volunteer overweight and obese Saudi women (body mass index (BMI) ≥25) who were older than 18 years, owned a mobile phone, and consented to taking part in the study. Women who were pregnant or were diagnosed with chronic diseases of lifestyle—such as cardiovascular disease, hypertension, diabetes mellitus, or cancer—were excluded from the study. Recruitment was done through posters placed around the King Saud University campuses (a public university located in Riyadh, Saudi Arabia, with free tuition that enables students from all segments of society to attend), local shopping malls, on social networks such as Twitter, as well as by word of mouth. The recruitment posters briefly described the study, eligibility criteria, and directed interested participants to complete an online screening questionnaire to ensure compliance with the inclusion and exclusion criteria. Ethical approval was obtained from the ethics committee of the College of Science Research Center of King Saud University. Written informed consent was obtained from all subjects.

Procedure

A total of 4 focus groups with 6-10 participants in each were planned, with the aim of reaching the point of data saturation until no new information was generated. Those complying with the eligibility criteria were invited to join 1 of 4 possible dates available for focus groups that suited them best. Prior to meeting, an information sheet and consent forms were sent to the participants. Each participant was randomly allocated to one of the most common Arabic/English weight loss apps and asked to download the app before the focus groups.

Focus groups were held in Arabic, within meeting rooms at the women's King Saud University campus. The focus groups started with a conversation about the participants' experiences of health-related mobile phone apps in general, following a topic guide that presented open-ended questions to encourage participants to express their opinions. Group management guidelines from Krueger and Casey [11] were followed. The facilitator provided time for the participants to try some of the weight loss apps that they had been asked to download to prompt further discussion. Participants were then asked to describe their thoughts and ideas about the features of these apps. They were asked what they liked, disliked, and would ideally like a weight loss app to have. Their experience of previous weight loss diets was also discussed to identify their opinions of barriers to weight loss.

Data Analysis

All focus groups were audio-recorded and an observer was available to take notes during the discussions. Refreshments were provided during the focus groups and participants were given a small gift voucher as a thank-you at the end of the focus group. All focus group discussions were transcribed verbatim in Arabic and translated to English. A subsample of the transcripts were then back translated to ensure accuracy. To initiate data analysis, the researchers familiarized themselves with the data by reading all translated transcriptions and identifying themes within each transcript. Thematic analysis was done independently by 2 researchers (AA and DM) for 1 transcript and then jointly for the other 3 transcripts. Once thematic analysis had been completed, specific quotations to support the themes were identified.

Results

Participant Characteristics

Four focus groups involving 39 women were conducted. Each group consisted of 7 to 12 participants with a mean age of 29 years (Table 1). Based on self-reported height and weight, the mean BMI was 29.1 with most (67%) of the sample classified

as overweight. Marital status was equally represented, and most women had been educated to a tertiary level (77%) and owned an Android mobile phone (56%).

Table 1. Demographic characteristics of Saudi women (N=39) involved in the 4 focus groups.

Demographic characteristic		Measurement
Age (years), mean (SD)		29.4 (8)
Height (m), mean (SD)		159.3 (6)
Weight (kg), mean (SD)		73.9 (14)
BMI (kg/m^2), mean (SD)		29.1 (5)
Weight classification, n (%)		
	Overweight (BMI 25-29.9)	26 (67%)
	Obese (BMI 30-34.9)	9 (23%)
	Severely obese (BMI >35)	4 (10%)
Marital status, n (%)		
	Single	19 (49%)
	Married	19 (49%)
	Divorced	1 (2%)
Education level, n (%)		
	High school	7 (18%)
	Diploma	2 (5%)
	4-year degree	26 (67%)
	Post-graduate	4 (10%)
Mobile phone brand, n (%)		
	iPhone	17 (44%)
	Samsung	14 (36%)
	BlackBerry	7 (18%)
	Sony	1 (2%)

Salient Themes

Overview

Themes were categorized for each of the main topics of discussion, namely the experience of apps, barriers to weight loss, and proposals for an ideal weight loss app.

Experience of Apps

Few participants were aware of the availability of health-related apps. Only a few participants had previously downloaded a diet app and there was a general lack of awareness of their availability.

> *I didn't expect there to be an app concerned about dieting.* [Participant 8; Focus Group (FG) 4]

Some women in the groups felt that raising awareness of the existence of weight-loss apps in Arabic by providing information or seminars at local clinics and schools would be essential to increasing the knowledge and use of the apps.

When women were asked if they felt they had previously benefited from any health app, similar responses of not knowing about them emerged. If they had downloaded them, it was usually for short periods of use.

> *I only had the application for 7 days.* [Participant 1; FG 3]

The most common reason stated for deleting apps was that it did not benefit them. Those that had downloaded weight loss apps indicated that most of them featured meal planners, calorie information, weight loss tips, a water reminder, and most were in Arabic. When women were asked about what health apps they had previously downloaded, they were unable to recall any names.

Barriers to Weight Loss

Although a number of women had tried different ways to lose weight, they identified a number of in-common barriers to weight loss including motivation, support (social and professional), boring diets, customs, lifestyle, and misinformation. They also reported poor motivation in initiating the diet, often delaying commencement.

> *Delaying. I decide to follow a diet on Saturday, and I buy all the diet things that I need on Friday, but I delay it for another 2 days. Then I tell myself that I will start it next Saturday. I have been in that situation for 7 months now.* [Participant 6; FG 2]

One of the most common issues for women seemed to be a lack of motivation due to inadequate social support.

> *If I see that no one is with me, I lose interest.* [Participant 2; FG 2]

They also identified a need for professional health care follow-up as those that had received a dietary prescription often described it as "too boring, restrictive, and frustrating" [Participant 1; FG 1].

Another commonly identified barrier was related to lifestyle, customs, and family obligations such as parties.

> *We don't have many entertainment places, so we meet in food places, restaurants, or friends' houses and everything calls you to eat, as well as girls today are good cooks, they make delicious sweets and you have to eat to keep up with them.* [Participant 2; FG 3]

From the discussions, it was clear that the majority of participants also had incorrect information or beliefs about weight loss.

> *You can put 3 bags of salt in your bathtub 3 times a week and that helps lose weight.* [Participant 1; FG 3]

Proposals for an Ideal Weight Loss App

Diverse themes emerged from the data as suggestions for an ideal weight loss app for these women. They were classified into 5 categories (Figure 1). The women overwhelmingly indicated that they would prefer an Arabic app.

> *I have difficulties that relate to the language (when trying the English diet apps).* [Participant 3; FG 2]

Indeed, it was not only about the ease of language, but also about the language in its social and cultural context. Most women that had previously tried commercial international diet programs felt it would be inconvenient to use them.

> *I have already subscribed to an American program that … the food was strange for me … we do not eat that because it is not in our culture.* [Participant 10; FG 2]
>
> *I tried to walk 7 days a week but the weather is bad in Riyadh; it is hot, along with shortness of breath and face cover.* [Participant 10; FG 2]

The most common theme seemed to be the emphasis on motivational support and social networking. The majority of all groups stated that "there has to be communication" [Participant 7; FG 2].

Numerous women indicated that having a social support network would be motivating. For example, one participant said, "The app itself should be like Instagram," in which women checked up on each other as a group motivator [Participant 1; FG 2].

Having illustrations was also suggested by some participants as a motivator. For example, a participant stated: "An image of a fat woman could appear when the calories consumed exceed the calories expended" [Participant 3; FG 4].

Some of the group members stated that having the ability to ask professionals questions and receive advice regarding their weight loss, in a time sensitive manner, would be greatly beneficial to achieving their weight loss goals.

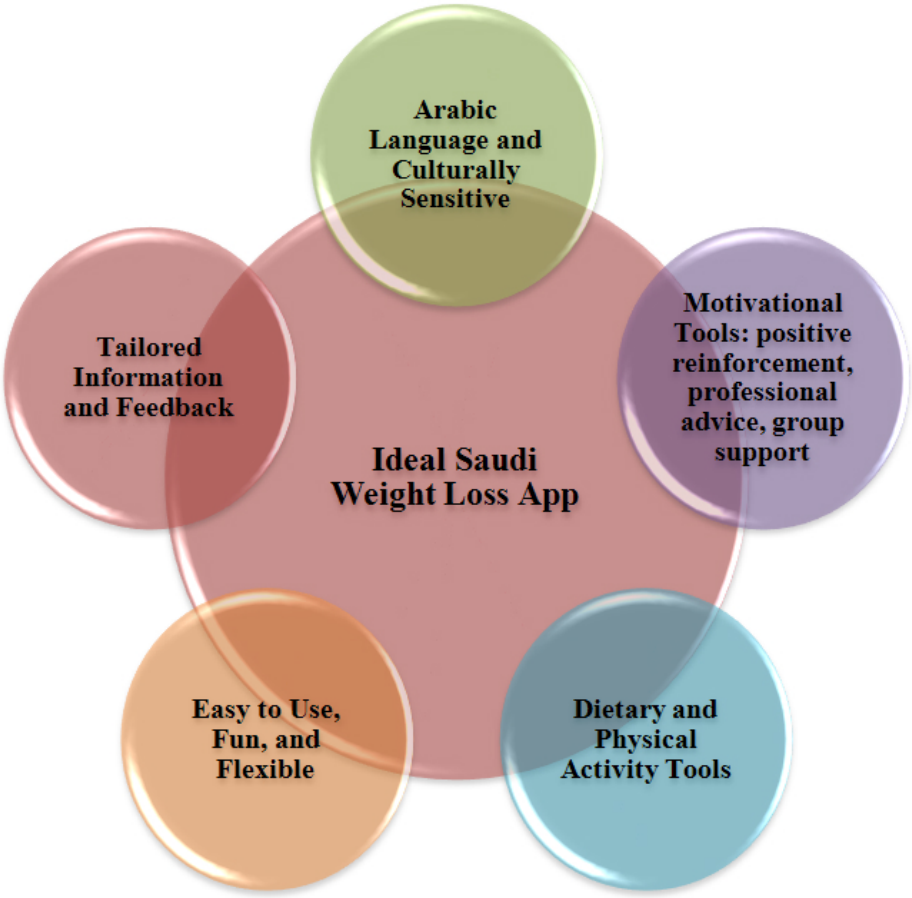
With regard to dietary and physical activity tools, group members mostly agreed that features such as a BMI calculator, calorie and physical activity counter, weight loss tracker, energy balance calculator, setting goals, vitamin information, and food product information would be useful. Opinions were mixed about planning meals. Some women felt that a meal planner is important and easier than counting calorie intake, but others did not want to follow specific meal plans. However, 2 groups reported that a diet score or competitive games would be encouraging and could be compared if shared with the community.

In terms of layout, participants generally agreed that "it should be simple" [Participant 1; FG 1] and easy to use. They suggested flexible search functions such as "autocomplete" when typing in a search term with results that show the most popular trends. In addition, a barcode reader to lower the burden of adding food items, and an autonomous sensor to monitor physical activity was proposed.

There were various suggestions about controlling the frequency and timing of notifications and tips, such as: "The app (that had been downloaded and used previously) was boring because it reminded me to drink water a lot" [Participant 8; FG 4]. Furthermore, most participants desired tailored notifications taking the users' previous information and preferences into consideration. Some of the women suggested screening for diseases, age, and dietary preferences (such as being vegetarian) before starting to use the app so that the information and tips provided could be user-specific.

Some expectations such as feedback requirements from the app developer and regular updating of the information were also discussed. Overall, it is clear that the demand for the development of an Arabic weight management app is high, with strong agreement in all 4 groups.

Figure 1. Proposed features for an ideal weight loss app.



Discussion

Principal Findings

Participants in the study were generally not aware of available health-related apps and had little experience in using them. This is not that surprising since 84% of mobile phone users report having downloaded apps, but only 19% have downloaded a health-related app [12]. Although the use of mobile phone apps in supporting health is promising [13], lack of knowledge of health-related apps might be a reason for the low download rate. Another reason for the low download rate might be the lack of features that Arabic health apps have when compared to English apps. It was found that English-language apps offered on average at least half of the investigated features such as tutorials on use, sharing capability of diet plans, and weight tracking. The Arabic apps only offered 4 of the 12 features mentioned, such as the collection of and information about diet plans and water consumption tracking, substantiating the absence of evidence-based practices. If the available features on the Arabic apps were more attractive to users, then the download rate would likely increase [14].

The Saudi women clearly expressed their preference for the app being in Arabic rather than in English, which is corroborated by the pilot study conducted among health care professionals in Qatar to develop an obesity management mobile phone app [15]. It is also evident that the users of these apps require culturally sensitive information such as locally available foods, physical activities that are possible in their environment, and advice that is specific to the Arabic culture and customs. Parker [16] has indicated that healthy behaviors such as physical activity are boosted or inhibited by custom and culture, so while it may be normal to see Arab men running or exercising at fitness clubs or outdoors, currently this is unusual among Arab women due to cultural restrictions on their activities [17]. A study on morbidly obese Saudi women reported that participants identified factors such as limited social relationships and not being able to practice exercise freely as being the main motivations for choosing surgery over a natural means of weight loss [18].

Most participants attributed negative weight loss experiences to a lack of motivation and social support. As suggested by participants, providing social media features such as Twitter and Instagram would enable participants to share interests, activities, and experiences and thereby provide motivation. Some studies have shown that engagement with Twitter results in greater weight loss [19], while others have found that participants disliked features that permit broadcasting of health-related goals or status updates to friends through social networks [9]. It is suggested that access to social networks be available, but that users have control over what is shared, such as weight loss goals reached or diet scores.

This control over social networking within a weight loss app links to participant suggestions for flexibility and tailoring of the app, where participants felt that they should be able to indicate their preferences for frequency and timing of reminders and notifications they receive, or what features they use. It was clear that participants wanted the app to be easy, user-friendly, and specific to their preferences. The features proposed, such as counters/trackers and goal setting, are consistent with the self-monitoring practices identified in previous studies and would achieve many of the evidence-informed weight loss practices [6-8]. Additional suggestions by participants for food product and vitamin information would address the advice-giving practices.

Many current English and Arabic apps can be described as advice-giving or self-monitoring but lack the behavior modification aspect of weight loss treatment [6-8]. It was interesting that participants proposed features that complied with many of the evidence-informed weight loss practices used by these authors as well as their emphasis on the importance of motivation and social support to evaluate weight loss apps. As many participants had misconceptions regarding weight control, providing evidence-based information or tips is suggested.

Conclusions

Saudi women were very enthusiastic about the development of a weight loss app that would address their needs. It is clear that cultural sensitivity and social support are seen as the most important aspects to them. This is extremely important as it strengthens the evidence that weight loss interventions need to include some aspect of behavior modification to be effective.

This paper provides unique insights into the views of overweight Saudi women regarding features that might support them in losing weight, giving a voice to the user and informing the development of a weight loss app for Saudi Arabian women. It is suggested, though, that these findings may be relevant to a broader community and should be considered as important aspects for consideration in any weight loss app development process.

Acknowledgments

This research project was supported by a grant from the Research Center of the Female Scientific and Medical Colleges, Deanship of Scientific Research, King Saud University.

Conflicts of Interest

As the focus groups were conducted by the researcher, who has a nutritional background, this may have influenced the interpretations of discussion points despite endeavoring to avoid any bias. The translation of focus group discussions from Arabic to English may also have influenced some subtle differences in meaning.

References

1. World Health Organization. 2014. Global Health Observatory URL: http://www.who.int/gho/ncd/risk_factors/obesity_text/en/ [accessed 2015-12-11] [WebCite Cache ID 6WfvvcucF]
2. Mansar S. A founding framework for addressing obesity in Qatar using mobile technologies. In: Cruz-Cunha MM, editor. ENTERprise Information Systems. Berlin, Germany: Springer Berlin Heidelberg; 2011:402-412.
3. Carter M. , Burley, V. , & Cade, J. Development of 'My Meal Mate'–A smartphone intervention for weight loss. Nutrition Bulletin; vol. 38, no. 1, pp 2013:80. [doi: 10.1111/nbu.12016]
4. Coons MJ, Demott A, Buscemi J, Duncan JM, Pellegrini CA, Steglitz J, et al. Technology Interventions to Curb Obesity: A Systematic Review of the Current Literature. Curr Cardiovasc Risk Rep 2012 Apr;6(2):120-134 [FREE Full text] [doi: 10.1007/s12170-012-0222-8] [Medline: 23082235]
5. Hijazi R, Ragon B. Exploring the Use of Smart Phones to Influence Healthy Behaviors. Walden University;: Doctoral dissertation; 2011.
6. Breton ER, Fuemmeler BF, Abroms LC. Weight loss-there is an app for that! But does it adhere to evidence-informed practices? Transl Behav Med 2011 Dec;1(4):523-529 [FREE Full text] [doi: 10.1007/s13142-011-0076-5] [Medline: 24073074]
7. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
8. Alnasser A, Sathiaseelan A, Al-Khalifa A, Marais D. Do Arabic weight loss mobile applications comply with evidence-based weight management guidelines? 2013 Presented at: 20th International Congress of Nutrition; 2013; Granada, Spin p. 1-1960.
9. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
10. Tang J, Abraham C, Stamp E, Greaves C. How can weight-loss app designers' best engage and support users? A qualitative investigation. Br J Health Psychol 2015 Feb;20(1):151-171. [doi: 10.1111/bjhp.12114] [Medline: 25130682]
11. Krueger R. Focus Groups: A Practical Guide for Applied Research. Focus groups: Sage Publications (CA); 2008.
12. Fox S, Duggan M. Mobile Health 2012. 2012 Nov 08. URL: http://www.pewinternet.org/files/old-media//Files/Reports/2012/PIP_MobileHealth2012_FINAL.pdf [accessed 2015-05-08] [WebCite Cache ID 6YNp4uS2O]

13. Klasnja P, Pratt W. Managing health with mobile technology. interactions 2014 Jan 01;21(1):66-69. [doi: 10.1145/2540992]
14. Kurdi HA, Alkhowaiter A, Al-Muaibed A, Alotaibi B, Alhaweal R, Alotaibi T. myPDA: A Mobile Healthcare Application for Personal Diet Assisting. 2012 Presented at: Advanced Computer Science Applications and Technologies (ACSAT), 2012 International Conference; 26-28 Nov. 2012; Kuala Lumpur p. 491-496. [doi: 10.1109/ACSAT.2012.17]
15. Limam MS, Jariwala S, Behih N, Shahzad M, Anggraini A. Adapting a database of text messages to a mobile-based weight loss program: the case of the middle East. Int J Telemed Appl 2014;2014:658149 [FREE Full text] [doi: 10.1155/2014/658149] [Medline: 24511311]
16. Parker AG. Designing for health activism. interactions 2013 Mar 01;20(2):22. [doi: 10.1145/2427076.2427082]
17. Ali H, Baynouna LM, Bernsen RM. Barriers and facilitators of weight management: perspectives of Arab women at risk for type 2 diabetes. Health Soc Care Community 2010 Mar;18(2):219-228. [doi: 10.1111/j.1365-2524.2009.00896.x] [Medline: 20059569]
18. Alqout O, Reynolds F. Experiences of obesity among Saudi Arabian women contemplating bariatric surgery: an interpretative phenomenological analysis. J Health Psychol 2014 May;19(5):664-677. [doi: 10.1177/1359105313476977] [Medline: 23479306]
19. Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: use of an online social network in a remotely delivered weight loss intervention. Transl Behav Med 2013 Sep;3(3):287-294 [FREE Full text] [doi: 10.1007/s13142-012-0183-y] [Medline: 24073180]

Abbreviations

BMI: body mass index
FG: focus group

Edited by G Eysenbach; submitted 06.03.15; peer-reviewed by J Sheats; comments to author 09.04.15; revised version received 22.04.15; accepted 23.04.15; published 20.05.15.

Please cite as:
Alnasser AA, Alkhalifa AS, Sathiaseelan A, Marais D
What Overweight Women Want From a Weight Loss App: A Qualitative Study on Arabic Women
JMIR mHealth uHealth 2015;3(2):e41
URL: http://mhealth.jmir.org/2015/2/e41/
doi:10.2196/mhealth.4409
PMID:25993907

Review

Public Health Guidelines for Physical Activity: Is There an App for That? A Review of Android and Apple App Stores

Emily Knight[1,2*], PhD; Melanie I Stuckey[3*], PhD; Harry Prapavessis[1*], PhD; Robert J Petrella[1,2,4,5*], MD, PhD

[1]University of Western Ontario, Faculty of Health Sciences, London, ON, Canada

[2]Aging, Rehabilitation and Geriatric Care Research Centre, Lawson Health Research Institute, London, ON, Canada

[3]Computer Science, University of Western Ontario, London, ON, Canada

[4]Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, Canada

[5]University of Western Ontario, Centre for Studies in Family Medicine, London, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Robert J Petrella, MD, PhD
University of Western Ontario
Centre for Studies in Family Medicine
1151 Richmond Street
Rm 2117
London, ON, N6A 3K7
Canada
Phone: 1 519 933 8455
Fax: 1 519 685 4071
Email: petrella@uwo.ca

Abstract

Background: Physical activity participation is an important behavior for modifying lifestyle-related disease risk. Mobile health apps for chronic disease management and prevention are being developed at a rapid rate. However, it is unclear whether these apps are evidence-based. Current public health recommendations for physical activity participation for adults highlight the importance of engaging in 150 minutes weekly of purposeful exercise, and muscle strengthening activities on at least 2 days of the week.

Objective: The aims of the present review were to (1) identify available evidence-based physical activity apps, and (2) identify technological features that could be leveraged to improve health outcomes.

Methods: iTunes and Google Play mobile app stores were searched using keyword and category searching during a single day (February 18, 2014) for physical activity apps available in English. The description pages of eligible apps were reviewed by 4 independent reviewers for evidence-based content, technological, and descriptive features. An a priori subset of apps was downloaded for further review (n=6 affiliated with a non-commercial agency; n=10 top rated; n=10 random selection), and developers were contacted for information regarding evidence-informed content.

Results: The initial search yielded 2400 apps, of which 379 apps (n=206 iTunes; n=173 Google Play) were eligible. Primary results demonstrated no apps (n=0) adhering to evidence-based guidelines for aerobic physical activity, and 7 out of 379 implementing evidence-based guidelines for resistance training physical activity. Technological features of apps included social networking (n=207), pairing with a peripheral health device (n=61), and measuring additional health parameters (n=139). Secondary results revealed 1 app that referenced physical activity guidelines (150 minutes/weekly of exercise), and demonstrated that apps were based on various physical activity reports (n=4) or personal expertise (n=2).

Conclusions: The present study demonstrated a shortage of evidence-based physical activity apps. This gap underscores the need for development of evidence-informed mobile apps. Results highlight the opportunity to develop evidence-informed mobile apps that can be used clinically to enhance health outcomes.

(JMIR mHealth uHealth 2015;3(2):e43) doi:10.2196/mhealth.4003

KEYWORDS

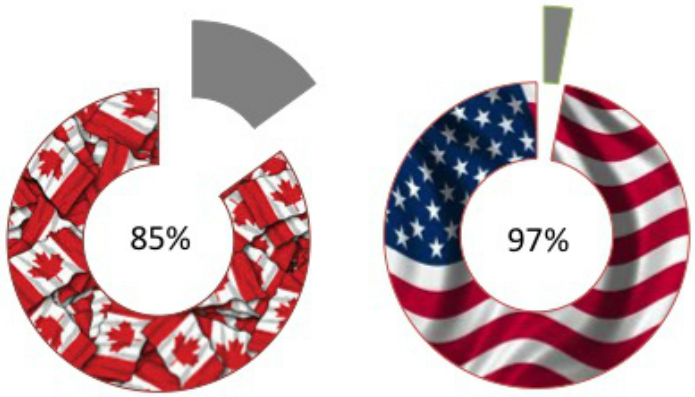
Mobile applications; Exercise; Public Health

Introduction

Background

Health systems worldwide are being increasingly challenged by care for chronic conditions and non-communicable diseases such as cardiovascular disease, cancer, and diabetes [1,2]. In North America, 89% of total mortality in Canada and 87% in the United States can be attributed to non-communicable disease [3]. Physical activity is an important determinant of health, including primary and secondary prevention of chronic and non-communicable diseases [4]. Physically inactive lifestyles are the fourth leading cause of death, and globally contribute to more than 3 million deaths per year [5]. In North America, the majority of adults in the United States and Canada are not meeting minimum public health recommendations for physical activity [6,7]. Engaging in unhealthy physical activity behaviors, such as a physically inactive lifestyle, has substantial negative consequences for public health including the economic burden for society [5] (see Figure 1).

Figure 1. Physical inactivity burden in North America. 85% of Canadian adults and 97% of American adults fail to meet public health guidelines for physical activity.



Physical Activity

Public health guidelines for physical activity are a summary of best available evidence [5,8-10]. For the general adult and older adult population (ie, ≥18 years), physical activity guidelines recommend engaging in at least 150 minutes of moderate- to vigorous-intensity physical activity weekly (ie, aerobic exercise), in bouts of at least 10 minutes, as well as whole body strength training activities for major muscle groups on at least 2 days per week (Textbox 1). Additionally the guidelines highlight that more physical activity is beneficial for health, and that older adults (≥65 years) benefit from balance training to reduce fall risk.

Textbox 1. Evidence-based physical activity guidelines for adults.

Aerobic Activity

- 150 minutes of moderate- to vigorous-intensity physical activity, accumulated in bouts ≥10 minutes
- More activity is beneficial for health

Strengthening Activity

- Resistance training to strengthen major muscle groups on ≥2 days/week

Prescribing Physical Activity

Health care providers prescribe treatment regimens to help clients manage their health. A written prescription holds symbolic meaning for clients, indicating that their health practitioner believes in the value of the behavior for managing or promoting health [11]. A health behavior message, such as physical activity, delivered by a health practitioner may be an important stimulus for individual change [12,13]. The evidence base demonstrates the efficacy of prescribing exercise behaviors through primary care for improving both physical activity levels and cardiovascular health [14]. However, the extent to which a client adheres to prescribed behaviors is highly variable, and clinicians may need to consider providing additional tools and interventions for clients to promote adoption of prescribed behaviors [13]. Client self-management and medical technologies can be leveraged to increase engagement in prescribed behaviors by attracting and involving patients in their own care [2,15].

Mobile Health Apps

The use of mobile health technologies involving smartphones (ie, broadband-enabled phones with the capacity to download and run apps) is a rapidly growing focus for chronic disease management and prevention [16]. Around the world, there are more than two billion smartphone subscriptions [17]. Moreover,

in 2014 there was an increase of 400 million subscriptions from the previous year [17]. In North America, 95% of Americans and 80% of Canadians have active smartphone subscriptions [17]. Mobile health technologies have demonstrated potential for engaging individuals in on-going self-management of prescribed physical activity behaviors for disease management and prevention [16,18-20]. It has been suggested that mobile health devices have the ability to deliver multifaceted behavior change interventions using health apps [21]. Moreover, technology-enhanced features included in apps have the potential to reduce user burden and facilitate behavior change through features such as integrated measurement of additional health parameters, social networking, reminders to engage in a behavior, calendar for scheduling behavior, and prompts for lapses in adherence with behaviors [22]. Smartphone subscriptions are pervasive around the world, and leveraging the accessibility of mobile health apps could hold promise for health practitioners and health behavior intervention.

Physical activity apps are abundant. On any given day, a category search (eg, "Health & Fitness") or keyword search (eg, "physical activity") of any platform will generate thousands of search results. A recent systematic review reported the use of apps to increase physical activity [23]. However, the authors noted that few apps have been examined through rigorous intervention [23], which underscores the challenge for users to select an evidence-based app. The challenge for clients and health practitioners is discerning which apps, if any, to use to promote prescribed health behaviors. Previous studies have explored the availability of mobile phone apps to change smoking behaviors, manage diabetes, and enhance weight loss outcomes [21,22,24]. These studies showed that available apps were generally not evidence-based, though some evidence-based features were included. While previous research has concluded that physical activity apps lack sufficient inclusion of evidenced-based health behavior change theories [25], it remains unclear whether available physical activity apps contain evidence-based physical activity content and could be used by health practitioners to counsel patients on healthy physical activity behaviors.

Purpose

The aim of the present review was to identify publically available mobile physical activity apps that represent the evidence-based public health guidelines for physical activity (Textbox 1). Additionally, technological features of apps that have previously been shown to improve utility as well as promote adherence and health outcomes [22] were identified.

Methods

Search Process

The search strategy was developed based on previously published studies examining evidence-based apps [22,24]. The mobile app stores for Apple (iTunes) and Android (Google Play) platforms were searched in February 2014 by 4 independent reviewers (2 per platform). Keyword ("physical activity", "fitness", "walking", and "pedometer") and category ("Health & Fitness" both free and paid) searching was conducted during a single day (February 18, 2014) to determine available apps. The search optimization was set to "relevance", meaning that results were presented in descending order of relevance using app store algorithms.

Primary Review

Building on the methods reported by Breton et al [24], the app description pages provided by each app store (iTunes and Google Play) of the first 100 results from each search (ie, 6 searches x 2 platforms) were screened for eligibility. Eligibility criteria included availability in English (demonstrated through text and/or screenshots provided in the description), primary aim of app was physical activity (eg, diet apps with secondary option to track physical activity, sleep tracking apps, or menstrual cycle tracking apps were ineligible), and the app tracked or measured physical activity. Data were extracted from the description page for all eligible apps, which included descriptive data for physical activity content, social behavior (eg, linking with social networks), and clinical utility (eg, cost, linking with peripheral health devices), as well as user ratings and number of downloads (available for Google Play only). While this search strategy is limited to information available on the description page, it represents the typical process a user (clinician or client) would follow when selecting an app for download.

Assessing Evidence-Based Content

The eligible apps were compared to physical activity guidelines to assess for evidence-based content. Specifically, app descriptions that included 150 minutes/weekly of moderate- to vigorous-intensity physical activity or ≥2 days/weekly of whole body strengthening activities were considered to reflect evidence-based guidelines. Adherence to public health recommendations for physical activity was coded as present/absent, and detailed description was added if present.

Assessing App Features

Features that could be important in designing future apps or in assisting clinicians and clients in selecting an app for use were identified. An a priori list of general and technological features was created to assist the review process. General features of eligible apps were recorded, including platform, cost, and user ratings. It was also noted if the app was endorsed or affiliated with an agency (eg, government, academic, commercial). Description of how physical activity was measured within the app, as well as any additional health parameters that were tracked using the app were recorded. Technological features of the app were also recorded, including capacity for reminders, calendars/scheduling, social networking, and connecting with other peripheral devices (eg, health devices, computers).

Secondary Review

Building on the methods of Pagoto et al [22], a subset of apps were downloaded for further review: (1) apps that were endorsed by a non-commercial agency (eg, a university or research group), (2) random selection of top-rated apps (ie, user ratings above 4.5 "stars"), and (3) random selection of remaining apps. Random selection was conducted using a Web-based randomization tool [26]. Additionally, publishers of the subset of downloaded apps were contacted via the email provided on

the app's description page to inquire about the physical activity evidence base that informed the app's development process.

Statistical Methods

Descriptive statistics were used to summarize the features available among all eligible apps. Data was managed in Microsoft Excel for Mac 2011. Data were collected in February 2014, and analyzed in March 2014.

Results

Overview

The search process is outlined in Figure 2. The initial search yielded 2400 apps. After removing duplicate results (n=1282) and ineligible (n=739) apps, descriptive review was conducted on 379 apps (n=206 Apple; n=173 Android). Descriptive results are reported in Table 1. Figure 3 displays examples of screenshots for individual apps from the app stores.

Table 1. Descriptive results from primary review (n=379).

Category	n (%)
Evidence base	
Includes public health recommendations for aerobic physical activity quantity	0 (0%)
Includes public health recommendations for resistance training	7 (1.8%)
Endorsement	
Reference to an affiliated agency	18 (4.7%)
Academic	5 (27.8%)
Commercial	13 (72.2%)
Technological features	
Calendar to schedule physical activity	93 (24.5%)
Reminder to engage in physical activity	45 (11.9%)
Pairs with a peripheral device	61 (16.1%)
Capacity for social networking	207 (54.6%)
Health features	
Includes a physical activity target (quantity)	117 (30.9%)
Records physical data (eg, blood pressure, heart rate, calories, body mass index, limb girths/circumferences)	139 (36.7%)

Figure 2. Flow of search results.

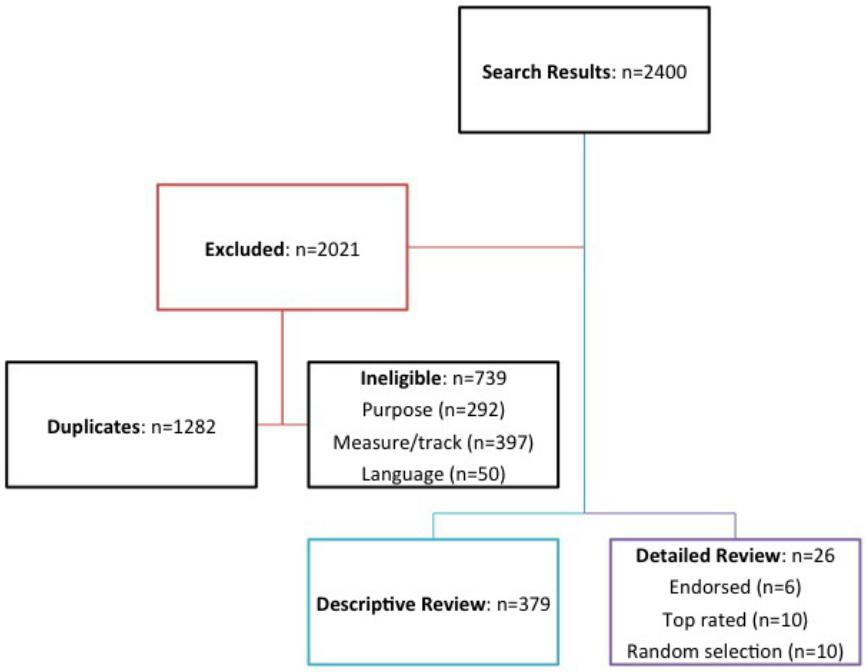
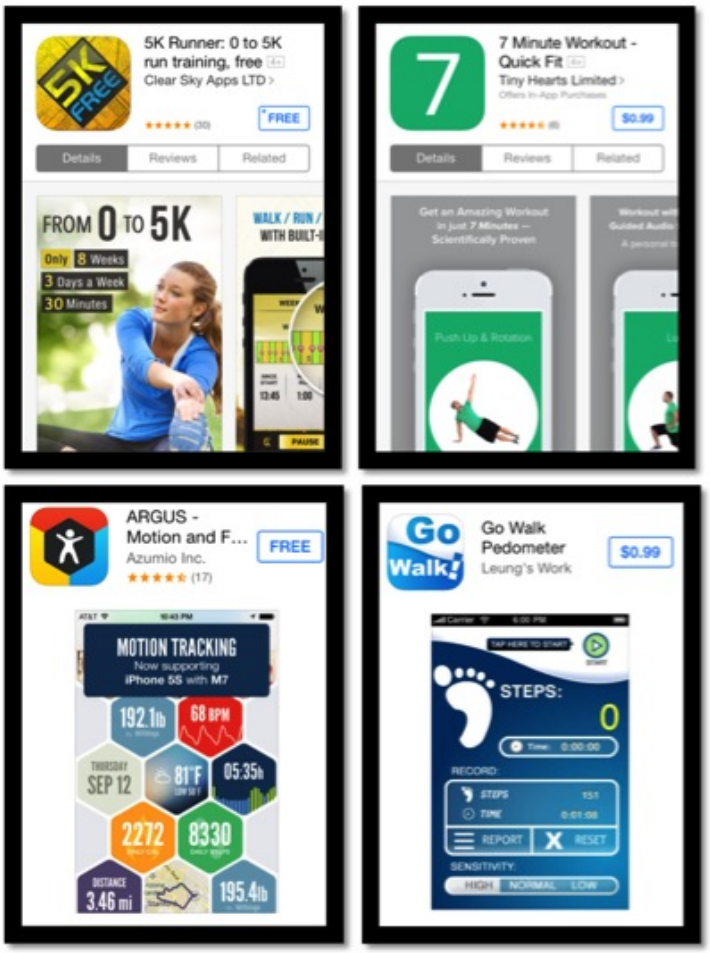


Figure 3. Example of app screenshots.



Primary Results

The majority of apps were free (237/379, 62.5%), with 142 out of 379 (37.5%) apps charging between CAD $0.99-9.87 to purchase. The most common app price was $0.99. A small portion of apps (18/379, 4.7%) were endorsed by or affiliated with an agency, of which 13 of 18 were commercial (such as corporate apparel and websites) and 5 of 18 were academic (such as research groups and universities).

Descriptive review revealed that no apps (n=0) included evidence-based public health targets for aerobic physical activity, while 7 out of 379 (1.8%) apps included evidence-based public health targets for resistance training. Less than one-third of apps (117/379, 30.9%) included a daily physical activity target, which ranged from pre-set targets (eg, 10,000 steps, 7 minutes, 100 push-ups) through user-defined targets. Methods for measuring physical activity included pedometers (140/379, 36.9%), self-report log (86/379, 22.7%), global positioning system (61/379, 16.1%), accelerometers (44/379, 11.6%), calories expended (89/379, 23.5%), distance (97/379, 25.6%), time (92/379, 24.3%), speed (23/379, 6.1%), and metabolic equivalents (2/379, 0.5%). A combination of these methods was present in nearly half of apps (181/379, 47.8%). In addition to physical activity, 139 out of 379 (36.7%) apps recorded other health data, including calories consumed/expended (52/379, 13.7%), heart rate (38/379, 10.0%), body weight (20/379, 5.3%), body mass index (19/379, 5.0%), limb girths/circumferences (10, 2.6%), and blood pressure (7/379, 1.8%). Similar to measuring physical activity, it was more common for apps to measure a combination of these health parameters (56/139, 40.3%) than a single additional health parameter.

Technological features of interest included capacity for social networking, scheduling features for planning physical activity, reminders to engage in physical activity, and capacity to pair with a peripheral device. Half of apps (207/379, 54.6%) had a social networking capacity; 93 out of 379 (24.5%) included a scheduling feature; 45 out of 379 (11.9%) included a reminder feature; and 61 out of 379 (16.1%) apps paired with a peripheral device, which included a proprietary device like step counters, or associated health devices like weigh scales, heart rate monitors, blood pressure monitors, and glucometers.

Secondary Results

Detailed review was conducted on a subset (n=26) of apps: the 6 from the initial search that were endorsed by a non-commercial agency (n=5 Apple, n=1 Android); 10 randomly selected from top user ratings (n=5 per platform); and 10 randomly selected from the remainder of the sample (n=5 per platform). In addition to information available from the description page, downloading apps for detailed review revealed that 1 of 26 (3%) apps included reference to public health guidelines for aerobic physical activity, and that 2 of 26 (7%) apps offered features such as social networking, tracking of multiple health parameters, and

ideas for physical activity programming available for additional costs (ie, purchase via subscription).

App developers from the subset of apps were contacted, of which 6 of 26 (23%) of developers responded. The physical activity content of 3 apps (coincidentally, 1 from each category selected for detailed review) was inspired by lay and peer-reviewed reports of high-intensity interval training circuits [27,28]; 1 app was also inspired by the physical training manuals for the United States army [29]. One app was inspired by a corporate report on health benefits of physical activity [30]. Two apps were inspired by personal expertise.

Discussion

Evidence-Based Content

The present study explored the presence of evidence-based content among physical activity apps marketed through iTunes and Google Play mobile app stores. Previous research examined evidence-based features of smoking cessation, weight loss, including pediatric obesity management, and diabetes management apps, and it was found that the majority of apps do not adhere to evidence-informed practices [21,24,31]. This research identified limitations in analyzing apps for evidence-based content. Review based solely on the description page may bias findings toward marketable features with limited explanation of evidence-based content [22,24]. Therefore, it has been suggested that downloading apps for detailed review may provide more robust data [22,24]. The current study combined both approaches by collecting data from the description page of 379 apps, and subsequently downloading a selected subset of apps based on a priori criteria for further review. Additionally, app developers were contacted to determine evidence-informed content. The combination of approaches for data collection may contribute to a more robust understanding of physical activity apps. Therefore, future research may wish to consider downloading apps for content review as well as communicating with app developers for comprehensive review.

Our primary results demonstrated that no apps included public health recommendations for aerobic physical activity (ie, 150 minutes of moderate- to vigorous-intensity physical activity, in bouts ≥10 minutes). However, detailed review of the subset of apps revealed that one app included reference to evidence-based recommendations for physical activity to inform users, and 4 apps were inspired by various documents about health benefits of physical activity. Additionally, a small proportion (2%) of available apps incorporated whole body strength training of major muscle groups in conjunction with aerobic intervals. The findings suggest limited use of evidence informed practices among physical activity apps. In order to assist users in selecting an app for health promotion, app developers may wish to reference the evidence-based content (eg, public health guidelines for physical activity) on the app's description page.

App Features

Descriptive review revealed a broad array of features available within apps, including pairing with peripheral devices to measure markers of health other than physical activity (eg, heart rate, blood pressure, blood glucose). Mobile health devices (such as peripheral devices to measure health markers) can be used to assist both clinicians and clients with evidence-based self-management care and chronic disease prevention [16]. Previous interventions combining mobile health devices and exercise prescription have demonstrated beneficial effects on clinical markers of cardiovascular health [20,32]. Therefore, the ability to connect with additional health devices may be of clinical benefit.

In addition to pairing with peripheral health devices, smartphones allow users to easily share information. Sharing information in a collaborative nature, such as through social media outlets, holds potential for engaging clients in prescribed health behaviors [15,33]. In turn, this may further enable clients and enhance treatment outcomes. Therefore, physical activity apps that include social features such as linking with other users, sharing information, as well as connecting with and building new social networks holds potential for engaging users in planned health behaviors such as physical activity. The present investigation demonstrated that approximately half (54.6%) of physical activity apps included a feature to allow for social networking, such as posting a workout, connecting with other users, and sharing physical activity information with other users. Clinical utility of physical activity apps may be enhanced through inclusion of social networking features.

Implications for Clinical Use

Direct-access health practitioners, such as primary care physicians, may be the initial health service access point for clients. Therefore, these health practitioners are perhaps ideal clinicians for prescribing physical activity behaviors to reduce economic and disease burden among clients. It has been suggested that a health message delivered in the primary care setting can be an important catalyst for change, in part by representing the value a health practitioner places on health behaviors such as physical activity [11,12,34]. Moreover, physical activity prescription through primary care can significantly increase self-reported physical activity levels, and positively impact cardiorespiratory fitness [14]. These health outcomes from prescribing physical activity may benefit public health, in part by reducing economic burden of unhealthy physical activity behaviors. However, regular visits for healthy clients are generally only one time per year and adherence to physical activity prescription tends to decrease over time. An evidence-based mobile health app could be one tool/strategy to assist clients with engaging in healthy physical activity behaviors during the gap between appointments. For example, recent evidence supports the use of an app through primary care to promote walking [35]. The ability to assess an app's adherence with evidence-informed practices is of use for both clinicians and clients [24]. Unfortunately, we are unaware of efficacy studies for publicly available physical activity apps available on smartphones to promote physical activity behaviors prescribed by clinicians.

Limitations

We are aware of limitations in our assessment of physical activity apps. Features available via additional subscription were not included in the present evaluation. For example, apps

that required download of additional features for cost, or subscription/membership once the app has been purchased were not considered in the descriptive analysis. Moreover, no rating was made on reliability or validity of physical activity measures within the apps, such as accuracy of step counters, body mass index calculator, heart rate monitors, etc. Some apps were specific to a proprietary device (eg, app that pairs with a brand-specific step counter). These apps were included in the present review; however, the cost for purchasing the proprietary device was not considered in the analysis of cost to download the app.

Moreover, the present review was limited to mobile apps on two of the leading platforms (ie, Apple, Android). Future research may benefit from including additional platforms. Additionally, Google Play search results generate a list of apps available on multiple devices (eg, computer, tablet, smartphone) while the iTunes mobile app store was limited to apps available on smartphone (ie, iPhone) only. Of particular interest for future investigation may be seeking apps available on various devices, including those that do not require a mobile (smart)phone subscription or monthly cellular data plan. This may enhance the applicability of results to allow for use by clients who do not have mobile data plans.

Previously it has been reported that physical activity apps lack sufficient inclusion of health behavior change theories [25]. While the present study is limited in assessing apps for inclusion of evidence-based physical activity guidelines, additional research may wish to explore adherence of physical activity apps to change user's activity behaviors. Previous studies have explored the availability of evidence-based mobile phone apps to change health-related behaviors such as smoking, chronic disease management, exercise, and obesity prevention [21-24,31]. More research is warranted that explores inclusion of techniques grounded in behavioral change theories to promote adoption of healthy physical activity behaviors (eg, motivational interviewing, transtheoretical model, social cognitive theory, health belief model, theory of planned behavior, behavior modification, social learning theory, and theory of reasoned action [14,36]).

Currently, there is no standardized tool to assess clinical features or evidence-based content of mobile apps. Future research may benefit from development of a tool to systematically assess apps. In the absence of such tool, the present review is limited to assessing evidence-based content in relation to established public health guidelines for physical activity.

The present investigation describes the current availability of physical activity apps. The rate of technological development far outpaces the research process. As such, findings should be considered in the context of selecting an app for use in practice or for developing an evidence-based physical activity app. For this purpose, reference to apps by proprietary name has been intentionally limited, as it was not the goal of this paper to recommend a specific app for use in clinical practice.

Conclusion

The present review demonstrates a shortage of evidence-based physical activity apps. This gap underscores the need for development of evidence-informed mobile apps. Results highlight the opportunity to develop evidence-informed mobile apps that can be used clinically to enhance health outcomes. Additionally, social integration features (eg, sharing and connecting with others) as well as technological features (eg, pairing with peripheral health devices) may offer the greatest potential to enhance health outcomes among clients prescribed healthy physical activity behaviors.

Acknowledgments

The authors wish to thank V Foo, A MacDougall, A Saffron, and E Shellington for their assistance with data collection. EK is supported by a doctoral research award from the Canadian Institutes of Health Research.

Conflicts of Interest

None declared.

References

1. Mathers C, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
2. WHO. Adherence to long-term therapies: Evidence for action. Geneva: World Health Organization; 2003. URL: http://whqlibdoc.who.int/publications/2003/9241545992.pdf [accessed 2015-05-15] [WebCite Cache ID 6YYDoEg60]
3. WHO. Noncommunicable disease country profiles. Geneva: World Health Organization; 2011. URL: http://www.who.int/nmh/publications/ncd_profiles2011/en/ [accessed 2015-05-15] [WebCite Cache ID 6YYDzGji2]
4. Centers for Disease Control and Prevention. 2011. Physical activity and health URL: http://www.cdc.gov/physicalactivity/everyone/health/ [accessed 2014-10-31] [WebCite Cache ID 6TjmRgy6n]
5. WHO. Global health recommendations on physical activity for health. Geneva: World Health Organization; 2010. URL: http://whqlibdoc.who.int/publications/2010/9789241599979_eng.pdf [accessed 2015-05-15] [WebCite Cache ID 6YYFyzDel]
6. Colley RC, Garriguet D, Janssen I, Craig CL, Clarke J, Tremblay MS. Physical activity of Canadian adults: accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. Health Rep 2011 Mar;22(1):7-14 [FREE Full text] [Medline: 21510585]

7. Tudor-Locke C, Brashear MM, Johnson WD, Katzmarzyk PT. Accelerometer profiles of physical activity and inactivity in normal weight, overweight, and obese U.S. men and women. Int J Behav Nutr Phys Act 2010;7:60 [FREE Full text] [doi: 10.1186/1479-5868-7-60] [Medline: 20682057]
8. CSEP. Canadian physical activity guidelines for adults.: Canadian Society for Exercise Physiology; 2011. URL: http://www.csep.ca/CMFiles/Guidelines/CanadianPhysicalActivityGuidelinesStatements_E.pdf [accessed 2015-05-15] [WebCite Cache ID 6YYGZ3mi7]
9. Tremblay MS, Warburton DER, Janssen I, Paterson D, Latimer A, Rhodes R, et al. New Canadian physical activity guidelines. Appl Physiol Nutr Metab 2011 Feb;36(1):36-46; 47 [FREE Full text] [doi: 10.1139/H11-009] [Medline: 21326376]
10. USDHHS. Physical activity guidelines for Americans.: United States Department of Health and Human Services; 2008. URL: http://www.health.gov/paguidelines/pdf/paguide.pdf [accessed 2015-05-15] [WebCite Cache ID 6YYVYTdVz]
11. Pfeiffer BA, Clay SW, Conatser RR. A green prescription study: does written exercise prescribed by a physician result in increased physical activity among older adults? J Aging Health 2001 Nov;13(4):527-538. [Medline: 11813739]
12. Petrella RJ, Koval JJ, Cunningham DA, Paterson DH. Can primary care doctors prescribe exercise to improve fitness? The Step Test Exercise Prescription (STEP) project. Am J Prev Med 2003 May;24(4):316-322. [Medline: 12726869]
13. Petrella RJ, Lattanzio CN, Overend TJ. Physical activity counseling and prescription among canadian primary care physicians. Arch Intern Med 2007 Sep 10;167(16):1774-1781. [doi: 10.1001/archinte.167.16.1774] [Medline: 17846397]
14. Orrow G, Kinmonth A, Sanderson S, Sutton S. Effectiveness of physical activity promotion based in primary care: systematic review and meta-analysis of randomised controlled trials. BMJ 2012;344:e1389 [FREE Full text] [Medline: 22451477]
15. Knight E. Leveraging the Power of Social Media to Engage Patients and Achieve Treatment Outcomes. Crit Rev Phys Rehabil Med 2012;24(3-4):169-177. [doi: 10.1615/CritRevPhysRehabilMed.2013007169]
16. Sieverdes JC, Treiber F, Jenkins C. Improving diabetes management with mobile health technology. Am J Med Sci 2013 Apr;345(4):289-295. [doi: 10.1097/MAJ.0b013e3182896cee] [Medline: 23531961]
17. International Telecommunications Union. ITU Committed to Connecting the World 2014 URL: http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx [accessed 2014-10-31] [WebCite Cache ID 6TjnqLahc]
18. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
19. Hjelm NM. Benefits and drawbacks of telemedicine. J Telemed Telecare 2005;11(2):60-70. [doi: 10.1258/1357633053499886] [Medline: 15829049]
20. Stuckey MI, Russell-Minda E, Read E, Munoz C, Shoemaker K, Kleinstiver P, et al. Diabetes and Technology for Increased Activity (DaTA) study: results of a remote monitoring intervention for prevention of metabolic syndrome. J Diabetes Sci Technol 2011 Jul;5(4):928-935 [FREE Full text] [Medline: 21880236]
21. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]
22. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
23. 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;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
24. Breton ER, Fuemmeler BF, Abroms LC. Weight loss-there is an app for that! But does it adhere to evidence-informed practices? Transl Behav Med 2011 Dec;1(4):523-529 [FREE Full text] [doi: 10.1007/s13142-011-0076-5] [Medline: 24073074]
25. 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]
26. Random lists. 2014. URL: http://www.randomlists.com/ [accessed 2014-10-12] [WebCite Cache ID 6TG1enaDt]
27. Comana F. A perspective on high-intensity interval training, in The National Academy of Sports Medicine (NASM) Blog. 2013. URL: http://blog.nasm.org/training-benefits/perspective-high-intensity-interval-training/ [accessed 2014-11-01] [WebCite Cache ID 6TlihusqY]
28. Kilka B, Jordan C. High-intensity circuit training using body weight: Maximum results with minimal investment. American College of Sports Medicine Health & Fitness Journal 2013;17(3):8-13. [doi: 10.1249/FIT.0b013e31828cb1e8]
29. United States Army. Army physical readiness training field manual. URL: http://armypubs.army.mil/doctrine/DR_pubs/dr_a/pdf/fm7_22.pdf [accessed 2014-11-01] [WebCite Cache ID 6TlirLoif]
30. Designed to move: A physical activity action agenda.: Nike; 2012. URL: http://www.designedtomove.org/en_US/?locale=en_US [accessed 2014-11-01] [WebCite Cache ID 6TljYwMUF]
31. Schoffman DE, Turner-McGrievy G, Jones SJ, Wilcox S. Mobile apps for pediatric obesity prevention and treatment, healthy eating, and physical activity promotion: just fun and games? Transl Behav Med 2013 Sep;3(3):320-325 [FREE Full text] [doi: 10.1007/s13142-013-0206-3] [Medline: 24073184]

32. Stuckey M, Fulkerson R, Read E, Russell-Minda E, Munoz C, Kleinstiver P, et al. Remote monitoring technologies for the prevention of metabolic syndrome: the Diabetes and Technology for Increased Activity (DaTA) study. J Diabetes Sci Technol 2011 Jul;5(4):936-944 [FREE Full text] [Medline: 21880237]
33. Knight E, Werstine RJ, Rasmussen-Pennington DM, Fitzsimmons D, Petrella R. Physical therapy 2.0: leveraging social media to engage patients in rehabilitation and health promotion. Phys Ther 2015 Mar;95(3):389-396. [doi: 10.2522/ptj.20130432] [Medline: 24627429]
34. Petrella R, Koval J, Cunningham D, Paterson D. A Self-Paced Step Test to Predict Aerobic Fitness in Older Adults in the Primary Care Clinic. Journal of the American Geriatrics Society 2001 May;49(5):632-638. [doi: 10.1046/j.1532-5415.2001.49124.x] [Medline: 11380757]
35. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):e384-e391. [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
36. Elder JP, Ayala GX, Harris S. Theories and intervention approaches to health-behavior change in primary care. Am J Prev Med 1999 Nov;17(4):275-284. [Medline: 10606196]

Edited by G Eysenbach; submitted 04.11.14; peer-reviewed by L Masse, A Direito; comments to author 20.02.15; revised version received 28.02.15; accepted 14.03.15; published 21.05.15.

Please cite as:
Knight E, Stuckey MI, Prapavessis H, Petrella RJ
Public Health Guidelines for Physical Activity: Is There an App for That? A Review of Android and Apple App Stores
JMIR mHealth uHealth 2015;3(2):e43
URL: http://mhealth.jmir.org/2015/2/e43/
doi:10.2196/mhealth.4003
PMID:25998158

Original Paper

How Willing Are Adolescents to Record Their Dietary Intake? The Mobile Food Record

Carol Jo Boushey[1], MPH, PhD; Amelia J Harray[2], Grad Dip Diet, BSc; Deborah Anne Kerr[2], BSc, MSc, PhD; TusaRebecca E Schap[3], MPH, PhD, RD; Stacey Paterson[4], AAS, BSc; Tanisha Aflague[1], MS; Marc Bosch Ruiz[5], MS Eng, PhD; Ziad Ahmad[6], BE; Edward J Delp[6], PhD

[1]University of Hawaii Cancer Center, University of Hawaii, Honolulu, HI, United States

[2]School of Public Health, Curtin University, Perth, Australia

[3]National Cancer Institute (NCI), National Institutes of Health (NIH), Cancer Prevention Fellowship Program, Division of Cancer Prevention, Division of Cancer Control and Population Sciences, Bethesda, MD, United States

[4]Department of Dietetics and Human Nutrition, University of Kentucky, Lexington, KY, United States

[5]Qualcomm, Inc, 10160 Pacific Mesa Boulevard, Office 605k, San Diego, CA, United States

[6]Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, United States

Corresponding Author:
Carol Jo Boushey, MPH, PhD
University of Hawaii Cancer Center
University of Hawaii
701 Ilalo Street,
Room 525,
Honolulu, HI, 96813
United States
Phone: 1 808 564 5915
Fax: 1 808 586 2982
Email: cjboushey@cc.hawaii.edu

Abstract

Background: Accurately assessing the diets of children and adolescents can be problematic. Use of technologies, such as mobile apps designed to capture food and beverages consumed at eating occasions with images taken using device-embedded cameras, may address many of the barriers to gathering accurate dietary intake data from adolescents.

Objective: The objectives of this study were to assess the willingness of adolescents to take images of food and beverages at their eating occasions using a novel mobile food record (mFR) and to evaluate the usability of the user confirmation component of the mFR app, referred to as the "review process."

Methods: Mixed methods combining quantitative and qualitative protocols were used in this study. Adolescents (11-15-year olds) attending a summer camp were recruited to participate in the study. First, the participants were asked to take images of foods and beverages consumed as meals and snacks for 2 consecutive days using the mFR app running on an iPhone and the number of images taken was noted. This was followed by focus group sessions to evaluate usability, which was analyzed by content and themes. After using the mFR, a think-aloud method was used to evaluate the usability of the mFR method for reviewing system-identified foods (ie, the review process). A usability questionnaire was administered at the end of all activities.

Results: The mFR was accepted by the majority of the 24 boys and 17 girls (n=41) but varied according to gender and eating occasion. Girls were significantly more likely than boys to capture images of their eating occasions (Fisher exact test, P=.03). Participants were more likely to take images of their breakfasts (90%, 36/40) and lunches (90%, 72/80) and least likely to capture afternoon and evening snacks, 54% (43/80) and 40% (32/80), respectively. The major themes from the focus groups with regard to using the mFR were games, rewards, and the need to know more about why they were using the app. Results of the usability questionnaire indicated that including a game component would be important to increase willingness to use the mFR, and a high majority of the participants indicated a willingness to use the mFR for 7 days or more. The image review process was found to be easy to use except for some confusion with overlapping markers on the screen.

Conclusions: The adolescents' experiences with and feedback about the mFR highlighted the importance of increased training, reminders, entertainment (eg, games), and training with practice in using the device to capture complete dietary intake as part of their active lifestyles.

(JMIR mHealth uHealth 2015;3(2):e47) doi:10.2196/mhealth.4087

KEYWORDS

adolescents; children; dietary assessment; mobile food record; novel technology

Introduction

Collecting information about dietary intake from children and adolescents is challenging. Developmental stages add to the complexity of deciding whether to gather information from the parent, the child, or both. Burrows and colleagues [1] performed a systematic review of validation studies that compared reported estimated energy intake from dietary assessment methods with the method of doubly labeled water as a biomarker for total energy intake among children aged 6 months to 21 years. The methods that provided the most accurate estimate of energy intake were the estimated dietary record as completed by the parent for 0.5-4-year olds, the multiple-pass 24-hour dietary recall as reported by the parent for 4-11-year olds, and the dietary history as reported by adolescents aged 16-21. However, a recommended method for 12-15-year olds did not emerge in their review. For early adolescents (11-14-year olds), the value of parental assistance remains equivocal. When girls without parental assistance completed dietary records at ages 10, 12, and 15, Bandini and colleagues [2] found that differences between estimated energy intakes compared with total energy expenditure calculated using the doubly labeled water method widened with increasing age. In a separate study, completion of dietary records with parental assistance was found to be associated with underreporting of energy intake by 20% and 33% at 11 and 12 years of age, respectively [3]. These latter results suggest that parental assistance for completion of dietary records among early adolescents does not enhance accuracy.

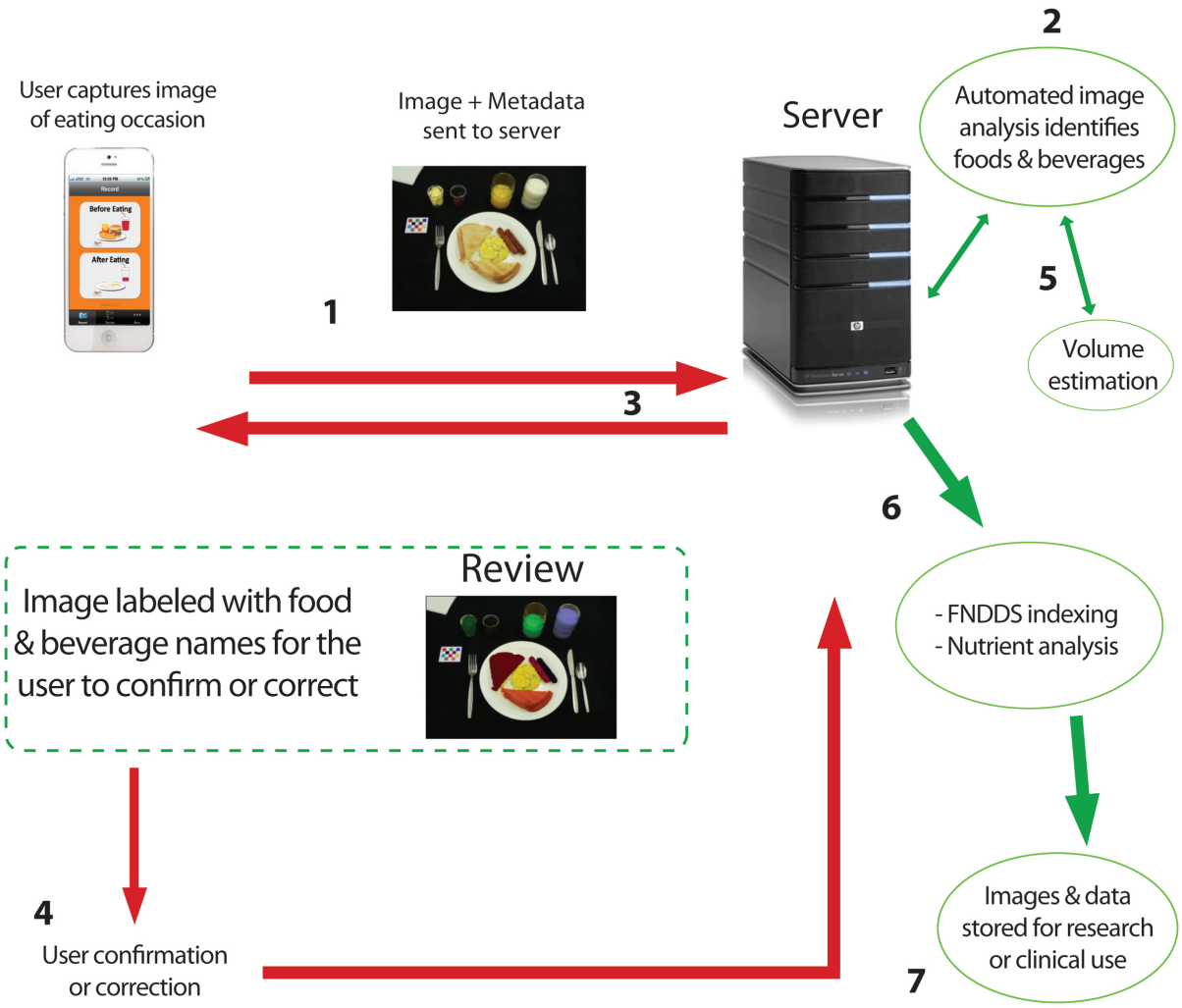
Livingstone et al [4] noted that assessing dietary intake in adolescents between 11 and 14 years of age is particularly problematic because the novelty of recording food wears off and the assistance from parents is no longer preferred. During focus group sessions [5], adolescents reported conventional dietary assessment methods, such as 24-hour dietary recalls, food frequency questionnaires, and dietary records, to be burdensome. Findings from the focus groups strengthen the need to explore more acceptable methods of assessing diet in adolescents in an effort to improve cooperation, which may lead to more accurate reports of dietary intake.

Taking images of eating occasions has been proposed as a method for reducing burden on adolescents participating in dietary studies [5]. When a group of 30 early adolescent boys and girls aged 10-14 years was asked to take photographs of their eating occasions for 1 day using disposable film cameras, all but 1 child complied; the majority took images for more than 1 day [5]. Of those taking images, 100% liked the method, compared with 35% liking the written dietary record and 52% liking the interviewer-administered 24-hour dietary recall. The children of this generation were born into a digital age, resulting in generational changes, such as difficulty sustaining attention to a specific task [6]. One barrier at the forefront of a mobile Internet world is that many children may attend a school that no longer teaches penmanship [7]. Digital technology, including social media, is shaping students' writing and may lead to an increased use of informal spelling and truncated expression (eg, tweets). Thus, methods that rely on handwriting may present barriers. Use of technologies, such as Web- and mobile-based apps, may address some of the burdens and barriers to gathering accurate dietary data from children using recording methods they are familiar with, such as taking images. The rapid advancement of technology and mobile phones has motivated researchers to develop a mobile food record (mFR) to address barriers reported by adolescents [8].

The mFR allows users to take images of their food and beverages at eating occasions, which could address the barriers of hand-written pen and paper or digital entry methods that still require spelling and focused entry. Image analysis methods have the potential to automatically identify foods and beverages in the image and estimate volumes [9-12] (Figure 1). For the captured images to be useful for analysis, the user must capture all foods and beverages before and, if applicable, after consumption. A fiducial marker, which resembles a checker board square, is included in the images as a color and size reference to help with the reconstruction of a three-dimensional environment that allows for estimation of the volume of the foods and beverages [13-16]. In addition, as shown in Figure 1, the system includes a user confirmation component referred to as the "review process" [17]. This step is important as a predominantly automated system requires a method to identify new foods (whether new to the system or new to market), correct foods that are misclassified, or identify attributes that cannot be captured in an image (eg, no salt, artificially sweetened). Figure 1 shows each of the steps involved in capturing an image: (1) A user captures an image of an eating occasion that is sent to a server. (2) The image undergoes image analysis to identify the foods and beverages. (3) The labeled image is returned to the user for the "review process" as shown in the dotted line. (4) The user confirms the automatic labels or corrects the labels. (5) The image is returned to the server for final identification and volume estimation. (6) Identified foods and amounts are matched to the Food and Nutrient Database for Dietary Studies for nutrient analysis. (7) Images and data are stored on a server for use by researchers or clinicians. Schap et al [18] showed that adolescents can look at images of their eating occasions and identify the foods in the images up to 14.5 hours postprandially. Thus, the utility of using images in dietary assessment appears to have great promise.

The Pew Research Center recently reported that as of 2013, 78% of adolescents aged 12-17 years in the United States have their own mobile phone and nearly half of those (47%) own smart phones [19,20]. These reports emphasize that mobile phones have become "indispensable modes of teen communication." The widespread penetration of mobile phones, particularly smart phones, with their increased computing power would suggest that these devices open up new opportunities for dietary data collection. They also have the potential to engage adolescents in the task. There is, however, a need to evaluate the usability and acceptability of mobile apps [21,22] and whether adolescents are willing to use a mobile phone to capture images of their eating occasions throughout the day. Therefore, the objectives of this study were to assess the willingness of younger adolescents to take images of food and beverages at eating occasions and whether willingness varied by gender, to evaluate the beta version of the review process component of the mFR and to assess acceptability and usability of the mFR.

Figure 1. Diagram of the Technology Assisted Dietary Assessment (TADA) system that starts with capturing an image with the mobile food record (mFR).



Methods

Study Participants

Data were collected from adolescents participating in a residential summer camp on the Purdue University campus in 2010. There were 41 adolescents (17 girls and 24 boys) between 11 and 15 years of age who participated in this study. The active camp environment included participation in organized sports, hiking, river rafting, rock-wall climbing, relay races, swimming, bowling, social events, classes, free time, and attending a minor league baseball game. The activities described in this study were an add-on to the original objective of the residential camp, which was to establish dietary calcium requirements for bone health [23]. Participants received an incentive of US $100/3 weeks of the residential camp. No additional incentive was provided for participation in the add-on study reported here. The study methods described here were approved by the Purdue University Institutional Review Board, and informed consent and assent were obtained from the adolescents and their parents, respectively.

Study Design

This mixed-methods study sequentially used quantitative and qualitative data-collection methods and analyses to better explore adolescents' willingness to use the mFR [22,24]. The camp was held for 3 weeks and included 2 sessions, with a single week at home between the sessions (Figure 2). An iPhone 3GS was distributed to each participant in 2-day cycles during the first session (3 weeks) of the camp. After breakfast on day 1 of the 2-day cycle, the participants were briefly introduced to iPhone 3GS and the mFR app Technology Assisted Dietary Assessment (TADA) installed on each mobile phone. Phone, data services, entertainment apps, and social media apps were preinstalled on each mobile phone. Internet and Internet-accessible apps were restricted to minors. They were instructed to take images of every eating occasion, including all meals, snacks, and beverages, with the fiducial marker. During the 2 days, the research staff did not remind the participants to capture images of their eating occasions. At the end of the first 3-week session of the camp, focus group sessions were conducted separately with boys and girls to obtain qualitative feedback about using the mFR.

During the second 3-week session of the camp, a subset of the same children used the iPhones again in groups of 8-11 at a time for breakfast and/or lunch and/or dinner. This data-collection protocol differed from the first session, in that the participants were monitored during the image capture process at each meal. At the end of the day, participants were provided instructions according to the think-aloud method [25], after which each participant used the beta version of the review process component on the mFR while a research staff observed (Figure 3). This process enabled the staff to follow the flow of task completion and take notes of the participants' dialog. At the end of the second 3-week session, participants completed a usability questionnaire to solicit objective feedback about their experiences using the mFR.

Figure 2. Mixed methods study design with quantitative and qualitative protocols.

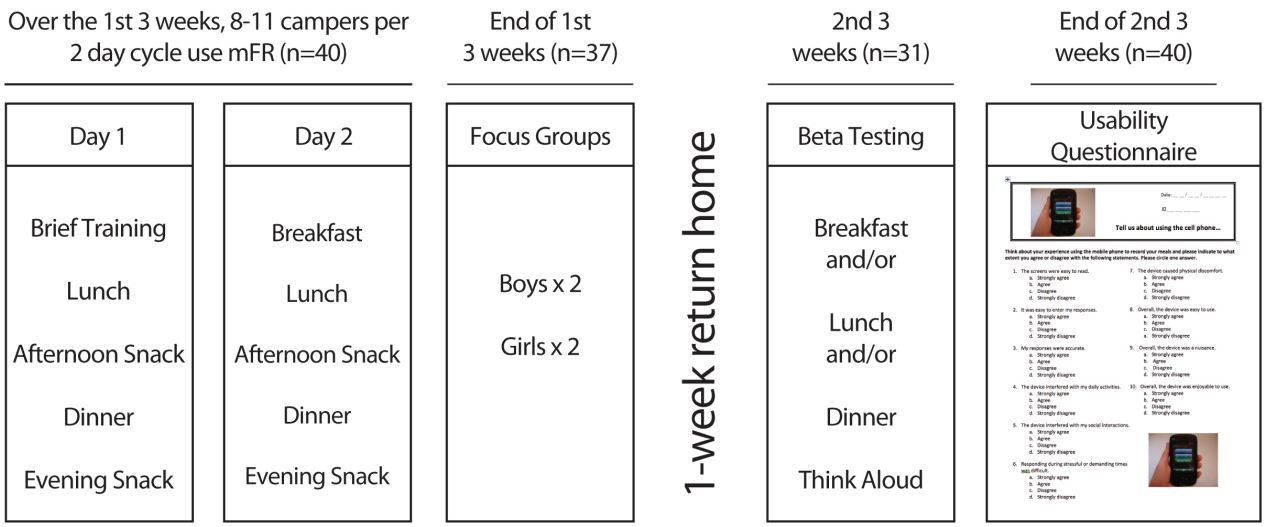
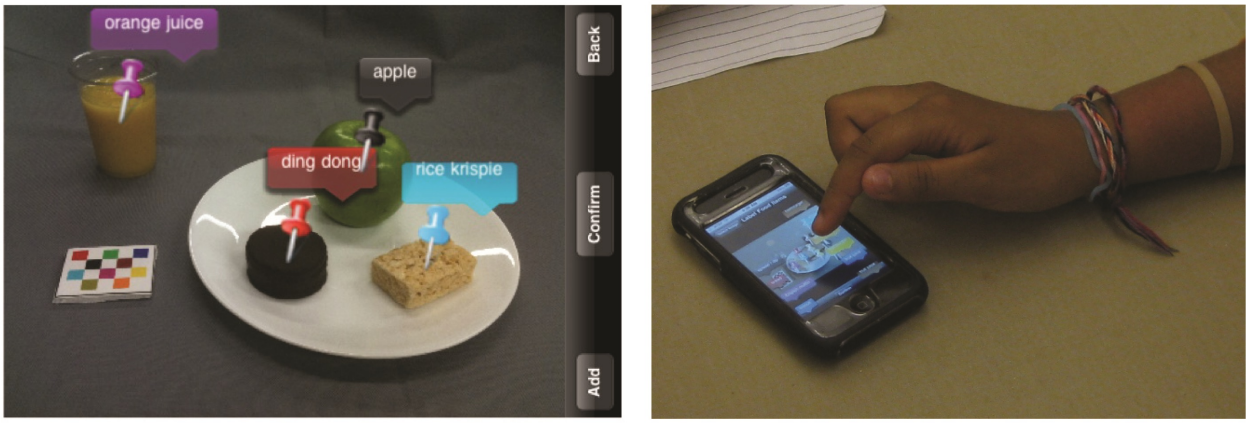


Figure 3. Beta testing the review component. The user selects an eating occasion to review. The before-eating image is in landscape view as displayed on the left. Foods are denoted with pins and the food label appears on a color coordinated bubble. To correct a misidentified food, the user taps the bubble associated with the food as shown on the right and selects the correct food from a list of foods (not shown).



Evaluation of Willingness to Take Images

The camp situation was unique in that the eating occasions were limited to breakfast, lunch, an afternoon snack, dinner, and an evening snack. A 4-day cyclic menu consisting of the aforementioned eating occasions was used. Each camper was provided with adequate food to maintain his/her weight, according to his/her estimated energy requirements, which was monitored on a daily basis. No other foods or beverages were allowed. Over the 2 days of using the mFR, each participant was expected to capture a before- and an after-eating image (image pair) for each eating occasion. It was acceptable for participants to also only capture a before-eating image because they were expected to finish all foods and beverages provided as part of the main study protocol. Thus, a usable image was considered to be either an image pair or just the before-eating image. Usable images were considered indicative of participants' willingness to comply with using the mFR. The set eating occasions allowed for the identification of the number of usable images and to classify participants as demonstrating high, moderate, or low willingness. One boy was unable to complete the task on both days due to reasons unrelated to this study and was excluded. For the remaining participants, the minimum expected total number of images was 360 (40 participants with 9 eating occasions, including 1 breakfast, 2 lunches, 2 dinners, 2 afternoon snacks, and 2 evening snacks). The boys and girls were classified as having taken 8-9 images (high willingness), 6-7 images (moderate willingness), and 3-5 images (poor willingness). At the end of each day, the research staff downloaded the mFR images captured using the TADA app to an Apple computer. The images were systematically reviewed by a trained analyst to enumerate only those containing images of the known meals or snacks.

Focus Group Sessions

Focus groups were convened to explore the issues influencing participants' willingness to capture images. The same script was used for all of the focus group sessions. All the girls attended 1 of the 2 girl-only focus group sessions. Three boys were unable to attend 1 of the 2 boy-only focus group sessions due to reasons unrelated to this study. Each focus group was led by 2 experienced moderators. Examples of questions relevant to this paper were as follows: "If we were to use these phones with other groups of teenagers, what could we do to help them remember that the main reason they have the phone is to take images of their meals and snacks?" Follow-up questions to this topic were "We have heard a lot of ideas about getting people to use the app, I just want to follow-up and see if anyone has more to add, what would motivate you to use the app each time you eat?" Another topic was "What did you like most about using the phones?" Copies of the entire script can be requested from the corresponding author of this paper. Two support staff hand-recorded each session and then pooled their notes. All staff present at a session reviewed the notes to achieve final consensus on the nature of the participants' responses and any unscripted probes from the moderators. Two staff, other than the recorders, independently reviewed and analyzed the notes, categorizing them according to content and common themes. After consensus, these were finalized.

Usability Testing a Beta Version of Image Review

Usability testing of a beta version of the user review process component of the app was used to identify potential improvements. Before the usability testing of the image review component, the children took images of at least 2 meals using the mFR. Thus, during the usability testing, all children were reviewing images of their own meals and were instructed to verbalize everything they were thinking as they decided which button to press first, and in every subsequent step, using the think-aloud method [25]. Participants were told that staff would not intervene or interrupt unless asked. Staff hand-recorded the key phrases verbalized by the participants. Standardized, nonthreatening prompts were used when a participant stopped verbalizing their thoughts before an interaction with the mFR. At the end of a session, the staff clarified their written notes with the participant. Two staff persons compiled the notes as related to the review process of the first image and subsequent image(s) and summarized common suggestions and common ease-of-use steps.

Usability Questionnaire

A usability questionnaire was used to further inform behaviors related to using the mFR. The questionnaire was mainly composed of forced choice questions with 5 responses of "strongly disagree" to "strongly agree" and several open-ended questions [26,27]. Examples of questions relevant to this paper were related to the use of fiducial marker (eg, "I found the fiducial marker easy to use.") and carrying 2 mobile phones. Open-ended questions included "How long would you record what you eat?" with response units of days, weeks, and months; and "What did you like the most about using the TADA iPhone?" Copies of the questionnaire can be requested from the corresponding author.

Data Analysis

Willingness to take images was evaluated objectively by enumerating images captured of the known eating occasions. Each image was classified as being taken of a particular meal or snack. Information about each image was entered into a database, including whether the image was the before- or after-eating image and useful metadata (ie, date and time stamps). Data from the image evaluation and the questionnaire were entered into a Microsoft Access database (Microsoft, Redmond, WA, USA) and then imported into an SPSS version 17 database (SPSS Inc, Chicago, IL, USA) for data analysis. Statistical examination to compare girls and boys included frequencies, two-by-two tables, Chi-square test, Fisher exact test, and independent samples *t* test. Statistical significance was set at $P<.05$.

Results

The classification of participants by ethnicity is shown in Table 1, and the characteristics of the boys and girls in this study are shown Table 2. During the first session, the participants used the mobile phones running the TADA app for 2 days and all mobile phones were returned with no damage. While in possession of the mobile phone, each participant had an opportunity to take images of 9 eating occasions. Of the 360

images expected as a minimum for the group of 40 participants, 246 (or 68.3%) of images were captured. There was no difference in the mean number of images captured by boys compared with girls (5.78 and 6.65 images, respectively). However, when examined by categories of high willingness (3 boys and 6 girls), moderate willingness (9 boys and 9 girls), or poor willingness (11 boys and 2 girls), girls were significantly more likely to capture images than boys (Fisher exact test P=.03). The participants were also more likely to take an image of the 1 breakfast (36/40, 90%) and the 2 lunches (72/80, 90%) than of the 2 dinners (63/80, 79%). Images of the afternoon and evening snacks were least likely to be captured at 54% (43/80) and 40% (32/80), respectively.

Table 1. Classification of study participants by ethnicity.

Ethnic group	Boys n (%)	Girls n (%)	Total[a] N (%)
Hispanic	22 (91.7)	17 (100)	39 (95.1)
Non-Hispanic	2 (8.3)	0 (0)	2 (4.9)

[a]There were a total of 24 boys (59%) and 17 girls (41%) in this study (N=41).

Table 2. Characteristics of adolescents using the mobile phone food record (mFR) while attending a summer camp.

Characteristics	Boys		Girls	
	Mean	SD	Mean	SD
Age, years	13.9	0.9	13.5	1.0
Weight, kg	73.0	25	67.6	20
Height, cm	167.1	9.1	158.6	5.1
Body mass index	25.7	7.2	26.9	8.0

In the focus groups, participants were asked what they liked best about using the mobile phones; the theme mentioned most often were games and other forms of entertainment installed on the device (Textboxes 1 and 2). In response to the question, "What would motivate you to use the app each time you eat?" the major themes were positive feedback or recognition and some humor for the girls, and rewards and reminders for the boys (Textboxes 1 and 2). Suggestions for improvement included incentives or rewards for every image taken (eg, tokens, points), reminders (eg, a screensaver with the fiducial marker, friendly alarms, or texts), "cooler" colors, money, or something appearing on the screen after taking an image (eg, fireworks, "Awesome!").

Textbox 1. Summaries of responses to some of the questions from the girls who participated in separate focus group sessions.

1. What did you like most about using the phones?
   - Games
   - Facebook
   - Internet
   - It had a case
   - Calling
   - Tracking device that provides information if people steal it.
2. What would motivate you to use the app each time you eat?
   - Knowing that others are using/doing the same thing as you.
   - Reward for a week straight with adding fun apps if we take all our pictures. Even a reward after 2 straight days of using the app.
   - Different funny pictures pop-up every time we take a picture. Maybe even saying "good job" and laughing.
   - Provide entertainment with pop-up pictures.

Textbox 2. Summaries of responses to some of the questions from the boys who participated in separate focus group sessions.

1. What did you like most about using the phones?
   - Games
   - YouTube
   - Music
   - Good for passing time
   - Apps
   - Did not get in trouble for using it
   - Keeping up with the World Cup
   - Internet
2. What would motivate you to use the app each time you eat?
   - Money
   - Extra money
   - A picture that gives an "awesome" after taking the image or "congratulations" with fireworks in the background.
   - A contest—whoever remembers to take the most photos of their meals wins something (eg, money).
   - If the phone had a bleep or vibrate around meal time.
   - Something pops up on the screen around meal time to remind you
   - Receiving a text message from TADA.
   - Alarms

The usability findings of the review process were useful (Figure 3). The most common reaction with regard to the review process was "confusing because there are too many pins overlapping." Some individuals wanted to tap the pin rather than the bubble containing the name to confirm or relabel the food. Many of the participants did not know what to do until the process was explained; however, most were able to figure out the process without instruction. The majority described the second time as "easier," and once the process was successfully mastered, many were observed as going so fast as they unintentionally deleted some pins.

The usability findings from the questionnaire pertaining to the app as a whole are shown in Table 3. With regard to using the fiducial marker, the majority of the adolescents (32/41, 78%) agreed that the fiducial marker was easy to use, carry around, and include in the images of their eating occasions. Some of the participants had their own mobile phones, and therefore, carrying the iPhone was an additional mobile phone to carry. However, there was a strong agreement that carrying 2 mobile phones was easy. Nearly all of the boys and girls agreed that understanding the purpose of the TADA app running on the iPhone would have been helpful. When asked "How long would you be willing to record what you eat using the TADA iPhone app?," 39% (16/41) wrote a value between 45 and 90 days, 32% (13/41) wrote 30 days, 10% (4/41) wrote a value between 10 and 15 days, 12% (5/41) wrote 7 days, and 7% (3/41) wrote 0-1 day. The most common response (30 of 40 responses) to the question "List what you liked most about using the TADA iPhone" was entertainment oriented (eg, games, Internet, taking pictures). Among those individuals classified as demonstrating high willingness to take images, all listed "games" in response to this question.

Table 3. Responses to statements in usability questionnaire among adolescents after using the Technology Assisted Dietary Assessment iPhone (n=40).

Statement, as presented	Agreed by boys[a] n (%)	Agreed by girls[b] n (%)
I found the fiducial marker easy to		
Use	21 (91)	11 (65)
Carry around	20 (87)	11 (65)
Include in the picture of my meals	20 (87)	12 (71)
Understanding the purpose of the TADA iPhone app would have been helpful	20 (87)	17 (100)
Carrying 2 telephones is easy (n=31[c])	16 (84)	12 (100)

[a]Boys (n) = 23

[b]Girls (n) = 17

[c]Refers to carrying TADA iPhone and the personal mobile phone. The numbers of boys and girls who were carrying both were 19 and 12, respectively.

Discussion

Principal Findings

This study assessed adolescents' willingness to take images of eating occasions using a mobile phone and observed their interactions with completing the beta version of the review process component of the app. A notable strength of this study was an objective marker for eating occasions. Just under half of the participants were moderately willing to capture images of their eating occasions indicating that some type of reminder system would be important for capturing total diet record. Girls appeared to be more willing to take images than boys, suggesting that more effort is needed to make this type of activity more compelling to boys. During the focus group sessions, the participants volunteered the importance of incentives and reminders to assist them in remembering to take images of every eating occasion. Whereas only basic instruction was provided about how to take the images, more extensive training may clarify the importance of taking images and might enhance willingness and cooperation [26]. Once the participants mastered the steps of the review process component, the majority were observed to quickly tap the labels on the screen much like approaching a game. The complaint of the overlapping pins encountered during the beta testing of the review process was addressed by adding a feature to enlarge the screen. To prevent accidental removal of a pin, a confirmation step was added.

Adolescents especially liked the game apps loaded on the iPhone. All of the individuals objectively classified as demonstrating high willingness also wrote "games" on the questionnaire as the item they liked most about the telephone. This would suggest that games may enhance the level of interaction with the telephone, which then transfers to remembering to take images of foods and beverages at eating occasions. Providing a mobile phone with preloaded games and apps or offering a selection of games or apps may provide additional motivation for cooperation. A study by Baranowski and colleagues [28] suggests that games themselves are a promising feature for enhancing behavior change in children. Overall, these results would support modifying the TADA app to incorporate aspects of gaming, which in turn may benefit usability and cooperation.

During the focus group sessions, both boys and girls expressed a desire to know why taking images of what they eat would be useful or important. Based on the questionnaire responses, nearly all participants agreed that understanding the purpose of the mFR would be helpful. Thus, full transparency would be important in gaining commitment and cooperation from adolescents. In this study, the amount of information initially shared with the participants was limited to determine salient issues. Thus, even with modern technology, the need for full disclosure remains important to adolescents.

Adherence was most influenced by the type of eating occasion (eg, snacks being the worst). Compliance with the dinner meals was substantially lower than the breakfast and lunches. The lower compliance with the dinner meals can partially be explained by the 1-dinner meal that was a picnic on the grass during a minor league baseball game. This particular meal was the most noncompliant meal and reinforces the importance of incorporating reminders to engage an individual when such occasions occur. The minimal training provided in this study was likely inadequate and highlights the need to develop appealing training to aid in the aforementioned situations [26].

Although carrying and using the fiducial marker and carrying 2 telephones intuitively seem burdensome, the majority of the participants agreed that dealing with the fiducial maker and 2 mobile phones was easy. Boys appeared to be more positive than girls with regard to using and carrying the fiducial marker. By contrast, the girls were more likely to state that the protective covers on the telephones were attractive, and if they also included space for the fiducial marker this may be additionally useful. Such a feature could promote a more positive preference among girls regarding the use of a fiducial marker. However, to improve the number of images including a fiducial marker, participants suggested incorporating an automated system into the app to alert the user when the fiducial marker was not visible. In response to this, the opening screen of mFR TADA app has been modified to include a cartoon of the fiducial marker as a reminder [29]. In addition, the app has been modified to detect the presence of the fiducial marker and improve clarity of the image [16,30]. Further, the user settings were modified to allow users to select favorite background colors, a suggestion raised in the focus group sessions.

Limitations

A limitation of this study was the hand recording of the think-aloud method, as well as questions and discussions during the focus group sessions being interpreted through the understanding of the recorders and not amenable to independent reanalysis by other coders. Another limitation was the time of year. The summer is an active period for adolescents and this could influence their ability to remember to take images of their eating occasions. During the school year, there may be other barriers, such as restriction of mobile devices in the school environment. By contrast, daily activities during the school year may be more routine than in an active camp environment. Nonetheless, an exploration of challenges with the community dwelling situation and the school environment is in order. Some additional next steps would include comparing results with biomarkers and determining the lowest age at which children are able to capture images either by themselves or with the aid of parents or other adults. Surrogate parent and care-provider reporting was found to be promising with few limitations [31]. Nonetheless, allowing young children to capture images of their eating occasions may ameliorate identified shortcomings.

Conclusions

The mFR was accepted by the majority of adolescents in the study, but varied according to gender and eating occasion. The adolescents' experience of using this novel method of technology to assess diet and their feedback highlighted the importance of reminders, games, and more training and practice in using the TADA app. Addressing these recommendations may assist adolescents with measuring their dietary intake as part of their active lifestyles.

Acknowledgments

We wish to thank Professor Connie Weaver and Dr Berdine Martin for facilitating access to "Camp Calcium." We also thank the study participants. This study was supported by grants from the National Cancer Institute (Grant No 1U01CA130784-01) and the National Institute of Diabetes, Digestive, and Kidney Disorders (Grant No 1R01-DK073711-01A1).

Authors' Contributions

The study was conceived and designed by CJB, DAK, TES, and EJD. Data were acquired by CJB, DAK, TES, SP, MB, and ZA. Data were analyzed and interpreted by CJB, DAK, TES, TA, MB, and ZA. The paper was drafted by CJB, TES, SP, and AJH. The paper was revised for intellectual content by CJB, DAK, TES, TA, and AJH. All authors approved the final content of the paper.

Conflicts of Interest

None declared.

References

1. Burrows TL, Martin RJ, Collins CE. A systematic review of the validity of dietary assessment methods in children when compared with the method of doubly labeled water. J Am Diet Assoc 2010 Oct;110(10):1501-1510. [doi: 10.1016/j.jada.2010.07.008] [Medline: 20869489]
2. Bandini LG, Must A, Cyr H, Anderson SE, Spadano JL, Dietz WH. Longitudinal changes in the accuracy of reported energy intake in girls 10-15 y of age. Am J Clin Nutr 2003 Sep;78(3):480-484 [FREE Full text] [Medline: 12936932]
3. Champagne CM, Baker NB, DeLany JP, Harsha DW, Bray GA. Assessment of energy intake underreporting by doubly labeled water and observations on reported nutrient intakes in children. J Am Diet Assoc 1998 Apr;98(4):426-433. [doi: 10.1016/S0002-8223(98)00097-2] [Medline: 9550166]
4. Livingstone MB, Prentice AM, Coward WA, Strain JJ, Black AE, Davies PS, et al. Validation of estimates of energy intake by weighed dietary record and diet history in children and adolescents. Am J Clin Nutr 1992 Jul;56(1):29-35 [FREE Full text] [Medline: 1609757]
5. Boushey CJ, Kerr DA, Wright J, Lutes JD, Ebert DS, Delp EJ. Use of technology in children's dietary assessment. Eur J Clin Nutr 2009 Feb;63(Suppl 1):S50-S57 [FREE Full text] [doi: 10.1038/ejcn.2008.65] [Medline: 19190645]
6. Anderson J, Raine L. Main findings: Teen, technology, and human potential in 2020. Washington, DC: Pew Research Center; 2012 Feb 29. URL: http://www.pewinternet.org/2012/02/29/main-findings-teens-technology-and-human-potential-in-2020/ [accessed 2014-11-27] [WebCite Cache ID 6UO0PC1wl]
7. Purcell K, Buchanan J, Friedrich L. The Impact of Digital Tools on Student Writing and How Writing is Taught in Schools. Washington, DC: Pew Research Center; 2013 Jul 16. URL: http://www.pewinternet.org/2013/07/16/the-impact-of-digital-tools-on-student-writing-and-how-writing-is-taught-in-schools/ [accessed 2014-11-14] [WebCite Cache ID 6U4a1F4rE]
8. Schap TE, Zhu F, Delp EJ, Boushey CJ. Merging dietary assessment with the adolescent lifestyle. J Hum Nutr Diet 2014 Jan;27(Suppl 1):82-88 [FREE Full text] [doi: 10.1111/jhn.12071] [Medline: 23489518]
9. Zhu F, Bosch M, Khanna N, Boushey CJ, Delp EJ. Multilevel segmentation for food classification in dietary assessment. Proc Int Symp Image Signal Process Anal 2011 Sep 04:337-342 [FREE Full text] [Medline: 22127051]

10. He Y, Xu C, Khanna N, Boushey CJ, Delp EJ. Analysis of Food Images: Features and Classification. New York: IEEE; 2014 Presented at: Proceedings of the IEEE International Conference on Image Processing October; Oct 27-30, 2014; Paris, France p. 2744-2748. [doi: 10.1109/ICIP.2014.7025555]
11. Zhu F, Bosch M, Woo I, Kim S, Boushey CJ, Ebert DS, et al. The use of mobile devices in aiding dietary assessment and evaluation. IEEE J Sel Top Signal Process 2010 Aug;4(4):756-766 [FREE Full text] [doi: 10.1109/JSTSP.2010.2051471] [Medline: 20862266]
12. Mariappan A, Bosch M, Zhu F, Boushey CJ, Kerr DA, Ebert DS, et al. Personal dietary assessment using mobile devices. Proc SPIE Int Soc Opt Eng 2009 Jan 1;7246:72460Z [FREE Full text] [doi: 10.1117/12.813556] [Medline: 21660219]
13. Lee CD, Chae J, Schap TE, Kerr DA, Delp EJ, Ebert DS, et al. Comparison of known food weights with image-based portion-size automated estimation and adolescents' self-reported portion size. J Diabetes Sci Technol 2012 Mar;6(2):428-434 [FREE Full text] [Medline: 22538157]
14. Xu C, He Y, Khanna N, Boushey CJ, Delp EJ. Model-based food volume estimation using 3d pose. New York: IEEE; 2013 Presented at: Proceedings of the IEEE International Conference on Image Processing; 2013; Melbourne, Victoria p. 2534-2538. [doi: 10.1109/ICIP.2013.6738522]
15. Chae J, Woo I, Kim S, Maciejewski R, Zhu F, Delp EJ, et al. Volume estimation using food specific shape templates in mobile image-based dietary assessment. Proc SPIE Int Soc Opt Eng 2011 Feb 7;7873:78730K [FREE Full text] [doi: 10.1117/12.876669] [Medline: 22025936]
16. Xu C, Zhu F, Khanna N, Boushey CJ, Delp EJ. Image enhancement and quality measures for dietary assessment using mobile devices. Bellingham, WA: SPIE; 2012 Feb 09 Presented at: Proceedings of the IS&T/SPIE Conference on Computational Imaging X; 9 Feb, 2012; Burlingame, CA. [doi: 10.1117/12.909949]
17. Boushey CJ, Kerr DA, Schap TE, Daugherty BL. Importance of user interaction with automated dietary assessment methods. Eur J Clin Nutr 2012 May;66(5):648. [doi: 10.1038/ejcn.2012.14] [Medline: 22333872]
18. Schap TE, Six BL, Delp EJ, Ebert DS, Kerr DA, Boushey CJ. Adolescents in the United States can identify familiar foods at the time of consumption and when prompted with an image 14 h postprandial, but poorly estimate portions. Public Health Nutr 2011 Jul;14(7):1184-1191 [FREE Full text] [doi: 10.1017/S1368980010003794] [Medline: 21324224]
19. Lenhart A, Ling R, Campbell S, Purcell K. Teens and Mobile Phones. Washington, DC: Pew Research Center; 2010 Apr 20. URL: http://www.pewinternet.org/2010/04/20/teens-and-mobile-phones/ [accessed 2014-11-14] [WebCite Cache ID 6U4c13lW2]
20. Madden M, Lenhart A, Duggan M, Cortesi S, Gasser U. Teens and Technology 2013. Washington, DC: Pew Research Center; 2013 Mar 13. URL: http://www.pewinternet.org/2013/03/13/teens-and-technology-2013/ [accessed 2014-11-14] [WebCite Cache ID 6U4cBs7bo]
21. 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;14(4):e111 [FREE Full text] [doi: 10.2196/jmir.2155] [Medline: 22868871]
22. 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]
23. Palacios C, Martin BR, McCabe GP, McCabe L, Peacock M, Weaver CM. Dietary calcium requirements do not differ between Mexican-American boys and girls. J Nutr 2014 Aug;144(8):1167-1173. [doi: 10.3945/jn.113.188318] [Medline: 24872223]
24. Creswell JW, Fetters MD, Ivankova NV. Designing a mixed methods study in primary care. Ann Fam Med 2004;2(1):7-12 [FREE Full text] [Medline: 15053277]
25. Henry H, Reicks M, Smith C, Reimer K, Atwell J, Thomas R. Identification of factors affecting purchasing and preparation of fruit and vegetables by stage of change for low-income African American mothers using the think-aloud method. J Am Diet Assoc 2003 Dec;103(12):1643-1646. [doi: 10.1016/j.jada.2003.09.039] [Medline: 14647093]
26. Six BL, Schap TE, Zhu FM, Mariappan A, Bosch M, Delp EJ, et al. Evidence-based development of a mobile telephone food record. J Am Diet Assoc 2010 Jan;110(1):74-79 [FREE Full text] [doi: 10.1016/j.jada.2009.10.010] [Medline: 20102830]
27. Daugherty BL, Schap TE, Ettienne-Gittens R, Zhu FM, Bosch M, Delp EJ, et al. Novel technologies for assessing dietary intake: Evaluating the usability of a mobile telephone food record among adults and adolescents. J Med Internet Res 2012;14(2):e58 [FREE Full text] [doi: 10.2196/jmir.1967] [Medline: 22504018]
28. Baranowski T, Baranowski J, Thompson D, Buday R, Jago R, Griffith MJ, et al. Video game play, child diet, and physical activity behavior change a randomized clinical trial. Am J Prev Med 2011 Jan;40(1):33-38 [FREE Full text] [doi: 10.1016/j.amepre.2010.09.029] [Medline: 21146765]
29. Kerr DA, Pollard CM, Howat P, Delp EJ, Pickering M, Kerr KR, et al. Connecting health and technology (CHAT): Protocol of a randomized controlled trial to improve nutrition behaviours using mobile devices and tailored text messaging in young adults. BMC Public Health 2012;12:477 [FREE Full text] [doi: 10.1186/1471-2458-12-477] [Medline: 22726532]
30. Xu C, Khanna N, Boushey CJ, Delp EJ. Low complexity image quality measures for dietary assessment using mobile devices. New York: IEEE; 2011 Presented at: 2011 IEEE International Symposium on Multimedia; 5-7 Dec, 2011; Dana Point, CA p. 351-356. [doi: 10.1109/ISM.2011.64]

31. Nicklas TA, O'Neil CE, Stuff J, Goodell LS, Liu Y, Martin CK. Validity and feasibility of a digital diet estimation method for use with preschool children: A pilot study. J Nutr Educ Behav 2012;44(6):618-623 [FREE Full text] [doi: 10.1016/j.jneb.2011.12.001] [Medline: 22727939]

Abbreviations

mFR: mobile food record
TADA: Technology Assisted Dietary Assessment

Edited by G Eysenbach; submitted 01.12.14; peer-reviewed by L Harnack, T Baranowski; comments to author 25.12.14; revised version received 26.02.15; accepted 16.03.15; published 29.05.15.

Please cite as:
Boushey CJ, Harray AJ, Kerr DA, Schap TE, Paterson S, Aflague T, Bosch Ruiz M, Ahmad Z, Delp EJ
How Willing Are Adolescents to Record Their Dietary Intake? The Mobile Food Record
JMIR mHealth uHealth 2015;3(2):e47
URL: http://mhealth.jmir.org/2015/2/e47/
doi:10.2196/mhealth.4087
PMID:26024996

Original Paper

Targeting Parents for Childhood Weight Management: Development of a Theory-Driven and User-Centered Healthy Eating App

Kristina Elizabeth Curtis[1], BSc (Psych)(Hons), MSc; Sudakshina Lahiri[1], BA, MSW, PhD; Katherine Elizabeth Brown[2], BSc (Psych)(Hons), MSc, PhD (Psych)

[1]The Institute of Digital Healthcare, Warwick Manufacturing Group, The University of Warwick, Coventry, United Kingdom

[2]Centre for Technology Enabled Health Research (CTEHR), Coventry University, Coventry, United Kingdom

Corresponding Author:
Kristina Elizabeth Curtis, BSc (Psych)(Hons), MSc
The Institute of Digital Healthcare
Warwick Manufacturing Group
The University of Warwick
Coventry
Coventry, CV4 7AL
United Kingdom
Phone: 44 24 7615 1632
Fax: 44 24 7615 1632
Email: k.e.curtis@warwick.ac.uk

Abstract

Background: The proliferation of health promotion apps along with mobile phones' array of features supporting health behavior change offers a new and innovative approach to childhood weight management. However, despite the critical role parents play in children's weight related behaviors, few industry-led apps aimed at childhood weight management target parents. Furthermore, industry-led apps have been shown to lack a basis in behavior change theory and evidence. Equally important remains the issue of how to maximize users' engagement with mobile health (mHealth) interventions where there is growing consensus that inputs from the commercial app industry and the target population should be an integral part of the development process.

Objective: The aim of this study is to systematically design and develop a theory and evidence-driven, user-centered healthy eating app targeting parents for childhood weight management, and clearly document this for the research and app development community.

Methods: The Behavior Change Wheel (BCW) framework, a theoretically-based approach for intervention development, along with a user-centered design (UCD) philosophy and collaboration with the commercial app industry, guided the development process. Current evidence, along with a series of 9 focus groups (total of 46 participants) comprised of family weight management case workers, parents with overweight and healthy weight children aged 5-11 years, and consultation with experts, provided data to inform the app development. Thematic analysis of focus groups helped to extract information related to relevant theoretical, user-centered, and technological components to underpin the design and development of the app.

Results: Inputs from parents and experts working in the area of childhood weight management helped to identify the main target behavior: to help parents provide appropriate food portion sizes for their children. To achieve this target behavior, the behavioral diagnosis revealed the need for eliciting change in parents' capability, motivation, and opportunity in 10-associated Theoretical Domains Framework (TDF) domains. Of the 9 possible intervention functions, 6 were selected to bring about this change which guided the selection of 21 behavior change techniques. Parents' preferences for healthy eating app features revolved around four main themes (app features, time saving and convenience, aesthetics, and gamification) whereupon a criterion was applied to guide the selection on which preferences should be integrated into the design of the app. Collaboration with the app company helped to build on users' preferences for elements of gamification such as points, quizzes, and levels to optimize user engagement. Feedback from parents on interactive mock-ups helped to inform the final development of the prototype app.

Conclusions: Here, we fully explicate a systematic approach applied in the development of a family-oriented, healthy eating health promotion app grounded in theory and evidence, and balanced with users' preferences to help maximize its engagement with the target population.

(JMIR mHealth uHealth 2015;3(2):e69) doi:10.2196/mhealth.3857

KEYWORDS

child; obesity; health behaviour; mHealth; healthy eating; evidenc-based; theory

Introduction

Background

Within the field of mobile health (mHealth), seen as mobile devices such as mobile phones, personal digital assistants (PDAs), and other wireless devices supporting a medical or public health practice [1], it is the advent of the mobile phone accompanied by an explosion of commercial mHealth apps that has gained the most attention [2]. Health promotion apps are by far the most commonly downloaded mHealth apps [3] and aim to support users to start or reinforce one or more health behaviors (eg, nutrition apps) and/or reduce risk behaviors (eg, smoking cessation apps) [4]. To date, nutrition and diet apps represent the fastest growing area of health promotion apps [2]. It is now well documented that mobile phones offer a number of attributes that maximize their potential for supporting health behavior change interventions including their accessibility (eg, global proliferation, widespread adoption across socioeconomic and demographic populations, and ubiquity) [2], personal nature (eg, always on the person, emotional attachment, and connectivity) [5], and programming flexibility (eg, information tailoring, context aware capabilities, and automated sensors) [6,7]. Additionally, mobile phones offer benefits for researchers regarding implementation (eg, low cost, rapid scalability, ease of use, zero-geography, and low participant burden) and real-time monitoring, data collection, and analysis [1,8].

The Potential for Mobile Health Apps in Childhood Weight Management

Mobile health (mHealth) tools are also particularly suitable when it comes to supporting parental involvement in childhood weight management interventions, where there is growing consensus among researchers and practitioners that novel approaches using the internet [9-11] and mHealth apps [8,12] should be explored. For example, their zero-geography feature means that access to apps is not restricted to locations and can be delivered directly to families in the comfort and privacy of their own home [8]. This is especially advantageous for a parent population that report lack of time, scheduling conflicts, and location difficulties as major barriers to attending childhood weight management programs [13]. Another benefit of mHealth apps is their 'glanceable displays' that can provide parents with a quick and coherent overview of their child's health information, potentially increasing their engagement with children's weight-related behaviors [5]. In addition, participants can continue to access an intervention long after completion, which is important in weight management where there are high rates of relapse [8].

Behavior change techniques (BCTs) are seen as the observable, replicable, and active ingredients in an intervention that directly bring about behavior change [14]. Certain BCTs, such as self-monitoring which have been shown to be effective for adult and childhood weight management [15-17], are optimized through this medium and continue to increase in their sophistication [7]. For example, in cases where parents report difficulties in monitoring children's dietary behaviors [18], mobile phone features such as cameras can be employed for children's dietary monitoring where parents and children can take pictures of their food [19]. This has been shown to be especially effective for helping users monitor and reflect on their eating and exercise behaviors [20]. Moreover, mHealth apps may offer a more detailed and accurate measure of dietary behaviors thus increasing the robustness of childhood weight management interventions where current studies are limited by self-report measures [7]. Additional techniques, such as role modeling behaviors, can also be effectively implemented through the use of games and health challenges that families can play together, allowing parents' behaviors to influence their children's behaviors [8].

Approaches to Mobile Health Development

With regards to mHealth app development, there is growing consensus that mHealth interventions should be based on evidence, behavior change theory, and formative research with the target audience [21]. Despite this, several reviews of commercial health promotion apps have revealed a significant lack of evidence-based guidelines [22-24] and health behavior change theory [25,26] in their development processes. With regards to childhood weight management, results from a recent review involving 57 pediatric weight management apps indicated that an overwhelming majority of the apps (61%) did not use any recommended strategies or behavioral targets. Moreover, few apps targeted parents/families [12], a vital element when managing children's weight [9,10,27,28]. However, evidence implies that mHealth apps with more evidence-based strategies are least popular amongst consumers [29]; suggesting commercial mHealth apps may be more engaging for consumers, despite their lack of theoretical content. Arguably, mHealth development would benefit from greater collaboration between experts in behavior change and the commercial app industry to help address these gaps [4,26].

In addition to theory, evidence, and engaging design principles, mHealth interventions should also have social validity with regards to acceptability amongst its stakeholders [30]. Consequently, there is a growing trend towards adopting a user-centered design (UCD) approach [31-33]. This is especially pertinent in the case of apps where approximately 26% of all apps downloaded are discarded after first use [34].

Theoretical Framework Guiding the Study

While theories and models of behavior change (eg, theory of planned behavior and transtheoretical model [35,36]) help guide intervention designers on which theoretical constructs to target in an intervention to elicit behavioral change, intervention development frameworks (eg, intervention mapping and/or Medical Research Council framework [37,38]) provide guidance on the development of a "coordinated set of activities" to help

translate theory into practice [39]. However, the majority of the prominent theories and models of behavior change fail to take into account the context in which a behavior occurs, fail to focus on reflective processes (eg, attitudes and intentions), are static in structure, and are unable to explicitly state how to bring about change [40]. As well, existing intervention development frameworks have been criticized for their lack of coherence, comprehensiveness (in terms of not offering the full range of intervention functions to change behavior), and grounding in a model of behavior change [39]. Therefore, this study applied a new framework, the *Behavior Change Wheel* (BCW) framework [14], underpinned by a new model of behavior change, the capability opportunity motivation behavior (COM-B) model [39], designed to incorporate existing theories of behavior change. The BCW incorporates a full range of intervention functions such as education, persuasion, and training that are likely to be effective in eliciting change in a specific target behavior. These intervention functions can be delineated into behavior change techniques (BCTs) using the BCT taxonomy: BCTT (V1) [41], which provides an extensive list of evidence-based BCTs.

The COM-B model defines behavior as part of a system where the three following psychological domains interact to enable a behavior to occur (1) capability (psychological and/or physical; eg, knowledge and skills), (2) motivation (reflective and/or automatic; eg, self-efficacy and emotion), and (3) opportunity (physical and/or social; eg environmental resources and social influences). The model helps to identify which components need to change in order for the target behavior to occur, thus supporting the design of behavior change interventions [40].

Within the BCW, the COM-B model can be further elaborated using the Theoretical Domains Framework (TDF), comprised of 14 theoretical domains drawn from a synthesis of 33 psychological theories and 128 key theoretical constructs relevant for behavior change [42].

Despite the major push to harness mobile phone features that support health behavior change, precisely how to develop theory-informed mHealth interventions that engage users remains a challenge and is rarely well documented in the literature. Therefore, this study provides a detailed outline of how the BCW has been applied in practice for the development of a theory and evidence-driven health promotion app within the context of childhood weight management, whilst also ensuring social validity and engagement amongst the target population.

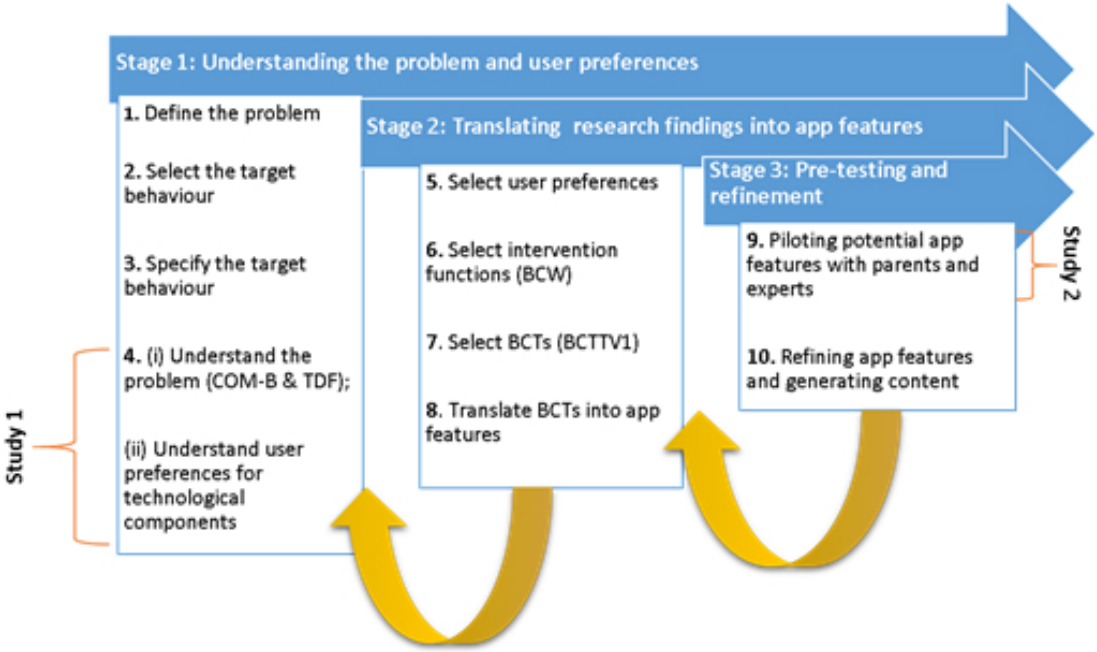
Methods

Overview

The mHealth app intervention development process followed these three stages (1) understanding the problem and user preferences, (2) translating research findings into app features, and (3) pre-testing the app features for further refinement (Figure 1).

Two empirical studies were conducted at stage 1 and stage 3 of the intervention development process. An iterative feedback loop approach was followed, wherein findings from each stage fed into the next stage of development and were also fed back to refine previous stages.

Figure 1. mHealth app intervention development process.



Stage 1: Understanding the Problem and User Preferences

Overview

The first stage comprised of 4 steps involved in defining the public health problem through to formative research, first with case workers and then parents, on the theoretical, user-centered, and technological components that should be considered in mHealth intervention development.

Step 1: Defining the Problem

Step 1 entailed defining the overall health problem in behavioral terms, taking the specific context into account. Hence, an extensive review of the evidence helped to make decisions on whether to focus on children's eating or exercise behaviors. Based upon this decision, all individuals, groups, and populations potentially contributing to this behavior were considered.

Step 2: Selecting the Target Behavior

The first activity at step 2 required a consideration of all the specific behaviors to potentially target in the intervention before narrowing these down to one or two. The BCW recommends a "less is more" approach whereby it is beneficial to start with small changes and build upon these incrementally [14]. Furthermore, discussions with the family weight management services commissioner indicated that the app is one element in a whole range of activities offered as part of the local weight management services. Therefore, it was not necessary to incorporate all possible weight related behaviors into the intervention.

Selecting the target behavior involved conducting empirical research with the target population (refer to step 4), along with consultations with a pediatrician, dietician, and two public health experts in the area of childhood weight management. Hence, focus groups with parents helped to identify the target behavior as well as explore the barriers and facilitators to parents' capability, opportunity, and motivation for enacting this behavior.

Step 3: Specify the Target Behavior

Upon selection of the target behavior, this step involved specifying the behavior and the context in which it occurs (eg, in the supermarket or at home).

Step 4: Understanding the Target Behavior and User Preferences

Step 4 involved conducting the first empirical study using a qualitative research design so that both the BCW and UCD methodologies could be simultaneously applied to app development. The former helped to explore barriers and facilitators to parents' capability, opportunity, and motivation in enacting the target behavior; whereas the latter helped to explore parents' preferences for app features. The qualitative research involved conducting 6 focus groups, one with family weight management case workers and five with parents.

An intervention mapping table produced from the behavioral diagnosis conducted in the first part of this step served as the basis for mapping theoretical components to app features. The table was reviewed by two health psychologists familiar with the BCW to ensure that the COM-B and TDF theoretical tools had been appopriately applied to the data.

Data and Sampling (Study 1)

An initial stakeholder meeting with the local public health family weight management commissioning team led to participant recruitment using a purposive sampling strategy. Emails were sent to case workers who were eligible to participate if they had been working with families with overweight children. Parents with overweight and very overweight children were recruited with the help of managers from two local weight management programs. Additionally, parents with healthy weight children were recruited via the university. Parents were eligible if they had a child of ≥5 years and owned a mobile phone.

A total of 48 participants were eligible to take part in this study. Of these, 5 were case workers and the remaining participants were parents. Among the case workers, 4 participated in the study. Of the 43 parents contacted, 22 agreed to participate, yielding a response rate of 51% (22/43) for this group. A total of 6 focus groups were then conducted; 1 with local case workers (4 participants), 4 with parents of overweight and very overweight children (3-4 participants), and 1 with parents of healthy weight children (8 participants). The parent sample comprised predominantly mothers (82%, 18/22), compared to fathers (18%, 4/22). With regards to mobile phone ownership, 77% (17/22) of the sample reported owning a mobile phone. Of those, 41% (7/17) were Android users, 29% (5/17) iPhone users, 18% (3/17) Blackberry users, and 12% (2/17) Windows users. Participants (n=15) recruited from the weight management program had children classified as very overweight (53%, 8/15), overweight (40%, 6/15), and healthy weight (7%, 1/15).

Focus groups were conducted using semi-structured questions developed from a review of existing evidence and structured around the COM-B model and TDF to explore barriers and facilitators to parents' capability, opportunity, and motivation to provide appropriate food portions for their children (see Textbox 1 for schedule of topics). Additional topics were explored with case workers to gain a deeper understanding of the context of childhood overweight. Topics also revolved around parents' preferences for healthy eating app features, representing the formative stage of the UCD approach. With the permission of participants, the focus groups were audio recorded and transcribed verbatim.

Transcripts were analyzed by two independent researchers using established principles for conducting thematic analysis [43]. This involved coding segments of data for their basic meaning and mapping these to the COM-B and TDF, as well as coding extracts that referred to users' preferences for healthy eating app features. An analysis of the data helped to identify which theoretical domains needed to change in order for the target behavior to occur. The BCW refers to this process as the behavioral diagnosis [14].

Textbox 1. Schedule of topics explored with case workers (topics 1-14) and parents (topics 6-14) for study 1.

Topic
1. Parent's recognition of children's overweight status.
2. Mothers' and fathers' roles in child feeding.
3. Parents' weight status.
4. Barriers to attending family weight management programs.
5. Issues that parents ask for help with.
6. Parents' knowledge of healthful foods and age appropriate portion sizes.
7. Parents' monitoring of children's eating habits.
8. Parents' interpersonal skills around healthy eating and weight issues.
9. Parents' beliefs about consequences of childhood overweight.
10. Parents' beliefs about capabilities of changing children's dietary habits.
11. Parents' emotions (limiting food, talking to children about weight, stress).
12. Other people in parents' environment (other people that make it difficult for parents to provide appropriate food portions).
13. Parents' use of existing technology (websites and apps).
14. Parents' preferences for app features.

Stage 2: Translating Research Findings Into App Features

Overview

Stage 2 was comprised of 3 steps including the selection of user preferences, intervention functions, and behavior change techniques. Collaboration with the commercial app company assisted in operationalising these components into app features.

Step 5: Select User Preferences

To help balance the theoretical findings with user preferences, consideration of whether to "reject" or "accept" each user preference was guided using the following criteria (1) relevance to the target behavior, (2) availability online, (3) ease of implementation, (4) alignment with usability and user experience recommendations, and (5) supported from theoretical findings and/or evidence. Consultations with the app company provided insights on the feasibility of user preferences with regards to their implementation (ie, development time and cost).

Step 6: Select Intervention Functions

Based on the results of the behavioral diagnosis conducted in step 4, the BCW guided designers on which types of interventions are likely to bring about change in these COM-B components and associated TDF domains. The 9 intervention functions to select from are education, persuasion, incentivisation, coercion, training, restriction, environmental restructuring, modeling, and enablement [14]. The selected intervention functions were then mapped to the intervention mapping table generated from the previous stage.

Step 7: Select Behavior Change Techniques

A behavior change technique (BCT) refers to an "active ingredient" and mechanism of change that is an observable, replicable, and irreducible component of a behavior change intervention [44]. Hence, the next step involved delineating intervention functions into BCTs whereupon a candidate list of BCTs was derived from the BCW, linking intervention functions with relevant BCTs [14]. A review of evidence on effective techniques for childhood weight management interventions allowed selection of potentially effective BCTs for use in the intervention such as goal setting [45-47], self-monitoring of behavior [45], and instruction on how to perform the behavior [48]. These were then mapped to the intervention mapping table and reviewed by two health psychologists for further verification.

Step 8: Translate Behavior Change Techniques Into App Features

Upon identification of BCTs, steps were taken to embed these as potential app features informed by the user preferences data retrieved from step 4. This involved liaising with the app company with regards to how BCTs could be implemented in the app. With respect to enhancing user experience and engagement, parents' preferences for app features were built on in consultations with the app company where elements of gamification, defined as the use of game design elements in non-game contexts [49], were applied. For example, progress bars, achievement badges, and points were identified as a way of providing feedback for parents on their children's eating behavior. In addition, consultations with a software engineer also helped to develop a flow chart of the user journey where BCTs and gamification techniques were linked to specific app features.

Stage 3: Pre-Testing and Refinement (Study 2)

Step 9: Piloting Potential App Features

This step encompassed the second empirical study and involved piloting the proposed features of the app, and seeking feedback for further refinement of app features. Focus groups (n=3) were conducted using similar recruitment strategies followed for study 1. A total of 21 parents were contacted to take part in this

study of which 20 took part in the pre-testing phase, yielding a response rate of 95% (20/21). Of the focus groups, 2 were recruited from local weight management programs (7 and 8 participants), consisting of mothers (87%, 13/15) and fathers (13%, 2/15). The third focus group recruited from the university (5 participants) consisted of mothers (60%, 3/5) and fathers (40%, 2/5). Of the participants, three quarters (75%, 15/20) reported owning a mobile phone. Of those, 53% (8/15) owned an Android, 27% (4/15) an iPhone, 13% (2/15) a Blackberry, and 7% (1/15) a Windows device. This study focused on parents' overall impressions of the app using interactive mock-ups. Focus group discussions were conducted using semi-structured, open-ended questions based on a schedule of topics presented in Textbox 2. The questions drew on a model of usability and user experience goals [50].

Textbox 2. Schedule of topics for study 2.

Topic
1. Overall impressions of the app
2. Overall helpfulness of app and app features
3. Whether it is perceived as fun and enjoyable
4. Whether it is perceived as satisfying (any features that are liked or not liked)
5. Whether it is perceived as entertaining (mainly referred to content in the quiz feature)
6. Whether certain features are perceived as motivating or not

Step 10: Refining App Features, Generating Content, and Developing the Prototype

The final stage involved refining app features based on feedback from parents and the dietary steering board, generating content, and development of the prototype app. The intervention mapping table was completed at this step, where both user preferences and final app features were mapped to relevant theoretical components. Text was used as a mode of delivery for several BCTs in the intervention in the form of two app features: within app text notifications (delivering tips and persuasive messages) and an interactive quiz.

The generation of text for the notifications required a review of the empirical evidence to guide the process of message framing which refers to whether health messages provide benefits of carrying out a behavior (gain-frame) or the consequences of not carrying out the behavior (loss-frame). Generally, gain-framed messages are shown to be more effective for preventative health behaviors, therefore, this framing was used to guide the persuasive messages, tips, and quiz questions in the app [51].

A laptop, tablet, and projector were used to present the interactive mock-up. The content of the interactive mock-up was refined after feedback from the first 2 focus groups and an updated version was presented to the last focus group. A thematic analysis similar to the process followed in study 1 was conducted on the data. The salient information from the analysis was then extracted and shared with the app company to help make further iterations to the app's functional specification.

Lastly, a dietary steering board comprised of two public health dieticians and a family weight management program manager was convened to provide support and feedback on the nutritional content of the app.

Results

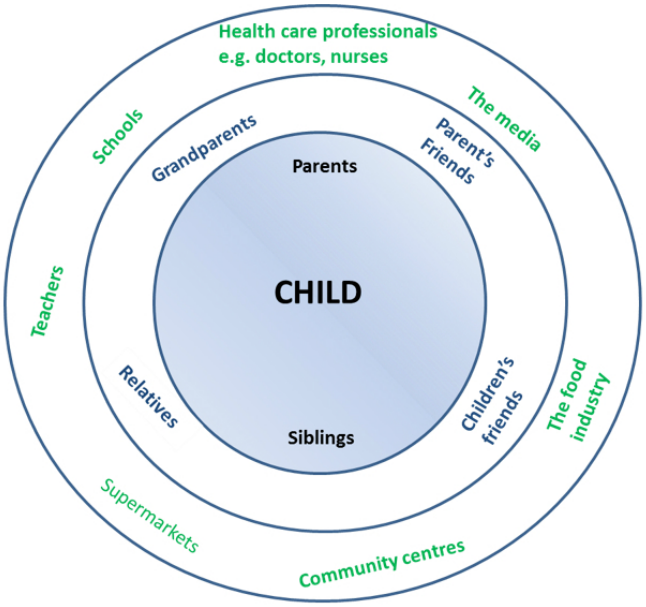
Stage 1: Understanding the Problem

Step 1: Define the Problem

Childhood overweight is a serious public health problem. In England, a third of 10-11 year olds and over a fifth of 4-5 year olds are reported as either overweight or obese [52]. A review of the evidence regarding the subsequent determinants of the energy balance equation provided greater support for focusing on improving children's diets with regards to reducing their overall energy intake [53-58]. Simultaneously, stakeholder meetings also identified a shortage of online resources in the local area targeting parents to help them improve their family's diets. Consideration of all the individuals, groups, populations, and sectors potentially contributing to children's energy intake are shown in Figure 2.

Evidence also highlighted the role of parents in children's energy intake and strongly supported their involvement in childhood weight management interventions [9,10,27,28], including direct involvement in the intervention development process [9]. Hence, this led to the decision to focus on parents as the main population to target in the intervention.

Figure 2. Relevant individuals, groups and populations involved in children's energy intake.



Step 2: Select the Target Behavior

Based on the decision to focus on reducing children's overall energy intake, a range of nutrition behaviors relating to achieving this overall behavior were considered including increasing intake of fruit and vegetables, and reducing intake of saturated fat, sugar, unhealthy snacks, portion sizes, and some carbohydrates. Next, focus groups with parents of overweight children helped to narrow down this list where the problem of children's consumption of large portion sizes was highlighted as an important behavior to target in the intervention, and one that appeared to be acceptable for parents in changing. Discussions with experts helped to confirm the decision to focus on supporting parents in providing appropriate portion sizes for their children.

Step 3: Specify the Target Behavior

Upon selecting the target population, behavior, and setting, the behavior was specified with regards to what needs to occur in order for the target behavior to be carried out (Textbox 3).

Textbox 3. How to specify the target behavior (adapted from [14]).

Specifications and details

- Target behavior
  - Parents providing appropriate portion sizes (and frequency of food) for their children across the five food groups: fruit and vegetables, protein, dairy, starchy foods, food, and drinks high in sugar and fat (as per the eatwell plate)
- Who needs to perform the behavior?
  - Parents
- What do they need to do differently to achieve the desired change?
  - Preparation, provision of portions (age appropriate portion sizes), and monitoring of food portions
- When do they need to do it?
  - At meal times and snack times
- Where do they need to do it?
  - At home
- How often do they need to do it?
  - Everyday
- With whom do they need to do it?
  - At home
- In what context do they need to do it?
  - The home environment

Step 4: Understanding the Target Behavoiur and User Preferences

Theoretical Analysis

The behavioral diagnosis revealed barriers to the target behavior (ie, parents providing appropriate portion sizes for their children) in all 3 COM-B components and 10 out of the 14 TDF domains. A full table of the results from the behavioral diagnosis supported with quotes from focus group participants is presented in Multimedia Appendix 1. This information laid the foundations for the beginning of an intervention mapping table used to link each of the components (eg, theory-user preferences in app features) in intervention development.

User Preferences for Healthy Eating App Features

Data collected in relation to parents' ideas and preferences for healthy eating app features formed the first stage of the UCD approach (an excerpt of the results is presented in Table 1). The following 4 key themes emerged from the analysis (1) parents' preferences for app features (eg, parents preference for the output of recipes from data input on household ingredients), (2) time saving and convenience (eg, parents specified that the app should be simple and quick to use), (3) aesthetics (eg, parents preference for visuals of food), and (4) gamification (eg parents preference for a point system for healthy eating behavior).

Table 1. Excerpt of parents preferences for app features.

Themes	Sub-themes	Quotes
App feature	Recipes of household ingredients	I would love to have an app on my phone that says, I have this food what can I do with it… (Parent, focus group 2)
Time saving and convenience	Simple to use	It would have to be quite simple I think. If it got too complicated I just wouldn't use it. (Parent, focus group 5)
Aesthetics	Visual aids for portion sizes	Pictures were given in the group of the portion sizes... we try to visualize that on the plate so we get roughly the amounts right…I think that would help (Parent, focus group 3)
Gamification	Points for healthy eating	Or you could have something that you could add, what have you had today? Yes I have had one of those, one of those right you get 50 points but I've also had one of those, deduct 20 points (Parent, focus group 5)

Stage 2: Translating Research Findings Into App Features

Step 5: Select User Preferences

Focus group discussions resulted in a total of 19 user preferences, wherein 3 were rejected, 5 were partly accepted and 11 were accepted. An excerpt of the user preferences along with reasons guiding decision making (using one or more of the criteria) is shown in Table 2.

Table 2. Excerpt of decisions for rejecting or accepting user preferences.

Main theme	User preference	Accept /reject	Reason
App features			
	Every family is different so they need to be able to choose their own goals	Accept	(v) Supported by literature (+)[a]
	Recipe output of household ingredients	Reject	(i) Aligned with target behavior (-)[b]; (ii) already apps and websites that provide this (-); (iii) not within budget (-)
Usability			
	Needs to be minimal data input	Accept	(iv) Aligned with recommended usability goals (+); (v) parents lack of time was identified as an important barrier to make changes (+)
Aesthetics			
	Visuals of food in the app	Accept	(iv) Aligned with recommended user experience goals (+)

[a](+) Aligned with criterion

[b](-) Not aligned with criterion

Step 6: Select Intervention Functions

A total of 6 out of the 9 possible intervention functions were selected: education, training, persuasion, environmental restructuring, enablement, and modeling (parent/child). Table 3 shows an excerpt of how these intervention functions were mapped to the corresponding COM-B and TDF components, along with examples of how they could be applied to supporting parents' portion control behaviors with their children.

Table 3. Excerpt of mapping intervention functions to COM-B and Theoretical Domains Framework (TDF) components.

COM-B	Theoretical Domains Framework (TDF)	Sub-themes	Intervention funtions	Example
Psychological capability	Skills (cognitive)	Parents have difficulty in measuring food portions	Training, Environmental restructuring	Train parents to measure portion sizes, provide a visual tool to help measure food
Reflective motivation	Beliefs about capabilities	Parents have a lack of confidence in their ability to make changes	Education, persuasion, Enablement	Educate, persuade and enable parents to increase their self-confidence in making changes to their children's eating habits.
Physical opportunity	Environmental context and resources	Parents' preferences for household objects such as plates to measure portion sizes	Environmental restructuring	Restructure the home environment to provide a tool for greater accuracy in measuring food portions

Step 7: Select Behavior Change Techniques

A total of 21 BCTs were selected at this step and mapped onto the intervention functions, TDF, and COM-B components as shown in Table 4.

Step 8: Translate Behavior Change Techniques Into App Features

Consulting with the app company facilitated the process of how BCTs identified in step 7 could be meaningfully combined with findings on user preferences (step 4) to create app features. Table 5 provides an example of this process where each BCT is mapped to each user preference and proposed app feature.

Further elements of gamification techniques, shown in Textbox 4, were recommended by the app company to increase parents' motivation in completing tasks, such as logging food and answering quiz questions. These also related to specific BCTs.

Consultations with a software engineer led to the development of a flow chart (Figure 3) to help map specific BCTs to app features, showing the sequence of intervention components delivered to parents. The diagram was further refined through parental feedback in the next stage of development.

Additionally, interactive mock-ups of the app were developed by the app company and used to pilot the proposed features with parents. An example of the home screen is shown in Figure 4.

Table 4. Mapping behavioral change techniques (BCTs) to intervention functions.

COM-B	Theoretical Domains Framework (TDF)	Intervention funtions	Behavioral change techniques (BCTs)
Psychological capability			
	Knowledge	Education	Instruction on how to perform the behavior, habit formation
	Memory, attention, and decision making skills	Training	Instruction on how to perform the behavior, behavioral practice/rehearsal/, habit formation
	Skills (cognitive and interpersonal)	Training, enablement	Instruction on how to perform the behavior, behavioral practice/rehearsal/, habit formation
	Behavioral regulation	Training, enablement, modelling	Monitoring of behavior by others without feedback, self-monitoring of behavior, feedback on behavior
Reflective motivation			
	Intentions	Persuasion	Commitment
	Social identity	Persuasion, modelling	Identification of self as role model, valued self-identity
	Beliefs about capabilities	Persuasion, training	Instruction on how to perform the behavior, goal setting, feedback on behavior, prompts/cues
	Beliefs about consequences	Education, persuasion, Training	Information about health consequences, information about social and environmental consequences
Automatic Motivation			
	Emotion	Persuasion	Social support (emotional), self-monitoring of behavior
Physical opportunity			
	Environmental context and resources	Environmental restructuring	Adding objects to the environment, restructuring the physical environment
Social opportunity			
	Social influences	Enablement	Social support (unspecified), social support (practical)

Table 5. Excerpt of examples of behavioral change technique (BCT) user-centered design (UCD) app translation.

Behavioral change techniques (BCTs)	User preferences	App features
Instruction on how to perform the behavior	Time saving and convenience, visual aids	Balance wheel, portion guide tool
Self-monitoring of the behavior	Gamification	Points for logging food
Feedback on the behavior	Time saving and convenience, visual aids,	Visual feedback of food groups to target in the following week
Non-specific reward	Gamification	Points and awards for completing activities

Textbox 4. Gamification techniques and behavioral change techniques (BCTs).

Technique

- Points
  - Non-specific reward
  - Feedback on the behavior
- Achievements
  - Non-specific reward
  - Feedback on the behavior
- Progress bars
  - Feedback on the behavior
- Quiz
  - Information provision
  - Instruction on how to perform the behavior
  - Information about health consequences

Figure 3. Intervention flow chart.

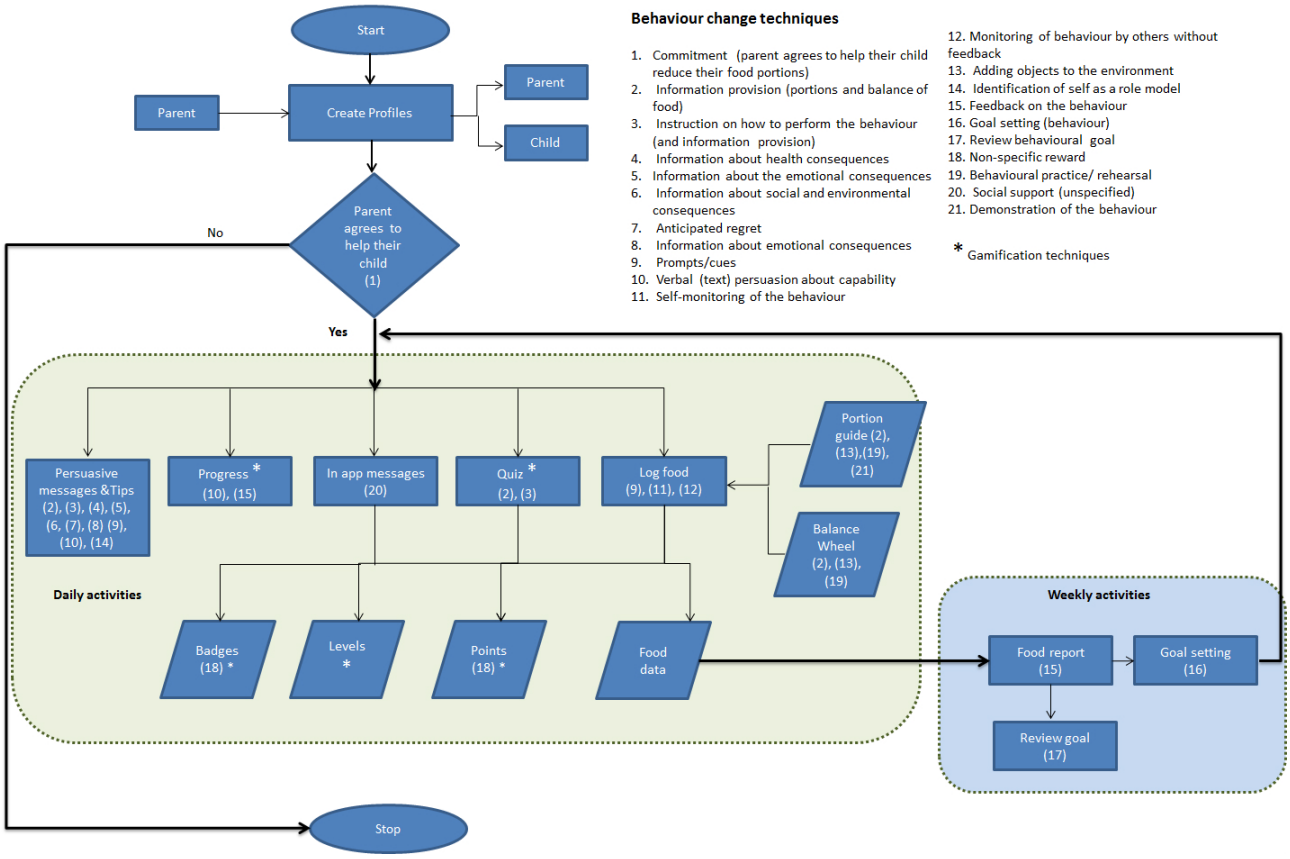
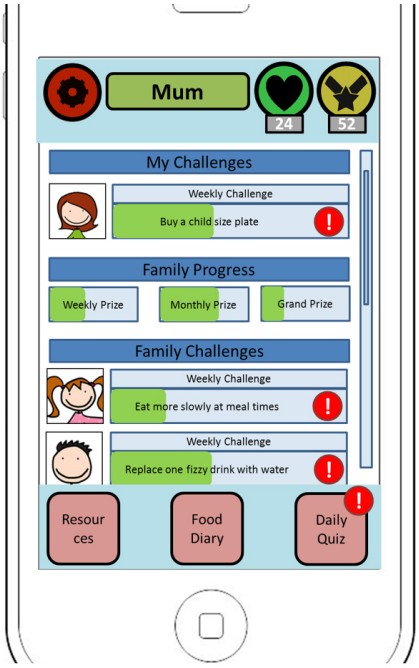


Figure 4. Interactive mock-up of homescreen.



Stage 3: Pre-Testing and Refinement of App Features

Overview

This stage involved the final 2 steps of the intervention development process where proposed app features were piloted with parents, content was generated and refined through consultations with experts, and the prototype app developed.

Step 9: Piloting Potential App Features

Feedback on the interactive mock-ups provided valuable insight into parents' impressions of the proposed app content. For example, the ideas identified around gamification were also piloted with parents to ensure that they were in line with parents' interests [59]. Hence, this stage focused on usability and user experience components such as whether the app features were perceived as fun, helpful, motivating, and aesthetically pleasing. This information was extracted and organized into the following three themes of feedback (1) app features (eg, parental feedback on the portion guide tool), (2) gamification (eg, parents feedback on gamification features), and (3) app positioning (eg, parents feedback on the app positioned as a healthy eating app as oppose to a weight management app) (Table 6).

Table 6. Excerpt of results from pilot testing interactive mock-ups with parents.

Theme	Sub-theme	U[a] codes	UX[b] codes	Example quote
App feature	Portion guide	Easy to remember how to use (+)[c], easy to learn (+),	Satisfying (+), helpful (+)	I think the bit with the hands and the portion sizes, I think that's really really good as it's so difficult to know what a portion size is and very easy to use (Parent, focus group 9)
Gamification	Competition against other families	Safe to use (-)[d]	Motivating (-)	Not too sure about that one as my son has a real complex about his weight so I think it would be tough on him to see other people and might get 'oh well they are doing better than me', do you know what I mean? It's like a confidence thing (Parent, focus group 8)
App positioning	Healthy eating app		Emotionally fulfilling (+), satisfying (+)	I like the idea that it's about healthy eating, you know, not weight control, I like the name as well (Parent, focus group 9)

[a]U: usability

[b]UX: user-experience

[c]Positively viewed (+)

[d]Negatively viewed (-)

Step 10 - Refining App Features, Generating Content and Developing the App Prototype

Following the aforementioned steps led to the development of a prototype app where the theoretical and user-centered components were systematically linked to app features. The overall concept of the app, final intervention mapping table, and examples of text content are presented below.

Overall Concept of the App

Once parents have downloaded the app onto their mobile phone and set up user profiles for family members, they must agree to help their children reduce their food portions before they can participate. All users are instructed to log their food using the camera function, indicating which food group they have eaten and how many portions, using the balance wheel and portion guide as references. The quiz feature offers users a new quiz question every day in relation to portion sizes and the balance of food groups. Once users have logged their food for one week, they will receive a visual report of their food portion intake, highlighting the food groups users may like to set goals in reducing portions in. Users can compose messages to send to other family members requesting help in achieving their weekly goals. Users will receive points for answering quiz questions, logging their food, and helping other family members. Feedback on users' progress towards their weekly goal will be shown visually in progress bars. Parents will receive daily notifications, within app text messages and feedback with regards to their child's progress towards their weekly goal. Parents are also signposted to local family dietician services, group weight management programs, and healthy recipes (see Multimedia Appendix 2).

The Final Intervention Mapping Table

The results of the final mapping table where each theoretical, user-centered, and app feature have been linked together can be viewed in Multimedia Appendix 3. It is important to consider that once app features have been developed, they may incorporate other BCTs that were not originally part of the mapping process which has been documented by other researchers in the field [60].

Content

Table 7 shows an example of the content that was generated for the persuasive messages (via within app messages and notifications), tips, and quizzes.

Table 7. Excerpt of content for within app messages, notifications, and quiz questions.

COM-B	Theoretical Domain Framework (TDF)	Behavioral change techniques (BCTs)	Example
Motivation	Beliefs about capabilities, Beliefs about consequences	Feedback on the behavior, Information about health consequences	Well done! By helping your child to maintain a healthy weight you will reduce their risk of becoming an overweight adult[a].
Capability	Knowledge	Instruction on how to perform the behavior	One glass of apple juice and one glass of orange juice count as how many portions of fruit[b]?

[a]Example of content for notifications and messages on loading screens

[b]Example of text content for quiz

Discussion

Principal Findings

Mobile phones possess a range of attributes that can facilitate health promotion apps to support behavior change; however, the development of a majority of these apps available to consumers has occurred in isolation of theory and evidence, resting mainly on developers' intuitions [60]. To date, few published research studies have provided enough detailed information about the steps involved in the development of a mHealth app that can be replicated [21]. This paper provides a step by step exemplar for how evidence, theory, and user-centered components were incorporated into a mHealth app.

A behavioral diagnosis using the BCW revealed that parents experienced barriers in their capability, opportunity, and motivation to provide appropriate portion sizes for their children. This led to the selection of 6 intervention functions and 21 behavior change techniques to bring about change in this target behavior. Findings with regards to parents' preferences for app features revolved around their desires for specific app features such as a recipe tool, simple and quick interactions with the app, visual aids, and elements of gamification such as scoring points for healthy eating. Techniques of gamification were further expanded to increase parents' engagement with the app and deliver specific BCTs.

A major strength of this study was the involvement of multiple stakeholders in the app development process including the local authority, family weight management service commissioners, community program managers (for recruitment of parents with overweight children and the implementation of the app), family case workers, parents of overweight and healthy weight children (for ensuring social validity amongst the target population), pediatricians, dieticians, psychologists (for the nutritional and psychological content), mobile app experts, and software engineers (for the translation of research findings into app features and technical development of the app). Hence, results can serve as a systematic framework for developers in terms of incorporating stakeholder-informed design elements in the development of health promotion apps.

An additional strength was the use of a comprehensive framework (BCW) where one of the major components differentiating it from other behavior change intervention frameworks is that it is underpinned by a model of behavior change. The COM-B model embodies a 360 degree

comprehensive inclusion of behavior change theories. It is a dynamic model that, unlike other theories of behavior change, takes account of the context of behavior, automatic processes (eg, habit, emotion), and environmental influences [61,62]. Furthermore, the BCW uses a standardized language of theoretical constructs and behavior change techniques which is essential for the replication and synthesis of research and evidence [63]. However, similar to other psychological models and health behavior change intervention frameworks, the BCW stops short of serving as a guide when it comes to translating behavior change techniques into mHealth app features due to the infancy of the mHealth field. Additionally, it's execution relied heavily on the expertise, creativity, and judicious decision making of the design team with regards to which components should actually be implemented in the app, as well as drawing on existing evidence, practical considerations, end-users' views, and expert advice. Thus, it is necessary to expand on the BCW using other disciplines in design and engineering and collaborate with the commercial app industry for the development of behavior change interventions that is relevant for the mobile app ecosystem. Fundamentally, the UCD approach yielded granular information on the relevance, acceptability, and preference of app features within the target group. Consultations with the app company combined with the research findings allowed for the application of *meaningful gamification* which can help parents to manage their children's eating habits as well as promote engagement [59].

Since conducting this study, a new framework, Behavioral Intervention Technology (BIT) model, has been published which attempts to integrate both conceptual and technological components of electronic health (eHealth) and mHealth interventions [60]. In particular, it offers a method for targeting distal clinical aims (eg, weight reduction) and translating behavior change strategies into an app features. However, in contrast, the BCW approach starts with a behavioral diagnosis of the target behavior, before moving onto the possible solutions. Furthermore, the BIT model does not integrate UCD elements, a crucial step that was followed in our work in ensuring the social validity of the app amongst the target population.

Limitations

There are several limitations with the current approach that must be acknowledged. Firstly, the time taken to develop a prototype app is an important consideration, a factor which has also been reported for other mHealth interventions [21,64]. Compared to intervention development, the associated technology development occurs at a much faster pace [65]. Consequently, by the time mHealth interventions are implemented and tested, the technology may have potentially moved on. Secondly, systematically developing a health promotion app intervention can also be resource intensive. In this study, incorporating UCD revealed the need for further refinement of app features which is a much needed step. At the same time, implementing the changes requires additional resources.

With regards to the specific target behavior, this paper focused on both actual portion sizes and frequencies of portions. Nevertheless, there are other nutrition behaviors that can be targeted in the app, along with exercise behaviors, which provides opportunities for further research. One solution may be to target behaviors at different periods of time. For example, after families have completed a 12-week intervention targeting portion control, they could then have the opportunity to move onto the next stage where a new behavior is targeted.

The empirical research used a qualitative study design, whereas quantitative surveys have typically been applied to research designs seeking the most appropriate targets for interventions to date [66,67]. However, the BCW approach does not require this, partly because the COM-B model includes factors that go beyond the sociocognitive spectrum (eg, opportunity) and questionnaires can only measure perception of opportunity rather than objectively assess this.

Future Research

In addition to applying techniques of gamification to the intervention as a way to help increase parents and families' engagement with the app, the study highlighted other important components that have the potential to increase (and decrease) parents' engagement with the app such as interactivity, novelty, and tailoring of app content. Engagement is a multidimensional construct; hence further research with the prototype app drawing on a validated model of user engagement [68] will be necessary. This would provide insight into which aspects are important for capturing parents' attention and encouraging their sustained use of the app.

The next stage in development will involve formal usability testing with parents which will result in further refinements prior to conducting an evaluation of the impact of the app on families' portion sizes. Lastly, the app is developed specifically for a parent population with young children. We encourage researchers to apply the developed methodology to other samples as this can help to refine and expand on the app intervention development process

Conclusions

Within the context of mHealth interventions, we cannot ignore the reality that theoretical, user-centered, and technological components are inexorably linked. Simultaneous consideration must therefore be afforded to them, following a systematic development process that draws on relevant theory, evidence, and research with the target population. In this paper it has been demonstrated how the BCW can serve as a systematic and comprehensive guide to ensure that a health promotion app is underpinned with relevant theory and evidence. Integrating this step by step approach with activities and methods from user-centered design and collaboration with the commercial app industry has also been clearly explicated. This work provides a template and practical guide for researchers and app developers looking to apply similarly systematic and rigorous approaches to content development of mHealth interventions in the future.

Acknowledgments

This research received support from an EPSRC Doctoral training grant awarded to the WMG at the University of Warwick [GR/T11371/01].

Conflicts of Interest

None declared.

Multimedia Appendix 1

COM-B behavioral diagnosis.

[PDF File (Adobe PDF File), 297KB - mhealth_v3i2e69_app1.pdf]

Multimedia Appendix 2

Health Heroes screenshots.

[PDF File (Adobe PDF File), 425KB - mhealth_v3i2e69_app2.pdf]

Multimedia Appendix 3

Final intervention mapping table.

[PDF File (Adobe PDF File), 98KB - mhealth_v3i2e69_app3.pdf]

References

1. mHealth: New horizons for health through mobile technologies.: World Health Organization; 2011. URL: http://www.who.int/ehealth/mhealth_summit.pdf [accessed 2015-06-02] [WebCite Cache ID 6YyrjVNAB]
2. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995. [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]
3. Butler C. How to pick useful health apps for mobile devices. 2012. URL: http://www.washingtonpost.com/national/health-science/how-to-pick-useful-health-apps-for-mobile-devices/2012/07/16/gJQAQ1uFpW_story.html [accessed 2015-06-02] [WebCite Cache ID 6YyrtNuv4]
4. Curtis K, Karasouli E. An assessment of the potential of health promotion apps to support health behaviour change. Health Psychology Update 2014;23(2).
5. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
6. Klasnja P, Pratt W. Managing health with mobile technology. Interactions 2014 Jan 01;21(1):66-69. [doi: 10.1145/2540992]
7. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
8. Tate EB, Spruijt-Metz D, O'Reilly G, Jordan-Marsh M, Gotsis M, Pentz MA, et al. mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. Transl Behav Med 2013 Dec;3(4):406-415 [FREE Full text] [doi: 10.1007/s13142-013-0222-3] [Medline: 24294329]
9. Hingle MD, O'Connor TM, Dave JM, Baranowski T. Parental involvement in interventions to improve child dietary intake: a systematic review. Prev Med 2010 Aug;51(2):103-111 [FREE Full text] [doi: 10.1016/j.ypmed.2010.04.014] [Medline: 20462509]
10. Knowlden AP, Sharma M. Systematic review of family and home-based interventions targeting paediatric overweight and obesity. Obes Rev 2012 Jun;13(6):499-508. [doi: 10.1111/j.1467-789X.2011.00976.x] [Medline: 22221298]
11. Staniford LJ, Breckon JD, Copeland RJ. Treatment of Childhood Obesity: A Systematic Review. J Child Fam Stud 2011 Jun 11;21(4):545-564. [doi: 10.1007/s10826-011-9507-7]
12. Schoffman DE, Turner-McGrievy G, Jones SJ, Wilcox S. Mobile apps for pediatric obesity prevention and treatment, healthy eating, and physical activity promotion: just fun and games? Transl Behav Med 2013 Sep;3(3):320-325 [FREE Full text] [doi: 10.1007/s13142-013-0206-3] [Medline: 24073184]
13. Barlow SE, Ohlemeyer CL. Parent Reasons for Nonreturn to a Pediatric Weight Management Program. Clinical Pediatrics 2006 May 01;45(4):355-360. [doi: 10.1177/000992280604500408]
14. Michie S, Atkins L, West R. The Behaviour Change Wheel A guide to designing interventions. London: Silverback; 2014:978-971.
15. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

16. Dombrowski SU, Sniehotta FF, Avenell A, Johnston M, MacLennan G, Araújo-Soares V. Identifying active ingredients in complex behavioural interventions for obese adults with obesity-related co-morbidities or additional risk factors for co-morbidities: a systematic review. Health Psychology Review 2012 Mar;6(1):7-32. [doi: 10.1080/17437199.2010.513298]
17. Mockus DS, Macera CA, Wingard DL, Peddecord M, Thomas RG, Wilfley DE. Dietary self-monitoring and its impact on weight loss in overweight children. Int J Pediatr Obes 2011 Aug;6(3-4):197-205 [FREE Full text] [doi: 10.3109/17477166.2011.590196] [Medline: 21722068]
18. Borra ST, Kelly L, Shirreffs MB, Neville K, Geiger CJ. Developing health messages: qualitative studies with children, parents, and teachers help identify communications opportunities for healthful lifestyles and the prevention of obesity. J Am Diet Assoc 2003 Jun;103(6):721-728. [doi: 10.1053/jada.2003.50140] [Medline: 12778044]
19. Boushey CJ, Kerr DA, Wright J, Lutes KD, Ebert DS, Delp EJ. Use of technology in children's dietary assessment. Eur J Clin Nutr 2009 Feb;63 Suppl 1:S50-S57 [FREE Full text] [doi: 10.1038/ejcn.2008.65] [Medline: 19190645]
20. Brown B, Chetty M, Grimes A, Harmon E. Reflecting on health: a system for students to monitor diet and exercise. In: CHI'06 Ext Abstr Hum factors Comput Syst. 298-4. New York: AMC; 2006:1807-1812.
21. Fjeldsoe BS, Miller YD, O'Brien JL, Marshall AL. Iterative development of MobileMums: a physical activity intervention for women with young children. Int J Behav Nutr Phys Act 2012;9:151 [FREE Full text] [doi: 10.1186/1479-5868-9-151] [Medline: 23256730]
22. Abroms L, Padmanabhan N, Thaweethai L, Phillips T. iPhone Apps for Smoking Cessation. American Journal of Preventive Medicine 2011 Mar;40(3):279-285. [doi: 10.1016/j.amepre.2010.10.032]
23. Breton ER, Fuemmeler BF, Abroms LC. Weight loss-there is an app for that! But does it adhere to evidence-informed practices? Transl Behav Med 2011 Dec;1(4):523-529 [FREE Full text] [doi: 10.1007/s13142-011-0076-5] [Medline: 24073074]
24. 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]
25. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012;14(3):e72 [FREE Full text] [doi: 10.2196/jmir.1977] [Medline: 22584372]
26. West JH, Hall PC, Arredondo V, Berrett B, Guerra B, Farrell J. Health Behavior Theories in Diet Apps. Journal of Consumer Health On the Internet 2013 Jan;17(1):10-24. [doi: 10.1080/15398285.2013.756343]
27. Tabacchi G, Giammanco S, La Guardia M, Giammanco M. A review of the literature and a new classification of the early determinants of childhood obesity: from pregnancy to the first years of life. Nutrition Research 2007 Oct;27(10):587-604. [doi: 10.1016/j.nutres.2007.06.001]
28. Birch LL. Child feeding practices and the etiology of obesity. Obesity (Silver Spring) 2006 Mar;14(3):343-344 [FREE Full text] [doi: 10.1038/oby.2006.45] [Medline: 16648602]
29. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
30. Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug;38(1):28-39 [FREE Full text] [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]
31. 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]
32. Carter MC, Burley VJ, Cade JE. Development of 'My Meal Mate' - A smartphone intervention for weight loss. Nutrition Bulletin 2013 Feb 07;38(1):80-84. [doi: 10.1111/nbu.12016]
33. Hebden L, Cook A, van der Ploeg HP, Allman-Farinelli M. Development of smartphone applications for nutrition and physical activity behavior change. JMIR Res Protoc 2012;1(2):e9 [FREE Full text] [doi: 10.2196/resprot.2205] [Medline: 23611892]
34. Localytics. 2012. URL: http://info.localytics.com/blog [accessed 2015-06-02] [WebCite Cache ID 6Yyq3kX0V]
35. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50:179-211.
36. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. Am Psychol 1992 Sep;47(9):1102-1114. [Medline: 1329589]
37. Bartholomew L, Parcel G, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. Health Educ Behav 1998 Oct;25(5):545-563. [Medline: 9768376]
38. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337(sep29 1):a1655-a1655. [doi: 10.1136/bmj.a1655]
39. 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]
40. Michie S, West R, Campbell R, Brown J, Gainforth H. ABC of behaviour change theories. London: Silverback; 2014:978-971.
41. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

42. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
43. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
44. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
45. Golley RK, Hendrie GA, Slater A, Corsini N. Interventions that involve parents to improve children's weight-related nutrition intake and activity patterns - what nutrition and activity targets and behaviour change techniques are associated with intervention effectiveness? Obes Rev 2011 Feb;12(2):114-130. [doi: 10.1111/j.1467-789X.2010.00745.x] [Medline: 20406416]
46. Hendrie GA, Brindal E, Corsini N, Gardner C, Baird D, Golley RK. Combined home and school obesity prevention interventions for children: what behavior change strategies and intervention characteristics are associated with effectiveness? Health Educ Behav 2012 Apr;39(2):159-171. [doi: 10.1177/1090198111420286] [Medline: 21984691]
47. Kamath CC, Vickers KS, Ehrlich A, McGovern L, Johnson J, Singhal V, et al. Clinical review: behavioral interventions to prevent childhood obesity: a systematic review and metaanalyses of randomized trials. J Clin Endocrinol Metab 2008 Dec;93(12):4606-4615. [doi: 10.1210/jc.2006-2411] [Medline: 18782880]
48. van der Kruk JJ, Kortekaas F, Lucas C, Jager-Wittenaar H. Obesity: a systematic review on parental involvement in long-term European childhood weight control interventions with a nutritional focus. Obes Rev 2013 Sep;14(9):745-760 [FREE Full text] [doi: 10.1111/obr.12046] [Medline: 23734625]
49. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: MindTrek '11 Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. New York, NY, USA: AMC; 2011:9-15.
50. Rogers Y, Sharp H, Preece J. Interaction Design: Beyond Human - Computer Interaction. New York: Wiley; 2002.
51. 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]
52. Child Obesity.: Public Health England URL: http://www.noo.org.uk/NOO_about_obesity/child_obesity [accessed 2015-06-02] [WebCite Cache ID 6YytNQTxL]
53. Lowe MR. Self-regulation of energy intake in the prevention and treatment of obesity: is it feasible? Obes Res 2003 Oct;11 Suppl:44S-59S. [doi: 10.1038/oby.2003.223] [Medline: 14569037]
54. Swinburn BA, Jolley D, Kremer PJ, Salbe AD, Ravussin E. Estimating the effects of energy imbalance on changes in body weight in children. Am J Clin Nutr 2006 Apr;83(4):859-863 [FREE Full text] [Medline: 16600939]
55. Katz DL, O'Connell M, Njike VY, Yeh M, Nawaz H. Strategies for the prevention and control of obesity in the school setting: systematic review and meta-analysis. Int J Obes (Lond) 2008 Dec;32(12):1780-1789. [doi: 10.1038/ijo.2008.158] [Medline: 19079319]
56. Cook-Cottone C, Casey CM, Feeley TH, Baran J. A meta-analytic review of obesity prevention in the schools: 1997-2008. Psychol. Schs 2009 Sep;46(8):695-719. [doi: 10.1002/pits.20409]
57. Katz DL. Unfattening our children: forks over feet. Int J Obes (Lond) 2011 Jan;35(1):33-37. [doi: 10.1038/ijo.2010.218] [Medline: 21060310]
58. Harris KC, Kuramoto LK, Schulzer M, Retallack JE. Effect of school-based physical activity interventions on body mass index in children: a meta-analysis. CMAJ 2009 Mar 31;180(7):719-726 [FREE Full text] [doi: 10.1503/cmaj.080966] [Medline: 19332753]
59. Nicholson S. A User-Centered Theoretical Framework for Meaningful Gamification.: Games+Learning+Society 8.0; 2012. URL: http://scottnicholson.com/pubs/meaningfulframework.pdf [accessed 2015-06-03] [WebCite Cache ID 6Z1N3cWpe]
60. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
61. Jackson C, Eliasson L, Barber N, Weinman J. Eur Heal Psychol. 2011. Applying COM-B to medication adherence work tended to focus on the role and its effects on patient URL: http://openhealthpsychology.net/ehp/issues/2014/v16iss1February2014/16_1_EHP_February2014%20Jackson%20et%20al.pdf [accessed 2015-06-10] [WebCite Cache ID 6ZBJSuGev]
62. Baranowski T, Cullen KW, Nicklas T, Thompson D, Baranowski J. Are current health behavioral change models helpful in guiding prevention of weight gain efforts? Obes Res 2003 Oct;11 Suppl:23S-43S. [doi: 10.1038/oby.2003.222] [Medline: 14569036]
63. 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]
64. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. J Health Commun 2012 May;17 Suppl 1:11-21. [doi: 10.1080/10810730.2011.649103] [Medline: 22548594]

65. Pagoto S, Bennett GG. How behavioral science can advance digital health. Transl Behav Med 2013 Sep;3(3):271-276 [FREE Full text] [doi: 10.1007/s13142-013-0234-z] [Medline: 24073178]
66. Birch LL, Ventura AK. Preventing childhood obesity: what works? Int J Obes (Lond) 2009 Apr;33 Suppl 1:S74-S81. [doi: 10.1038/ijo.2009.22] [Medline: 19363514]
67. Brown R, Ogden J. Children's eating attitudes and behaviour: a study of the modelling and control theories of parental influence. Health Educ Res 2004 Jun;19(3):261-271 [FREE Full text] [doi: 10.1093/her/cyg040] [Medline: 15140846]
68. O'Brien H, Toms EG. The development and evaluation of a survey to measure user engagement. J Am So Inf Sci 2009 Oct 19;61(1):50-69. [doi: 10.1002/asi.21229]

Abbreviations

BCT: behavioral change technique
BCW: behavioral change wheel
mHealth: mobile health
PDA: personal digital assistants
TDF: Theoretical Domains Framework
UCD: user-centered design

Edited by G Eysenbach; submitted 11.09.14; peer-reviewed by L Yardley, A DeSmet, SH Tee, E Norris; comments to author 22.01.15; accepted 02.03.15; published 18.06.15.

Please cite as:
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
URL: http://mhealth.jmir.org/2015/2/e69/
doi:10.2196/mhealth.3857
PMID:26088692

Original Paper

Uptake and Acceptability of Information and Communication Technology in a Community-Based Cohort of People Who Inject Drugs: Implications for Mobile Health Interventions

Andrew Genz[1], MPH; Gregory Kirk[1,2], MD, PhD, MPH; Damani Piggott[1,2], MD, PhD; Shruti H Mehta[1], PhD, MPH; Beth S Linas[1], PhD, MHS; Ryan P Westergaard[3], MD, PhD, MPH

[1]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[3]Department of Medicine, University of Wisconsin-Madison, Madison, WI, United States

Corresponding Author:
Ryan P Westergaard, MD, PhD, MPH
Department of Medicine
University of Wisconsin-Madison
1685 Highland Ave
MFCB 5220
Madison, WI, 53705
United States
Phone: 1 6082657927
Fax: 1 6082634464
Email: rpw@medicine.wisc.edu

Abstract

Background: Mobile phone and Internet-based technologies are increasingly used to disseminate health information and facilitate delivery of medical care. While these strategies hold promise for reducing barriers to care for medically-underserved populations, their acceptability among marginalized populations such as people who inject drugs is not well-understood.

Objective: To understand patterns of mobile phone ownership, Internet use and willingness to receive health information via mobile devices among people who inject drugs.

Methods: We surveyed current and former drug injectors participating in a longitudinal cohort study in Baltimore, Maryland, USA. Respondents completed a 12-item, interviewer-administered questionnaire during a regular semi-annual study visit that assessed their use of mobile technology and preferred modalities of receiving health information. Using data from the parent study, we used logistic regression to evaluate associations among participants' demographic and clinical characteristics and their mobile phone and Internet use.

Results: The survey was completed by 845 individuals, who had a median age of 51 years. The sample was 89% African-American, 65% male, and 33% HIV-positive. Participants were generally of low education and income levels. Fewer than half of respondents (40%) indicated they had ever used the Internet. Mobile phones were used by 86% of respondents. Among mobile phone owners, 46% had used their phone for text messaging and 25% had accessed the Internet on their phone. A minority of respondents (42%) indicated they would be interested in receiving health information via phone or Internet. Of those receptive to receiving health information, a mobile phone call was the most favored modality (66%) followed by text messaging (58%) and Internet (51%).

Conclusions: Utilization of information and communication technology among this cohort of people who inject drugs was reported at a lower level than what has been estimated for the general U.S. population. Our findings identify a potential barrier to successful implementation of mobile health and Internet-based interventions for people who inject drugs, particularly those who are older and have lower levels of income and educational attainment. As mobile communication technology continues to expand, future studies should re-examine whether mHealth applications become more accessible and accepted by socioeconomically disadvantaged groups.

(JMIR mHealth uHealth 2015;3(2):e70) doi:10.2196/mhealth.3437

KEYWORDS

substance abuse; intravenous; Internet; cellular phone; text messaging; telemedicine

Introduction

Utilization of mobile phones for communication and access to information has become nearly ubiquitous in both low- and high-income countries. At the end of 2013, the number of active mobile phone subscriptions worldwide was estimated at 6.8 billion, or approximately 96 subscriptions for every 100 inhabitants of the world [1]. The global availability of mobile broadband services is rapidly increasing. In the foreseeable future, the majority of the world's population will have access to the Web via a mobile device. With this expanded access to increasingly powerful handheld devices has come recognition of potential applications for improving health and health care [2].

Persons affected by substance abuse and poverty face substantial challenges in consistently accessing and utilizing health care for chronic medical conditions. Successful chronic disease management typically requires consistent health-promoting behaviors such as daily medication administration, periodic monitoring of disease-specific biomarkers, and clinical assessments by medical providers. Maladaptive, drug-seeking behaviors and material deprivation that are frequently associated with substance use disorders may interfere with all of these steps. For example, people who inject drugs (PWID) have been shown to have inferior medication adherence and more frequent interruptions in therapy when treated for human immunodeficiency virus (HIV) [3,4], hepatitis C virus (HCV)[5,6], and tuberculosis[7,8]. PWID also have high rates of depression [9-11] and alcohol dependence [11,12], chronic disorders for which regular clinical monitoring is important for preventing disability and poor social functioning. In view of these challenges, strategies that leverage information and communication technology (ICT) to enhance care delivery and support treatment adherence among PWID could play an important role in reducing health disparities.

Growth in the number of health-related apps and steady reduction in costs for mobile devices has created the potential for mobile communication technology to add considerable value to medical care and public health programs, particularly among populations with limited resources. Mobile health, or mHealth, has been broadly defined as medical or public health practice supported by mobile devices. If demonstrated to be effective, affordable and scalable, mHealth approaches could be uniquely beneficial to medically underserved or marginalized patient groups, who may encounter geographic or socioeconomic barriers to health care or have limited access to timely and relevant health information.

If mobile health apps tailored to the needs of PWID are demonstrated to be feasible and efficacious, additional translational research will be necessary in order to understand whether and how they can be successfully implemented in clinic and community settings. A recent cross-sectional survey of 100 consecutive HIV-infected patients seen at an urban clinic in Baltimore, Maryland demonstrated near-universal mobile phone ownership (96%), and a high level of willingness to use mobile phones for medication adherence support [13]. However, relatively few respondents owned a smartphone (28%) or had used a mobile phone to view the Web content (34%). These proportions are somewhat lower than estimates that 51% of Americans owned a smartphone in 2012 [14].

The goal of the present study was to gain insight into the real-world feasibility and acceptance of mHealth approaches among a cohort of older, predominantly African American PWID in a largely poor, urban community in the United States. We conducted a cross-sectional survey among volunteers in a long-running community-based cohort study in order to evaluate mobile phone ownership and use of the Web for health-related purposes. Based on the research team's experience and personal interaction with the study participants and the high median age of the active cohort (51.8 years), we hypothesized that the adoption of ICT among our study sample would be lower than the levels reported in previous studies.

Methods

Recruitment

The study sample was comprised of participants in the AIDS Linked to the Intravenous Experience (ALIVE) cohort who were active in the study during 2011. ALIVE is a community-based cohort study that has continuously followed current and former PWID in Baltimore since 1988 [15]. The population has a high prevalence of HIV infection, viral hepatitis, and substance use disorders. Previous research with this cohort has demonstrated suboptimal levels of treatment utilization [16,17] and high levels of multimorbid chronic conditions [18,19], characteristics that make this group an attractive target for mHealth-based enhancements to routine clinical care.

Data Collection

Between March and September of 2011, all ALIVE participants attending a regularly scheduled, semiannual study assessment were invited to complete an additional 12-item, interviewer-administered questionnaire. The questionnaire was developed by investigators specifically for this sub-study, and contained items assessing ownership of mobile phones, and utilization of phones for voice calls, texting or data download. Individuals' lifetime history of Web access and patterns of recent Web use were assessed with multiple choice questions.

Study participants who volunteered to complete the supplemental ICT questionnaire were notified that the ALIVE investigators were exploring different ways to collect and disseminate health information among PWID. In this context, all participants were asked whether they would be willing to receive health information via voice calls on a mobile phone, text messages on a mobile phone, or using the Internet. The full text of the questionnaire is available in Multimedia Appendix 1.

Baseline sociodemographic characteristics (e.g., race, sex, age, education) were taken from participants' baseline assessments conducted at enrollment in the parent study. For the subset of respondents who were HIV-positive, clinical parameters reflecting effective treatment (e.g., CD4+ cell count, HIV viral load, use of antiretroviral therapy) were captured from the most recent ALIVE study visit.

Statistical Analysis

For the present analysis, the two questionnaire items of foremost concern were "Do you currently own a cell phone?" and "Have you ever used the Internet?" Responses to these items comprised the primary outcome variables. We compared the frequency of "yes" responses to these items across categories of sociodemographic and clinical characteristics. Chi-squared tests were used to assess whether any individual characteristics were associated with phone ownership and Web use, assuming that a p-value less than or equal to 0.05 indicated a statistically significant bivariate association.

Using logistic regression, we calculated adjusted odds ratios to estimate the independent association between individual characteristics and the two main outcomes. Covariates included in the model were age, race/ethnicity, gender, education level and amount of legal income in the past six months. These variables were selected based on prior research suggesting that patterns of Web and mobile phone use vary across categories of these demographic characteristics. All analyses were performed using STATA version 11. The Institutional Review Board at the Johns Hopkins Bloomberg School of Public Health reviewed and approved the study protocol.

Results

Sample Characteristics

Of 1,024 individuals invited to complete the questionnaire between March 7, 2011 and September 29, 2011, 845 agreed to participate, yielding a response rate of 82.5% (Table 1). The median age of participants was 51.8 years (IQR=47.0-56.7). Of these, 89.2% were African American, 65.1% were male, and 40.6% had a high school diploma or GED. A majority of respondents (55.9%) reported that their legal income during the prior six months was less than $5000.

Approximately one third of respondents (275/845, 32.5%) were known to be HIV seropositive at the time of the survey, and 84.0% (710/845) were seropositive for hepatitis C virus. Of the HIV-infected subgroup, 76.0% (209/275) were receiving antiretroviral therapy and 50.2% (138/275) of these had an undetectable HIV viral load, representing fully effective HIV treatment.

Table 1. Participant characteristics (N=845).

Characteristics	n (%)[a]
Age (median, IQR)[b]	51.8 (46.9-56.6)
Female	295 (34.9)
Male	550 (65.1)
African American	754 (89.2)
Finished high school or GED	342 (40.6)
Legal income during past 6 months	
None	157 (18.6)
$0 - $4,999	473 (55.9)
$5,000 or higher	215 (25.4)
Homeless in past 6 months	58 (6.9)
Current smoker	660 (78.2)
Alcohol use in past 6 months	407 (47.2)
Injected drugs in past 6 months	207 (24.5)
HIV positive	275 (32.5)
HCV[c] positive	710 (84.0)
HIV viral load undetectable[d]	138 (50.4)
CD4+ cell count (median, IQR)[d]	408 (255-659)
Currently taking ART[e]	209 (76.0)

[a] All values presented are N(%) unless otherwise specified

[b] IQR=interquartile range

[c] HCV=hepatitis C virus

[d] Clinical variables presented only for 275 HIV-infected respondents

[e] ART=antiretroviral therapy

Mobile Phone and Web Use

Responses to questionnaire items assessing mobile phone ownership, use of phones for texting and Web use are summarized in Table 2. Of the 845 respondents asked, 86.0% (727) reported owning a mobile phone at the time of the survey. Over half of these (56.7%, 412/727) subscribed to a monthly payment plan through a wireless carrier; 25.7% (187/727) utilized a prepaid or pay-as-you-go payment system; 45.9% (334/727) used a free phone provided by a government program. Most respondents (64.9%, 549/845) reported having a single phone number (including mobile and land lines) during the three months preceding the survey, but 9.9% (84/845) used three or more numbers and 0.8% (7/845) used 10 or more numbers during that time. Nearly all respondents (92.2%, 779/845) had heard of free government phone programs (e.g. Safelink or “Obamaphone”) and about half (50.1%, 423/845) had utilized such a program. See Multimedia Appendix 2 for these details on participants’ phone use practices.

Four participants had incomplete information on both phone use questions.

Approximately half of participants (46.2%, 334/723) who owned a mobile phone reported sending or receiving text messages. Fewer (18.5%, 134/723) reported they use a mobile phone to access the Web. Lifetime Web use was low in this cohort. Overall, 40.5% (342/845) reported ever using the Web. Table 3 presents a detailed account of responses to these questions by demographic characteristics and HIV infection status.

Table 2. Participant responses to key questions.

	Owns a mobile phone (n=845)[a] n (%)	Uses a mobile phone for texting (n=723)[b] n (%)	Uses a mobile phone for Web (n=723)[b] n (%)	Ever used Web (n=845)[a] n (%)
Overall	727 (86.0)	334 (46.2)	134(18.5)	342 (40.5)

[a] N=all 845 participants surveyed

[b] n=only the 723 participants who owned a mobile phone

Table 3. Breakdown of mobile phone ownership, text messaging, and Web use by demographic characteristics and HIV status.

	Owns a mobile phone (n=845) n (%)	Uses a mobile phone for texting (n=723) n (%)	Uses a mobile phone for Web (n=723) n (%)	Ever used Web (n=845) n (%)
Gender				
Male	463 (84.2)	193 (41.9)	79 (17.2)	228 (41.5)
Female	264 (89.5)	141(53.8)	55 (20.9)	114 (38.6)
Race				
Non-AA	73 (80.2)	39 (53.4)	23 (31.5)	58 (63.7)
AA	654 (86.7)	295 (45.4)	111 (17.1)	284 (37.7)
Age				
<40	63 (88.7)	42 (66.7)	23 (36.5)	46 (64.8)
41-50	222 (84.1)	118 (53.1)	54 (24.3)	126 (47.8)
51-60	353 (86.3)	141 (40.4)	45 (12.9)	136 (33.3)
60+	89 (86.0)	33 (37.1)	12 (13.5)	34 (33.7)
Education				
No HS/GED	435 (87.0)	182 (42.0)	70 (16.2)	165 (33.0)
HS/GED	289 (84.5)	151 (52.6)	64 (22.2)	175 (51.2)
Legal income [b]				
$0	126 (80.3)	45 (35.7)	14 (11.1)	44 (28.0)
<$5000	409 (86.5)	195 (47.8)	76 (18.6)	184 (38.9)
>$5000	192 (89.0)	94 (49.7)	44 (23.3)	114 (53.0)
HIV status				
HIV-negative	487 (85.4)	218 (44.9)	93 (19.1)	240 (42.1)
HIV-positive	240 (87.0)	116 (49.0)	41 (17.3)	102 (37.1)

[a] Percent is for each row, for each question. (e.g. 84.2% of men had a phone and 15.8% of men did not)

[b] During six months prior to questionnaire

There were numerous disparities in mobile phone and Web use across demographic and socioeconomic strata. Mobile phone ownership was more prevalent among women, African Americans, and those with higher income (Table 4). Reporting any Web use was independently associated with younger age, completion of high school and higher income (Table 5). African American respondents were half as likely to report ever using the Web, an association that remained statistically significant after adjusting for age, education and income (adjusted OR 0.5, 95% CI 0.3 – 0.9).

Among the 310 Web-using respondents, there were no differences by race or gender in the proportion who reported accessing the Web using a mobile phone. Based on the multivariate model, the only factor associated with using phones to access the Web was age. Respondents who were under 50 were more than twice as likely to report mobile Web use as those who were older than 50.

Table 4. Associations among mobile phone use and selected characteristics.

	Owns mobile phone (n=842)		Uses phone for text messaging (n=720)	
	Adjusted OR[a]	95% CI [a]	Adjusted OR [a]	95% CI [a]
Gender				
Male	reference		reference	
Female	1.6	1.0 – 2.5	1.5	1.1 – 2.1
Race				
Non-African American	reference		reference	
African American	1.8	1.0 – 3.4	1.2	0.7 – 2.1
Age				
<40	reference		reference	
41-50	0.5	0.2 – 1.2	0.4	0.2 – 0.9
51-60	0.6	0.2 – 1.4	0.3	0.1 – 0.5
60+	0.7	0.2 – 1.9	0.2	0.1 – 0.5
Education				
Less than HS/GED	reference		reference	
Completed HS/GED	0.8	0.5 – 1.2	1.6	1.2 – 2.3
Legal income, past 6 months				
$0	reference		reference	
<$5000	1.5	1.0 – 2.5	1.7	1.1 – 2.6
>$5000	2.1	1.1 – 3.8	2.0	1.2 – 3.2
HIV status				
HIV-negative	reference		reference	
HIV-positive	1.1	0.7 – 1.7	1.2	0.9 – 1.7
Taking ART [b]				
Yes	reference		reference	
No	1.1	0.5 – 2.7	0.7	0.4 – 1.3
HIV viral load [b]				
Undetectable	reference		reference	
Detectable	0.4	0.2 – 1.0	0.8	0.4 – 1.4

[a] Adjusted for gender, race, age, educational level, and income

[b] Model limited to 275 HIV-positive participants

Table 5. Associations among Web use and selected characteristics.

	Ever used Web (n=842)		Accessed Web using mobile phone (n=310)	
	Adjusted OR[a]	95% CI [a]	Adjusted OR [a]	95% CI [a]
Gender				
Male	reference		reference	
Female	0.9	0.6 – 1.2	1.4	0.9 – 2.3
Race				
Non-African American	reference		reference	
African American	0.5	0.3 – 0.9	1.1	0.5 – 2.2
Age				
<40	reference		reference	
41-50	0.6	0.3 – 1.0	0.7	0.3 – 1.6
51-60	0.3	0.1 – 0.5	0.4	0.2 – 0.9
60+	0.2	0.1 – 0.5	0.4	0.2 – 1.5
Education				
Less than HS/GED	reference		reference	
Completed HS/GED	2.1	1.5 – 2.8	1.0	0.6 – 1.5
Legal income, past 6 months				
$0	reference		reference	
<$5000	1.8	1.2 – 2.8	1.7	0.8 – 3.8
>$5000	3.4	2.1 – 5.5	1.8	0.7 – 4.1
HIV status				
HIV-negative	reference		reference	
HIV-positive	0.8	0.6 – 1.1	0.9	0.6 – 1.5
Taking ART [b]				
Yes	reference		reference	
No	0.5	0.2 – 0.9	0.6	0.2 – 1.4
HIV viral load [b]				
Undetectable	reference			
Detectable	0.4	0.2 – 0.8	0.6	0.3 – 1.3

[a] Adjusted for gender, race, age, educational level, and income

[b] Model limited to 275 HIV-positive participants

Responses Among HIV-Infected PWID

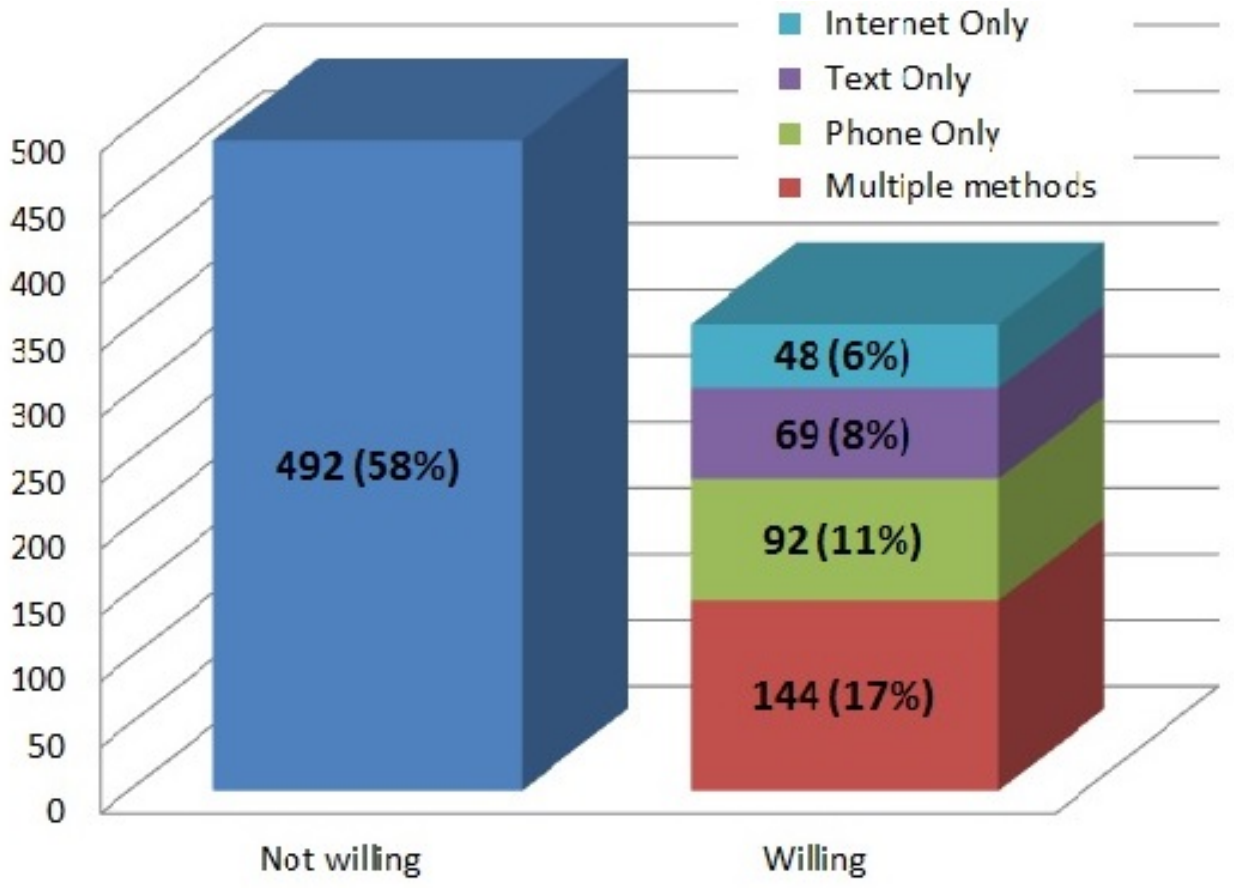
HIV-positive and HIV-negative respondents had similar levels of mobile phone ownership and Web use. Among the HIV-positive cohort, 87.3% of participants (240/275) reported owning a mobile phone, and this proportion did not vary by disease stage or self-reported use of antiretroviral therapy. However, compared with HIV-positive respondents who successfully achieved viral suppression, respondents who had a detectable HIV viral load at the most recent ALIVE visit reported lower levels of ICT adoption. Fewer participants with detectable HIV viremia owned mobile phones (83.0%, 224/270 vs. 90.6%, 125/138) and used the Web (31.9%, 86/270 vs. 42.8%, 59/138) than their virally suppressed counterparts. After adjusting for gender, race, age, education and income, PWID with uncontrolled HIV infection were significantly less likely to own a mobile phone (adjusted OR 0.4, 95% CI 0.2 – 1.0) and to use the Web (adjusted OR 0.4, 95% CI 0.2 – 0.8).

Willingness to Receive Health Information Via Mobile Phone and Web

The majority of respondents (58.2%, 492/845) indicated they would not like to receive health information via mobile phone, text message, or Web. Of the 353 respondents who expressed willingness to receive health information via one or more ICT modalities, 65.7% (232/353) indicated they would like to receive health information by phone, 57.5% (203/353) were willing to receive health-related text messages, and 50.7% (179/353) would use the Web to receive health information (Figure 1). When queried about the type of health information they would be

willing to receive, 62.3% (220/353) indicated they would be interested in information about smoking cessation and 58.4% (206/353) would utilize ICT for medication reminders. Most indicated they would prefer health-related communications to be infrequent. Only 15.6% (55/353) preferred daily communication, while larger proportions favored weekly (25.5%, 90/353) or monthly (42.2%, 149/353) communication.

Figure 1. Willingness to receive health information via mobile phone, text message or internet (N=845).



Discussion

Principal Findings

In this cross-sectional study of former and current PWID, we observed a high prevalence of mobile phone ownership, but lower-than-expected levels of Web use. Relatively few PWID in this cohort expressed interest in using information and communication technology for monitoring or improving health. Taken together, these findings suggest that at the time of the survey, PWID in Baltimore may not be highly receptive to health promotion interventions featuring mobile phones or Web-based communication. As of late 2011, about 60% of the ALIVE cohort reported they had never used the Web, and only about 1 in 5 considered the Web a desirable means of receiving health information.

Whether these results should temper enthusiasm about mHealth approaches to chronic disease management among PWID depends on several factors. The ALIVE cohort consists of low income, inner city residents who are generally poorly educated and therefore have limited employment opportunities. Over half (58%; 631/1088) of all ALIVE participants seen in 2011 had not graduated from high school. It is possible that members of this cohort will be “late adopters” of mobile technology because they have been economically disadvantaged, but will become increasingly receptive to mHealth strategies as mobile devices and data subscriptions become more accessible and affordable. Our results could alternatively reflect structural barriers based on local economic factors and service availability, or even more deeply seated resistance to adopting technology related to cultural norms. Future surveys to monitor changes in technology use among this population will help clarify whether the limited uptake of ICT persists.

Comparison With Prior Work

Prior surveys performed among similar urban-dwelling African American individuals agreed with our findings that mobile phone ownership and use are commonplace. There is evidence, however, that despite widespread use of mobile phones, the use of mobile technology for communicating health information has not been adopted equally among subgroups with older age and less education. A recent survey among women attending a sexually transmitted diseases clinic in Baltimore showed high levels of mobile phone ownership (93%), text messaging (79%), and Web use (80%) [20]. Among women in this study, those who were older and had lower levels of education were significantly less willing to receive health information via mobile phone. A survey among ethnic minority parents attending an urban pediatric clinic similarly showed near-universal mobile

phone ownership, but found that only 17% of respondents ever shared or received health information via text messaging [21].

While adoption of ICT may be lagging in the communities comprising our study population, there is growing evidence that mobile devices and Web-based apps are feasible and acceptable to PWID in research settings. For example, heroin- and cocaine-dependent patients have effectively used handheld devices to monitor real-time experiences of stress, drug cravings, and drug use [22,23]. As previously described by our research group, mobile phone-based ecologic momentary assessment (EMA) methods are also being evaluated for real-time collection of data relevant to both drug use and antiretroviral treatment adherence [24].

Our study found significant disparities in ICT adoption by race. African American respondents were more likely to own a mobile phone than white or Hispanic participants, but were substantially less likely to have used the Web. Low-income, educationally disadvantaged minority communities have been considered a "digitally underserved" population based on previous surveys [25]. Whether race or ethnicity is independently predictive of low ICT uptake is less clear. Contrary to our findings, a 2012 report published by the National Urban League found evidence that African Americans used broadband Internet to apply for jobs more often than white Americans, and that this discrepancy was most pronounced among those without a high school diploma and an annual income of less than $20,000 [26].

To our knowledge, the finding of an association between Web use and successful HIV treatment has not been previously reported. Among HIV-infected respondents in our study, those who received ART and achieved an undetectable HIV viral load had more than twice the odds of reporting they used the Web. While this association is not likely causal in nature, it suggests that patients at the highest risk for suboptimal treatment outcomes are least likely to be current adopters of ICT. This represents a potentially important barrier to implementing mHealth solutions among patients who stand to benefit from them most. Treatment of HIV infection exemplifies the complex requirements for successful long-term disease management. Without high levels of medication adherence, viral replication may continue unchecked, leading to immune system dysfunction, elevated risk of antiretroviral drug resistance, and higher likelihood of HIV transmission. If mHealth modalities are to have a potentially beneficial role in facilitating treatment engagement for high-risk patients, implementation strategies may need to address the same social and structural barriers that tend to limit the effectiveness of medical care.

Limitations

The usual limitations of cross-sectional research apply to our results. The landscapes of mobile technology and patterns of Web use are continually changing. In this study sample with relatively low levels of ICT adoption, it is to be expected that individuals' use of technology has evolved since the time the data were collected. Because this survey was conducted among a socioeconomically homogenous sample of PWID in a single U.S. city, our findings cannot likely be generalized to many other contexts. Related strengths of the study are its large sample size and moderately high response rate among a population that is not well represented in previous research on this topic.

Conclusions

As behavioral determinants of health play a central role in chronic disease management for patients who use drugs, there may be uniquely beneficial applications of mHealth technology for supporting the special needs and vulnerabilities of drug-using patients. Moreover, the capacity for mobile data collection and processing, combined with a growing marketplace for software apps has fostered the development of innovative and sophisticated approaches to monitoring symptoms and even promoting and facilitating behavior change. Successful implementation of these concepts outside the research setting will require an accurate understanding of the adoption of ICT among targeted populations. Strategies to eliminate barriers to ICT utilization may be essential components of mHealth interventions aimed at improving health among PWID and other marginalized groups.

Acknowledgments

Funding for the study was provided through the National Institutes of Health/National Institute on Drug Abuse (Grants: K23DA032306; U01DA023832; R01DA04334; and R01DA12568).

Conflicts of Interest

None declared.

Multimedia Appendix 1

The ALIVE ICT questionnaire.

[PDF File (Adobe PDF File), 85KB - mhealth_v3i2e70_app1.pdf]

Multimedia Appendix 2

Frequencies of mobile phone ownership among ALIVE participants.

[PDF File (Adobe PDF File), 3KB - mhealth_v3i2e70_app2.pdf]

References

1. International Telecommunication Union (ITU). Geneva, Switzerland; 2013. Measuring the Information Society URL: http://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2013/MIS2013_without_Annex_4.pdf [accessed 2014-02-05] [WebCite Cache ID 6NA8baDVm]
2. Collins F. How to fulfill the true promise of “mHealth”. Sci Am 2012 Jun 19;307(1):16. [doi: 10.1038/scientificamerican0712-16]
3. Mugavero MJ, Lin HY, Allison JJ, Giordano TP, Willig JH, Raper JL, et al. Racial disparities in HIV virologic failure: do missed visits matter? J Acquir Immune Defic Syndr 2009 Jan 1;50(1):100-108 [FREE Full text] [doi: 10.1097/QAI.0b013e31818d5c37] [Medline: 19295340]
4. Yehia BR, Fleishman JA, Metlay JP, Moore RD, Gebo KA. Sustained viral suppression in HIV-infected patients receiving antiretroviral therapy. JAMA 2012 Jul 25;308(4):339-342 [FREE Full text] [doi: 10.1001/jama.2012.5927] [Medline: 22820781]
5. Mehta SH, Lucas GM, Mirel LB, Torbenson M, Higgins Y, Moore RD, et al. Limited effectiveness of antiviral treatment for hepatitis C in an urban HIV clinic. AIDS 2006 Nov 28;20(18):2361-2369. [doi: 10.1097/QAD.0b013e32801086da] [Medline: 17117023]
6. Mehta SH, Genberg BL, Astemborski J, Kavasery R, Kirk GD, Vlahov D, et al. Limited uptake of hepatitis C treatment among injection drug users. J Community Health 2008 Jun;33(3):126-133 [FREE Full text] [doi: 10.1007/s10900-007-9083-3] [Medline: 18165889]
7. Horsburgh CR, Goldberg S, Bethel J, Chen S, Colson PW, Hirsch-Moverman Y, Tuberculosis Epidemiologic Studies Consortium. Latent TB infection treatment acceptance and completion in the United States and Canada. Chest 2010 Feb;137(2):401-409. [doi: 10.1378/chest.09-0394] [Medline: 19793865]
8. Caylà JA, Rodrigo T, Ruiz-Manzano J, Caminero JA, Vidal R, García JM, Working Group on Completion of Tuberculosis Treatment in Spain (Study ECUTTE). Tuberculosis treatment adherence and fatality in Spain. Respir Res 2009;10:121 [FREE Full text] [doi: 10.1186/1465-9921-10-121] [Medline: 19951437]
9. Springer SA. High rates of depressive symptomatology among injecting drug users in Saskatoon, Canada. Evid Based Ment Health 2012 Feb;15(1):9. [doi: 10.1136/ebmh.2011.100199] [Medline: 22044870]
10. Pilowsky DJ, Wu LT, Burchett B, Blazer DG, Ling W. Depressive symptoms, substance use, and HIV-related high-risk behaviors among opioid-dependent individuals: results from the Clinical Trials Network. Subst Use Misuse 2011;46(14):1716-1725 [FREE Full text] [doi: 10.3109/10826084.2011.611960] [Medline: 21973307]
11. Mackesy-Amiti ME, Donenberg GR, Ouellet LJ. Prevalence of psychiatric disorders among young injection drug users. Drug Alcohol Depend 2012 Jul 1;124(1-2):70-78 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.12.012] [Medline: 22226707]
12. Kidorf M, Disney ER, King VL, Neufeld K, Beilenson PL, Brooner RK. Prevalence of psychiatric and substance use disorders in opioid abusers in a community syringe exchange program. Drug Alcohol Depend 2004 May 10;74(2):115-122. [doi: 10.1016/j.drugalcdep.2003.11.014] [Medline: 15099655]
13. Miller CWT, Himelhoch S. Acceptability of Mobile Phone Technology for Medication Adherence Interventions among HIV-Positive Patients at an Urban Clinic. AIDS Res Treat 2013;2013:1-6 [FREE Full text] [doi: 10.1155/2013/670525] [Medline: 23997948]
14. Duggan M. Rainie, Cell Phone Activities. Washington, DC, USA: Pew Research Center’s Internet & Life Project; 2012. L URL: http://pewinternet.org/Reports/2012/Cell-Activities/Main-Findings.aspx [accessed 2014-02-12] [WebCite Cache ID 6NKuvCBW4]
15. Vlahov D, Anthony JC, Munoz A, Margolick J, Nelson KE, Celentano DD, et al. The ALIVE study, a longitudinal study of HIV-1 infection in intravenous drug users: description of methods and characteristics of participants. NIDA Res Monogr 1991;109:75-100. [Medline: 1661376]
16. Kavasery R, Galai N, Astemborski J, Lucas GM, Celentano DD, Kirk GD, et al. Nonstructured treatment interruptions among injection drug users in Baltimore, MD. J Acquir Immune Defic Syndr 2009 Apr 1;50(4):360-366 [FREE Full text] [doi: 10.1097/QAI.0b013e318198a800] [Medline: 19214124]
17. Westergaard RP, Hess T, Astemborski J, Mehta SH, Kirk GD. Longitudinal changes in engagement in care and viral suppression for HIV-infected injection drug users. AIDS 2013 Oct 23;27(16):2559-2566 [FREE Full text] [doi: 10.1097/QAD.0b013e328363bff2] [Medline: 23770493]
18. Piggott DA, Muzaale AD, Mehta SH, Brown TT, Patel KV, Leng SX, et al. Frailty, HIV infection, and mortality in an aging cohort of injection drug users. PLoS One 2013;8(1):e54910 [FREE Full text] [doi: 10.1371/journal.pone.0054910] [Medline: 23382997]
19. Salter ML, Lau B, Go VF, Mehta SH, Kirk GD. HIV infection, immune suppression, and uncontrolled viremia are associated with increased multimorbidity among aging injection drug users. Clin Infect Dis 2011 Dec;53(12):1256-1264 [FREE Full text] [doi: 10.1093/cid/cir673] [Medline: 21976463]
20. Samal L, Hutton HE, Erbelding EJ, Brandon ES, Finkelstein J, Chander G. Digital divide: variation in internet and cellular phone use among women attending an urban sexually transmitted infections clinic. J Urban Health 2010 Jan;87(1):122-128 [FREE Full text] [doi: 10.1007/s11524-009-9415-y] [Medline: 19941085]

21. Mitchell S, Godoy L, Shabazz K, Horn IB. Internet and mobile technology use among urban African American parents: survey study of a clinical population. J Med Internet Res 2014;16(1):e9 [FREE Full text] [doi: 10.2196/jmir.2673] [Medline: 24418967]
22. Epstein DH, Willner-Reid J, Vahabzadeh M, Mezghanni M, Lin JL, Preston KL. Real-time electronic diary reports of cue exposure and mood in the hours before cocaine and heroin craving and use. Arch Gen Psychiatry 2009 Jan;66(1):88-94 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2008.509] [Medline: 19124692]
23. Preston KL, Epstein DH. Stress in the daily lives of cocaine and heroin users: relationship to mood, craving, relapse triggers, and cocaine use. Psychopharmacology (Berl) 2011 Nov;218(1):29-37 [FREE Full text] [doi: 10.1007/s00213-011-2183-x] [Medline: 21336579]
24. Kirk G, Linas BS, Westergaard RP, Piggott D, Bollinger RC, Chang LW, et al. The exposure assessment in current time study: implementation, feasibility, and acceptability of real-time data collection in a community cohort of illicit drug users. AIDS Res Treat 2013;2013:594671 [FREE Full text] [doi: 10.1155/2013/594671] [Medline: 24307943]
25. Bates K, Malakoff L, Kane S, Pulidindi J. National Leage of Cities. 2012. Closing the digital divide: promoting broadband adoption among underserved populations URL: http://www.nlc.org/Documents/Find%20City%20Solutions/Research%20Innovation/Infrastructure/Closing_Digital_Divide_Promoting_Broadband_Adoption_Underserved_Populations.pdf [accessed 2014-02-27] [WebCite Cache ID 6NhiihiAE]
26. Wijewardena M, Hardy V. National Urban League Policy Institute. 2012. Connecting the dots: linking broadband adoption to job creation and job competitiveness URL: http://www.twcresearchprogram.com/pdf/TWC_WijewardenaReport.pdf [accessed 2014-02-25] [WebCite Cache ID 6NhjylsBj]

Abbreviations

ALIVE: AIDS linked to the intravenous experience
EMA: ecologic momentary assessment
HIV: human immunodeficiency virus
ICT: information and communication technology
PWID: people who inject drugs

Edited by G Eysenbach; submitted 31.03.14; peer-reviewed by K Phillips, JM Suelves; comments to author 27.08.14; revised version received 04.12.14; accepted 14.12.14; published 25.06.15.

Please cite as:
Genz A, Kirk G, Piggott D, Mehta SH, Linas BS, Westergaard RP
Uptake and Acceptability of Information and Communication Technology in a Community-Based Cohort of People Who Inject Drugs: Implications for Mobile Health Interventions
JMIR mHealth uHealth 2015;3(2):e70
URL: http://mhealth.jmir.org/2015/2/e70/
doi:10.2196/mhealth.3437
PMID:26111915

Original Paper

Identification of Behavior Change Techniques and Engagement Strategies to Design a Smartphone App to Reduce Alcohol Consumption Using a Formal Consensus Method

Claire Garnett[1], MSc; David Crane[1], MSc; Robert West[2], PhD; Jamie Brown[1,2], PhD; Susan Michie[1], DPhil

[1]Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom

[2]Cancer Research UK, Health Behaviour Research Centre, University College London, London, United Kingdom

Corresponding Author:
Claire Garnett, MSc
Department of Clinical, Educational and Health Psychology
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 2076795682
Email: c.garnett.12@ucl.ac.uk

Abstract

Background: Digital interventions to reduce excessive alcohol consumption have the potential to have a broader reach and be more cost-effective than traditional brief interventions. However, there is not yet strong evidence for their ability to engage users or their effectiveness.

Objective: This study aimed to identify the behavior change techniques (BCTs) and engagement strategies most worthy of further study by inclusion in a smartphone app to reduce alcohol consumption, using formal expert consensus methods.

Methods: The first phase of the study consisted of a Delphi exercise with three rounds. It was conducted with 7 international experts in the field of alcohol and/or behavior change. In the first round, experts identified BCTs most likely to be effective at reducing alcohol consumption and strategies most likely to engage users with an app; these were rated in the second round; and those rated as effective by at least four out of seven participants were ranked in the third round. The rankings were analyzed using Kendall's W coefficient of concordance, which indicates consensus between participants. The second phase consisted of a new, independent group of experts (n=43) ranking the BCTs that were identified in the first phase. The correlation between the rankings of the two groups was assessed using Spearman's rank correlation coefficient.

Results: Twelve BCTs were identified as likely to be effective. There was moderate agreement among the experts over their ranking (W=.465, χ^2_{11}=35.8, *P*<.001) and the BCTs receiving the highest mean rankings were self-monitoring, goal-setting, action planning, and feedback in relation to goals. There was a significant correlation between the ranking of the BCTs by the group of experts who identified them and a second independent group of experts (Spearman's rho=.690, *P*=.01). Seventeen responses were generated for strategies likely to engage users. There was moderate agreement among experts on the ranking of these engagement strategies (W=.563, χ^2_{15}=59.2, *P*<.001) and those with the highest mean rankings were ease of use, design – aesthetic, feedback, function, design – ability to change design to suit own preferences, tailored information, and unique smartphone features.

Conclusions: The BCTs with greatest potential to include in a smartphone app to reduce alcohol consumption were judged by experts to be self-monitoring, goal-setting, action planning, and feedback in relation to goals. The strategies most likely to engage users were ease of use, design, tailoring of design and information, and unique smartphone features.

(JMIR mHealth uHealth 2015;3(2):e73) doi:10.2196/mhealth.3895

KEYWORDS

smartphone apps; alcohol consumption; consensus; Delphi technique; behavior change techniques

Introduction

Excessive alcohol consumption is a serious problem for population health [1,2]. Brief interventions to address this are time limited interventions delivered by health care workers targeting heavier drinkers and can be effective at reducing alcohol consumption [3]. There are substantial barriers to their delivery such as lack of time, training, and financial resources. These barriers can perhaps be avoided by delivering an intervention via a digital platform. While digital interventions have not been found to be as effective as face-to-face brief interventions [4], they may be more effective than no intervention [4-13], and have the advantage of being cost effective, avoid the stigma associated with help-seeking in person [10], and have greater reach than traditional health services. Smartphone applications or 'apps' have the additional advantage of being with the individual almost all of the time, which offers the potential to engage users in real time and in their everyday situations. Apps also have the ability to sense and report locations and events (in conjunction with calendar function) to provide moment-to-moment support when it is needed unlike traditional interventions. Despite a large number of apps to reduce excessive alcohol consumption in the general population, none, to our knowledge, have been rigorously evaluated. There has been a recent trial of an app on the related issue of recovery from alcoholism [14] that showed a reduction in the number of risky drinking days and therefore of probable benefit to patients in continuing care for alcohol dependence.

Reviews of digital interventions (not apps) suggest they can be effective, but there is substantial heterogeneity between different interventions [4,7,8,11,12]. Moreover, interventions have many components and their evaluations have rarely specified content in a way that would allow identification of the components responsible for the variation (e.g. [4,8,11,12]). A reliable method for specifying content and evaluating the effectiveness of complex behavior change interventions is to identify behavior change techniques (BCTs) [15]. BCTs are defined as the smallest, observable, replicable components with the potential to bring about change in behavior [16].

In order for an alcohol reduction app to be effective, it must be engaging for users, thus allowing them to be exposed to its active components. It is well established that a large proportion of users of digital interventions in health trials do not maintain engagement [17]. This degree of attrition undermines the potential of apps to be effective, and generalizable evaluation is made difficult when a large proportion of users cannot be recontacted due to disengagement with the intervention [18]. Engagement in Web-based interventions is increased by use of prompts [19-21], peer support [19], counselor support [19], and the combination of tailored communication with the use of reminders and incentives [22]. However, these have only been examined in the context of websites and there is a need to identify the most effective strategies for engagement with apps.

In sum, there is not yet an established evidence base to draw on to inform the selection of BCTs or engagement strategies in developing apps aimed at reducing alcohol consumption amongst the general population. In areas of research where there is a lack of, inconsistent, or contradictory scientific evidence, formal consensus methods have been used to guide action [23,24]. This study used a formal consensus building methodology with a small group of world-class experts in the field of alcohol and/or behavior change to identify intervention components judged to be the 'best bets' to reduce alcohol consumption (in general and in the context of an app) and to maintain engagement with an app, and then compared the original expert group's ranking of intervention components with a new, broader expert review.

This study addressed the following research questions:

1. What BCTs do experts in the field of alcohol research agree are most likely to be effective in general and when delivered by an app?
2. What engagement strategies do experts believe are most likely to be effective initially and over time?

Methods

First Phase: 3-Round Consensus Exercise

Study Design

A Delphi-style methodology was used to generate consensus among experts about what intervention components are likely to be the most effective at reducing alcohol consumption, and what strategies are most likely to improve engagement with an app. Experts were asked to generate a list of 'best bet' intervention components and engagement strategies which were subsequently rated and ranked.

The Delphi method of generating consensus was selected as a formal, systematic and reproducible method of arriving at a consensus. It was conducted anonymously to avoid biases produced by perceived authority, persuasion or bandwagon effects [23,25].

Participants

Seven international academic experts (six male) were purposively identified from a range of scientific networks and backgrounds (health psychology, biological psychology, developmental psychopathology and addiction research) on the basis of their knowledge of the alcohol literature, and/or experience of designing or delivering behavior change interventions. Seven participants are considered sufficient for reliable group judgment [24,25]. None of the experts were identified based on any user experience expertise. The authorship team used their experience to judge the suitability of invited experts. Once the experts were identified, each was formally approached by an email invitation. All the experts who were approached agreed to take part. Experts were from the UK (n=6) and the Netherlands (n=1). Six were professors and one was a senior research fellow.

Measures

Round 1:

Participants were asked to provide between three and five responses to each of three questions:

1. What intervention components do you believe would be the best bets for helping people reduce their alcohol consumption?
2. What intervention components do you believe would be the best bets for helping people to reduce their alcohol consumption when delivered by a smartphone app?
3. What do you think are the best strategies or techniques for maintaining engagement with an app aiming to help people reduce their alcohol consumption?

Each question was preceded by the statement: "Please answer the following questions based on your knowledge of the research literature, relevant theory and your clinical experience. Please also provide the reason behind your choice." For question 2, participants were given the option to indicate that their answers were the same as for question 1.

Round 2:

Participants were provided with an alphabetical list of the responses generated in the first round for each of the questions. They were instructed "Please rate your agreement with each of these techniques for the three different questions on the five-point Likert scales provided". The scale ranged from 1 (*strongly disagree*), 2 (*disagree*), 3 (*neither agree nor disagree*), 4 (*agree*) to 5 (*strongly agree*). Participants were given the option to make comments on their rating.

Round 3:

The *n* responses were listed alphabetically with the mean agreement rating and rationale provided for each response. Participants were asked to rank the *n* responses from 1 (*most likely to be a best bet*) to *n* (*least likely to be a best bet*) for each of the questions. At this stage, participants were only asked to rank responses about which there had been broad agreement in the previous round, defined as a minimum of four out of seven of the participants agreeing (i.e., rating of 4 or above) that the technique was likely to be either effective or engaging (depending upon the question) [23]. The reason for removing responses about which there was little agreement was to improve responding by minimizing the time required to complete the survey [23]. There was the option to make any final comments at this point.

Procedure

This study was conducted using the online survey tool Qualtrics. A link to the survey for each of the three rounds was emailed to the participants and they were given between one and two weeks to complete it. Non-responders were sent reminders until all participants had completed each round. Participants provided informed consent.

Analysis

Round 1:

For each question, similar responses were summarized and combined. For question 1, a BCT was selected from one of two taxonomies [15,26] to describe each response for the intervention components, where appropriate. The summarizing, combining and coding of responses was conducted by CG & SM.

Round 2:

The mean, standard deviation (SD), and mode of the agreement ratings for each response to each of the three questions were calculated.

Round 3:

The final rankings were analyzed by calculating Kendall's W coefficient of concordance [27], which measures the extent to which judges agree on their rankings of items. The value of W ranges from 0 (indicating no consensus) to 1 (indicating perfect consensus) between participants. A value of .1 corresponds to very weak agreement, .3 to weak agreement, .5 to moderate agreement, .7 to strong agreement and .9 to unusually strong agreement [28]. The Kendall's W statistic uses the χ^2test to test the independence of the ranking of the components.

Second Phase: External Validation

Study design

The intervention components generated and ranked in the first phase of the study were also ranked by a second group of experts in the field of alcohol.

Participants

Assistant and Senior Editors (n=179) from the journal Addiction were invited to take part in the study if they believed they had a sufficiently informed 'opinion on interventions that might help people who drink more alcohol than is good for them to reduce or quit'. This invitation yielded 43 participants.

Measures

Participants were asked to rank from 1 (highest) to 12 (lowest), the value of 12 responses generated in the first phase of the study by the original group of experts, in response to the question "What intervention components do you believe would be the best bets for helping people reduce their alcohol consumption?"

Procedure

An email was circulated to all the assistant and senior editors at the journal of Addiction with an alphabetical list of the "best bet" intervention component responses. If they wished to take part in the study, they were asked to reply (via email) with a ranking for each of the intervention components. Participants were given one week to reply before the study closed.

Analysis

The correlation between the rankings of the original and the new independent group of experts was assessed using Spearman's rank correlation coefficient. The new rankings were also analyzed using Kendall's W coefficient of concordance [27] to assess the extent to which this second group agreed with each other.

Results

First Phase: 3-Round Consensus Exercise

In response to the question of what intervention components are likely to be the most effective at reducing alcohol consumption, 24 responses were recorded in round 1. Eighteen

of these responses were similar to at least one other, resulting in 12 components (see Multimedia Appendix 1), of which 11 corresponded directly with a BCT (see Table 1). Six of the 7 participants thought that intervention components likely to be effective in general would be the same as in an app. The other participant generated one suggestion to do with the intervention modality itself and how to present the intervention in a unique way. The response was therefore included with the responses to the question regarding engagement strategies.

Four of the 12 components (self monitoring, goal setting, action planning, and feedback in relation to goals) had a mean ranking score greater than the average rank (6 out of 12) and the lowest mean agreement rating for these four BCTs was 4.3 (see Table 1). Overall the original group of experts displayed moderate agreement (Kendall's W=.465) in their ranking of intervention components ($\chi^2_{(11)}$=35.77, *P*<.001).

Table 1. Responses generated by the expert group on effective behaviour change techniques to reduce alcohol consumption.[a]

		Agreement rating[b]			Ranking score[c]	
Responses generated	Equivalent BCTs	Mean (SD)	Mode	Agree : Disagree[d]	Mean (SD)	Mode
Self monitoring	Self monitoring of behavior[e]	4.6 (.54)	5	7:0	2.4 (1.81)	1
Goal setting	Goal setting (behavior)[e]	4.7 (.049)	5	7:0	2.6 (1.51)	1, 2
Action planning	Action planning[e]	4.3 (.49)	4	7:0	4.3 (.95)	4
Feedback in relation to goals	Provide feedback on performance[f]	4.6 (.54)	5	7:0	4.43 (2.70)	3
Behavior substitution	Behavior substitution[f]	4.1 (.38)	4	7:0	6.3 (2.06)	5, 7
Environmental triggers and drivers	Advise on environmental restructuring[f]	3.9 (.69)	4	5:2	7.3 (4.07)	2, 9
Provide information	Provide information on consequences of excessive alcohol consumption & reducing excessive alcohol consumption[f]	4.0 (.58)	4	6:1	7.4 (4.47)	12
Feedback in relation to people	Provide normative information about others' behavior and experiences[f]	4.0 (.58)	4	6:1	8.4 (1.90)	7
Motivational interviewing	Conduct motivational interviewing[f]	3.9 (1.07)	4	5:2	8.4 (3.41)	12
Inhibition training		3.6 (.54)	4	4:3	8.4 (3.51)	10
Reward	Provide rewards contingent on successfully reducing excessive alcohol consumption[f]	3.9 (.69)	4	5:2	8.9 (2.12)	11
Habit reversal	Habit reversal[f]	3.4 (.79)	4	4:3	9.1 (1.68)	10

[a]Responses ordered in terms of mean ranking score (from round 3).

[b]Agreement rating (1: *strongly disagree*, 5: *strongly agree*).

[c]Ranking score (1: *highest*, 12: *lowest*).

[d]Agree:Disagree (ratio of (*agree/strongly agree*): (*neither/disagree/strongly disagree*) used as inclusion criteria for round 3.

[e]BCTs as referred to in the 93-item BCT Taxonomy v1 [15]

[f]BCTs as referred to in the 42-item excessive alcohol reduction specific taxonomy [26]

Of the 20 engagement strategies generated, six were similar to at least one other and thus were combined, which resulted in 17 unique strategies (see Multimedia Appendix 2 for the rationale for each of the 17 responses). Seven strategies (ease of use, design aesthetic, feedback, function, ability to change design to suit own preferences, tailored information and unique smartphone features) had a mean ranking score greater than average rank (8 out of 16) and the lowest mean agreement rating for these strategies was 3.6 (see Table 2). Overall the experts showed a moderate degree of consensus in their ranking of the strategies (Kendall's W=.563, χ^2_{15}=59.2, *P*<.001).

Table 2. Responses generated by the expert group on engagement strategies.[a]

	Agreement rating[b]			Ranking score[c]	
Responses	Mean (SD)	Mode	Agree:Disagree[d]	Mean (SD)	Mode
Ease of use	4.9 (.38)	5	7:0	1.4 (.79)	1
Design – aesthetic	4.6 (.54)	5	7:0	3.1 (1.57)	2, 5
Feedback	4.6 (.54)	5	7:0	3.9 (1.68)	4
Function	4.0 (.82)	4	5:2	6.6 (3.60)	11
Design – ability to change design to suit own preferences	3.6 (.79)	4	5:2	6.9 (4.74)	3
Tailored information	4.3 (.76)	4, 5	6:1	7.9 (3.39)	6, 7
Unique smartphone features	4.4 (.54)	4	7:0	7.9 (5.79)	6
Prompts	4.1 (.38)	4	7:0	8.4 (2.44)	8
Graded tasks	4.0 (.82)	4	5:2	8.7 (3.50)	12
Gamification	4.1 (.69)	4	6:1	8.9 (5.30)	10
Social comparison	3.9 (.69)	4	5:2	10.4 (3.36)	9
Reward type Novelty	4.0 (.82)	4	5:2	11.6 (2.23)	12
Reward type Games	3.7 (.49)	4	5:2	11.9 (2.97)	11, 15
Reward type Positive messages	4.0 (.58)	4	6:1	12.1 (2.79)	8, 10, 11, 12, 13, 15, 16
Reward type Financial	3.6 (.98)	4	4:3	12.3 (1.98)	13
Social connectivity	4.0 (.58)	4	6:1	14.1 (1.95)	15, 16
Reward type- cue signaling reward[e]	3.4 (.98)	3	3:4	-	

[a]Responses ordered in terms of mean ranking score (from round 3).

[b]Agreement rating (1: *strongly disagree*, 5: *strongly agree*).

[c]Ranking score (1: *highest*, 16: *lowest*).

[d]Agree:Disagree (ratio of (*agree/strongly agree*): (*neither/disagree/strongly disagree*) used as inclusion criteria for round 3.

[e]This response was not included in round 3 because there was not substantive agreement that it would be an effective engagement strategy in round 2 (defined as a minimum of 4 out of 7 of the participants agreeing (i.e., rating of 4 or above) that the technique was likely to be engaging).

Second Phase: External Validation

The ranking of the BCTs by the original group was validated by an independent group of experts: there was a significant correlation between their two rankings (see Table 3; ρ=.69, *P*=.01). Table 3 shows the ranking by the independent group of experts of the intervention components generated and agreed by the original group. There was modest but significant agreement amongst the broader group of experts (Kendall's W=.320, χ^2_{11}=151.52, *P*<.001).

Table 3. Comparison between rankings of phase 1 expert group and larger expert group of effective behavior change techniques for alcohol use reduction.[a]

Responses	Phase 1 experts N=7 Mean Rank (SD)	Phase 2 experts N=43 Mean Rank (SD)
Self monitoring	2.4 (1.81)	3.4 (2.88)
Goal setting	2.6 (1.51)	3.8 (3.00)
Action planning	4.3 (.95)	6.4 (2.72)
Feedback in relation to goals	4.4 (2.70)	4.1 (2.28)
Behavior substitution	6.3 (2.06)	7.6 (2.51)
Environmental triggers and drivers	7.3 (4.07)	5.1 (2.72)
Provide information	7.4 (4.47)	9.5 (2.87)
Feedback in relation to people	8.4 (1.90)	7.4 (3.27)
Motivational interviewing	8.4 (3.41)	7.2 (2.82)
Inhibition training	8.4 (3.51)	8.8 (2.15)
Reward	8.9 (2.12)	6.8 (3.44)
Habit reversal	9.1 (1.68)	7.9 (2.69)

[a]Responses ordered in terms of mean ranking score for the original experts (from round 3)

Discussion

BCTs of self monitoring, goal setting, action planning, and feedback in relation to goals were ranked most likely to be effective for reducing alcohol use by a group of international experts in the field of alcohol or behavior change or both. This finding was validated by a larger independent group of alcohol experts. None of the experts thought that the BCTs likely to be effective in general would differ from those in an app, though one participant suggested presenting information in a way that was unique to an app. The most highly ranked engagement strategies were ease of use, design-aesthetic, feedback, function, design-ability to change design to suit own preferences, tailored information and unique smartphone features.

There is empirical evidence for the effectiveness of some of the BCTs identified in this study for reducing excessive alcohol consumption. Self monitoring has been found to be effective in brief interventions [26], and is also used in a number of apps to reduce alcohol consumption [29] though none of these have been evaluated. The BCT 'feedback in relation to people' is often referred to as normative feedback in the alcohol behavior change literature. There is evidence to suggest that this BCT may have a small effect by several different modes of delivery: face-to-face [30], via phone [31], mailed [32,33] and via digital platforms [30,34,35]. However, this research is often limited to college and university students [30,32,34,35]. The highest priority engagement strategies of prompts, social connectivity and tailored information have all been shown to result in increased use of Web-based interventions [19-22].

The use of a Delphi approach to selecting intervention components is clearly not guaranteed to result in the best choices, but on a priori grounds it seems preferable to the more usual practice of drawing on expertise and interest within a single research team. It may have been that no consensus would be achieved so, while the level of agreement within each group of experts was modest, the fact that the aggregate rankings of the two expert groups showed a high level of concordance was reassuring that the study tapped into a shared perspective on the existing evidence.

It is possible that the results of the Delphi exercise could have been biased by choosing an expert group with similar backgrounds to those of the research team. Therefore, the use of a second group of experts to validate the rankings provided important support for this not being the case. The journal Addiction has a very large pool of international experts on its editorial team and arguably includes most of the leading researchers in the field covering a wide range of expertise. The question regarding user engagement was included for exploratory purposes. As shown in this study, experts in the academic field of research did not identify any BCTs as being effective for an app compared with a traditional intervention. This may be because they are not aware of the additional functions an app can provide in terms of a behavior change intervention. Future research is planned to compare the views of experts in the relevant academic field with that of user experience experts to see if there are any discrepancies between these groups and if so, how their opinions differ.

The results of this study will be used to inform the building of a prototype app that will be evaluated in a field experiment. Following the principle of optimization [36] each component will be included in a full form or minimal form using a factorial design so that its effect can be assessed. The findings should also be useful to other research teams considering developing and evaluating apps in this area.

Acknowledgments

CG is funded by the UK Centre for Tobacco and Alcohol Studies (UKCTAS). DC is funded by the National Institute of Health Research, School for Public Health Research (NIHR SPHR). JB's post is funded by a fellowship from the UK Society for the Study of Addiction (SSA). RW is funded by Cancer Research UK. The research team is part of the UKCTAS and we are grateful to UKCTAS, NIHR SPHR and SSA for funding the study.

Conflicts of Interest

JB has received an unrestricted research grant from Pfizer related to the surveillance of smoking cessation trends. RW has received research funding and undertaken consultancy for companies that manufacture smoking cessation medications. CG, DC and SM have no declared conflicts of interest.

Multimedia Appendix 1

Intervention components generated by the experts in the first round.

[PDF File (Adobe PDF File), 37KB - mhealth_v3i2e73_app1.pdf]

Multimedia Appendix 2

Engagement strategies generated by the experts in the first round.

[PDF File (Adobe PDF File), 34KB - mhealth_v3i2e73_app2.pdf]

References

1. World Health Organization. World Health Organization. 2011. Global status report on alcohol and health URL: http://www.who.int/substance_abuse/publications/global_alcohol_report/msbgsruprofiles.pdf [accessed 2015-06-07] [WebCite Cache ID 6Z76jjvHt]
2. Balakrishnan R, Allender S, Scarborough P, Webster P, Rayner M. The burden of alcohol-related ill health in the United Kingdom. J Public Health (Oxf) 2009 Sep;31(3):366-373 [FREE Full text] [doi: 10.1093/pubmed/fdp051] [Medline: 19493915]
3. Kaner EFS, Dickinson HO, Beyer F, Pienaar E, Schlesinger C, Campbell F, et al. The effectiveness of brief alcohol interventions in primary care settings: a systematic review. Drug Alcohol Rev 2009 May;28(3):301-323. [doi: 10.1111/j.1465-3362.2009.00071.x] [Medline: 19489992]
4. Elliott JC, Carey KB, Bolles JR. Computer-based interventions for college drinking: a qualitative review. Addict Behav 2008 Aug;33(8):994-1005 [FREE Full text] [doi: 10.1016/j.addbeh.2008.03.006] [Medline: 18538484]
5. Cunningham JA, Humphreys K, Koski-Jännes A, Cordingley J. Internet and paper self-help materials for problem drinking: is there an additive effect? Addict Behav 2005 Sep;30(8):1517-1523. [doi: 10.1016/j.addbeh.2005.03.003] [Medline: 15893433]
6. Bewick BM, Trusler K, Barkham M, Hill AJ, Cahill J, Mulhern B. The effectiveness of web-based interventions designed to decrease alcohol consumption--a systematic review. Prev Med 2008 Jul;47(1):17-26. [doi: 10.1016/j.ypmed.2008.01.005] [Medline: 18302970]
7. Carey KB, Scott-Sheldon LAJ, Elliott JC, Bolles JR, Carey MP. Computer-delivered interventions to reduce college student drinking: a meta-analysis. Addiction 2009 Nov;104(11):1807-1819 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02691.x] [Medline: 19744139]
8. Carey KB, Scott-Sheldon LAJ, Elliott JC, Garey L, Carey MP. Face-to-face versus computer-delivered alcohol interventions for college drinkers: a meta-analytic review, 1998 to 2010. Clin Psychol Rev 2012 Dec;32(8):690-703 [FREE Full text] [doi: 10.1016/j.cpr.2012.08.001] [Medline: 23022767]
9. Khadjesari Z, Murray E, Hewitt C, Hartley S, Godfrey C. Can stand-alone computer-based interventions reduce alcohol consumption? A systematic review. Addiction 2011 Feb;106(2):267-282 [FREE Full text] [doi: 10.1111/j.1360-0443.2010.03214.x] [Medline: 21083832]
10. Riper H, Spek V, Boon B, Conijn B, Kramer J, Martin-Abello K, et al. Effectiveness of E-self-help interventions for curbing adult problem drinking: a meta-analysis. J Med Internet Res 2011;13(2):e42 [FREE Full text] [doi: 10.2196/jmir.1691] [Medline: 21719411]
11. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. Addiction 2010 Aug;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x] [Medline: 20528806]
12. Vernon ML. A review of computer-based alcohol problem services designed for the general public. J Subst Abuse Treat 2010 Apr;38(3):203-211 [FREE Full text] [doi: 10.1016/j.jsat.2009.11.001] [Medline: 20015607]

13. 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]
14. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
15. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
16. Michie S, Abraham C, Eccles MP, Francis JJ, Hardeman W, Johnston M. Strengthening evaluation and implementation by specifying components of behaviour change interventions: a study protocol. Implement Sci 2011 Jan;6(1):10 [FREE Full text] [doi: 10.1186/1748-5908-6-10] [Medline: 21299860]
17. Eysenbach G. The law of attrition. J Med Internet Res 2005 Jan;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
18. Eysenbach G. Issues in evaluating health websites in an Internet-based randomized controlled trial. J Med Internet Res 2002 Dec;4(3):E17 [FREE Full text] [doi: 10.2196/jmir.4.3.e17] [Medline: 12554548]
19. Brouwer W, Kroeze W, Crutzen R, de NJ, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
20. Schneider F, van Osch L, Schulz DN, Kremers SP, de Vries H. The influence of user characteristics and a periodic email prompt on exposure to an internet-delivered computer-tailored lifestyle program. J Med Internet Res 2012;14(2):e40 [FREE Full text] [doi: 10.2196/jmir.1939] [Medline: 22382037]
21. Svensson M, Svensson T, Hansen AW, Lagerros YT. The effect of reminders in a web-based intervention study. Eur J Epidemiol 2012 May;27(5):333-340. [doi: 10.1007/s10654-012-9687-5] [Medline: 22531973]
22. Crutzen R, de Nooijer J, 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]
23. Osborne J, Collins S, Ratcliffe M, Millar R, Duschl R. What "ideas-about-science" should be taught in school science? A Delphi study of the expert community. J Res Sci Teach 2003 Aug 25;40(7):692-720. [doi: 10.1002/tea.10105]
24. Black N, Murphy M, Lamping D, McKee M, Sanderson C, Askham J, et al. Consensus development methods: a review of best practice in creating clinical guidelines. J Health Serv Res Policy 1999 Oct;4(4):236-248. [Medline: 10623041]
25. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. Health Technol Assess 1998;2(3):1-88 [FREE Full text] [Medline: 9561895]
26. Michie S, Whittington C, Hamoudi Z, Zarnani F, Tober G, West R. Identification of behaviour change techniques to reduce excessive alcohol consumption. Addiction 2012 Aug;107(8):1431-1440. [doi: 10.1111/j.1360-0443.2012.03845.x] [Medline: 22340523]
27. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. Information & Management 2004 Dec;42(1):15-29. [doi: 10.1016/j.im.2003.11.002]
28. Schmidt RC. Managing Delphi Surveys Using Nonparametric Statistical Techniques. Decision Sciences 1997 Jul;28(3):763-774. [doi: 10.1111/j.1540-5915.1997.tb01330.x]
29. 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]
30. Foxcroft DR, Moreira MT, Santimano NMLA, Smith LA. Social norms information for alcohol misuse in university and college students. Cochrane Database Syst Rev 2015;1:CD006748. [doi: 10.1002/14651858.CD006748.pub3] [Medline: 25622306]
31. Cunningham JA, Neighbors C, Wild TC, Humphreys K. Normative misperceptions about alcohol use in a general population sample of problem drinkers from a large metropolitan city. Alcohol Alcohol 2012;47(1):63-66 [FREE Full text] [doi: 10.1093/alcalc/agr125] [Medline: 22028458]
32. Collins SE, Carey KB, Sliwinski MJ. Mailed personalized normative feedback as a brief intervention for at-risk college drinkers. J Stud Alcohol 2002 Sep;63(5):559-567. [Medline: 12380852]
33. Cunningham JA, Wild TC, Bondy SJ, Lin E. Impact of normative feedback on problem drinkers: a small-area population study. J Stud Alcohol 2001 Mar;62(2):228-233. [Medline: 11327189]
34. Kypri K, Langley JD. Perceived social norms and their relation to university student drinking. J Stud Alcohol 2003 Nov;64(6):829-834. [Medline: 14743946]
35. Neighbors C, Larimer ME, Lewis MA. Targeting misperceptions of descriptive drinking norms: efficacy of a computer-delivered personalized normative feedback intervention. J Consult Clin Psychol 2004 Jun;72(3):434-447. [doi: 10.1037/0022-006X.72.3.434] [Medline: 15279527]

36. 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]

Abbreviations

BCTs: behavior change techniques
NIHR SPHR: National Institute of Health Research, School for Public Health Research
SSA: UK Society for the Study of Addiction
UKCTAS: UK Centre for Tobacco and Alcohol Studies

Edited by G Eysenbach; submitted 26.09.14; peer-reviewed by D Spruijt-Metz, K Donoghue; comments to author 20.02.15; revised version received 20.03.15; accepted 23.03.15; published 29.06.15.

Please cite as:
Garnett C, Crane D, West R, Brown J, Michie S
Identification of Behavior Change Techniques and Engagement Strategies to Design a Smartphone App to Reduce Alcohol Consumption Using a Formal Consensus Method
JMIR mHealth uHealth 2015;3(2):e73
URL: http://mhealth.jmir.org/2015/2/e73/
doi:10.2196/mhealth.3895
PMID:26123578

Original Paper

Effectiveness of a mHealth Lifestyle Program With Telephone Support (TXT2BFiT) to Prevent Unhealthy Weight Gain in Young Adults: Randomized Controlled Trial

Stephanie R Partridge[1], BSc (Hons), APD; Kevin McGeechan[2], PhD; Lana Hebden[1], PhD; Kate Balestracci[1], RD, MS; Annette TY Wong[1], PhD; Elizabeth Denney-Wilson[3], PhD; Mark F Harris[4], PhD, FRACGP; Philayrath Phongsavan[2], PhD; Adrian Bauman[2], PhD, FAFPH; Margaret Allman-Farinelli[1], PhD, FDAA

[1]School of Molecular Bioscience, Charles Perkins Centre, University of Sydney, Sydney, Australia

[2]Sydney School of Public Health, Charles Perkins Centre, University of Sydney, Sydney, Australia

[3]Faculty of Health, University of Technology Sydney, Sydney, Australia

[4]Centre for Primary Health Care and Equity, University of New South Wales, Sydney, Australia

Corresponding Author:
Margaret Allman-Farinelli, PhD, FDAA
School of Molecular Bioscience
Charles Perkins Centre
University of Sydney
Level 4 East
Charles Perkins Centre
Sydney, 2006
Australia
Phone: 61 2 9036 7045
Fax: 61 2 8627 1605
Email: margaret.allman-farinelli@sydney.edu.au

Abstract

Background: Weight gained in young adulthood often persists throughout later life with associated chronic disease risk. Despite this, current population prevention strategies are not specifically designed for young adults.

Objective: We designed and assessed the efficacy of an mHealth prevention program, TXT2BFiT, in preventing excess weight gain and improving dietary and physical activity behaviors in young adults at increased risk of obesity and unhealthy lifestyle choices.

Methods: A two-arm, parallel-group randomized controlled trial was conducted. Subjects and analyzing researchers were blinded. A total of 250 18- to 35-year-olds with a high risk of weight gain, a body mass index (BMI) of 23.0 to 24.9 kg/m^2 with at least 2 kg of weight gain in the previous 12 months, or a BMI of 25.0 to 31.9 kg/m^2 were randomized to the intervention or control group. In the 12-week intervention period, the intervention group received 8 text messages weekly based on the transtheoretical model of behavior change, 1 email weekly, 5 personalized coaching calls, a diet booklet, and access to resources and mobile phone apps on a website. Control group participants received only 4 text messages and printed dietary and physical activity guidelines. Measured body weight and height were collected at baseline and at 12 weeks. Outcomes were assessed via online surveys at baseline and at 12 weeks, including self-reported weight and dietary and physical activity measures.

Results: A total of 214 participants—110 intervention and 104 control—completed the 12-week intervention period. A total of 10 participants out of 250 (4.0%)—10 intervention and 0 control—dropped out, and 26 participants (10.4%)—5 intervention and 21 control—did not complete postintervention online surveys. Adherence to coaching calls and delivery of text messages was over 90%. At 12 weeks, the intervention group were 2.2 kg (95% CI 0.8-3.6) lighter than controls (*P*=.005). Intervention participants consumed more vegetables (*P*=.009), fewer sugary soft drinks (*P*=.002), and fewer energy-dense takeout meals (*P*=.001) compared to controls. They also increased their total physical activity by 252.5 MET-minutes (95% CI 1.2-503.8, *P*=.05) and total physical activity by 1.3 days (95% CI 0.5-2.2, *P*=.003) compared to controls.

Conclusions: The TXT2BFiT low-intensity intervention was successful in preventing weight gain with modest weight loss and improvement in lifestyle behaviors among overweight young adults. The short-term success of the 12-week intervention period shows potential. Maintenance of the behavior change will be monitored at 9 months.

Trial Registration: Trial Registration: The Australian New Zealand Clinical Trials Registry ACTRN12612000924853; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=12612000924853 (Archived by WebCite at http://www.webcitation.org/6Z6w9LlS9).

(JMIR mHealth uHealth 2015;3(2):e66) doi:10.2196/mhealth.4530

KEYWORDS

young adults; weight gain prevention; lifestyle behavior; mHealth

Introduction

More than 35% of adults globally are overweight or obese, and in developed countries the peak prevalence of obesity is moving to younger ages [1]. For example, younger Americans and Australians are gaining more weight than any other adult age group [2-4]. As body mass index (BMI) exceeds 23 kg/m^2, risks of cardiovascular disease, certain cancers, diabetes, osteoarthritis, and chronic kidney disease increase [1]. The Coronary Artery Risk Development in Young Adults (CARDIA) cohort study reported that weight maintenance over time (both normal weight and overweight) in young adults protects against cardiovascular risk, but weight gain increases the risk [5]. Thus, interventions focused on prevention of weight gain in overweight young adults may help prevent obesity and its associated health consequences [6].

Coordinated prevention approaches aimed at improving detrimental lifestyle behaviors have been proposed to prevent obesity [7,8]. Compared with other age groups, young adults eat the least amount of fruits and vegetables [9,10], drink the most sugar-sweetened beverages (SSB) [11], more frequently eat food prepared outside the home (ie, takeout food) [12], and demonstrate declines in physical activity [13-15]. These adverse behavioral lifestyle choices predict excessive weight gain and increased risk of chronic disease later in life [16].

Several recent prevention programs have shown short-term efficacy in young adults to prevent further weight gain [17], but few investigated the use of mHealth (ie, mobile or cellular phone) technology. Advantages of such technology include its wide reach and, once created, its low costs compared with health professional time. The 18- to 29-year-old age group is also the most likely age group to own a mobile phone, with 83% ownership in the US [18]. Interventions delivered via short message service (SMS) text messaging show promise in positively impacting health-related behavior change [19,20]. Our previous pilot study demonstrated the feasibility of delivering an mHealth lifestyle program [21]. Participants in the intervention group decreased their body weight and SSB intake and increased their physical activity and vegetable consumption, although changes were not significant. Qualitative feedback facilitated improvements to the program and informed the development of the TXT2BFiT mHealth program aimed at improving weight management and weight-related dietary and physical activity behaviors among young adults.

Here we report on the efficacy of a randomized controlled trial (RCT) of a larger mHealth lifestyle program, TXT2BFiT, among young adults deemed at high risk for development of obesity. We hypothesized that compared with young adults assigned to a control condition, those who received the TXT2BFiT mHealth intervention would maintain or lose a modest amount of weight and improve lifestyle behaviors.

Methods

Overview

The TXT2BFiT study is a two-arm, parallel-design RCT in 18- to 35-year-olds recruited from the Greater Sydney Area, NSW, Australia, between November 2012 and July 2014. All study materials were designed specifically for use in this study only. The trial was approved by the University Human Research Ethics Committee in September 2012 (approval number 15226) and all the participants gave written informed consent. The trial is registered with the Australian New Zealand Clinical Trials Registry (ACTRN12612000924853). Both the protocol and recruitment methods have been previously published [22]. A concise description appears below.

Subjects

Participants who responded to recruitment materials were directed to complete an online screener survey. Eligible participants were deemed at risk of excess weight gain if they met the following inclusion criteria: had a BMI of 25.0 to 31.9 kg/m^2, or 23.0 to 24.9 kg/m^{2}with reported weight gain greater than 2 kg over the previous 12 months; had a fruit intake of less than two servings daily; had a vegetable intake of less than five servings daily; had an SSB intake of at least 1 L weekly; had energy-dense meals prepared away from home (ie, takeout food) more than once per week; and/or engaged in moderate-intensity physical activity of less than 60 minutes daily. Individuals were excluded if they were pregnant or planning to fall pregnant within the study period, were enrolled in an alternate weight loss program, had lost greater than 10 kg in the past 3 months, taken medications that have caused weight gain of greater than 2 kg, had medical conditions that precluded following dietary or physical activity recommendations, and/or did not speak English. Participants were also required to have a mobile phone capable of receiving text messages and accessing the Internet at least once a week.

Based on our previous meta-analysis [6], it appeared that a difference of 1.7 kg could be expected. The sample size required

for detection of a difference of 2 kg with 80% power, significance level of .05, 10 kg standard deviation, and a correlation between baseline and final weight of .8, was 354 subjects after allowing for a 20% dropout rate. Due to a slower-than-expected recruitment rate, and with time and funding constraints, recruitment was stopped at 250 participants.

Recruitment

Recruitment occurred via letters of invitation from participating general practitioners (GPs) (ie, primary care physicians) in two Medicare Locals—Australian primary health care services units responsible for coordinating care over specified geographic areas—or via electronic or print advertisements, including Facebook and Google (ie, social media and advertising), university electronic newsletters, printed posters, mailbox drops, and newspapers. Participants provided informed written consent. Young adults were compensated for their participation by receiving gift vouchers for completing a 12-week online survey and attending an in-person weigh-in.

Randomization

A random sequence was generated by an independent researcher and concealed from those responsible for enrolling participants into the intervention arm. Eligible participants were randomized in a 1:1 ratio into intervention and control arms. Randomization was based on a stratified randomized block design, where the strata were the GP clinic and participant gender. While participants were aware of another arm to the trial, every attempt was made to ensure that the nature of this other arm was not revealed.

Measurements

Demographic characteristics were collected by online survey and included age, gender, postcode (for categorizing socioeconomic status [23]), ethnicity (language spoken at home [24]), education level [24], and income in Australian Dollars (AUD) [24]. Body weight (kg) and height (cm) data were collected to calculate BMI (kg/m^2) at baseline via both measured and self-report methods. Participants' GPs used a standardized protocol to measure body weight to the nearest 0.1 kg and height to the nearest 0.1 cm at baseline [25]. Participants in both arms were invited for an optional in-person body weight (kg) and height (cm) measurement at the University Metabolic Facility within a 2-week window following the 12-week intervention completion (ie, weeks 13 and 14). Measures were taken by two higher-degree research students blinded to participant allocation.

Online surveys were administered at baseline and within a 2-week window following the 12-week intervention completion (ie, weeks 13 and 14). Data collected included self-reported weight (kg) and height (cm); short categorical questions to assess usual weekly intake of SSB [26], daily intake of fruits and vegetables [26], and weekly takeout meals [27]; and questions about physical activity in the previous 7 days using the short-form International Physical Activity Questionnaire (IPAQ) [28]. The IPAQ was scored using established methods [29] and data were reported as a continuous measure in metabolic equivalent of task (MET)-minutes per week. All data were reported by participants via the online deidentified survey website, SurveyMonkey, from which data were downloaded for analysis.

Engagement with the intervention was assessed using text message replies and number of coaching calls completed. Intervention participants were asked to reply "OK" to 16 messages in the 12 weeks and control participants were asked to reply "OK" to all 4 text messages. Text message delivery reports were created from the text message service provider, My MessageMedia, for delivery status and replies. Detailed records of all coaching calls were collated in a database. The 12-week postintervention survey also asked participants about their access to, and use of, program materials.

TXT2BFiT Program

The 12-week intervention program comprised the following: 8 weekly motivational text messages based on the transtheoretical model of behavior change, whereby messages were matched to stage-of-change for each of the individual lifestyle behaviors; 5 personalized coaching calls; weekly emails; and password-protected access to purpose-designed mobile phone apps that provided education and allowed self-monitoring [30], community blog, and support resources available on a password-protected website designed for the study [31] (see Figure 1). Support resources included “easy, healthy eating on a budget,” “emergency meal tool kit,” “meal planning worksheet,” “commit yourself: physical activity planner,” “tips for take-out meals,” “seasonal guide to fruit and vegetables,” and “staying healthy over the holidays.” Text messages were scheduled by two higher-degree research students. The text messages, based on the transtheoretical model of behavior change [32], consisted of 2 per week for each of the four behaviors—SSB, fruits and vegetables, physical activity, and food prepared away from home/takeout—for a total of 8 messages, weekly, tailored to the participant's stage of readiness to change [22] and sent using the My MessageMedia program. Two accredited practicing dietitians conducted the coaching calls according to a standardized protocol and allowed the participants to set goals, and to discuss barriers, enablers, and their progress. Each call lasted approximately 10 to 15 minutes, with 25 minutes allocated for the initial coaching call. The mobile phone apps were educational, for example, providing nutritional information on SSB and takeout meals, providing serving sizes for fruits and vegetables, and allowing self-monitoring of participants' behavior. One email was sent each week reiterating the information in the text messages and included links to the mobile phone apps to remind participants.

Intervention participants were also mailed a printed 18-page booklet containing the two-page control handout summarizing the Australian National Dietary and Physical Activity Guidelines [33,34]. Additional information included sample meal plans, recommendations for daily servings from the core food groups with example serving sizes [34], and information about the four target behaviors addressed by the program—physical activity and sedentary behavior, intake of fruits and vegetables, intake of energy-dense takeout meals prepared away from home, and SSB intake.

Control participants received the mailed two-page handout, the introductory call at week 0 to introduce the program (no

coaching given), 4 text messages (one every 3 weeks, during weeks 1 to 12) that restated information in the handout, and access to a website with only electronic versions of the two-page handout, consent form, study information statement, and contact information.

Figure 1. TXT2BFiT program screenshots.

Statistical Analysis

The primary outcomes, body weight (kg) and BMI (kg/m^2) at 12 weeks, were compared between the two groups using analysis of covariance models adjusting for baseline values, GP clinic, and gender. Secondary outcomes that were continuous—physical activity MET-minutes and physical activity days—were also analyzed using analysis of covariance models. Robust regression models were used for analyses where residuals indicated nonnormality. Secondary outcomes that were categorical—fruit and vegetable servings per day, SSB consumption per week, and energy-dense takeout meal intake per week—were analyzed using Mantel-Haenszel chi-square tests stratified by GP clinic and gender. The analysis used the "intention-to-treat" principle with multiple imputations to account for missing data. Five imputed datasets were created and the results for continuous outcomes pooled using Rubin's rules. Chi-square statistics were pooled, and *P* values estimated, using the method described by Li et al [35]. A *P* value <.05 was considered statistically significant. Researchers analyzing participant outcomes were blinded to participant allocation.

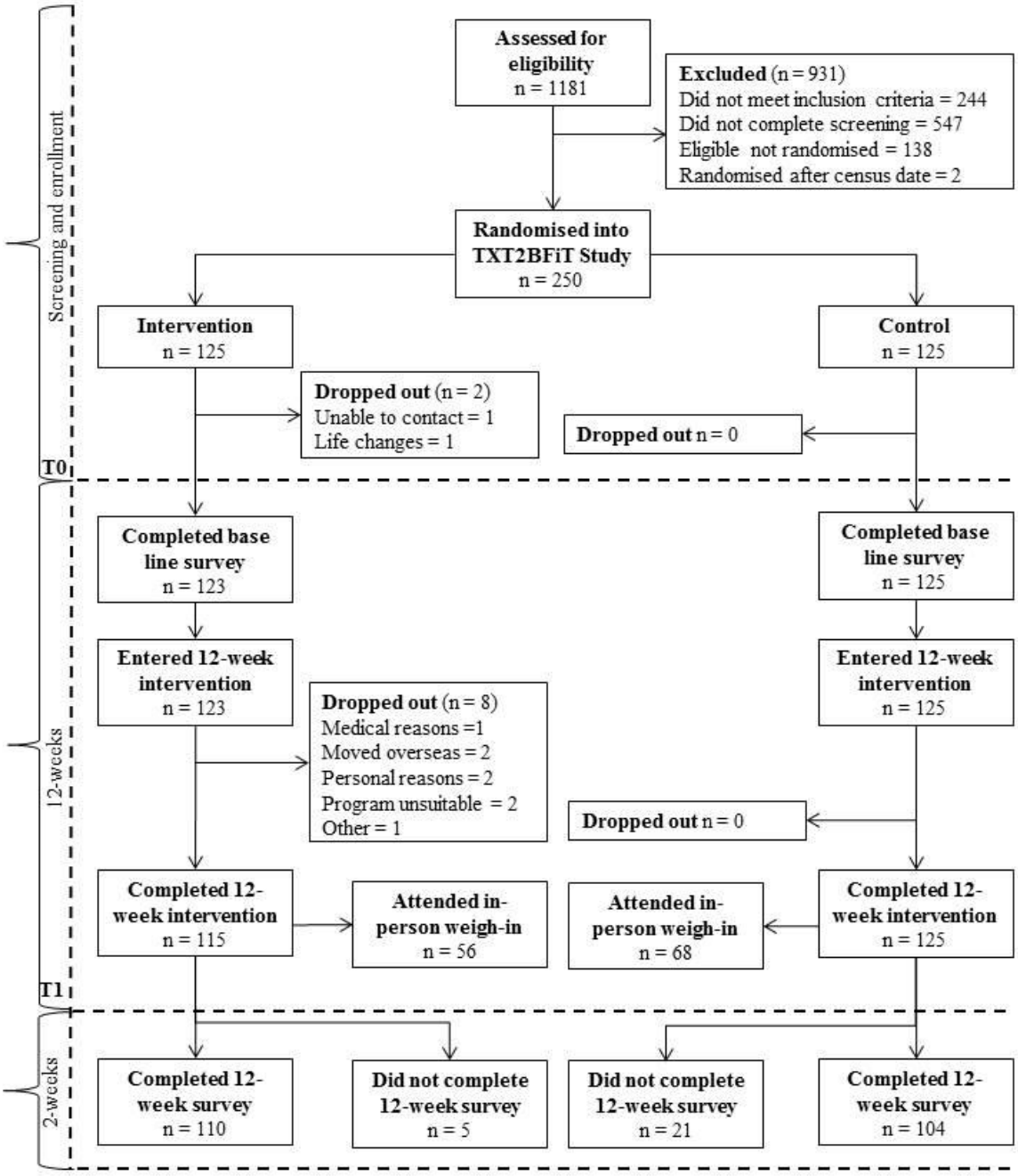
We also compared baseline characteristics and baseline primary or secondary outcomes between completers and noncompleters, and between in-person weigh-in attenders and nonattenders, using chi-square tests for categorical variables and independent-sample *t* tests for continuous variables. We compared self-reported weight and BMI with measured values using paired *t* tests. Analyses were performed using SPSS version 22.0 (IBM Corp, Armonk, NY, USA), Stata Statistical Software: Release 13 (Stata Corp, College Station, TX, USA), and SAS version 9.2 (SAS Institute Inc, Cary, NC, USA) on the full intention-to-treat sample.

Results

Participant Flow and Attrition

Recruitment resulted in 1181 enquires, of which 78.83% (931/1181) were excluded or failed to complete screening requirements (see Figure 2). A total of 250 young adults were randomly assigned to the intervention or control group. A total of 10 participants out of 125 (8.0%) dropped out of the intervention group during the 12-week intervention. Reasons for dropping out were as follows: 1 for lack of contact, 1 for life changes, 1 for medical reasons, 2 for personal reasons, 2 moved overseas, 2 found the program unsuitable, and 1 for other reasons not stated. An additional 5 participants out of 125 (4.0%) failed to complete the postintervention online surveys in the intervention group (see Figure 2). No participants dropped out of the control group, and 21 participants out of 125 in the control group (16.8%) did not complete the postintervention online surveys. Completers and noncompleters did not differ significantly in allocation, baseline demographic characteristics, or baseline primary or secondary outcomes (*P*>.11), except for noncompleters who consumed more takeout meals at baseline (*P*=.004). Nearly half of all participants (124/250, 49.6%) accepted the invitation for in-person weight and height measurements—intervention, 56/125, 44.8%; control, 68/125, 54.4%. There were no significant differences in baseline characteristics between participants that attended the in-person weight and height measurements and those that did not (*P*>.34), except that those attending ate less fruit at baseline (*P*=.03).

Figure 2. Flow diagram of participants in the TXT2BFiT study from week 0 to week 12.



Baseline Characteristics

Baseline characteristics of participants are shown in Table 1. Participants in the total randomized sample were mostly older (30 years or older, 107/248, 43.1%), female (152/248, 61.3%), English-speaking only (172/248, 69.4%), highly educated (153/248, 61.7%), and living in a socioeconomically advantaged area (187/248, 75.4%). Participants were overweight on the basis of BMI classification (intervention, 27.3 kg/m^2; control, 27.1 kg/m^2) (see Tables 2 and 3). Table 4 shows that, by design, most participants did not meet the recommended servings of fruit (intervention, 82/123, 66.7%; control, 77/125, 61.6%) or servings of vegetables (intervention, 116/123, 94.3%; control, 121/125, 96.8%) [34]; consumed more than 1 L of SSB per week (intervention, 16/123, 13.0%; control, 22/125, 17.6%); and consumed two or more takeout meals per week (intervention, 75/123, 61.0%; control, 79/125, 63.2%). All participants reported above-average levels of recommended physical activity [36] (intervention, 1619.9 MET-minutes per week; control, 1646.8 MET-minutes per week) (see Table 5).

Table 1. Baseline demographic characteristics for all randomized participants in the TXT2BFiT study by allocation (n=248)[a].

Characteristic	Intervention group (n=123)[a], mean (SD) or n (%)	Control group (n=125), mean (SD) or n (%)
Age in years, mean (SD)	28.1 (4.9)	27.2 (4.9)
Gender, n (%)		
Male	50 (40.7)	46 (36.8)
Female	73 (59.3)	79 (63.2)
SES [b] quintile, n (%)		
0-60[c]	8 (6.5)	7 (5.6)
61-80	28 (22.8)	17 (13.6)
81-100 (highest)	87 (70.7)	101 (80.8)
Ethnicity, n (%)		
English speaking	82 (66.7)	90 (72.0)
European	14 (11.4)	11 (8.8)
Asian	19 (15.4)	19 (15.2)
Other[d]	8 (6.5)	5 (4.0)
Education level, n (%)		
High school or below	27 (22.0)	21 (16.8)
Some university or technical school	22 (17.8)	25 (20.0)
University bachelor degree or higher	74 (60.2)	79 (63.2)
Weekly income (AUD [e]), n (%)		
Nil or negative	9 (7.3)	13 (10.4)
$1-499	36 (29.3)	30 (24.0)
$500-999	19 (15.4)	25 (20.0)
$1000-1499	36 (29.3)	26 (20.8)
$1500-1999	14 (11.4)	22 (17.6)
≥ $2000	9 (7.3)	9 (7.2)

[a]All participants had measured variables including 2 participants who did not complete baseline self-report surveys.

[b]Socioeconomic status (SES).

[c]Bottom-three SES quintiles collapsed.

[d]Pacific Islander and Arabic ethnicities collapsed.

[e]Australian Dollar (AUD).

Table 2. Effect of the TXT2BFiT program on measured weight and BMI outcomes for all randomized participants in the study by allocation (intention-to-treat analysis) (n=250).

Measured variable	Intervention group (n=125)[a], mean (SD)		Control group (n=125), mean (SD)		Model β[b] (95% CI)	*P*
	Baseline	12 weeks	Baseline	12 weeks		
Body weight, measured in kg	78.3 (11.4)	76.4 (11.1)	79.3 (12.7)	79.5 (12.4)	2.2 (0.8-3.6)	.005
BMI[c], measured in kg/m^2	27.3 (2.4)	26.4 (1.9)	27.1 (2.7)	26.8 (2.2)	0.5 (0.1-1.0)	.02

[a]All participants had measured variables including 2 participants who did not complete baseline self-report surveys.

[b]Model coefficients and *P* values were obtained from analysis of covariance models adjusting for baseline values, general practitioner clinic, and gender. Missing baseline and follow-up values were imputed to create five datasets and results were pooled using Rubin's rules.

[c]Body mass index (BMI).

Table 3. Effect of the TXT2BFiT program on self-reported weight and BMI outcomes for all randomized participants in the study by allocation (intention-to-treat analysis) (n=248)[a].

Self-reported variable	Intervention group (n=123)[a], mean (SD)		Control group (n=125), mean (SD)		Model β[b] (95% CI)	*P*
	Baseline	12 weeks	Baseline	12 weeks		
Body weight, self-reported in kg	78.4 (11.2)	76.2 (10.7)	79.3 (12.6)	79.1 (12.8)	2.1 (1.4-2.8)	<.001
BMI[c], self-reported in kg/m^2	27.3 (2.3)	26.5 (2.3)	27.0 (2.7)	26.9 (2.5)	0.6 (0.3-1.0)	<.001

[a]All participants had measured variables including 2 participants who did not complete baseline self-report surveys.

[b]Model coefficients and *P* values were obtained from analysis of covariance models adjusting for baseline values, general practitioner clinic, and gender. Missing baseline and follow-up values were imputed to create five datasets and results were pooled using Rubin's rules.

[c]Body mass index (BMI).

Table 4. Effect of the TXT2BFiT program on secondary outcomes for diet for all randomized participants in the study by allocation (intention-to-treat analysis) (n=248)[a].

Variable[b]	Intervention group (n=123)[a], n (%)		Control group (n=125), n (%)		*P*
	Baseline	12 weeks	Baseline	12 weeks	
Fruit servings [c] per day					
≤1	82 (66.7)	30 (24.4)	77 (61.6)	50 (40.0)	.18
2	31 (25.2)	75 (61.0)	31 (24.8)	55 (44.0)	
≥3	10 (8.1)	18 (14.6)	17 (13.6)	20 (16.0)	
Vegetable servings [d] per day					
≤1	35 (28.5)	12 (9.8)	34 (27.2)	25 (20.0)	.009
2	46 (37.4)	32 (26.0)	46 (36.8)	40 (32.0)	
3	23 (18.7)	36 (29.3)	27 (21.6)	32 (25.6)	
≥4	19 (15.4)	43 (35.0)	18 (14.4)	28 (22.4)	
SSB [e] intake per week in mL					
Nil	22 (17.9)	37 (30.1)	33 (26.4)	32 (25.6)	.002
ASD[f]	27 (22.0)	32 (26.0)	17 (13.6)	15 (12.0)	
<500	37 (30.1)	45 (36.6)	31 (24.8)	43 (34.4)	
500-999	21 (17.1)	8 (6.5)	22 (17.6)	26 (20.8)	
≥1000	16 (13.0)	1 (0.8)	22 (17.6)	9 (7.2)	
Takeout meal intake per week					
Nil	3 (2.4)	3 (2.4)	2 (1.6)	8 (6.4)	.01
≤1	45 (36.6)	85 (69.1)	44 (35.2)	60 (48.0)	
2-3	58 (47.2)	28 (22.8)	53 (42.4)	37 (29.6)	
4-5	11 (8.9)	5 (4.1)	21 (16.8)	17 (13.6)	
6-7	6 (4.9)	2 (1.6)	5 (4.0)	3 (2.4)	

[a]All participants had measured variables including 2 participants who did not complete baseline self-report surveys.

[b]All questions were asked for average daily or weekly intake over the previous month. *P* values were adjusted for practice and gender. All variables were analyzed using Mantel-Haenszel chi-square tests stratified by general practitioner clinic and gender. Five imputed datasets were created and the results for the chi-square statistics were pooled, and *P* values estimated, using the method described by Li et al [35].

[c]One serving of fruit is equivalent to one medium piece (eg, one apple or one orange), two small pieces (eg, two plums), or one cup of diced pieces (fresh or canned).

[d]One serving of vegetables is equivalent to half a cup of cooked vegetables (fresh, frozen, or canned) or one cup of raw salad vegetables.

[e]Sugar-sweetened beverages (SSB).

[f]Artificially sweetened drinks (ASD).

Table 5. Effect of the TXT2BFiT program on secondary physical activity outcomes from the IPAQ[a]for all randomized young adults in the study by allocation (intention-to-treat analysis) (n=248)[b].

Variable	Intervention group (n=123)[b], mean (SD)		Control group (n=125), mean (SD)		Model β[c](95% CI)	*P*
	Baseline	12 weeks	Baseline	12 weeks		
Vigorous physical activity						
MET[d]-minutes per week	758.6 (1112.0)	1006.1 (1463.6)	840.0 (1072.1)	944.1 (958.3)	-20.0 (-195.9 to 155.9)	.80
Days per week	1.5 (1.6)	2.1 (1.7)	1.8 (1.8)	2.0 (1.7)	-0.3 (-0.7 to 0.2)	.20
Walking physical activity						
MET-minutes per week	691.9 (867.5)	927.3 (1163.0)	630.0 (595.3)	777.3 (828.7)	-69.8 (-180.2 to 40.6)	.20
Days per week	4.3 (2.0)	5.2 (1.9)	4.6 (2.2)	4.7 (2.2)	-0.6 (-1.1 to -0.1)	.02
Moderate physical activity						
MET-minutes per week	169.4 (359.8)	258.7 (417.9)	176.8 (393.9)	170.8 (222.4)	8.0 (-34.3 to 50.5)	.70
Days per week	0.8 (1.2)	1.4 (1.6)	0.9 (1.3)	1.0 (1.3)	-0.4 (-0.7 to 0.1)	.10
Total physical activity						
MET-minutes per week	1619.9 (1581.1)	2192.4 (2133.1)	1646.8 (1474.6)	1892.7 (1539.3)	-252.5 (-503.8 to -1.2)	.05
Days per week	6.6 (3.3)	8.8 (3.6)	7.4 (3.8)	7.7 (3.6)	-1.3 (-2.2 to -0.5)	.003

[a]International Physical Activity Questionnaire (IPAQ).

[b]All participants had measured variables including 2 participants who did not complete baseline self-report surveys.

[c]Model coefficients and *P* values were obtained from analysis of covariance models adjusting for baseline values, general practitioner clinic, and gender. Robust regression models were used for analyses where residuals indicated nonnormality. Missing baseline and follow-up values were imputed to create five datasets and results were pooled using Rubin's rules.

[d]Metabolic equivalent of task (MET).

Engagement With the Program

The mean number of coaching calls completed in the intervention group was 4.6 (SD 1.1) out of 5 (82.4% overall completed all 5). All participants who completed the postintervention survey reported engaging with coaching calls. Of the 12,308 text messages sent during the 12-week intervention (control, 500; intervention 11,808), only 2.27% (280) were not delivered (control, 15/500, 3.0%; intervention, 265/11,808, 2.24%). Over half (66/123, 53.7%) of the intervention participants replied to 8 or more of the 16 SMS text messages with a requested response, with 25 of the 123 participants (20.3%) replying to all. Most control participants replied to 2 or more of the 4 text messages (114/125, 91.2%), with 62.4% (78/125) replying to all 4 of them. A total of 100 of the 110 (90.9%) intervention participants who completed the follow-up survey self-reported that they used the SMS text messages. Email delivery was 100%, with 84 of 110 (76.4%) participants reporting that they used the email messages during the study. A total of 82 out of 110 (74.5%) intervention participants reported that they did not access the mobile phone apps during the study. The mailed booklet was used by 72 of the 110 (65.5%) intervention participants and only 7 out of 110 (6.4%) used the blog. Most intervention participants (65/110, 59.1%) did not use the resources available on the website. Of those that did, the takeout meal planner was reported as most used by the intervention participants (28/110, 25.5%).

Body Weight (kg) and BMI (kg/m^2)

Young adults in the intervention group were 2.2 kg lighter at 12 weeks compared to the control group using measured body weight after adjusting for baseline-measured body weight (95% CI 0.8-3.6, *P*=.005) (see Table 2). A similar pattern was observed with BMI, which was 0.5 kg/m^{2}less at 12 weeks (95% CI 0.1-1.0, *P*=.02) for the intervention group compared to the control group using measured BMI.

Using self-reported body weight measures, intervention participants were 2.1 kg (95% CI 1.4-2.8, *P*<.001) and 0.6 BMI units (kg/m^2) (95% CI 0.3-1.0, *P*<.001) lighter than control participants at 12 weeks (see Table 3).

At baseline, there was no significant difference between measured and self-reported weight and BMI (248/250, 99.2%) (*P*>.11). At 12 weeks, among participants with a measured weight (124/250, 49.6%), average self-reported weight was 0.7 kg (SD 1.3) less than the measured weight (*P*<.001). However, there was no difference between intervention (56/125, 44.8%; 0.8 kg, SD 1.2) and control groups (68/125, 54.4%; 0.6 kg, SD 1.4) (*P*=.44). There was no difference between measured and self-reported BMI at 12 weeks (*P*=.26).

Fruit and Vegetable Intake

The majority of participants reported consuming the recommended two servings of fruit per day or more after 12 weeks (see Table 4), with a nonsignificant difference between intervention group and control group (*P*=.18). Intervention participants were more likely to consume greater quantities of vegetables after 12 weeks compared to control participants (*P*=.009). For example, 35.0% (43/123) of intervention participants consumed four or more servings of vegetables compared to 22.4% (28/125) of control participants.

Sugar-Sweetened Beverage and Takeout Meal Intake

Intervention participants consumed SSB less frequently after 12 weeks compared with the control participants (*P*=.002) (see Table 4). For example, 92.7% (114/123) of intervention participants consumed 500 mL or less of SSB compared to 72.0% (90/125) of control participants at 12 weeks.

After 12 weeks, intervention participants reported consuming energy-dense takeout meals less frequently during the week compared with the control participants—54.4% (68/125) of intervention participants compared to 71.5% (88/123) of control participants consumed one or fewer energy-dense takeout meals per week (*P*=.01) (see Table 4).

Physical Activity

Intervention participants reported a mean increase of 563.1 (SD 1983.6) MET-minutes per week after 12 weeks. Control participants reported a mean increase of 244.4 (SD 1510.6) MET-minutes per week (see Table 5). These observed increases in energy expenditure were predominantly due to increased reported vigorous and walking activities, which increased by an average 243.0 (SD 1073.3) and 231.8 (SD 1313.9) MET-minutes per week among intervention participants, respectively, and 102.5 (SD 1148.6) and 148.1 (SD 747.3) MET-minutes per week among control participants, respectively. After adjusting for baseline MET-minutes per week, GP clinic, and gender there was a significant effect of the intervention on average MET-minutes per week at 12 weeks (95% CI -503.8 to -1.2, *P*=.05). Total and walking physical activity days increased more in the intervention group (95% CI -2.2 to -0.5, *P*=.003) compared to the control group (95% CI -1.1 to -0.1, *P*=.02).

Discussion

Principal Findings

This 12-week TXT2BFiT mHealth intervention was effective in preventing unhealthy weight gain, resulting in modest weight loss and improvement in lifestyle behaviors. Compared with control participants, intervention participants consumed more vegetables and less SSB, consumed fewer energy-dense meals prepared away from home, and increased their physical activity, with increased total and walking days of physical activity. As far as we are aware, this is the first reported trial of a multi-component mobile phone-based program conducted in young adults.

Participants in the intervention program weighed 2.2 kg less than control participants at 12 weeks. The prevention of weight gain is an important public health priority for this population, given the likelihood of weight gain reported by prior observational studies in young adult populations [37,38]. Furthermore, young adults have been born into an increasingly "obesogenic environment" and are at a greater risk of becoming obese [4]. If there is no effort to change these behavioral patterns, it is likely that young adults and subsequent generations will have a higher incidence of overweight and obesity.

A greater number of intervention participants reported increasing vegetable servings compared with controls. While recommendations for increased fruits and vegetables alone may not prevent weight gain [39], intervention participants also reported reductions in energy-dense meals prepared away from home and in SSB intake, and further increased their physical activity.

Most of the recent weight gain prevention interventions in young adults have targeted improvements in healthy eating and physical activity through in-person group interventions [17]. Two previous intervention studies in young adults were conducted via an online tutorial-style platform [40,41]. Green et al conducted a 3-month online curriculum-designed program based on nondiet principles, with weekly goal setting to increase fruit and vegetable intake and increase physical activity [40]. Intervention participants increased fruit and vegetable intake and increased physical activity compared to controls, however, no significant change in weight outcomes resulted. A shorter social cognitive theory-based intervention of 6 weeks by Gow et al focused on diet and physical activity habits pertinent to the transition period to college [41]. BMI was lower in the intervention group participating in the online activities and receiving weekly emails, with no effects on diet and physical activity outcomes. Other studies investigating the use of technology in the prevention of weight gain for young adults have published protocol papers, but there have been no reports of efficacy to date [42].

This study was innovative in the use of text messages, already demonstrated as successful in older adults in combination with coaching telephone calls [43]. This study design did not test the efficacy of the individual components, but the engagement data suggested that the coaching calls and text messages were useful to participants, with 100% and 90.9% reporting having used these components, respectively. A text message intervention in normal-weight young adults showed messages based on a habit framework can improve fruit consumption, and simply reminding young adults to be conscious of their food choices may be sufficient to improve their overall vegetable consumption [44]. Costs of delivering a mobile program could be reduced without coaching calls, but our previous pilot intervention did not detect effectiveness in dietary change with text messages alone, without coaching calls.

An important strength of this study was low attrition (14.4% at 12 weeks), and the interventions were delivered according to protocol, with 92% of coaching calls completed and only 2.3% of SMS text messages failing to send. Primary health care facilities and public advertisements increased reach in recruitment. This study also recruited more males than is often expected in studies of this type [45]. Another strength included

analyzed outcomes being blinded to treatment allocation and using intention-to-treat principles. The use of GPs to measure height and weight on their scales could introduce measurement error. However, GP clinic was one of the strata for participant randomization, and observer and equipment error would have been distributed equally across groups. Therefore, measurement bias should not have impacted the results from the analysis of covariance. It is acknowledged that using GP scales for baseline weight and clinic scales with a trained dietitian for follow-up measures was not ideal. As self-reported measures were used for all studied outcomes, the data may be biased. Self-report may underestimate weight, but has been shown to accurately identify overweight and/or obesity in the majority of a sample of young people [46]. An element of social desirability might influence reporting of lifestyle behaviors. Both groups were provided dietary and physical activity guidelines, however, greater significant improvements in intervention participants were seen in this study. Further, the sample was mostly well educated and from higher socioeconomic areas, which may influence the generalizability of the results [47].

Conclusions

In conclusion, intervening in the lives of young adults with unhealthy lifestyle behaviors, who have an increased risk of weight gain and developing obesity, appears to have a beneficial impact on preventing weight gain. While the short-term efficacy of the 12-week TXT2BFiT intervention program is promising, maintenance of outcomes in the longer term will be evaluated at 9 months. The potentially wide reach and low delivery costs of using mHealth, coupled with the growing problem of obesity in younger adulthood, means translation and implementation of this program to the community at large also warrants further consideration.

Acknowledgments

We wish to thank Ms Sally Jo-Hsin Juan for her contribution to the recruitment efforts in the study, and the partnering general practices and Medicare Local offices for their time and support in this research.

Funding to conduct this trial was received from the Hospitals Contribution Fund (HCF) Medical Research Foundation (reference number MAUsyd1008201111), listed on the Australian Competitive Research Grants Register Category 1. This work was also supported by the Commonwealth Government of Australia via an Australian Postgraduate Award Scholarship to SRP, and LH was supported by a National Health and Medical Research Council (NHMRC) Scholarship between 2011 and 2013. MFH was partly funded by an NHMRC Senior Research Fellowship.

Authors' Contributions

KM, LH, KB, AW, EDW, MFH, PP, AB, and MAF designed the research; SRP, LH, KB, AW, and MAF conducted the research; SRP, KM, and MAF analyzed the data; SRP and MAF interpreted the study findings and wrote the paper; KM, MFH, AB, and PP assisted in drafting the manuscript and in the interpretation of study findings; and MAF had primary responsibility for the final content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.1 [48].

[PDF File (Adobe PDF File), 1MB - mhealth_v3i2e66_app1.pdf]

References

1. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014 Aug 30;384(9945):766-781. [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
2. Tanamas SK, Magliano DJ, Lynch B, Sethi P, Willenberg L, Polkinghorne K, et al. AusDiab 2012: The Australian Diabetes, Obesity and Lifestyle Study. Melbourne, Australia: Baker IDI Heart and Diabetes Institute; 2013. Chapter 3: Obesity URL: http://www.bakeridi.edu.au/Assets/Files/Baker%20IDI%20Ausdiab%20Report_interactive_FINAL.pdf [accessed 2015-06-01] [WebCite Cache ID 6Ywr3YytG]
3. Reither EN, Hauser RM, Yang Y. Do birth cohorts matter? Age-period-cohort analyses of the obesity epidemic in the United States. Soc Sci Med 2009 Nov;69(10):1439-1448 [FREE Full text] [doi: 10.1016/j.socscimed.2009.08.040] [Medline: 19773107]
4. Allman-Farinelli MA, Chey T, Bauman AE, Gill T, James WP. Age, period and birth cohort effects on prevalence of overweight and obesity in Australian adults from 1990 to 2000. Eur J Clin Nutr 2008 Jul;62(7):898-907. [doi: 10.1038/sj.ejcn.1602769] [Medline: 17440514]

5. Truesdale KP, Stevens J, Lewis CE, Schreiner PJ, Loria CM, Cai J. Changes in risk factors for cardiovascular disease by baseline weight status in young adults who maintain or gain weight over 15 years: the CARDIA study. Int J Obes (Lond) 2006 Sep;30(9):1397-1407 [FREE Full text] [doi: 10.1038/sj.ijo.0803307] [Medline: 16534519]
6. Hebden L, Chey T, Allman-Farinelli M. Lifestyle intervention for preventing weight gain in young adults: a systematic review and meta-analysis of RCTs. Obes Rev 2012 Aug;13(8):692-710. [doi: 10.1111/j.1467-789X.2012.00990.x] [Medline: 22413804]
7. Strong KA, Parks SL, Anderson E, Winett R, Davy BM. Weight gain prevention: identifying theory-based targets for health behavior change in young adults. J Am Diet Assoc 2008 Oct;108(10):1708-1715 [FREE Full text] [doi: 10.1016/j.jada.2008.07.007] [Medline: 18926139]
8. Forget G, Doyon M, Lacerte G, Labonté M, Brown C, Carpentier AC, et al. Adoption of American Heart Association 2020 ideal healthy diet recommendations prevents weight gain in young adults. J Acad Nutr Diet 2013 Nov;113(11):1517-1522. [doi: 10.1016/j.jand.2013.06.346] [Medline: 23988512]
9. Kimmons J, Gillespie C, Seymour J, Serdula M, Blanck HM. Fruit and vegetable intake among adolescents and adults in the United States: percentage meeting individualized recommendations. Medscape J Med 2009;11(1):26. [Medline: 19295947]
10. Boeing H, Bechthold A, Bub A, Ellinger S, Haller D, Kroke A, et al. Critical review: vegetables and fruit in the prevention of chronic diseases. Eur J Nutr 2012 Sep;51(6):637-663 [FREE Full text] [doi: 10.1007/s00394-012-0380-y] [Medline: 22684631]
11. Duffey KJ, Gordon-Larsen P, Steffen LM, Jacobs DR, Popkin BM. Drinking caloric beverages increases the risk of adverse cardiometabolic outcomes in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. Am J Clin Nutr 2010 Oct;92(4):954-959 [FREE Full text] [doi: 10.3945/ajcn.2010.29478] [Medline: 20702604]
12. Larson N, Neumark-Sztainer D, Laska MN, Story M. Young adults and eating away from home: associations with dietary intake patterns and weight status differ by choice of restaurant. J Am Diet Assoc 2011 Nov;111(11):1696-1703 [FREE Full text] [doi: 10.1016/j.jada.2011.08.007] [Medline: 22027052]
13. Hankinson AL, Daviglus ML, Bouchard C, Carnethon M, Lewis CE, Schreiner PJ, et al. Maintaining a high physical activity level over 20 years and weight gain. JAMA 2010 Dec 15;304(23):2603-2610 [FREE Full text] [doi: 10.1001/jama.2010.1843] [Medline: 21156948]
14. De Cocker KA, van Uffelen JG, Brown WJ. Associations between sitting time and weight in young adult Australian women. Prev Med 2010 Nov;51(5):361-367. [doi: 10.1016/j.ypmed.2010.07.009] [Medline: 20638408]
15. Allman-Farinelli MA, Chey T, Merom D, Bowles H, Bauman AE. The effects of age, birth cohort and survey period on leisure-time physical activity by Australian adults: 1990-2005. Br J Nutr 2009 Feb;101(4):609-617. [doi: 10.1017/S0007114508019879] [Medline: 18588734]
16. Spring B, Moller AC, Colangelo LA, Siddique J, Roehrig M, Daviglus ML, et al. Healthy lifestyle change and subclinical atherosclerosis in young adults: Coronary Artery Risk Development in Young Adults (CARDIA) study. Circulation 2014 Jul 1;130(1):10-17. [doi: 10.1161/CIRCULATIONAHA.113.005445] [Medline: 24982115]
17. Partridge SR, Juan SJ, McGeechan K, Bauman A, Allman-Farinelli M. Poor quality of external validity reporting limits generalizability of overweight and/or obesity lifestyle prevention interventions in young adults: a systematic review. Obes Rev 2015 Jan;16(1):13-31. [doi: 10.1111/obr.12233] [Medline: 25407633]
18. Pew Research Centre. Washington, DC: Pew Internet & American Life Project; 2014. Mobile technology fact sheet URL: http://www.pewinternet.org/fact-sheets/mobile-technology-fact-sheet/ [accessed 2015-04-14] [WebCite Cache ID 6XlyWn7Ky]
19. Shaw R, Bosworth H. Short message service (SMS) text messaging as an intervention medium for weight loss: A literature review. Health Informatics J 2012 Dec;18(4):235-250 [FREE Full text] [doi: 10.1177/1460458212442422] [Medline: 23257055]
20. Siopis G, Chey T, Allman-Farinelli M. A systematic review and meta-analysis of interventions for weight management using text messaging. J Hum Nutr Diet 2015 Feb;28 Suppl 2:1-15. [doi: 10.1111/jhn.12207] [Medline: 24480032]
21. Hebden L, Cook A, van der Ploeg HP, King L, Bauman A, Allman-Farinelli M. A mobile health intervention for weight management among young adults: a pilot randomised controlled trial. J Hum Nutr Diet 2014 Aug;27(4):322-332. [doi: 10.1111/jhn.12155] [Medline: 23992038]
22. Hebden L, Balestracci K, McGeechan K, Denney-Wilson E, Harris M, Bauman A, et al. 'TXT2BFiT' a mobile phone-based healthy lifestyle program for preventing unhealthy weight gain in young adults: study protocol for a randomized controlled trial. Trials 2013;14:75 [FREE Full text] [doi: 10.1186/1745-6215-14-75] [Medline: 23506013]
23. Pink B. Socio-Economic Indexes for Areas (SEIFA). Canberra, Australia: Australia Bureau of Statistics; 2011. URL: http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/22CEDA8038AF7A0DCA257B3B00116E34/$File/2033.0.55.001%20seifa%202011%20technical%20paper.pdf [WebCite Cache ID 6UhadJnhd]
24. Australian Bureau of Statistics. Canberra, Australia: Australian Bureau of Statistics; 2011. Census dictionary URL: http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/2901.0Main%20Features12011?opendocument&tabname=Summary&prodno=2901.0&issue=2011&num=&view= [WebCite Cache ID 6Xlyp0wbi]

25. World Health Organization. Obesity: Preventing and Managing the Global Epidemic. Geneva, Switzerland: World Health Organization; 2000.
26. Rutishauser IHE, Webb K, Abraham B, Allsopp R. Evaluation of Short Dietary Questions From the National Nutrition Survey. 2001. URL: https://www.health.gov.au/internet/main/publishing.nsf/Content/17B241DC3956A205CA257BF00020A773/$File/evaluation.pdf [accessed 2015-04-14] [WebCite Cache ID 6Xm004UsN]
27. Smith KJ, McNaughton SA, Gall SL, Blizzard L, Dwyer T, Venn AJ. Takeaway food consumption and its associations with diet quality and abdominal obesity: a cross-sectional study of young adults. Int J Behav Nutr Phys Act 2009;6:29 [FREE Full text] [doi: 10.1186/1479-5868-6-29] [Medline: 19473547]
28. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. Int J Behav Nutr Phys Act 2011;8:115 [FREE Full text] [doi: 10.1186/1479-5868-8-115] [Medline: 22018588]
29. Patterson E. International Physical Activity Questionnaire. 2010 Oct 04. IPAQ scoring protocol URL: https://sites.google.com/site/theipaq/scoring-protocol [accessed 2015-04-14] [WebCite Cache ID 6XlwFAOEO]
30. Hebden L, Cook A, van der Ploeg HP, Allman-Farinelli M. Development of smartphone applications for nutrition and physical activity behavior change. JMIR Res Protoc 2012;1(2):e9 [FREE Full text] [doi: 10.2196/resprot.2205] [Medline: 23611892]
31. TXT2BFiT. 2013. URL: http://txt2bfit.com/ [accessed 2015-04-14] [WebCite Cache ID 6Xlz0UTLW]
32. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. Am Psychol 1992 Sep;47(9):1102-1114. [Medline: 1329589]
33. National Physical Activity Guidelines for Adults. Canberra, Australia: Department of Health and Ageing; 1999. URL: http://fulltext.ausport.gov.au/fulltext/1999/feddep/physguide.pdf [accessed 2014-12-09] [WebCite Cache ID 6UgO0DcHr]
34. Australian Dietary Guidelines: Incorporating the Australian Guide to Healthy Eating (draft for public consultation). Canberra, Australia: National Health and Medical Research Council; 2011 Dec. URL: https://www.eatforhealth.gov.au/sites/default/files/files/public_consultation/n55_draft_australian_dietary_guidelines_consultation_111212.pdf [accessed 2015-04-14] [WebCite Cache ID 6Xm0A0GWp]
35. Li KH, Meng XL, Raghunathan TE, Rubin DB. Significance levels from repeated p-values with multiply-imputed data. Stat Sin 1991;1:65-92.
36. 2008 Physical Activity Guidelines for Americans. Washington, DC: The US Department of Health and Human Services; 2008. URL: http://www.health.gov/paguidelines/pdf/paguide.pdf [accessed 2015-04-14] [WebCite Cache ID 6Xm0L72ev]
37. Gordon-Larsen P, Adair LS, Nelson MC, Popkin BM. Five-year obesity incidence in the transition period between adolescence and adulthood: the National Longitudinal Study of Adolescent Health. Am J Clin Nutr 2004 Sep;80(3):569-575 [FREE Full text] [Medline: 15321794]
38. Lewis CE, Jacobs DR, McCreath H, Kiefe CI, Schreiner PJ, Smith DE, et al. Weight gain continues in the 1990s: 10-year trends in weight and overweight from the CARDIA study. Coronary Artery Risk Development in Young Adults. Am J Epidemiol 2000 Jun 15;151(12):1172-1181 [FREE Full text] [Medline: 10905529]
39. Kaiser KA, Brown AW, Bohan Brown MM, Shikany JM, Mattes RD, Allison DB. Increased fruit and vegetable intake has no discernible effect on weight loss: a systematic review and meta-analysis. Am J Clin Nutr 2014 Aug;100(2):567-576. [doi: 10.3945/ajcn.114.090548] [Medline: 24965308]
40. Greene GW, White AA, Hoerr SL, Lohse B, Schembre SM, Riebe D, et al. Impact of an online healthful eating and physical activity program for college students. Am J Health Promot 2012;27(2):e47-e58. [doi: 10.4278/ajhp.110606-QUAN-239] [Medline: 23113786]
41. Gow RW, Trace SE, Mazzeo SE. Preventing weight gain in first year college students: an online intervention to prevent the "freshman fifteen". Eat Behav 2010 Jan;11(1):33-39 [FREE Full text] [doi: 10.1016/j.eatbeh.2009.08.005] [Medline: 19962118]
42. Lytle LA, Svetkey LP, Patrick K, Belle SH, Fernandez ID, Jakicic JM, et al. The EARLY trials: a consortium of studies targeting weight control in young adults. Transl Behav Med 2014 Sep;4(3):304-313 [FREE Full text] [doi: 10.1007/s13142-014-0252-5] [Medline: 25264469]
43. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. J Cardiovasc Nurs 2013;28(4):320-329 [FREE Full text] [doi: 10.1097/JCN.0b013e318250a3e7] [Medline: 22635061]
44. Rompotis CJ, Grove JR, Byrne SM. Benefits of habit-based informational interventions: a randomised controlled trial of fruit and vegetable consumption. Aust N Z J Public Health 2014 Jun;38(3):247-252. [doi: 10.1111/1753-6405.12232] [Medline: 24890483]
45. O'Hara BJ, Phongsavan P, Venugopal K, Bauman AE. Characteristics of participants in Australia's Get Healthy telephone-based lifestyle information and coaching service: reaching disadvantaged communities and those most at need. Health Educ Res 2011 Dec;26(6):1097-1106 [FREE Full text] [doi: 10.1093/her/cyr091] [Medline: 21987479]
46. Bowring AL, Peeters A, Freak-Poli R, Lim MS, Gouillou M, Hellard M. Measuring the accuracy of self-reported height and weight in a community-based sample of young people. BMC Med Res Methodol 2012;12:175 [FREE Full text] [doi: 10.1186/1471-2288-12-175] [Medline: 23170838]

47. 2011 Census of Population and Housing: Basic Community Profile (New South Wales). Canberra, Australia: Australian Bureau of Statistics URL: http://www.censusdata.abs.gov.au/CensusOutput/copsub.NSF/All%20docs%20by%20catNo/2011~Community%20Profile~1/$File/BCP_1.zip?OpenElement [accessed 2015-04-14] [WebCite Cache ID 6Xm0RN3kF]
48. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

ASD: artificially sweetened drinks
AUD: Australian Dollar
BMI: body mass index
CARDIA: Coronary Artery Risk Development in Young Adults
GP: general practitioner
HCF: Hospitals Contribution Fund
IPAQ: International Physical Activity Questionnaire
MET: metabolic equivalent of task
NHMRC: National Health and Medical Research Council
RCT: randomized controlled trial
SES: socioeconomic status
SMS: short messages service
SSB: sugar-sweetened beverages

Edited by G Eysenbach; submitted 14.04.15; peer-reviewed by L Lytle; comments to author 10.05.15; accepted 24.05.15; published 15.06.15.

Please cite as:
Partridge SR, McGeechan K, Hebden L, Balestracci K, Wong ATY, Denney-Wilson E, Harris MF, Phongsavan P, Bauman A, Allman-Farinelli M
Effectiveness of a mHealth Lifestyle Program With Telephone Support (TXT2BFiT) to Prevent Unhealthy Weight Gain in Young Adults: Randomized Controlled Trial
JMIR mHealth uHealth 2015;3(2):e66
URL: http://mhealth.jmir.org/2015/2/e66/
doi:10.2196/mhealth.4530
PMID:26076688

Original Paper

Valuable Features in Mobile Health Apps for Patients and Consumers: Content Analysis of Apps and User Ratings

Martin F Mendiola[1*], MD MPH; Miriam Kalnicki[1*], MBA; Sarah Lindenauer[1*], MPH

SocialWellth, Inc, Las Vegas, NV, United States

[*]all authors contributed equally

Corresponding Author:
Martin F Mendiola, MD MPH
SocialWellth, Inc
231 S. 3rd Street, Suite 100
Las Vegas, NV, 89101
United States
Phone: 1 702 821 0818
Fax: 1 702 821 1032
Email: mfmmdmph@gmail.com

Abstract

Background: The explosion of mobile phones with app capabilities coupled with increased expectations of the patient-consumers' role in managing their care presents a unique opportunity to use mobile health (mHealth) apps.

Objectives: The aim of this paper is to identify the features and characteristics most-valued by patient-consumers ("users") that contribute positively to the rating of an app.

Methods: A collection of 234 apps associated with reputable health organizations found in the medical, health, and fitness categories of the Apple iTunes store and Google Play marketplace was assessed manually for the presence of 12 app features and characteristics. Regression analysis was used to determine which, if any, contributed positively to a user's rating of the app.

Results: Analysis of these 12 features explained 9.3% (R^2=.093 n=234, P<.001) of the variation in an app's rating, with only 5 reaching statistical significance. Of the 5 reaching statistical significance, plan or orders, export of data, usability, and cost contributed positively to a user's rating, while the tracker feature detracted from it.

Conclusions: These findings suggest that users appreciate features that save time over current methods and identify an app as valuable when it is simple and intuitive to use, provides specific instructions to better manage a condition, and shares data with designated individuals. Although tracking is a core function of most health apps, this feature may detract from a user's experience when not executed properly. Further investigation into mHealth app features is worthwhile given the inability of the most common features to explain a large portion of an app's rating. In the future, studies should focus on one category in the app store, specific diseases, or desired behavior change, and methods should include measuring the quality of each feature, both through manual assessment and evaluation of user reviews. Additional investigations into understanding the impact of synergistic features, incentives, social media, and gamification are also warranted to identify possible future trends.

(JMIR mHealth uHealth 2015;3(2):e40) doi:10.2196/mhealth.4283

KEYWORDS

mHealth; mobile apps; consumer preference; Affordable Care Act

Introduction

The impact of recent health reform efforts are far-reaching, with perhaps one of the biggest shifts occurring in the convergence of clinical care delivery and consumer health. The Patient Protection and Affordable Care Act (PPACA) is a health care reform measure enacted in 2010 by the US Congress under President Barack Obama that seeks to make access to health care more affordable, efficient, and comprehensive for Americans. New mandates and incentives within PPACA call for a larger role for patients in health care, presenting an opportunity to incorporate and integrate digital health (wireless sensors, social networking, and mobile connectivity) into what Americans consider institutional health care (hospitals, physicians, and insurance plans). This has the potential to fundamentally change how patients manage their own health.

PPACA was intended to be the catalyst, and technology the facilitator, that empowered patients to be more prominent participants in their own health management.

Expectations of increased patient enablement through the use of technology seem much more attainable since the introduction and proliferation of the mobile phone. Today, 90% of American adults own a cellular phone, and more than 70% own a mobile phone with app capabilities [1]. Of those with mobile phones, approximately 20% already use a mobile health (mHealth) app, and by the end of 2015 that number is expected to rise to 33% [2]. Although promising, mobile phones continue to present an untapped opportunity to reach more patient consumers ("users"), especially since 70% of adults track some sort of health indicator for themselves or a loved one through paper logs or other means [3].

The explosion of mobile phone ownership coupled with increased expectations of the users' role in managing their care presents a unique opportunity for mHealth apps. Chronic diseases account for 75% of the $2.7 trillion in annual US health care spending [4], therefore tools like apps to help prevent, slow the progression of, or manage chronic disease are seen as valuable in helping to lower health care costs. Currently, identifying which health apps would be most effective for a population seems daunting, even more so for a specific individual [5]. Few research efforts have focused on understanding what a user values in a health app, a timely question considering that there are over 100,000 apps in the medical, health, and fitness sections of the Apple iTunes store and Google Play marketplace [6]. Additionally, the sustained use of any one app is low, where 68% of mobile phone users open ≤5 apps at least once a week [7] while 80%-90% percent of apps are used just once and then later deleted.

Early research in the digital health industry focused on identifying and listing mobile app features and characteristics for easy cataloguing [8,9], however, if users are expected to use mHealth apps regularly in order to play a larger role in their own care, then it is essential to gain a better understanding of the qualities and approaches that lead to long-term app usage. Arnhold et al examined the usability and functionality of a subset of mHealth apps specifically designed for diabetics [10]. They evaluated the impact of different app functions on user experience, and showed that although there are many apps that address the diabetes condition, very few offer more than just one or two basic functions. Those that offered more functions fared considerably worse when assessed for usability.

In order to promote sustained usage and positively influence health outcomes, it is critical that mHealth apps have a basis in behavioral science. Several studies have evaluated mHealth apps using frameworks rooted in behavior theory, with many focusing on identifying features that facilitate one of the following three basic psychological needs of the Self-Determination Theory (SDT) (1) autonomy, (2) competence, and (3) relatedness [11]. Autonomy refers to individuals' desires to regulate behavior based on their own values and interests. In order for an mHealth app to encourage autonomous motivation, it must accurately portray the value of the associated behavior change, give users a choice in their interaction with the app, acknowledge users' perspectives, and provide an action plan to support individuals and their needs. Competence refers to an individual's need to feel capable and confident to change his/her behavior, and can be facilitated by providing relevant information, tools, resources, and feedback throughout the health journey. Relatedness refers to the degree of connectedness that an individual feels with others, and it can be fostered through an mHealth app by creating virtual communities, connecting to social media sites, or actively attempting to better understand the individual user.

A recent study by Choi et al [12] evaluated smoking cessation apps for both content and functionality using tenets from SDT. Findings revealed that most of the smoking cessation apps (94.3%) had at least one feature that employed one of the three basic SDT needs, but few (10.3%) addressed all three. While not specific to one theory, Dahlke et al recently investigated the number of health behavior and communication constructs applied in mobile phone cancer survivorship apps, and found that while some of the apps utilized theoretical elements of behavior change, there remains an overall need for more theory-based apps in the mHealth space [13]. However, neither of these studies examined the relative influence of the various features on user satisfaction or sustained use.

Whereas most digital health studies, including those mentioned above, have focused on apps targeting just one condition, Payne et al systematically reviewed 24 studies that utilized a variety of mHealth apps across a range of health behavior interventions [14]. These apps were examined to identify features and functions central to behavior change, and the findings suggested that all of the apps included some element that addressed a behavior change theory or strategy, although their relative influence was not reported.

The foundation for digital health studies has been built on cataloguing the number and types of features in apps addressing a particular disease state, determining the theoretical impact of certain features on user experience, and identifying the behavioral theories expressed in a subset of mHealth apps. Additional studies should explore specific app features across a breadth of wellness and medical apps that lead to a positive user experience and ultimately long-term behavior change. The features investigated should closely align with SDT and common usability principles, since these are two prominent components of long-term engagement. Through analysis of a subset of mHealth apps, this paper aimed to identify those features that are aligned with SDT and common usability principles, and are most-valued and have contributed positively to a user's rating of the app in order to ultimately provide a roadmap for future mHealth app development.

Methods

General App Inclusion Criteria

First, a set of inclusion criteria was established to limit the scope of apps being evaluated prior to attempting to identify app features deemed valuable to a user. To create a group of comparable apps from the more than 100,000 mHealth apps in the Apple iTunes store and Google Play marketplace, the dataset

was limited to apps that were associated with reputable health organizations: these were defined as developers or evaluators with relationships with content-credible health care entities (Textbox 1).

Textbox 1. Definitions of reputable health organization inclusion criteria.

Requirement and description

- Clinical trial/research study
  - Have undergone a research study or clinical trial and had results published
- Food and Drug Administration (FDA)
  - Have been approved by the US FDA
- Government-approved
  - Have been developed or endorsed by a non-FDA government agency (eg, US Department of Veterans Affairs and Centers for Disease Control and Prevention)
- US hospital system-approved
  - Have been developed or endorsed by a US hospital system (eg, Cleveland Clinic and Carolinas Health System)
- US academic medical institution-approved
  - Have been developed or endorsed by a US academic medical institution (eg, Harvard Medical School and Vanderbilt University)
- Medical specialty society-approved
  - Have been developed or endorsed by a medical specialty society (eg, American College of Cardiology and American Society of Clinical Oncology)
- Non-profit health care organization-approved
  - Have been developed or endorsed by a non-profit health care organization (eg, American Diabetes Association and National Breast Cancer Foundation)
- Consumer organization with health focus-approved
  - Have been developed or endorsed by a national consumer company focused on health (eg, WebMD and Walgreens)
- US physician-approved
  - Have been developed or endorsed by a board-certified US physician
- Third-party payer-approved
  - Have been developed or endorsed by a private third-party insurance payer (eg, Aetna)
- Pharmaceutical or medical technology company-approved
  - Have been developed by a pharmaceutical or medical technology company (eg, Novartis Consumer Health and Medtronic)

Between March 19 and April 8, 2014, a list of the apps that met the inclusion criteria was compiled using information available from several systems. Using the PubMed and mHealth Evidence websites [15,16], the terms “iPhone,” “Android,” “Apple,” and “Google Play” were used separately as search queries to identify apps that had undergone a clinical trial or research study with published results. A relational database was created using the services of 42matters, a privately-held technology company that provides services for app discovery and analytics. This database contained the names, developers, and descriptions of the 100,000 mHealth apps in the Apple iTunes store and Google Play marketplace, and was used to identify any apps that referenced the name of any federal government agency, US hospital system, US academic medical institution, medical specialty society, private third-party insurance payer, pharmaceutical company, or medical technology company, as well as any notable non-profit health care organizations, national consumer companies focused on health, and board-certified US physicians that have a strong presence in mHealth. The apps that satisfied the reputable health organization criteria were then subjected to additional criteria related to purpose and functionality (Textbox 2).

Textbox 2. Additional purpose and functionality criteria.

Criteria

- Individual user ratings (>25) exist for the app in the respective app store
- Health consumers are its primary audience, as determined through active use and exploration of the app. Some apps, particularly in the "Medical" category of the app stores, target health care professionals and students as a means of reference or supplementary training and were omitted from analysis
- Created for US audiences, sold in US app stores, and contain an English-language user interface (UI)
- Ability to function independently of a medical device
- No required special passwords or access codes associated with a provider or payer program (since access to these apps require the user to be a patient of a particular provider or member of a payer program in order to obtain an access code, and without it, the app's features and functions could not be assessed)

Since the dataset used for analysis was pulled from two sources (Apple iTunes store and Google Play marketplace, duplicate apps (ie, identical features in the same app listed independently in both app stores) were eliminated. However, apps by the same developer, similarly named, but not having identical feature sets were treated as two different apps. Furthermore, the ratings were adjusted (Y2) to reflect a consistency in the presentation of the ratings, (ie, Google Play has continuous values to one tenth of a point, whereas Apple iTunes rounds ratings to the nearest half point). The app store ratings were converted into Bayesian Ratings (Y3) since there is growing support for the use of Bayesian analysis to assess any user-generated content, such as ratings, games, and cases that take into account individual judgment [17]. Lastly, with a recent spotlight shone on questionable techniques for increasing an app's number of ratings [18], any apps that received an unusually large number of ratings (ie, >3000 individual user ratings) were deemed outliers and eliminated from consideration.

Analysis of App Features and Characteristics

The remaining apps were downloaded and manually assessed for the presence of certain features or characteristics that have been studied in other published research [14] as a means to engage and change behavior (Table 1).

Using a binary system, apps were assigned a "1" to indicate the presence of a particular feature, or a "0" to indicate the absence. Only one attribute, cost, was assessed on 3 parameters because of the mutually exclusive cost options of free, free with in-app purchases, and paid, which were assigned values of "0," "0.5," and "1," respectively. Usability, being a more subjective and complex characteristic, required additional analysis before being assigned a score. Each app was downloaded and functionality explored before being rated against five of Jakob Nielsen's general principles for interaction design (Textbox 3) [19]. Apps that met a majority of the usability principles received a score of "1," otherwise a "0" was assigned.

Regression analysis was performed using Microsoft Excel to investigate the features that influenced an individual user's rating of an app. Since users are most often asked to rate an app only after they have begun using it regularly, ratings found in the app stores can serve as a proxy for assessing an app's value to its users. In the end, all regressions were executed against the following 3 separate dependent (Y) variables (1) Y1: ratings, (2) Y2: adjusted ratings, and (3) Y3: Bayesian ratings.

The inclusion criteria variables were also evaluated as independent variables to confirm that they were not confounders and could be eliminated from consideration for further analysis. At this point, multiple regression analysis could not be performed because the number of variables under consideration exceeded the maximum capacity of Microsoft Excel. Therefore, simple linear regression analysis was performed with each independent variable against Y1, Y2, and Y3 to gauge if any app feature or inclusion criterion independently influenced the dependent variables.

Using the independent variables that exhibited at least minimal influence, multiple regression analysis was performed against the same dependent variables to determine whether a combination of features could explain a user's rating of an app. Using a 95% confidence level, independent variables were eliminated based on *P* values, and the model was assessed for accuracy based on *F* statistic (primarily) and R-squared values. Correlation analysis was conducted to assess whether there were any pairwise associations between variables. Finally, the user reviews of a random sample of apps (10.3% of the total dataset (n=24) were assessed to determine whether users focused on the app features and characteristics addressed in the study.

Table 1. App features and characteristics.

App feature or characteristic	Description of feature or characteristic	Relevant construct(s) & principle(s)
Export of Data	Feature that allows the user to communicate or send information/ data to a health care provider (eg, email and EHR/PHR)	Relatedness (care team collaboration and support)
Gamification	Feature that offers points, badges, or movement through levels as a health objective is achieved or the more a patient is engaged (see Figure 1)	Autonomy (extrinsic motivation, engagement)
General education	Feature that provides basic educational material about a disease/condition, including causes, treatment, or management	Autonomy (intrinsic motivation); competence (knowledge)
Plan or orders	Feature that provides a plan of action for reaching target goal, including specific, executable steps to guide the process (see Figure 2)	Competence [actionable insights]; Autonomy [goal-setting]
Reminder	Feature that prompts the user to partake in a specific behavior through the use of a predetermined alert (see Figure 3)	Competence [cue to action]
Community forum	Feature that functions as a message board or chat room and allows likeminded individuals, whether patients with similar health conditions or their caregivers, the opportunity to share questions and experiences	Relatedness (social support, social norms); autonomy (acknowledging individual perspectives)
Social media	Feature that connects the user to Facebook, Twitter, or other social media platforms, thereby allowing the user to communicate progress with family, friends, colleagues, or others with ties to the user	Relatedness (social support, contextualization)
Addresses symptoms	Feature that addresses and assists in managing a disease that is associated with pain or other noticeable symptom(s)	Competence (educate, inform); autonomy (self-monitoring)
Tailored education	Feature that offers patient-specific education tailored to a person's needs, interests and usage depending on his/her stage or progression of disease (eg, week of pregnancy)	Relatedness (personalization); competence (knowledge, skill development)
Tracker	Feature that allows for self-monitoring by recording information in order to modify personal attitudes or behaviors to achieve a predetermined goal or outcome (see Figure 4)	Autonomy (self-monitoring, self-regulation)
Cost	Identification of cost of the app (free, upfront payment, and/or in-app purchases)	N/A
Usability	Identification of satisfactory usability based on compliance with five interface design heuristics	Nielsen's usability heuristics for user interface design

Textbox 3. Jakob Nielsen's five general principles for interaction design.

Principle

1. Visibility of system status: app's ability to keep users informed about what is going on and/or how they are progressing toward a goal.
2. User control and freedom: app provides the ability to easily control interactions, such as exit, save, go back, or edit.
3. Flexibility and efficiency of use: app provides the ability to accomplish intended tasks (eg, logging a meal or tracking blood pressure) quickly and efficiently.
4. Aesthetic and minimalist design: app is pleasant to look at and not overcrowded with irrelevant information.
5. Help users recognize, diagnose, and recover from errors: error messages within the app use plain language, simply state the problem, and outline steps to fixing it.

Figure 1. Gamification feature in the NFL PLAY 60 app.

Figure 2. Plan or orders feature in the Couch-to-5K app.

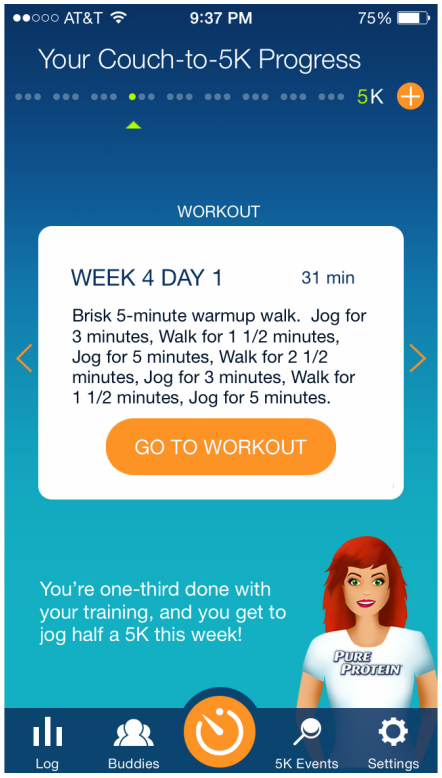


Figure 3. Reminder feature in the Glucose Buddy Diabetes Log app.

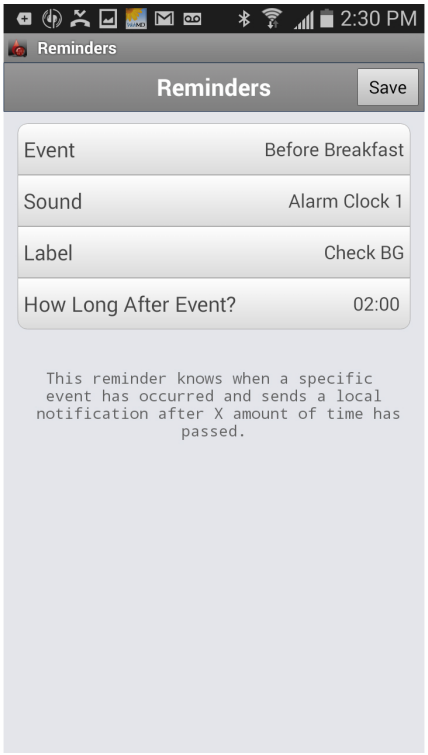
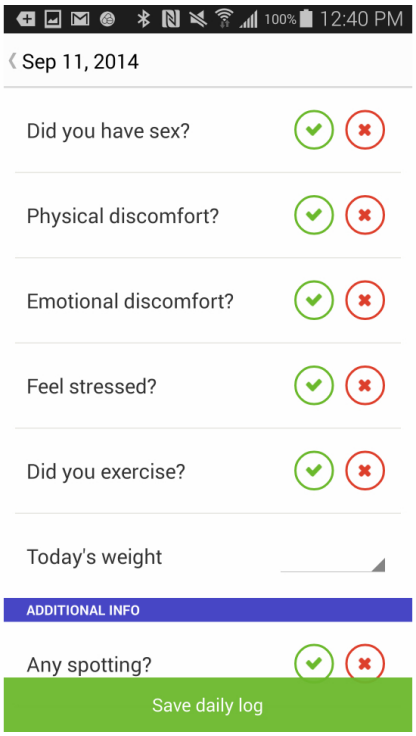


Figure 4. Tracking feature in the Glow Fertility & Ovulation app.



Results

General App Inclusion Criteria

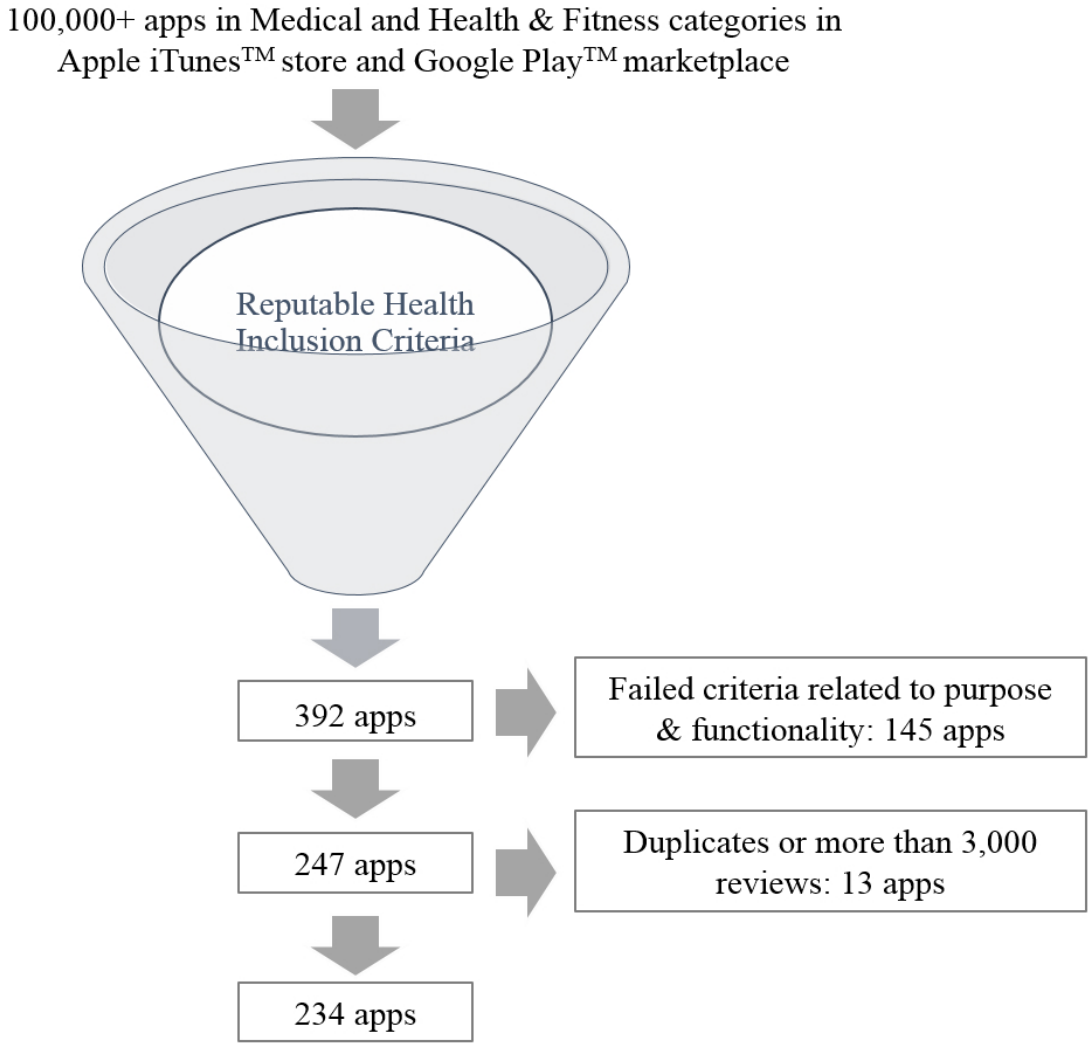
Initially, 392 apps were identified by PubMed and mHealth Evidence, and the relational database provided by 42matters as having met the reputable health organization inclusion criteria. Figure 5 shows the elimination of apps at various stages throughout the initial app inclusion evaluation. Of the 392 apps, 145 were eliminated after not meeting the inclusion criteria related to purpose and functionality. Another 13 apps were then eliminated from consideration either because they were duplicates, or because of their unusually large number of reviews (ie, >3000 individual user ratings). Eliminating those with an unusually large number of reviews resulted in a slightly more explanatory model (R^2=.093 vs R^2=.090), with little change to the strength of the model (F=4.667 vs F=4.769) (Multimedia Appendix 1). The number of apps included in the final analysis

was 234 (Multimedia Appendix 2). A breakdown of the total number and percentage of the 234 apps that met each of the inclusion criteria is displayed in Table 2. The reputable health organization inclusion criteria were each also treated as independent variables to confirm that they were not confounders, and once the analysis of the inclusion criteria showed no explanatory power (ie, <0.5%), those variables were eliminated from further analysis.

Table 2. Reputable health organization inclusion criterion totals (N=234).

Inclusion criteria	Apps, n (%)
Clinical trial/ research study	45 (19.2)
FDA-approved	8 (3.4)
Government-approved	16 (6.8)
US hospital system-approved	12 (5.1)
US academic medical institution-approved	21 (9.0)
Medical specialty society-approved	7 (3.0)
Non-profit health care organization-approved	39 (16.7)
Consumer organization with health focus-approved	68 (29.1)
US physician-approved	49 (20.9)
Third-party payer-approved	2 (0.9)
Pharmaceutical or medical technology company-approved	14 (6.0)

Figure 5. Flow Diagram of the app inclusion process.



Analysis of App Features and Characteristics

The 234 apps that remained were downloaded and manually assessed for the presence of certain features or characteristics. The findings of the app feature and characteristic assessment are shown in Table 3.

Simple linear regression analysis did not show that any single independent variable significantly impacted a user's app rating. However, the features plan or orders, usability, cost, tracker, and gamification influenced the dependent variables to some degree (R^2 approximating ≥1% at a 95% significance level) (Table 4).

Multiple regression analysis was performed to investigate whether a combination of features, in particular, the aforementioned five features, could explain a user's rating of an app. The same dependent and independent variables were used. The best model showed that 9.3% of the adjusted ratings (Y2) could be explained by plan or orders, usability, cost, tracker, and export of data (Tables 5 and 6).

The results of the correlation analysis are displayed in Table 7. Findings show a moderate positive correlation between the export of data and tracker features (.48), and a slight positive correlation between the tracker and usability features (.36). Any output <.2 or >-.2 was not considered significant.

Table 3. Assessment of the app features and characteristics (N=234).

App feature or characteristic	Apps, n (%)
Export of data	108 (46.2)
Gamification	27 (11.5)
General education	82 (35.0)
Plan or orders	41 (17.5)
Reminder	74 (31.6)
Community forum	46 (19.7)
Social media	61 (26.1)
Addresses symptoms	79 (33.8)
Tailored education	34 (14.5)
Tracker	170 (72.6)
Cost (free)	151 (64.5)
Usability	190 (81.2)

Table 4. R^2 and *P* value results for the simple linear regression analysis of individual app features at 95% significance level.

	Y1[a]		Y2[b]		Y3[c]	
App feature	R^2, %	*P* values	R^2, %	*P* values	R^2, %	*P* values
Plan or orders	4.2	<.001	4.3	<.001	4.0	<.001
Usability	1.3	.07	1.3	.07	1.7	.04
Cost	1.2	.08	1.0	.11	1.4	.06
Tracker	0.8	.18	0.8	.15	0.2	.48
Gamification	0.6	.21	0.7	.17	1.1	.09
Tailored education	0.4	.35	0.5	.28	0.5	.28
Addresses symptoms	0.3	.38	0.2	.48	0.3	.41
Reminder	0.3	.42	0.2	.45	0.3	.42
Community forum	0.2	.48	0.2	.47	0.4	.34
Export of data	0.2	.41	0.1	.51	0.3	.27
Social media	0.1	.65	0.1	.64	0.2	.53
General education	0.0	.98	0.0	.98	0.0	.75

[a]App store ratings

[b]Adjusted ratings

[c]Bayesian ratings

Table 5. Analysis of variance table with significance at the $P<.05$ level (N=234).

Source[a]	df[b]	*F*	MS[c]	*P*
Regression	5	4.667	(2.347)	*P*<.001
Residual	229		(.503)	

[a]Microsoft Excel
[b]Degrees of freedom
[c]Mean square

Table 6. Summary output for the multiple regression analysis to explain users' app ratings (R^2= 0.093, 95% confidence level).

Source[a]	B[b]	SE B[c]	β[d]
Cost	.172	0.111	.103
Usability	.279[e]	0.130	.154[e]
Plan or Orders	.357[e]	0.127	.184[f]
Tracker	-.373[f]	0.125	-.226[f]
Export of Data	.226[e]	0.109	.151[e]

[a]Microsoft Excel
[b]Regression coefficient (beta)
[c]Standard Error of beta
[d]Standardized beta
[e]*P*<.05
[f]*P*<.01

Table 7. Correlation analysis of variables in the best model.

	Cost	Usability	Plan or orders	Tracker	Export of data
Cost	1				
Usability	-.06	1			
Plan or orders	.17	.02	1		
Tracker	.09	.36	-.10	1	
Export of data	.02	.22	-.19	.48	1

Discussion

Principal Findings

It was found that 9.3% of a user's rating of an app can be explained by 5 app features or characteristics. Of these, plan or orders, export of data, usability, and cost contributed positively to a user's rating, while the tracker feature impacted it negatively. Users value an app that is simple and intuitive to use, which aligns with Nielsen's findings on usability [19]. Furthermore, users value tailored information and actionable insights regarding their condition and its management. This touches on both the autonomy and competence needs associated with SDT. Lastly, users want to be able to share their data with designated individuals, supporting the last basic psychological need of SDT, relatedness. In addition, the 4 app features that contributed positively to a user's rating share one common theme: each provides a mechanism for care management that would appear to be less time-consuming and more efficient than current methods (Textbox 4).

Textbox 4. App features that contributed positively to a user's rating share a common theme.

Feature

- Plan or orders
  - Users can save time by not having to investigate, decipher, and interpret the steps required to achieve a desired health goal, and in the process, appreciate immediate access to viewing their progress.
- Export of data
  - Users understand the value of sharing their progress (and setbacks) with their health care provider, and appreciate the time saved by not needing to input data into an email or having a member of the health care provider's staff copy the results into a health record.
- Usability
  - Users value the layout of an app that is efficient, intuitive, and allows for easy input of information.
- Cost
  - Users rate paid apps consistently higher than free apps, presumably because paid apps are usually void of advertisements, (ie, the main revenue source for most free apps), which can lead to a more efficient experience.

Although the fifth feature, tracker, returned a negative coefficient, further analysis revealed that the tracker feature is positively correlated with the export of data and the usability features. The moderate positive correlation between the tracker and export of data features (.48) may indicate that the ability to track progress isn't valuable to the user without the ability to transfer the data collected. Since a large majority of apps are able to both track and export data, an app that doesn't have both components is likely outdated or lacks sophistication, and thus may not be rated highly. It was determined that of the 234 apps studied as part of this research, 180 (76.9%, 180/234) contained a tracker feature. Of those, 62 (34.4%, 62/180) did not provide a method to export the data to a website, email, or electronic health record.

A moderate positive correlation between the tracker and the usability features (.36) strengthens the argument that the process of entering information into the tracker function of an app, as well as the value of the output display of the data collected, may be of great importance when a user assesses the tracker feature. Any further research to better understand the relationship between the overall user experience and tracking should begin by focusing on the differences between active tracking through the manual input of information and passive tracking where data is collected through sensors or devices.

Interestingly, popular and well-studied features such as gamification and the ability to connect to social media did not appear to influence a user's rating of apps in this analysis. These results were unexpected, particularly since social media and gaming apps are consistently the most downloaded and used apps on mobile phones [20]. The findings do not conclude that the aforementioned features are not valuable in engaging a patient, changing behavior, or improving outcomes, but solely that they do not seem to factor in the rating of the apps reviewed. However, these features seem poised to play a pivotal role in the future of digital health.

Limitations

Overview

Although the list of app features and characteristics compiled in order to explain user ratings was fairly exhaustive, this analysis does not account for >90% of an app rating. Potential reasons for this discrepancy are discussed in the following sections.

One-Size Fits All

Similar to other solutions in health care, apps are not a one-size-fits-all answer. Different users will value different features, layouts, and approaches.

Combining Apps

This analysis intentionally combined and analyzed apps from the medical, health, and fitness categories of the app stores. It is likely that users may rate a feature as valuable for one category of apps that may be irrelevant or detrimental to another, thereby negating its value in the overall analysis. For example, a reminder feature is essential for medication trackers, but it may be counterproductive in a smoking cessation app. Additional research is needed to focus on one category in the app store, specific diseases [8,21], or the desired behavior change; eventually, it may be determined that different evaluation criteria are needed for different types of apps.

Quality

This investigation focused on the presence or absence of most of the app features without evaluating the quality of the feature. A brief, informal examination of app store reviews for 10.3% (24/234) of the apps analyzed in this study (chosen randomly) showed that users often expressed displeasure with features of poor quality. The mere presence of a feature does not assure its value to the user; future app feature research should likely include a qualitative component, and overall user experience should be taken into account.

Rating Systems

The process of rating an app in the Apple iTunes store is more complex than in the Google Play marketplace [22], which may explain why the Google Play marketplace had, on average, a higher number of user ratings for the same app. An attempt to address this issue was made by omitting apps with <25 user ratings in the app store, but it is possible that the differences in rating processes may have impacted the results.

Patient-Consumers

The only apps included in this analysis were those intended for use by patient-consumers. This determination was made by a single reviewer, who downloaded each app, explored the features and functions, and subjectively determined the intended audience. Having had multiple reviewers participate in this assessment would have helped the process be more objective.

Usability

Although rooted in Nielsen's usability heuristics, some of the usability principles assessed are subjective by nature (eg, aesthetic and minimalist design). Having multiple reviewers participate in the assessment would have helped the process be more objective.

Number of Ratings

Reports have raised some concerns as to the legitimacy of the quantity of ratings and reviews for some mobile apps, which may confound results. Although precautions were taken by omitting apps from the analysis that contained an unusually large number of ratings (ie, >3000 individual user ratings) in order to reduce the potential of confounding, some illegitimate app ratings may have gone undetected and altered the findings.

Future Research

Gamification deserves more attention and study, particularly as a method to engage adults with chronic conditions. One-third of adults between 30-49 years old have at least one chronic condition compared with 60% of adults aged 50-64 years [23], with the majority being female [24]. This would indicate that mHealth app features that assist in disease management should be tailored to an older demographic and slightly more to women. Although some may postulate that gamification would not appeal to older generations of adults, McKinsey's Global iConsumer research found that approximately 50% of casual gamers are between 35-64 years of age, 54% are female, and the majority will stick to the same casual games for >6 months [25]. Lastly, at the time of the study, the app developers in the examined dataset did not offer incentives or rewards for reaching milestones. As such, further investigation into gamification coupled with a rewards and/or incentive program is warranted.

Similarly, social media seems ideally suited for mHealth engagement, even though the ability to access social media through an app did not impact the app's rating. Research by the Pew Internet Project indicated that as of January 2014, nearly 75% of those accessing the Internet also use social media [26], and a 2012 PwCHealth study showed that nearly one-third of those surveyed would be interested in having their social media conversations monitored if it would help them improve their health or better coordinate care [27]. Even at its lowest levels of adoption, approximately 65% of individuals between the ages of 50-65 use some form of social media (compared to nearly 90% in younger demographics) [26]. Therefore, age does not seem to be a limiting factor for integrating social media into health.

However, social media's role in this analysis may not be representative of its true value. Until recently, third-party private health insurance plans could deny coverage to patients with pre-existing conditions or insure them at significantly higher premiums. Sharing personal health information publicly carried financial concerns related to insurance status, potentially explaining why people would be more hesitant to share their health information in the same way that they share other personal information. With the implementation of PPACA and the elimination of pre-existing condition exclusions, social media may yet play a much larger role in transforming and managing care. Undoubtedly, deeper exploration is necessary to gain a better understanding of the roles that social media and gamification can play in unlocking the true potential of mHealth solutions for better health management.

Clearly, additional research must be conducted to expand the scope of mHealth apps reviewed and better understand what aspects and features are most valuable to its users. Deeper investigations and varied approaches are necessary to determine the roles of future versions of gamification, incentive programs, social media, and trackers within defined app categories in the app store, specific diseases, and desired behavior changes. Furthermore, similar analyses examining apps within only one condition or wellness category (eg, asthma management or nutrition) is necessary. This would allow for the investigation of more specific app features and has the potential to yield stronger findings. Lastly, a deeper dive into the impact of usability on user ratings is warranted. Since this study chose to focus largely on behavioral science features, it may also be useful to better understand the relative impact of each of Nielsen's 10 usability principles on a user's rating of health apps.

Implications and Recommendations

The field must keep working to move toward developing more sophisticated and better integrated digital tools in order to gain overall user acceptance, sustained engagement, and ultimately, clinical value and behavior change. As the digital health industry evolves, users will be able to collect more data and achieve better results while having to actively coordinate, input, and transmit less activity. Based on ratings of apps associated with reputable health organizations, users find some value in apps and features that save them time and effort, but additional research is critical in order to maximize digital health's potential while advancing the triple aim of health care to improve access and increase patient satisfaction while lowering overall costs.

Acknowledgments

We are grateful to Stuart Cohen, MD, Montefiore Medical Center, Leslie Isenegger, GNYHA Ventures, and Ismail Sirtalan, Greater New York Hospital Association for their counsel and expertise. We are also appreciative of our science and clinical advisers, Janet Greenhut, MD and Kevin Wildenhaus, PhD. Lastly, we would like to acknowledge David Vinson, Corey Ackerman, Adam Hurwitz, and John Scherer for their guidance and continued support in advancing the mobile health industry and promoting the role of technology in the future of health care.

Conflicts of Interest

At the time of publication, the authors were employed by SocialWellth, Inc, a software solution that engages payers, providers, and employers by delivering white label experiences while enabling patients to better manage their care using digital health tools.

Multimedia Appendix 1

Side-by-side comparison of multiple regression models to explain users' app ratings.

[PDF File (Adobe PDF File), 24KB - mhealth_v3i2e40_app1.pdf]

Multimedia Appendix 2

The 234 mobile health (mHealth) apps that met the inclusion criteria and guidelines.

[PDF File (Adobe PDF File), 44KB - mhealth_v3i2e40_app2.pdf]

References

1. Nielsen Reports. 2014 Apr 22. Smartphone Milestone: Half of Mobile Subscribers Ages 55+ Own Smartphones URL: http://www.nielsen.com/us/en/insights/news/2014/smartphone-milestone-half-of-americans-ages-55-own-smartphones.html [accessed 2015-01-21] [WebCite Cache ID 6VkbX5OFa]
2. Fox S, Duggan M. Pew Internet. 2012 Nov 08. Mobile Health 2012 URL: http://www.pewinternet.org/2012/11/08/mobile-health-2012/ [accessed 2015-01-21] [WebCite Cache ID 6Vkboyhgr]
3. Pew Internet Research. 2014 Jan. Mobile Technology Fact Sheet URL: http://www.pewinternet.org/fact-sheets/mobile-technology-fact-sheet/ [accessed 2015-01-21] [WebCite Cache ID 6VkbzlBLd]
4. Harris JR, Wallace RB. The Institute of Medicine's New Report on Living Well With Chronic Illness. Preventing Chronic Disease 9.120126 (2012) 2012.
5. Pagoto SL, Schneider KL, Jojic M, DeBiasse MA, Mann DM. Weight loss using evidence-based strategies in mobile apps. Authors' response. Am J Prev Med 2014 Jul;47(1):e2. [doi: 10.1016/j.amepre.2014.03.010] [Medline: 24952888]
6. Jahns R. research2guidance. 2014 May 06. mHealth App Developer Economics: Fourth Annual Study on mHealth App Publishing URL: http://research2guidance.com/r2g/mHealth-App-Developer-Economics-2014.pdf [accessed 2015-01-21] [WebCite Cache ID 6VkdYLI3q]
7. Purcell K. Pew Internet Research. 2011 Nov 02. Half of Adult Cell Phone Owners have Apps on Their Phones URL: http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP_Apps-Update-2011.pdf [accessed 2015-01-21] [WebCite Cache ID 6VkdhzwwS]
8. Brewer AC, Endly DC, Henley J, Amir M, Sampson BP, Moreau JF, et al. Mobile applications in dermatology. JAMA Dermatol 2013 Nov;149(11):1300-1304. [doi: 10.1001/jamadermatol.2013.5517] [Medline: 24067948]
9. Stevens DJ, Jackson JA, Howes N, Morgan J. Obesity surgery smartphone apps: a review. Obes Surg 2014 Jan;24(1):32-36. [doi: 10.1007/s11695-013-1010-3] [Medline: 23749609]
10. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
11. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [Medline: 11392867]
12. Choi J, Noh GY, Park DJ. Smoking cessation apps for smartphones: content analysis with the self-determination theory. J Med Internet Res 2014;16(2):e44 [FREE Full text] [doi: 10.2196/jmir.3061] [Medline: 24521881]
13. Dahlke DV, Fair K, Hong YA, Beaudoin CE, Pulczinski J, Ory MG. Apps Seeking Theories: Results of a Study on the Use of Health Behavior Change Theories in Cancer Survivorship Mobile Apps. JMIR mHealth uHealth (2015) 2015 Mar 27;3(1):e31. [doi: 10.2196/mhealth.3861] [Medline: 25830810]
14. 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;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
15. PubMed. URL: http://www.ncbi.nlm.nih.gov/pubmed/ [WebCite Cache ID 6Y9Sf7My7]

16. mHealth Evidence. URL: https://www.mhealthevidence.org/ [accessed 2015-04-29] [WebCite Cache ID 6Y9ShgqCm]
17. Hewett P, Logan P, Mulhausen J, Ramachandran G, Banerjee S. Rating exposure control using Bayesian decision analysis. J Occup Environ Hyg 2006 Oct;3(10):568-581. [doi: 10.1080/15459620600914641] [Medline: 16998991]
18. Siegel E. Apptentive. 2014 May 27. Fake Reviews in Google Play and Apple App Store URL: http://www.apptentive.com/blog/fake-reviews-google-play-apple-app-store/ [accessed 2015-03-23] [WebCite Cache ID 6XFcPC3ah]
19. Nielsen J, Mack RL. Usability inspection methods. New York: Wiley; 1994.
20. Ruder Finn. 2013 Mar. mHealth Report URL: http://www.ruderfinn.com/pdf/Ruder%20Finn%20US%20mHealth%20report%20FINAL.pdf [accessed 2015-01-21] [WebCite Cache ID 6Vke6hgF8]
21. Hundert AS, Huguet A, McGrath PJ, Stinson JN, Wheaton M. Commercially available mobile phone headache diary apps: a systematic review. JMIR Mhealth Uhealth 2014;2(3):e36 [FREE Full text] [doi: 10.2196/mhealth.3452] [Medline: 25138438]
22. Xu W, Liu Y. mHealthApps: A Repository and Database of Mobile Health Apps. JMIR Mhealth Uhealth 2015;3(1):e28 [FREE Full text] [doi: 10.2196/mhealth.4026] [Medline: 25786060]
23. Pew Internet Research. 2014 Oct 24. Health Fact Sheet URL: http://www.pewinternet.org/fact-sheets/health-fact-sheet/ [accessed 2015-01-21] [WebCite Cache ID 6VkeDDnrE]
24. Ward BW, Schiller JS. Centers for Disease Control and Prevention. 2013 Dec 12. Prevalence of Multiple Chronic Conditions Among U.S. Adults: Estimates From the National Health Interview Survey URL: http://www.cdc.gov/pcd/issues/2013/12_0203.htm [accessed 2015-01-21] [WebCite Cache ID 6Vkdz3cD2]
25. McKinsey's Global iConsumer Research. 2011 Nov. Gaming Expands Its Presence in the Digital Universe: Insights from McKinsey's Global iConsumer Research URL: https://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/High%20Tech/PDFs/Gamings_Expanded_Presence_in_the_Digital_Universe.ashx [WebCite Cache ID 6Vkdqf1iy]
26. Pew Internet Research. 2014 Jan. Social Networking Fact Sheet URL: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/ [accessed 2015-01-21] [WebCite Cache ID 6VkeHiY2O]
27. PwC Health. 2012 Apr. Social Media "likes" Healthcare: From Marketing to Social Business URL: http://pwchealth.com/cgi-local/hregister.cgi/reg/health-care-social-media-report.pdf [accessed 2015-01-21] [WebCite Cache ID 6VkeRUlYW]

Abbreviations

mHealth: Mobile health
PPACA: Patient Protection and Affordable Care Act
SDT: Self-Determination Theory

Edited by G Eysenbach; submitted 26.01.15; peer-reviewed by M Scheibe, L van Velsen; comments to author 16.02.15; revised version received 07.04.15; accepted 22.04.15; published 13.05.15.

Please cite as:
Mendiola MF, Kalnicki M, Lindenauer S
Valuable Features in Mobile Health Apps for Patients and Consumers: Content Analysis of Apps and User Ratings
JMIR mHealth uHealth 2015;3(2):e40
URL: http://mhealth.jmir.org/2015/2/e40/
doi:10.2196/mhealth.4283
PMID:25972309

Original Paper

Patient Engagement With a Mobile Web-Based Telemonitoring System for Heart Failure Self-Management: A Pilot Study

Shiyi Zan[1,2*], BS; Stephen Agboola[1,2,3*], MD, MPH; Stephanie A Moore[2,3], MD; Kimberly A Parks[2,3], DO; Joseph C Kvedar[1,2,3], MD; Kamal Jethwani[1,2,3], MD, MPH

[1]Center for Connected Health, Partners HealthCare, Boston, MA, United States

[2]Harvard Medical School, Boston, MA, United States

[3]Massachusetts General Hospital, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Kamal Jethwani, MD, MPH
Center for Connected Health
Partners HealthCare
25 New Chardon Street
Suite 300
Boston, MA,
United States
Phone: 1 617 724 3410
Fax: 1 617 228 4634
Email: kjethwani@partners.org

Abstract

Background: Intensive remote monitoring programs for congestive heart failure have been successful in reducing costly readmissions, but may not be appropriate for all patients. There is an opportunity to leverage the increasing accessibility of mobile technologies and consumer-facing digital devices to empower patients in monitoring their own health outside of the hospital setting. The iGetBetter system, a secure Web- and telephone-based heart failure remote monitoring program, which leverages mobile technology and portable digital devices, offers a creative solution at lower cost.

Objective: The objective of this pilot study was to evaluate the feasibility of using the iGetBetter system for disease self-management in patients with heart failure.

Methods: This was a single-arm prospective study in which 21 ambulatory, adult heart failure patients used the intervention for heart failure self-management over a 90-day study period. Patients were instructed to take their weight, blood pressure, and heart rate measurements each morning using a WS-30 bluetooth weight scale, a self-inflating blood pressure cuff (Withings LLC, Issy les Moulineaux, France), and an iPad Mini tablet computer (Apple Inc, Cupertino, CA, USA) equipped with cellular Internet connectivity to view their measurements on the Internet. Outcomes assessed included usability and satisfaction, engagement with the intervention, hospital resource utilization, and heart failure-related quality of life. Descriptive statistics were used to summarize data, and matched controls identified from the electronic medical record were used as comparison for evaluating hospitalizations.

Results: There were 20 participants (mean age 53 years) that completed the study. Almost all participants (19/20, 95%) reported feeling more connected to their health care team and more confident in performing care plan activities, and 18/20 (90%) felt better prepared to start discussions about their health with their doctor. Although heart failure-related quality of life improved from baseline, it was not statistically significant (P=.55). Over half of the participants had greater than 80% (72/90 days) weekly and overall engagement with the program, and 15% (3/20) used the interactive voice response telephone system exclusively for managing their care plan. Hospital utilization did not differ in the intervention group compared to the control group (planned hospitalizations P=.23, and unplanned hospitalizations P=.99). Intervention participants recorded shorter average length of hospital stay, but no significant differences were observed between intervention and control groups (P=.30).

Conclusions: This pilot study demonstrated the feasibility of a low-intensive remote monitoring program leveraging commonly used mobile and portable consumer devices in augmenting care for a fairly young population of ambulatory patients with heart failure. Further prospective studies with a larger sample size and within more diverse patient populations is necessary to determine the effect of mobile-based remote monitoring programs such as the iGetBetter system on clinical outcomes in heart failure.

(JMIR mHealth uHealth 2015;3(2):e33) doi:10.2196/mhealth.3789

KEYWORDS

heart failure; disease self-management; remote monitoring; telemonitoring; interactive voice response system; mobile health; Web portal; patient engagement; quality of life

Introduction

The Burden of Congestive Heart Failure

Congestive heart failure is a chronic condition that is associated with significant morbidity, mortality, and reductions in quality of life, particularly among older adults ≥ 65 years of age. Hospital readmission rates for heart failure are among the highest of any chronic disease, and account for much of the financial burdens on the health care system. Upon discharge from the hospital, half of heart failure patients experience rehospitalization within 6 months [1]. In 2009, the total cost of heart failure-related treatment in the United States was about US $39 billion; by 2030, this number is projected to double as a result of our aging population [2].

Remote Monitoring for Congestive Heart Failure

Remote monitoring by structured telephone support or telemonitoring has been commonly explored as a promising strategy for improving heart failure outcomes [3]. These programs offer the potential to provide access to specialist care for a much larger number of patients across a much greater geography, assist health care providers in patient management, and effectively lower the burden of care from providers by engaging and supporting patients in self-care practices [3]. Nevertheless, to date, remote monitoring programs that have achieved success in improving clinical outcomes and reducing hospital readmissions are highly intensive (ie, requiring close clinical oversight and follow-up), and may therefore not be appropriate for all patient groups, especially those who have less severe disease. Furthermore, the scalability of remote monitoring programs has often been limited by high equipment costs and the logistics and time delay associated with initial patient set-up on these programs. Considering that a high proportion of 30-day readmissions occur relatively soon after patients are discharged, there is a need for a solution that can help facilitate smoother transitions in care from the hospital to the home environment [4].

It is widely accepted in the heart failure literature that patients who are actively involved in their own care and adhere to treatment regimens are more likely to have improved survival, decreased readmission rates, and experience better quality of life [5,6]. The nearly ubiquitous penetration of wireless Internet, adoption of mobile phones, and availability of portable and affordable consumer-facing personal health monitoring devices offer a potential means of addressing the demands and burdens associated with disease self-management, while reducing heart failure-associated health care costs. To date, a number of studies have evaluated mobile device-based telemonitoring programs for heart failure, but no existing system that we are aware of has incorporated a patient-facing Web portal and leveraged consumer-facing digital devices to engage and empower patients in disease self-management [7-10].

Our Aims

In this study, we pilot-tested a Web- and telephone-based self-management intervention that leverages personal digital health monitoring devices in a population of ambulatory, adult heart failure patients. Our primary aim was to evaluate the feasibility and acceptability of use of the intervention in patients with heart failure. In addition, we assessed patient engagement with the program, and its impact on patients' quality of life. We also examined the effect of the intervention on hospital utilization by comparing patients who used the program with a matched set of passive controls obtained from the electronic medical record (EMR) system.

Methods

Study Design

This feasibility pilot was designed as a single-arm prospective study in which participants used the intervention for heart failure self-management over a 90-day study period.

Study Participants

Participants were recruited from the outpatient clinic of two cardiologists of the Massachusetts General Hospital (MGH) Heart Center's Heart Failure and Cardiac Transplant Program. As this was a pilot study, a convenience sample was used, consisting of patients who had scheduled outpatient visits over the course of a 4-month period between February and May 2013.

Patients were deemed eligible if they were ambulatory, English-speaking adults ≥18 years of age with a current diagnosis of heart failure. Patients were required to have regular access to a telephone and be able to navigate a simple website, understand the scope of the study, and provide written informed consent. Patients who were admitted to the hospital and/or enrolled in another remote monitoring program, as well as those with significant visual, hearing and/or cognitive impairments, and those with significant medical or psychiatric comorbidities were excluded. The Partners HealthCare Institutional Review Board approved all study procedures.

Study Procedures

After obtaining informed consent at an initial study visit, patients were set up with an account on the Web platform (their own "patient" portal) and were trained on how to use all components of the intervention.At the visit, demographic and other baseline data were collected using an enrollment questionnaire and digital devices were provided as part of the study. In addition to baseline demographics, the questionnaire also captured patients' baseline technology use and baseline social health using the emotional, informational, and instrumental support domains of the Patient Reported Outcomes Measurement Information System measures, and heart failure-related quality of life using the Minnesota Living with Heart Failure Questionnaire

(MLHFQ). Given that depression is a known risk factor of poor outcomes in heart failure, in addition, we also assessed levels of baseline depression using the 8-item Patient Health Questionnaire (PHQ-8) [11].

Patients were provided with a step-by-step reference guide, and instructed to take their measurements and log their care plan activities daily through the patient portal and/or the interactive voice response (IVR) telephone system. They were also instructed to notify study staff in advance if they expected any interruptions in study procedures (ie, planned hospital admissions) so that their study timeline could be adjusted appropriately. Following completion of the 3-month study, patients completed a closeout questionnaire by mail. All other relevant study data were collected from the EMR system and the Web platform.

Matched controls were identified from the EMR system, via the Partners HealthCare Research Patient Data Repository, for comparison with study participants on hospital resource utilization. Patients receiving the intervention were matched 1:1 with controls by age (±2 years), gender, race, and diagnosis.

Study Intervention

The iGetBetter system is comprised of a Web platform, IVR system, and portable consumer-facing digital devices that measure and collect key vital signs. Patients were enrolled in the intervention following a regular scheduled outpatient appointment with their cardiologist at the MGH Heart Failure Clinic, and began using the system at home the following morning.

Patients were instructed to take their weight, blood pressure, and heart rate measurements each morning using the provided devices (see Figure 1), which included a WS-30 bluetooth weight scale and self-inflating blood pressure cuff (Withings LLC, Issy les Moulineaux, France). Patients were also provided with an iPad Mini tablet computer (Apple Inc, Cupertino, CA, USA) equipped with cellular Internet connectivity for the accessibility and convenience of being able to view their measurements online.

Measurements taken by patients were transmitted onto the Web platform, where they were stored and displayed in a graphical fashion (see Figure 2). Patients were instructed to log into the patient portal (see Figure 3) to check off their care plan activities using the iPad Mini provided or their own computer at home. Patients who did not complete one or more of the listed steps prior to their self-selected morning reminder call time would receive a reminder phone call from the IVR system prompting them to log their care activities and/or manually enter in their measurements using their phone keypad. The IVR system essentially provided patients an alternative, manual means of recording their measurements and care activities, which were similarly uploaded and stored on the centralized Web platform.

Study participants' cardiologists had access to the Web platform through a "clinician" portal to view their patients' progress over time, but were not required to perform any active role in the study aside from recommending a predefined range of acceptable values for vital signs, which were input into the system for each patient during enrollment. During the study, if a patients' measurement fell outside of the preset range, a system alert would be triggered; research study staff monitored the Web platform for these alerts during regular business hours. Study staff made follow-up calls to patients if their measurements fell outside of their predefined range, and all clinically relevant alerts were routed to a patient's cardiologist and their care team. We emphasized to patients that the system was not to be used for reporting emergencies and did not serve as a substitute to their usual care regimen and clinic visits.

Figure 1. Portable devices provided to participants as part of the intervention (from left: iPad Mini, bluetooth weight scale, auto-inflating blood pressure cuff).

Figure 2. Screenshot of a sample progress report on the Web portal displaying a graphical representation of a patient's vitals taken using study devices.

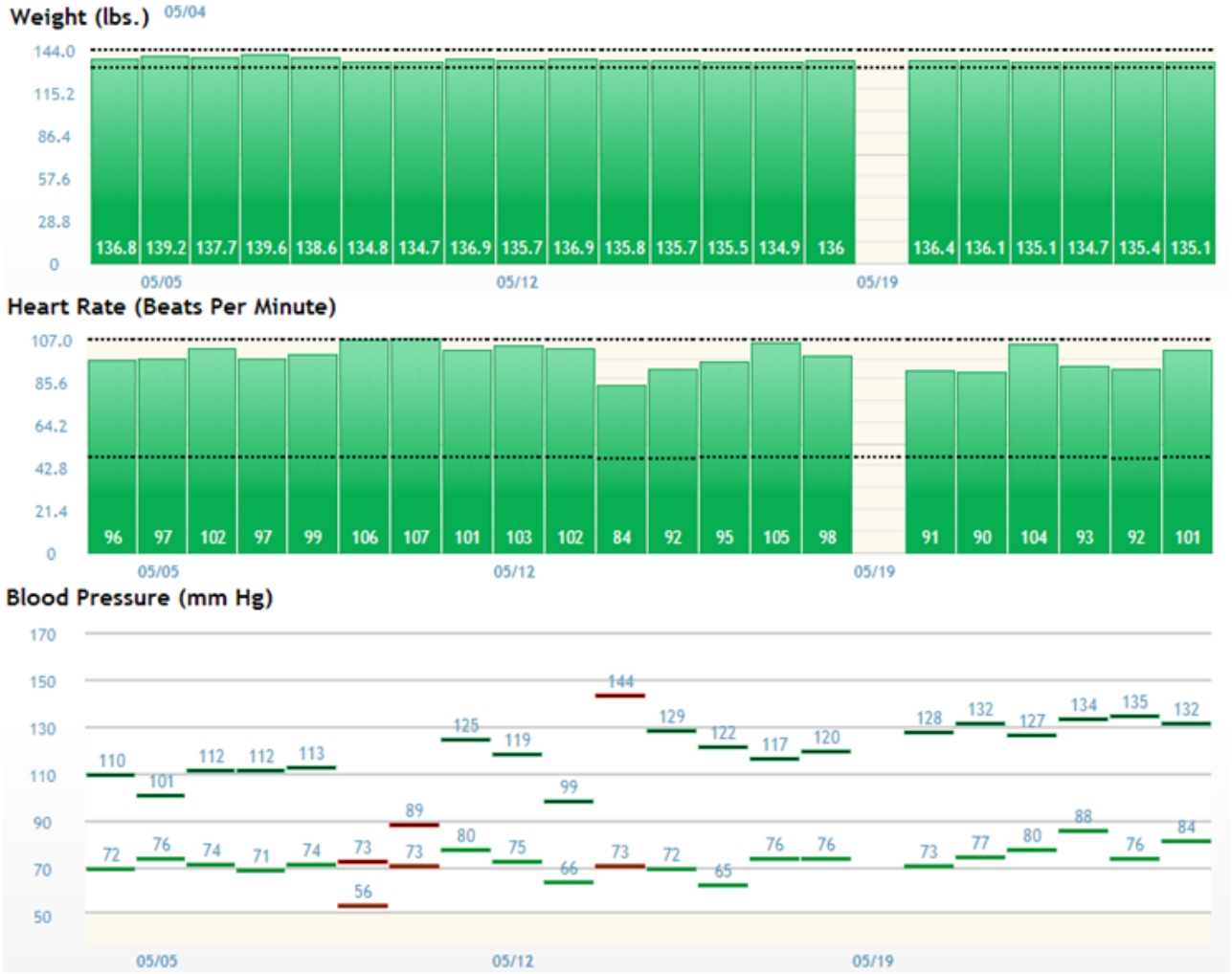


Figure 3. Screenshot of the patient-facing Web portal upon logging in.

Outcome Measures

Primary Outcome

Usability and satisfaction with the intervention was assessed through self-reported patient questionnaires.

Secondary Outcomes

Hospital resource utilization encompassed visits to the emergency room/urgent care clinic, as well as heart failure-related inpatient hospitalizations. The primary source of data for this measure was the EMR system for both the intervention group and their matched controls, but the intervention group was asked to self-report any additional out-of-system hospitalizations in the study closeout questionnaire. Heart failure-related quality of life was assessed pre and postintervention using the MLHFQ [12]. Previous studies have shown this questionnaire to be sensitive to quality of life changes in the heart failure patient population [13,14]. Engagement with the intervention was assessed objectively via daily care plan logging and Web portal log-ins.

Data Analysis

Baseline characteristics for the intervention group and EMR-matched control group were summarized with percentages for categorical variables and means and SD for continuous variables. The frequency and duration of hospitalizations for patients receiving the intervention were summarized and compared with EMR-matched control data. Descriptive statistics were used to analyze and summarize usability and satisfaction measures, heart failure-related quality of life measures, and overall engagement indices. Data analysis was performed using Stata version 12 with an alpha of .05 set *a priori*.

Results

Study Recruitment and Baseline Characterisitics

Of 32 patients assessed for eligibility, 21 agreed to participate and were enrolled into the study. There were 11 patients that were excluded from the study: 1 did not meet eligibility criteria, as they were no longer receiving care from an MGH cardiologist; 8 declined to participate due to either transportation limitations or lack of interest in the study; and 2 were unreachable (see Figure 4). Of the 21 enrolled, 20 completed the study.

The mean age of study participants was 53 years (SD 17); these patients were predominantly male, white, married, and had received higher education (ie, > 1 year of college-level education). One-third of participants were employed full-time at the time of enrollment (see Table 1). EMR-matched control demographics were similar to that of study participants and are reported in Table 1.

Of the 21 enrolled participants, 11 (52%) reported no depression as measured by the PHQ-8, while 3 (14%) reported moderate-severe levels of depression (see Table 1). The vast majority of participants scored well on social health; all reported adequate informational support, while the mean scores for emotional support and instrumental support were 96% and 94%, respectively. Participants' baseline technology use is displayed in Table 2.

Table 1. Baseline demographic characteristics of study participants.

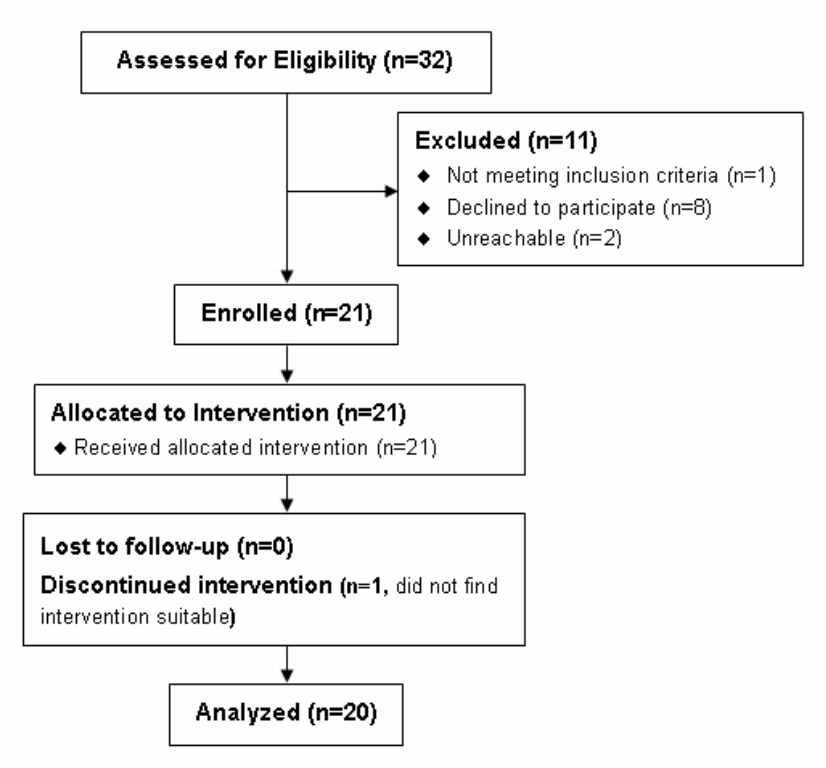
Baseline demographic characteristics		Intervention group, N=21	Matched control group, N=20
Age, years (SD)			
	Mean	53 (17)	53 (17)
	Range	21-81	22-81
Average # in household		2.4	-
Male gender, n (%)		15 (71)	14 (70)
Marital status, n (%)			
	Married	19 (90)	7 (35)
	Single	1 (5)	6 (30)
	Divorced	1 (5)	2 (10)
New York Heart Association class, n (%)			
	1	5 (24)	-
	2	9 (43)	-
	3	7 (33)	-
Left ventricular ejection fraction, mean (SD)		34.6 (14.9)	-
Education, n (%)			
	4+ yrs of college	12 (57)	-
	1-3 yrs of college	3 (14)	-
	12th grade, GED	3 (14)	-
	9th-11th grade	1 (5)	-
	1st-8th grade	1 (5)	-
Race, n (%)			
	White	19 (90)	16 (80)
	Black/African American	2 (10)	2 (10)
	Other	0 (0)	2 (10)
Employment status, n (%)			
	Employed full-time (includes self-employment)	7 (33)	-
	Retired	7 (33)	-
	Disabled	4 (19)	-
	Employed part-time (includes self-employment)	1 (5)	-
	Homemaker	1 (5)	-
	Unemployed	1 (5)	-
PHQ-8, n (%)			
	None (0-4)	11 (52)	-
	Mild (5-9)	5 (24)	-
	Moderate (10-14)	2 (10)	-
	Moderate-severe (15-19)	3 (14)	-
	Severe (20-24)	0 (0)	-
Social support (mean score %)			
	Emotional	96	-
	Informational	100	-

Baseline demographic characteristics		Intervention group, N=21	Matched control group, N=20
	Instrumental	94	-

Table 2. Baseline technology use of study participants (N=21).

Baseline technology use		n (%)
Method of accessing the Internet		
	Broadband	12 (60)
	Cellular network	10 (50)
	Wireless network	10 (50)
	Dial-up telephone	2 (10)
Have used the Internet for...		
	Email	19 (95)
	Looking for health/medical information	17 (85)
	Banking	17 (85)
	Sharing photos	16 (80)
	Instant messaging/online chat	15 (75)
	Accessing social networking sites	9 (45)
	Tracking weight, diet, or exercise routine	4 (20)
	Tracking other health indicators (eg, blood pressure, sleep, headaches)	4 (20)
Computer ownership		
	Laptop computer	15 (75)
	Desktop computer	14 (70)
	Tablet computer	8 (40)
Telephone ownership		
	Landline phone	15 (75)
	Cellular phone	17 (85)
	Mobile phone	15 (75)
Have used a cellular phone for...		
	Text messaging	16 (80)
	Sharing photos	15 (75)
	Accessing the Internet	14 (70)
	Email	14 (70)
	Banking	12 (60)
	Looking for health/medical information on the Internet	10 (50)
	Accessing social networking sites	6 (30)
	Tracking weight, diet, or exercise routine	3 (15)
	Tracking other health indicators (eg, blood pressure, sleep, headaches)	2 (10)

Figure 4. Participant enrollment and inclusion in to the study.



Usability and Satisfaction With the Intervention

Overall, all 20 study participants reported high levels of satisfaction with the intervention. Of these participants, 19 (95%) agreed "mostly" if not "definitely" with statements evaluating whether they felt comfortable using the intervention, whether they were satisfied with how easy it was to use the intervention, and whether they would recommend the intervention to friends and family members (see Table 3).

All participants also rated the intervention highly on usability, with the majority expressing definite agreement to statements regarding how easy it was to learn the system and carry out their home care activities using the system (see Table 4).

Table 3. Overall usability and satisfaction with the intervention as reported by study participants (N=20).

Survey question	Likert response, n (%)			
	Definitely true	Mostly true	A little bit true	Definitely not true
I felt comfortable using this system	18 (90)	2 (10)		
I was able to easily perform my home care activities using this system	16 (80)	4 (20)		
It was easy learning to use this system	15 (75)	4 (20)	1 (5)	
I am satisfied with how easy it was to use this system	13 (65)	7 (35)		
I would recommend the system to a friend or family member	12 (60)	7 (35)	1 (5)	

Table 4. Usability and satisfaction with specific components of the intervention as reported by study participants (N=20).

Survey question	Likert response, n (%)			
	Definitely true	Mostly true	A little bit true	Definitely not true
The measurement devices were easy to use	14 (70)	6 (30)		
The iPad Mini was easy to use	14 (70)	5 (25)	1 (5)	
It was easy to enter my data and log care plan activities on the website	7 (35)	6 (30)	2 (10)	3 (15)
The alert function was very useful when my measurements were out of range	7 (35)	4 (20)	3 (15)	4 (20)

Perceived Effect of the Intervention on Care

All but one of the final 20 participants (95%) reported feeling more confident in performing home care activities, and more connected to their health care team. Of these participants, 18 (90%) felt that the system helped them in starting discussions about their health with their doctor, and 16 (80%) also believed

that their disease was better controlled as a result of using the intervention (see Table 5).

Table 5. Patients' perceived effect of the intervention on heart failure related care (N=20).

Survey question	Likert response, n (%)			
	Definitely true	Mostly true	A little bit true	Definitely not true
The intervention helped me...				
Feel more confident in performing my home care activities	12 (60)	5 (25)	2 (10)	1 (5)
Feel more connected to my care team	10 (50)	4 (20)	5 (25)	1 (5)
Start discussions about my health with my doctor	8 (40)	4 (20)	5 (25)	2 (10)
Better control my disease	6 (30)	7 (35)	3 (15)	4 (20)
Remember to take my medications more regularly	8 (40)	1 (5)	5 (25)	6 (30)

Feedback on Intervention Components

All participants "liked" the measurement devices (ie, blood pressure cuff and weight scale) that were part of the intervention, and reported that they found the devices easy to use. With respect to the telephone component, 7 of the 20 participants (35%) "liked" the automated IVR reminder phone calls, and 14 (70%) reported that they found the follow-up phone calls from study staff helpful whenever they recorded an out-of-parameter measurement (see Table 6).

Table 6. Participants' (N=20) ratings of statements regarding use of the intervention components.

Survey question	Likert response, n (%), N=20			
	I like it very much	I like it	I like it a little bit	Not at all
Checking weight	13 (65)	5 (25)	2 (10)	
Blood pressure monitoring	12 (60)	6 (30)	2 (10)	
Out-of-range alerts for measurements	8 (40)	5 (25)	1 (5)	5 (25)
Viewing measurements on the website	7 (35)	3 (15)	5 (25)	3 (15)
IVR reminder phone calls	3 (15)	1 (5)	3 (15)	13 (65)

Use of the Web Platform

Of the 20 participants, 11 (55%) reported that they logged in to the patient portal to view their measurements. Of the 11 individuals, 7 (64%) viewed their measurements daily, and 8 (73%) believed that being able to view their measurements inspired more interest in their own health, and helped them better manage their health condition. All but one of the 11 participants (91%) reported discussing the system with others, while only 6 (55%) reported discussing the system with their doctor.

Engagement With the Intervention

The overall engagement with the two main components of the intervention (ie, the IVR system and the Web platform) was assessed by participants' daily care plan logging trends. Over half of study participants had 80% or greater adherence to care plan logging over the course of the study (ie, logged data for more than 72/90 days). Although most participants engaged through the patient portal, 3 participants (15%) who displayed high adherence to care plan logging were found to have used the IVR system exclusively. Upon breaking down patient engagement with the intervention by study week, we found that overall care plan logging engagement decreased following the first 4 weeks, but 15 of the 20 participants continued to engage with care plan logging throughout each week of the study.

The median number of Web portal log-ins was 28 log-ins per day in the first week of the study; log-ins decreased in subsequent weeks, but remained above 12 log-ins per day for the duration of the study period. Although 4 participants (20%) were found to have stopped logging in to their patient portal after the first week, 14 (70%) appeared to have consistently accessed the patient portal for the duration of the study.

Hospital Resource Utilization

The breakdown of all hospital encounters for the intervention group and EMR-matched control group are displayed in Table 7. Within the intervention group of 20 participants, 2 patients (10%) recorded unplanned hospital visits, and 3 (15%) recorded planned hospital admissions during the study. Only one of these hospital encounters was found to be a 30-day readmission. In addition, 3 patients (15%) from the intervention group were admitted to the hospital for planned procedures during the study period. Within the EMR-matched control group, one patient (5%) recorded hospitalizations within the study timeframe; this individual was found to have been hospitalized 3 times, and recorded two 30-day readmissions (see Table 7).

The mean unplanned length of hospital stay for the intervention group was 3.4 days compared to 8.7 days in the EMR-matched control group, although no significant differences were observed between the two groups (P=.30). Overall, no significant differences were observed between the intervention group and

control group in the number of admissions, length of hospital stay, or frequency of 30-day readmissions.

Table 7. Aggregate hospital encounter data for study participants and EMR-matched controls.

	Intervention, N=20	Control, N=20	*P* value
Inpatient hospital admissions, # of encounters (# of patients)			
Planned	5 (3)	0	.23
Unplanned	2 (2)	3 (1)	.99
Emergency room/urgent care visits, # of encounters (# of patients)	2 (2)	1 (1)	.99

Change in Self-Reported Quality of Life and Health

On the MLHFQ, a lower score indicates improvement in quality of life. Study participants' score on the MLHFQ decreased by approximately 4 points from pre to postintervention (see Table 8). This difference did not reflect a clinically meaningful change, and was not found to be statistically significant (*P*=.55).

Participants' score on the general self-rated health question item increased by 0.05 points from pre to postintervention, but this difference was not found to be statistically significant (*P*=.83) (see Table 8).

In addition to measures of heart failure-related quality of life, patients were also asked validated question items evaluating one's confidence in their ability to perform home care activities, take medications, and attend scheduled medical appointments. We found that the responses of patients to these questions did not change as a result of the intervention, and no statistically significant changes were detected.

Table 8. Quality of life measures pre and postintervention for study participants (N=20).

Self-reported health measure	Preintervention score, mean (SD)	Postintervention score, mean (SD)	Difference, mean (SD)	*P* value
Minnesota Living with Heart Failure Questionnaire Scale				
Overall	43 (26)	39 (27)	-4 (31)	.55
Physical	19 (12)	17 (13)	-2 (13)	.46
Emotional	9 (7)	8 (7)	-1.7 (9.0)	.44
General self rated health item				
Overall	2.70 (1.22)	2.75 (1.21)	0.05 (1.05)	.83

Discussion

Principal Findings

We report the results of a formative evaluation of a Web- and telephone-based intervention, leveraging portable consumer health technologies, for low intensive heart failure self-monitoring in a population of ambulatory, adult heart failure patients recruited from the cardiology clinic of an academic medical center. To the best of our knowledge, this is the first study evaluating the use of such a program incorporating portable consumer-facing digital devices to engage heart failure patients in self-managing their disease outside of the hospital setting.

Overall, patients reported high acceptability of the iGetBetter system, and found the intervention highly feasible and applicable to their care. Although the intervention encouraged patients to take their vitals, which were automatically transferred to a centralized Web portal, we also offered patients the option of using an IVR telephone system to report data to the Web platform (albeit by manual input of measurements), and set automated daily reminder calls for their care plan items. Our most significant finding was that 80% of patients (16/20) maintained a consistent pattern of reporting and viewing their data over the course of the 90-day follow-up period. In contrast, other studies have found engagement levels to be lower, and also taper off over time [7-10,15]. A possible explanation for our observed findings is that the portability and user-friendliness of study devices played an important role in fostering patient engagement compared to often outdated conventional remote monitoring devices (which rely on telephone lines to transmit data) that have been used in previous studies.

To gain further understanding of patients' use of the system, we examined the means by which patients engaged with different components of the intervention. In doing this, we found that 55% of study participants (11/20) consistently logged in to their patient portal to view their data; and that among these individuals, 64% (7/11) logged in using the iPad Mini provided and viewed their data daily. In addition, 73% of these patients (8/11) believed that being able to view their measurements inspired more interest in their own health, and that this in turn helped them better manage their health condition. These numbers show that despite low overall engagement, all patients who did engage appeared to have reported some benefit as a result of using the system. Furthermore, although only 35% of patients (7/20) reported finding the IVR system helpful, 15% (3/20) continued to use the IVR to log their data during the study. Given that the majority of the population today owns a mobile phone and mobile phone adoption rates are on the rise in older adults, the development of a Web platform that is compatible across multiple devices including a mobile phone interface may prove to be more useful and engaging for a wider spectrum of patient populations [16].

With respect to hospital resource utilization, although the total number of unplanned inpatient and emergency room/urgent care visits was comparable between the two groups, the mean duration of hospital stay for unplanned admissions in the EMR-matched control group was over twice as long as that of patients using the intervention. Furthermore, patients using the intervention also recorded fewer 30-day readmissions than controls, although these differences were not statistically significant. Nevertheless, a number of patients visited the hospital for scheduled procedures during the study period. It is possible that as a result of being more connected with their care team, the intervention helped facilitate better communication between patients and their care providers, resulting in more timely care management decisions. It is also possible that the ability to view patients' physiologic trends over time, in addition to alert notifications, could have provided clinicians with more data that allowed them to determine when an intervention or follow-up with a patient was warranted. In accordance with our findings, a previous study by Wu et al has also reported trends toward increased hospitalizations for planned procedures among users of a Web-based heart failure management platform compared to nonusers receiving standard care [17].

Quality of life, a major issue in heart failure and a key focus of treatment, is affected by the functional capabilities, symptoms, and psychosocial perceptions of patients [5]. Although not statistically significant, the final quality of life measure improved from baseline, suggesting that the intervention may have had a positive effect on patients' self-perceived disease burden. Eighty-percent of patients (16/20) believed that their heart failure was better controlled as a result of using the intervention; this, along with other feedback we received from patients, suggests that the majority felt more empowered to carry out their self-care activities at home with the help of the system. Although more impressive and significant improvements in quality of life have been observed in studies using Web-based heart failure self-management programs, in our study, we enrolled patients in the outpatient setting while they were clinically stable, and patients were also only followed up for a 90-day study period, so it is possible that observing a change in quality of life under these circumstances was more difficult [15,18,19].

The fact that over half of our study population had greater than 80% adherence to daily care plan logging throughout the study period is encouraging and somewhat unprecedented when compared with previous Web-based self-management programs for heart failure. Nevertheless, a few patients did not engage with the system for reasons that may range from a lack of insight regarding their condition, not being sufficiently "activated" for self-management, or potentially, if they were already satisfied with their level of knowledge regarding self-care.

Limitations

A limitation to our study is in the small sample size. Because we designed the study as a feasibility pilot, we did not conduct a formal sample size calculation, and our study was not powered to detect the effects of the intervention on clinical outcomes. Additionally, since patients were identified through purposeful sampling from two cardiologists to participate in the study, selection bias cannot be ruled out. Although one patient reported not having an Internet connection at home, the majority of patients were more likely to be familiar with technology at baseline; it may well be the case that patients who may not have been as tech savvy, but who were at least more open to trying out a new program were more likely to enroll and engage with the system. Furthermore, those who comprised our study sample were predominantly married, white male, ambulatory heart failure patients with varying socioeconomic and educational backgrounds; this limits the generalizability of our findings and precludes us from extrapolating findings from this study to other populations. Finally, due to the relatively low intensity of the remote monitoring component, the current system may not be suitable for patients with more severe disease; further investigation involving a more diverse heart failure patient population with varying degrees of disease severity is necessary to evaluate the true impact of the intervention.

Conclusions

Our study demonstrates the feasibility and acceptability of a Web-based telemonitoring system that incorporates personal digital health monitoring devices in a group of ambulatory heart failure patients. Although this study is a small sample pilot not powered for statistical inference, our findings suggest the potential for improved health outcomes in similar patient populations who use the system. The portability and convenience offered by the consumer-facing digital devices provided as part of the remote monitoring system likely contributed to patient satisfaction and high engagement levels, while dissatisfaction with the IVR system and technical difficulties likely affected adoption and engagement in certain patients. Further study within a larger sample size is needed to determine the extent of the benefits of the system on heart failure outcomes.

Acknowledgments

This study was funded by iGetBetter, Inc (Waltham, MA). The authors would like to express sincere thanks to Patrick Wyld, Laura Peterson, and the rest of the clinic staff and participating patients at the MGH Heart Center's Heart Failure and Cardiac Transplant Program; Timothy Hale, PhD; Ellen Murachver, MSW; Brian Hammond; Meaghan Doherty; John Coco; Suzanne Grant; and the rest of the research team at the Center for Connected Health for making this study possible.

Conflicts of Interest

None declared.

References

1. Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, American Heart Association Statistics CommitteeStroke Statistics Subcommittee. Heart disease and stroke statistics--2012 update: A report from the American Heart Association. Circulation 2012 Jan 3;125(1):e2-e220 [FREE Full text] [doi: 10.1161/CIR.0b013e31823ac046] [Medline: 22179539]
2. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, American Heart Association Advocacy Coordinating Committee, Council on Arteriosclerosis, ThrombosisVascular Biology, Council on Cardiovascular RadiologyIntervention, Council on Clinical Cardiology, et al. Forecasting the impact of heart failure in the United States: A policy statement from the American Heart Association. Circ Heart Fail 2013 May;6(3):606-619 [FREE Full text] [doi: 10.1161/HHF.0b013e318291329a] [Medline: 23616602]
3. Inglis SC, Clark RA, McAlister FA, Stewart S, Cleland John G F. Which components of heart failure programmes are effective? A systematic review and meta-analysis of the outcomes of structured telephone support or telemonitoring as the primary component of chronic heart failure management in 8323 patients: Abridged Cochrane Review. Eur J Heart Fail 2011 Sep;13(9):1028-1040. [doi: 10.1093/eurjhf/hfr039] [Medline: 21733889]
4. Hernandez AF, Greiner MA, Fonarow GC, Hammill BG, Heidenreich PA, Yancy CW, et al. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. JAMA 2010 May 5;303(17):1716-1722. [doi: 10.1001/jama.2010.533] [Medline: 20442387]
5. Lainscak M, Blue L, Clark AL, Dahlström U, Dickstein K, Ekman I, et al. Self-care management of heart failure: Practical recommendations from the Patient Care Committee of the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 2011 Feb;13(2):115-126. [doi: 10.1093/eurjhf/hfq219] [Medline: 21148593]
6. van der Wal Martje H L, van Veldhuisen Dirk J, Veeger Nic J G M, Rutten FH, Jaarsma T. Compliance with non-pharmacological recommendations and outcome in heart failure patients. Eur Heart J 2010 Jun;31(12):1486-1493 [FREE Full text] [doi: 10.1093/eurheartj/ehq091] [Medline: 20436049]
7. Nundy S, Razi RR, Dick JJ, Smith B, Mayo A, O'Connor A, et al. A text messaging intervention to improve heart failure self-management after hospital discharge in a largely African-American population: Before-after study. J Med Internet Res 2013;15(3):e53 [FREE Full text] [doi: 10.2196/jmir.2317] [Medline: 23478028]
8. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: A randomized controlled trial. J Med Internet Res 2012;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]
9. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, Telemedical Interventional Monitoring in Heart Failure Investigators. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: The telemedical interventional monitoring in heart failure study. Circulation 2011 May 3;123(17):1873-1880 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.018473]
10. Scherr D, Kastner P, Kollmann A, Hallas A, Auer J, Krappinger H, MOBITEL Investigators. Effect of home-based telemonitoring using mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation: Randomized controlled trial. J Med Internet Res 2009;11(3):e34 [FREE Full text] [doi: 10.2196/jmir.1252]
11. Jankowska EA, Drohomirecka A, Ponikowska B, Witkowska A, Lopuszanska M, Szklarska A, et al. Deficiencies in circulating testosterone and dehydroepiandrosterone sulphate, and depression in men with systolic chronic heart failure. Eur J Heart Fail 2010 Sep;12(9):966-973. [doi: 10.1093/eurjhf/hfq108]
12. Rector TS, Cohn JN. Assessment of patient outcome with the Minnesota Living with Heart Failure questionnaire: Reliability and validity during a randomized, double-blind, placebo-controlled trial of pimobendan. Pimobendan Multicenter Research Group. Am Heart J 1992 Oct;124(4):1017-1025. [Medline: 1529875]
13. Guyatt GH. Measurement of health-related quality of life in heart failure. J Am Coll Cardiol 1993 Oct;22(4 Suppl A):185A-191A. [Medline: 8376691]
14. Guyatt GH, Nogradi S, Halcrow S, Singer J, Sullivan MJ, Fallen EL. Development and testing of a new measure of health status for clinical trials in heart failure. J Gen Intern Med 1989;4(2):101-107. [Medline: 2709167]
15. Maric B, Kaan A, Araki Y, Ignaszewski A, Lear SA. The use of the Internet to remotely monitor patients with heart failure. Telemed J E Health 2010;16(1):26-33. [doi: 10.1089/tmj.2009.0094] [Medline: 20070163]
16. Smith A. Pew Research Center. 2013. Smartphone ownership – 2013 update URL: http://www.pewinternet.org/files/old-media//Files/Reports/2013/PIP_Smartphone_adoption_2013_PDF.pdf [accessed 2015-02-15] [WebCite Cache ID 6WMgfVXdO]
17. Wu RC, Delgado D, Costigan J, Maciver J, Ross H. Pilot study of an Internet patient-physician communication tool for heart failure disease management. J Med Internet Res 2005;7(1):e8 [FREE Full text] [doi: 10.2196/jmir.7.1.e8] [Medline: 15829480]
18. Kashem A, Droogan MT, Santamore WP, Wald JW, Bove AA. Managing heart failure care using an internet-based telemedicine system. J Card Fail 2008 Mar;14(2):121-126. [doi: 10.1016/j.cardfail.2007.10.014] [Medline: 18325458]
19. Delgado DH, Costigan J, Wu R, Ross HJ. An interactive Internet site for the management of patients with congestive heart failure. Can J Cardiol 2003 Nov;19(12):1381-1385. [Medline: 14631472]

Abbreviations

EMR: electronic medical record
IVR: interactive voice response
MGH: Massachusetts General Hospital
MLHFQ: Minnesota Living with Heart Failure Questionnaire
PHQ-8: Patient Health Questionnare-8

Edited by G Eysenbach; submitted 15.08.14; peer-reviewed by P Kastner; comments to author 21.09.14; revised version received 04.11.14; accepted 16.12.14; published 01.04.15.

Please cite as:
Zan S, Agboola S, Moore SA, Parks KA, Kvedar JC, Jethwani K
Patient Engagement With a Mobile Web-Based Telemonitoring System for Heart Failure Self-Management: A Pilot Study
JMIR mHealth uHealth 2015;3(2):e33
URL: http://mhealth.jmir.org/2015/2/e33/
doi:10.2196/mhealth.3789
PMID:25842282

Original Paper

Identifying Factors Associated With Dropout During Prerandomization Run-in Period From an mHealth Physical Activity Education Study: The mPED Trial

Yoshimi Fukuoka[1], RN, PhD; Caryl Gay[2], PhD; William Haskell[3], PhD; Shoshana Arai[4], RN, PhD; Eric Vittinghoff[5], PhD

[1]Institute for Health & Aging, Department of Social and Behavioral Sciences, University of California, San Francisco, San Francisco, CA, United States
[2]Institute for Health & Aging, Department of Family Health Care Nursing, University of California, San Francisco, San Francisco, CA, United States
[3]Stanford Prevention Research Center, Stanford University, Palo Alto, CA, United States
[4]Institute for Health & Aging, University of California, San Francisco, San Francisco, CA, United States
[5]Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Yoshimi Fukuoka, RN, PhD
Institute for Health & Aging
Department of Social and Behavioral Sciences
University of California, San Francisco
#LHts-340
3333 California St
San Francisco, CA, 94143-0646
United States
Phone: 1 415 476 8419
Fax: 1 415 502 5208
Email: Yoshimi.Fukuoka@ucsf.edu

Abstract

Background: The mobile phone-based physical activity education (mPED) trial is a randomized controlled trial (RCT) evaluating a mobile phone-delivered physical activity intervention for women. The study includes a run-in period to maximize the internal validity of the intervention trial, but little is known about factors related to successful run-in completion, and thus about potential threats to external validity.

Objective: Objectives of this study are (1) to determine the timing of dropout during the run-in period, reasons for dropout, optimum run-in duration, and relevant run-in components, and (2) to identify predictors of failure to complete the run-in period.

Methods: A total of 318 physically inactive women met preliminary eligibility criteria and were enrolled in the study between May 2011 and April 2014. A 3-week run-in period was required prior to randomization and included using a mobile phone app and wearing a pedometer. Cross-sectional analysis identified predictors of dropout.

Results: Out of 318 participants, 108 (34.0%) dropped out prior to randomization, with poor adherence using the study equipment being the most common reason. Median failure time was 17 days into the run-in period. In univariate analyses, nonrandomized participants were younger, had lower income, were less likely to drive regularly, were less likely to have used a pedometer prior to the study, were generally less healthy, had less self-efficacy for physical activity, and reported more depressive symptoms than randomized participants. In multivariate competing risks models, not driving regularly in the past month and not having used a pedometer prior to the study were significantly associated with failure to be randomized (P=.04 and .006, respectively), controlling for age, race/ethnicity, education, shift work, and use of a study-provided mobile phone.

Conclusions: Regular driving and past pedometer use were associated with reduced dropout during the prerandomization run-in period. Understanding these characteristics is important for identifying higher-risk participants, and implementing additional help strategies may be useful for reducing dropout.

Trial Registration: ClinicalTrials.gov NCT01280812; https://clinicaltrials.gov/ct2/show/NCT01280812 (Archived by WebCite at http://www.webcitation.org/6XFC5wvrP).

(JMIR mHealth uHealth 2015;3(2):e34) doi:10.2196/mhealth.3928

KEYWORDS

run-in period; eligibility; randomized controlled trial; pedometer; mobile phone; mHealth

Introduction

Given the exponential growth of mobile phone use—both basic- and advanced-feature mobile phones—across all age groups [1], mobile health technology has become a popular way to deliver physical activity interventions and monitor physical activity. Applying mobile phone technologies to a randomized controlled trial (RCT) has great potential to improve measurement and intervention methodologies. However, several important methodological questions, such as the value and limitations of run-in procedures for mobile technology-based RCTs, have not been adequately addressed [2], particularly for physical activity interventions.

An RCT is the gold standard for examining intervention efficacy and effectiveness. RCTs tend to focus on internal validity at the expense of external validity, and thus, study findings can have limited external validity (ie, generalizability) [3]. A run-in procedure has been proposed for RCTs [4] to minimize the challenges of attrition and nonadherence to the intervention being evaluated, which can be significant threats to trials' internal validity [4]. Several RCTs have used this design [5,6]. Using a run-in period as part of mobile app-based physical activity intervention trials allows researchers to screen out ineligible (eg, already active) or noncompliant (eg, low adherence to app use) research participants prior to randomization, and thereby improve the internal validity of these RCTs. Run-in periods can be especially useful with technology-based interventions, since individuals often adopt and discontinue technology use at different speeds [4].

The mobile phone-based physical activity education (mPED) study [7] is an RCT with a run-in procedure and is designed to evaluate the efficacy of a mobile app-delivered physical activity intervention—for both basic- and advanced-feature mobile phones—for physically inactive women. An overall goal of this paper was to describe the process of selecting mPED participants prior to randomization and their characteristics. Generally, factors affecting run-in attrition include environmental factors (ie, physical and social environment, such as social support and program location), program factors (eg, design, recruitment processes, and eligibility criteria), and person-based factors (eg, demographics and beliefs about exercise) [8]. Understanding such characteristics among the physically inactive women who were randomized and those who were not is a critical part of evaluating the mPED study's external validity, and could also yield useful information for guiding the implementation and possible dissemination of this mobile phone-delivered physical activity intervention. In addition, evaluating the timing and patterns of ineligibility can inform the design of run-in periods in future RCTs by determining optimum duration and relevant components.

Thus, the aims of this study were to (1) describe the timing of, and reasons for, nonrandomization in the mPED trial and (2) identify predictors of failure to complete the run-in period.

Methods

Study Design

The mPED study is an RCT that included a preliminary telephone screening call, a screening/baseline study visit, and a 3-week run-in period to determine participants' eligibility for randomization. In this paper, mPED data from this prerandomization phase of the study were analyzed to compare the characteristics of randomized versus nonrandomized participants and to evaluate the timing of ineligibility during the 3-week run-in period. The study protocol was approved by the University of California, San Francisco Committee on Human Research and the mPED Data and Safety Monitoring Board. The study protocol was published in 2011 [7]. All potential participants received a copy of the informed consent form electronically or by mail after completion of telephone screening and were asked to review it before the screening/baseline visit. All participants provided written consent prior to study enrollment. This RCT was registered at ClinicalTrials.gov (NCT01280812).

Subject Recruitment

Physically inactive women were recruited from the San Francisco Bay Area from May 2011 to April 2014. With the aim of recruiting a diverse and representative sample, four broad types of subject recruitment strategies were used: (1) media advertising (eg, newspaper, radio, Craigslist and Facebook ads, commercial email distribution lists, and study, clinic, and ClinicalTrials.gov websites), (2) posting fliers in the community (eg, stores, bus stops, medical and dental clinics, community centers, university campuses, and churches), (3) random mailing of the study announcement to women aged 25 to 69 who live in San Francisco, and (4) referral from friends, family members, health care providers, or others contacts.

Inclusion and Exclusion Criteria

Preliminary inclusion criteria were assessed during the telephone screening and included the following: (1) female, aged 25 to 69 years, (2) body mass index (BMI) of 18.5-43.0 kg/m^2, (3) physically inactive lifestyle as indicated by a Stanford Brief Activity Survey [9] score indicating inactivity or light activity during leisure time and at work, if employed, (4) intent to become physically active, (5) willingness to use the pedometer and intervention app every day for 9 months, (6) access to a home telephone or mobile phone, and (7) ability to speak and read English. Preliminary exclusion criteria included the following: (1) known medical conditions or physical problems that require special attention in an exercise program, (2) planning an international trip during the next 4 months, which could interfere with daily server uploads of mobile phone data,

(3) pregnant/gave birth during the past 6 months, (4) severe hearing or speech problem, (5) history of an eating disorder, (6) current substance abuse, (7) current participation in lifestyle modification programs or research studies that may confound study results, and (8) history of bariatric surgery or plans for bariatric surgery in the next 12 months. Women who had never used a mobile phone or were not current mobile phone users were not excluded.

Additional inclusion criteria were assessed at the screening/baseline visit or during/after the run-in period and included the following: (1) a physical exam confirming BMI and medical eligibility information obtained during the initial telephone screening (eg, height, weight, resting blood pressure), (2) a fasting blood test confirming medical eligibility, (3) a baseline average of <8500 daily steps measured during the run-in period, and (4) at least 80% adherence to all run-in activities (described below). Participants were also assessed using the Mini-Cog test [10,11] and were excluded if there was evidence of mild cognitive impairment.

Telephone Screening

During the initial screening call, a trained study staff member screened potential participants for preliminary eligibility. Potential participants who met preliminary eligibility criteria were invited to attend a screening/baseline visit and were sent the study consent form, public transportation and parking information, directions to the research office, and a list of study requirements, which included a picture of the pedometer they would be asked to wear.

Screening/Baseline Visit

The screening/baseline visit was scheduled approximately 1 week after completion of the telephone screening. The primary goal of the screening/baseline visit was to further determine the participant's eligibility. Once written informed consent was obtained, participants were screened for mild cognitive impairment using the Mini-Cog test and were asked to complete baseline questionnaires (described in the Measures section below) and a physical exam. Eligible participants were issued a mobile phone and a pedometer (described below)—training was provided to ensure participants could successfully use both devices. The mean training time for the use of the mobile phone and pedometer was 14.1 (SD 6.3) minutes.

Run-in Period

Overview

The run-in period lasted approximately 3 weeks. During that time, participants were asked to wear the pedometer, use the mobile phone app, and have a fasting blood test.

Pedometer

The Omron Active Style Pro HJA-350IT with triaxial accelerometer was selected for this trial because it has well-established reliability and validity, has been used in similar studies, and records 150 days of activity data. Its dimensions are 74x46x34 mm (width/height/depth) including the clip, and it weighs 60 grams (2.1 oz), including batteries. The pedometer was set to record and store physical activity (eg, steps), but not to display the step counts—only date and time were displayed to limit reactivity. Participants were asked to wear the pedometer all day, except when showering/bathing, swimming, or sleeping, from the time they got up in the morning until they went to bed at night for the duration of the run-in period. Participants were asked to wear the pedometer on their waist, aligned with their dominant knee.

Mobile Phone App

A run-in mobile phone app was created specifically for this phase of the study—it was designed to mimic the intervention app without any content to encourage or support increasing physical activity. A Java 2 Platform Micro Edition (J2ME) app for basic-feature mobile phones was initially developed for the study, and later iPhone (iOS) and Android apps were developed. None of the delivery platforms (ie, J2ME, iOS, and Android) required Internet connectivity. The run-in app sent daily messages unrelated to physical activity throughout the run-in period (eg, "Did you eat breakfast today?"), and participants were instructed to respond to each message. In addition, participants were instructed to enter an estimate of their daily step count into the app's daily activity diary every day of the run-in period. Adherence to these instructions was monitored remotely, and participants with low adherence were contacted by project staff to troubleshoot problems and alert them that they were at risk for not meeting the run-in criteria. The app could be installed on a participant's personal phone if they had a compatible mobile phone, and the study reimbursed the cost of upgrading their text and data plans to cover study-related use. Alternatively, participants were provided with a mobile phone for the purpose of the study. Study-issued mobile phones had unlimited data and text messaging and 70 minutes of voice calls per month, and no restrictions on personal use.

Fasting Blood Test

Participants were also asked to complete a fasting blood test in a clinical research lab during the first week of the run-in period. Blood samples were assayed to obtain a lipid profile, fasting plasma glucose, and hemoglobin A1c (HbA1c) values.

The run-in period had three purposes: (1) to determine a baseline average number of daily steps, (2) to determine the participant's level of adherence to the study procedures, and (3) to obtain baseline fasting blood test results to evaluate the potential risks or benefits of participating in the study. If a participant had five or more cardiovascular risk factors, research staff, with the participant's permission, sent a study enrollment notification letter to their health care provider.

Randomization Visit and Adherence Indicators

At the conclusion of the run-in period, participants were scheduled for a randomization visit. At this visit, the pedometer data were downloaded and reviewed to ensure that the participant met the criterion for a daily average of <8500 steps across the run-in period. Adherence indicators for both the pedometer and mobile phone app were also reviewed to ensure 80% adherence to each of the following criteria: (1) pedometer-wearing time of at least 8 hours per day, (2) responding to the app's daily messages, and (3) using the app's daily activity diary. In addition to consistently using the pedometer and mobile phone app throughout the run-in period,

participants were required to successfully complete the fasting blood test to be eligible for randomization.

Participant Payments

Participants ineligible to start the run-in were paid US $10 in cash at the screening/baseline visit. Participants who completed the run-in, but were not randomized were paid US $20 in cash at the randomization visit. Randomized participants received a US $40 check when they completed the 3-month study visit. Parking was provided at the research office, and participants were reimbursed for parking expenses for the blood draw. Those who took public transportation did not receive specific reimbursement for their transportation.

Measures

Overview

Sociodemographics, lifestyle and health characteristics, and past digital technology use were assessed during the telephone screening or screening/baseline visit and are summarized in Table 1. Total television and computer usage (hours per week) was assessed using a questionnaire developed by the research team based on a thorough review of the literature. In addition, the following four standardized and validated questionnaires were administered at the screening/baseline visit.

Self-Efficacy for Physical Activity Scale

A 6-item modified version of the original 5-item Self-Efficacy for Physical Activity Scale [12] was used to assess confidence in one's ability to exercise, an important determinant of the stages of change for exercise behavior. The scale assesses one's perceived ability to exercise despite common challenges (ie, bad weather, limited time, feeling tired, bad mood, or being on vacation). Based on a pilot study [13], the scale was modified to include a sixth item assessing one's ability to exercise during times of stress. Total scores can range from 6 to 30, with higher scores indicating greater self-efficacy for physical activity.

Social Support for Exercise Survey

The Social Support for Exercise Survey consists of 13 items assessing the level of perceived support from family and friends for behavior changes related to exercise [14]. Each item is scored separately for family and friends, and scores can range from 13 to 65, with higher scores indicating greater support.

Barriers to Being Active Quiz

The Barriers to Being Active Quiz consists of 21 items assessing seven types of barriers to physical activity: lack of time, lack of social influence, lack of energy, lack of willpower, fear of injury, lack of skill, and lack of resources [15]. Scores can range from 0 to 63, with higher scores indicating more barriers to physical activity.

Center for Epidemiological Studies Depression Scale

The Center for Epidemiological Studies Depression Scale (CES-D) is a 20-item questionnaire widely used for assessing symptoms of depression [16]. Scores can range from 0 to 60, with higher scores indicating more depressive symptoms.

Definition of Time to Drop Out

The primary outcome was time to drop out, defined as the number of days between the screening/baseline visit and the earliest indicator of run-in failure. This outcome was measured using the following: (1) the date a participant called or emailed to withdraw from the study, or (2) the date a participant entirely stopped wearing the pedometer or using the mobile phone app. For participants who completed the run-in period, the time to drop out was defined as the number of days between the screening/baseline visit and the randomization visit.

Statistical Analysis

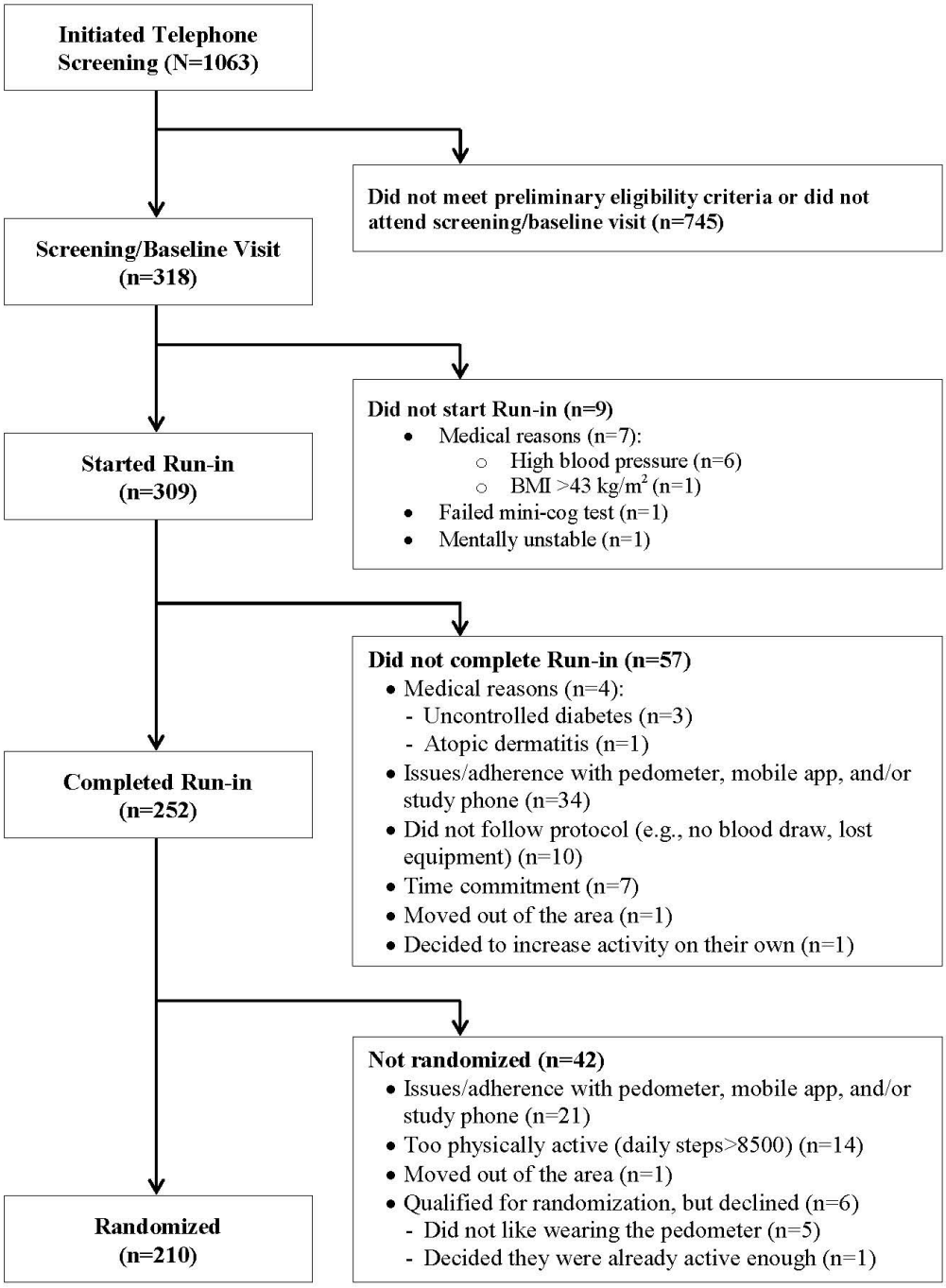
Sample size was based on the primary outcomes of the RCT [7]. Descriptive statistics were used to summarize participant characteristics. Comparisons of randomized and nonrandomized participants were conducted using chi-square tests or independent sample *t* tests, as appropriate. Fine-Gray competing risk models [17] were used to estimate adjusted covariate effects on dropout during the run-in period, accounting for successful randomization as a competing risk. All variables in Table 1 were evaluated as potential predictors. The final adjusted model was obtained using forward selection with an entry criterion of $P<.10$ except that age, race/ethnicity, education, shift work, and use of a study-provided mobile phone were included and retained by default for face validity. Cumulative incidence of dropout was estimated by the baseline cumulative incidence function of a simple Fine-Gray model with no covariates. Analyses were conducted using Stata version 13.1 (Stata Corp, College Station, TX).

Results

Subject Enrollment and Dropout

A total of 1063 potential participants were screened by telephone for initial eligibility. Of these, 745 (70.08%) did not meet the initial eligibility criteria or did not attend the screening/baseline visit, and the remaining 318 (29.92%) women were initially eligible, enrolled in the study, and completed the screening/baseline visit. Of the 318 participants enrolled in the study, 210 (66.0%) successfully completed the run-in period and were randomized to one of the three groups, and the remaining 108 (34.0%) were not randomized for various reasons, as listed in Figure 1. The most common reasons for nonrandomization were issues related to the pedometer, study phone, and/or mobile app (55/108, 50.9%). Among the 108 participants who were not randomized, the median failure time was 17 days into the run-in period. Issues/adherence problems with the pedometer included not wanting to wear it because it was bulky, uncomfortable, or difficult to wear and failing to wear it 8 hours per day on at least 80% of the run-in days. Issues/adherence problems with the mobile app included not liking aspects of it (eg, the timing of the messages and diary), not understanding the instructions, not meeting the 80% adherence rate, as well as mobile app glitches. Issues with the phone included not liking the study-provided mobile phone and finding it difficult to use.

Figure 1. Flow diagram of the mPED run-in period.



Comparisons of Randomized and Nonrandomized Participants

Table 1 summarizes univariate comparisons between the 210 randomized participants and 108 nonrandomized participants. Compared to randomized participants, nonrandomized participants were younger, had lower income, had a lower self-rating of general health, reported less self-efficacy for physical activity and more depressive symptoms, were less likely to drive on a regular basis, and were less likely to have used a pedometer prior to the study. Downloading the mPED app to one's own phone was not associated with randomization or age (P=.33). Although not statistically significant, nonrandomized participants were slightly less likely to have participated in a weight-loss or diet program and had slightly higher scores on the Barriers to Being Active scale.

Table 1. Univariate comparisons of randomized and nonrandomized participants (n=318).

Measure	Nonrandomized participants (n=108), n (%) or mean (SD)	Randomized participants (n=210), n (%) or mean (SD)	*P*
Sociodemographics and lifestyle characteristics			
Age (years), mean (SD)	49.0 (12.4)	52.4 (11.1)	*.01* [a]
Race/ethnicity, n (%)			.69
Native Hawaiian/Pacific Islander	1 (0.9)	0 (0.0)	
Black/African American	9 (8.3)	17 (8.1)	
Hispanic/Latino	7 (6.5)	13 (6.2)	
Asian	24 (22.2)	41 (19.5)	
White (non-Hispanic)	60 (55.6)	119 (56.7)	
More than one race/ethnicity	7 (6.5)	20 (9.5)	
Education, n (%)			.81
Completed high school or some college	25 (23.1)	52 (24.8)	
Completed college	42 (38.9)	86 (41.0)	
Graduate school	41 (38.0)	72 (34.3)	
Household income (annual), n (%)			*.03*
≤ US $40,000	30 (27.8)	32 (15.2)	
US $40,001 to $75,000	27 (25.0)	50 (23.8)	
> US $75,000	42 (38.9)	111 (52.9)	
Don’t know/declined to state	9 (8.3)	17 (8.1)	
Marital status, n (%)			.82
Never married	35 (32.4)	64 (30.5)	
Currently married/cohabitating	51 (47.2)	107 (51.0)	
Divorced/widowed	22 (20.4)	39 (18.6)	
Employment and shift work, n (%)			.66
Full- or part-time job with no shift work	52 (48.1)	108 (51.4)	
Full- or part-time job with shift work	23 (21.3)	48 (22.9)	
No paid employment	33 (30.6)	54 (25.7)	
Has a dog, n (%)	22 (20.4)	44 (21.0)	.90
Smoked a cigarette during the past 7 days, n (%)	3 (2.8)	4 (1.9)	.62
Drove at least once a week during the last month, n (%)	80 (74.1)	176 (83.8)	*.04*
Prior weight-loss or diet program participation, n (%)	56 (51.9)	132 (62.9)	.06
Used a pedometer prior to the study, n (%)	38 (35.2)	109 (51.9)	*.005*
Technology use, n (%)			
Used a mobile phone at least once a week during the last month	104 (96.3)	200 (95.2)	.66
Used a computer or accessed the Internet at least once a week during the last month	106 (98.1)	208 (99.0)	.49
Owns advanced mobile phone	70 (64.8)	125 (59.5)	.36
Subscribed to text messaging plan (n=316)[b]	88/108 (81.5)	158/208 (76.0)	.66
Used Facebook during the last month (n=303)[b]	64/98 (65.3)	140/205 (68.3)	.60
Used their own phone for the study (n=313)[c]	36/103 (35.0)	70/210 (33.3)	.78
Type of mobile phone used during the study (n=313) [c]			.31

Measure	Nonrandomized participants (n=108), n (%) or mean (SD)	Randomized participants (n=210), n (%) or mean (SD)	*P*
Motorola	12/103 (11.7)	16/210 (7.6)	
Pantech	31/103 (30.1)	51/210 (24.3)	
iPhone	59/103 (57.3)	142/210 (67.6)	
Android	1/103 (1.0)	1/210 (0.5)	
Health characteristics			
Overall rating of general health (scale 1-7), mean (SD)	4.74 (1.14)	5.00 (1.05)	*.01*
Measured body mass index (kg/m^2), mean (SD)	29.2 (6.1)	29.9 (6.2)	.34
Self-reported high blood pressure, n (%)	28 (25.9)	52 (24.8)	.98
Self-reported high cholesterol, n (%)	26 (24.1)	71 (33.8)	.16
Self-reported prediabetes or type 2 diabetes, n (%)	7 (6.5)	16 (7.6)	.71
Self-report questionnaires, mean (SD)			
Self-Efficacy for Physical Activity Scale	18.0 (5.1)	19.2 (4.6)	*.04*
Social Support for Exercise Survey			
Family	31.8 (9.7)	32.1 (9.7)	.81
Friend	32.2 (7.8)	31.5 (8.4)	.43
Barriers to Being Active: total score	25.5 (9.7)	23.4 (10.1)	.08
Depressive symptoms (CES-D[d])	12.1 (9.6)	9.7 (7.6)	*.02*
Total television and computer usage (hours per week), mean (SD)	28.5 (20.2)	27.5 (18.5)	.67
Participant recruitment strategies, n (%)			.67
Media advertising	29 (26.9)	66 (31.4)	
Posting fliers in the community	36 (33.3)	62 (29.5)	
Selective mailing	31 (28.7)	53 (25.2)	
Referral from friends, family members, health care providers, or other contacts	12 (11.1)	29 (13.8)	

[a]*P* values <.05 appear in italics.

[b]Missing data.

[c]Study mobile phones were not issued to 5 excluded participants at the start of the run-in period.

[d]Center for Epidemiological Studies Depression Scale (CES-D).

Predictors of Time to Failure During the Run-in Period

Table 2 summarizes the adjusted subdistribution hazard ratios (SHRs) for covariate effects on failure during the run-in period, accounting for the competing risk of successful randomization. In models adjusting for all variables in Table 2, not driving and not having used a pedometer before were the only predictors that remained statistically significant ($P<.05$). Sensitivity analyses excluding the 9 participants ineligible to start the run-in period indicated no meaningful change in the results reported in Table 2.

Table 2. Multivariate subdistribution hazard models predicting time to failure during the run-in period (n=313).

Predictors		SHR[a] (95% CI)	*P*
Face validity predictors			
Age		0.98 (0.97-1.00)	.09
Race/ethnicity			
	Other[b] (reference)	1	
	White (non-Hispanic)	1.12 (0.74-1.70)	.60
Education			
	Completed high school (reference)	1	
	Completed college	1.14 (0.66-1.96)	.65
	Completed graduate school	1.58 (0.91-2.76)	.11
Employment status			
	Employed, day shift (reference)	1	
	Employed with shift work	1.23 (0.72-2.12)	.45
	Unemployed	1.52 (0.97-2.37)	.07
Used their own phone for the study		1.05 (0.70-1.59)	.81
Other predictors			
Drove at least once a week in the past month		0.61 (0.38-0.98)	*.04* [c]
Used a pedometer prior to the study		0.56 (0.37-0.85)	*.006*
Self-efficacy for physical activity		0.96 (0.91-1.00)	.05

[a]Subdistribution hazard ratio (SHR).

[b]Includes Native Hawaiian/Pacific Islander, black/African American, Hispanic/Latino, Asian, and more than one race/ethnicity.

[c]*P* values <.05 appear in italics.

Discussion

Principal Findings

To our knowledge, this study represents the first report to examine participant characteristics in relation to the timing of dropout during the run-in period of an intervention trial utilizing both mobile phone and pedometer technology. Overall, 34.0% (108/318) of the women who successfully completed the screening/baseline visit were not randomized, and the median time from the screening/baseline visit to dropout was 17 days. Equipment issues were the most common reasons for not being randomized, specifically trouble using the study phone or mobile phone app and refusal or failure to consistently wear the pedometer. Furthermore, women who had never used a pedometer prior to the trial were more likely not to be randomized, compared to those with a prior history of pedometer use. This finding could be explained by the study requirement of wearing a large pedometer all day throughout the 3-week run-in period. Wearing a relatively large pedometer every day for 9 months can be a considerable challenge for some women, particularly those who wear dresses or do not like the appearance of a large device clipped to their clothing. Although a picture of the pedometer was sent to participants immediately after the telephone screening, those who had never used a pedometer before may not have realized what they looked like or what it would be like to wear one every day. This finding highlights the importance of assessing past pedometer usage and helping potential participants understand all study requirements before the screening/baseline visit.

Women who drove regularly in the month prior to study enrollment were more likely to be randomized compared to those who did not. This finding could be explained by the fact that the study provided parking stickers to participants who drove to the study office, but no reimbursement was provided for public transportation. As reported in other studies [3,18], providing transportation/parking to study participants appeared to be critical to participant retention in this physical activity trial. In addition, receiving a separate reimbursement for travel may increase study retention, even when the total amount is the same as when provided in a single payment [19]. Providing door-to-door service or no-cost transportation appears to be important, particularly for nonwhite women. Basing studies in the communities they aim to serve may also reduce transportation barriers for women who do not drive regularly. The participant retention rate in relation to the research costs associated with these services needs to be evaluated when designing a study.

In contrast, neither older age nor use of one's own mobile phone for the study had a significant effect on attrition during the run-in period, even though a small proportion 14/318 (4.4%) of enrolled women did not regularly use a mobile phone. Researchers often assume that older participants and subjects who do not download the study app onto their own mobile phone

will have high attrition rates, but neither of these assumptions was supported by the findings of this study. Other recent papers on weight-loss interventions also concluded that younger women were more likely to drop out from weight-loss clinical trials compared to older women [20,21]. In this study, a similar trend was observed. We believe that providing a brief mPED app training session, as well as the simple app design and use of the app prior to the run-in period, helped older women to start and continue the run-in period.

Mobile phone technologies are evolving at an exponential rate to better meet the needs of consumers, and these technological improvements can also be incorporated into clinical trials to improve the experience of study participants. However, rapid adaptation to these changing technologies in the middle of a clinical trial can be challenging, if not impossible. Despite this challenge, three versions of the mPED app were developed and used over the 3-year study period. Yet, only 106/313 (33.9%) of the participants used the mPED app on their own mobile phone during the run-in period, while the remaining participants used an mPED study mobile phone.

Strengths and Limitations

The findings from this study have methodological implications for researchers who design clinical trials involving mobile phone and pedometer technologies. The inclusion of subjects with diverse characteristics and the extensive technology information collected allowed us to explore these factors in relation to attrition during the run-in period. However, several study limitations need to be acknowledged. Only women aged 25 to 69 years were included, and thus, the findings may not generalize to men or children. Also, the study provided free parking, but did not provide a separate reimbursement for public transportation costs, which may have decreased retention, particularly among lower-income participants who are less likely to drive. In addition, San Francisco Bay Area's extensive public transportation network makes it possible to access the research office without a car, and thus, the effect of driving status on attrition might be different in areas with more limited public transportation. Finally, the study was not able to distinguish between the different types of nonadherence because participants who struggled with one form of technology often gave up on the others as well. Future studies should develop more rigorous ways of independently assessing adherence to different technologies.

Conclusions

In conclusion, regular driving and pedometer use prior to the study were associated with reduced dropout from the mPED trial's prerandomization run-in period. Understanding these characteristics is important when interpreting results of the mPED trial or when designing a similar physical activity trial in the future.

Acknowledgments

This project was supported by a grant (R01HL104147) from the National Heart, Lung, and Blood Institute and by the American Heart Association. The study sponsors had no role in the study design, the collection, analysis, or interpretation of data, writing the report, or the decision to submit the report for publication.

Conflicts of Interest

None declared.

References

1. Pew Research Institute. Washington, DC: Pew Internet & American Life Project; 2014. Mobile technology fact sheet URL: http://www.pewinternet.org/fact-sheets/mobile-technology-fact-sheet/ [accessed 2014-03-15] [WebCite Cache ID 6O6HTLIjx]
2. Affuso O, Kaiser KA, Carson TL, Ingram KH, Schwiers M, Robertson H, et al. Association of run-in periods with weight loss in obesity randomized controlled trials. Obes Rev 2014 Jan;15(1):68-73 [FREE Full text] [doi: 10.1111/obr.12111] [Medline: 24118736]
3. Coday M, Boutin-Foster C, Goldman ST, Tennant J, Greaney ML, Saunders SD, et al. Strategies for retaining study participants in behavioral intervention trials: retention experiences of the NIH Behavior Change Consortium. Ann Behav Med 2005 Apr;29 Suppl:55-65. [doi: 10.1207/s15324796abm2902s_9] [Medline: 15921490]
4. 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]
5. Nang EE, Gitau Ngunjiri SA, Wu Y, Salim A, Tai ES, Lee J, et al. Validity of the International Physical Activity Questionnaire and the Singapore Prospective Study Program physical activity questionnaire in a multiethnic urban Asian population. BMC Med Res Methodol 2011;11:141 [FREE Full text] [doi: 10.1186/1471-2288-11-141] [Medline: 21995825]
6. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. Int J Behav Nutr Phys Act 2011;8:115 [FREE Full text] [doi: 10.1186/1479-5868-8-115] [Medline: 22018588]
7. Fukuoka Y, Komatsu J, Suarez L, Vittinghoff E, Haskell W, Noorishad T, et al. The mPED randomized controlled clinical trial: applying mobile persuasive technologies to increase physical activity in sedentary women protocol. BMC Public Health 2011;11:933 [FREE Full text] [doi: 10.1186/1471-2458-11-933] [Medline: 22168267]

8. King AC, Blair SN, Bild DE, Dishman RK, Dubbert PM, Marcus BH, et al. Determinants of physical activity and interventions in adults. Med Sci Sports Exerc 1992 Jun;24(6 Suppl):S221-S236. [Medline: 1625548]
9. Taylor-Piliae RE, Norton LC, Haskell WL, Mahbouda MH, Fair JM, Iribarren C, et al. Validation of a new brief physical activity survey among men and women aged 60-69 years. Am J Epidemiol 2006 Sep 15;164(6):598-606 [FREE Full text] [doi: 10.1093/aje/kwj248] [Medline: 16840522]
10. Borson S, Scanlan J, Brush M, Vitaliano P, Dokmak A. The mini-cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. Int J Geriatr Psychiatry 2000 Nov;15(11):1021-1027. [Medline: 11113982]
11. Borson S, Scanlan JM, Chen P, Ganguli M. The Mini-Cog as a screen for dementia: validation in a population-based sample. J Am Geriatr Soc 2003 Oct;51(10):1451-1454. [Medline: 14511167]
12. Marcus BH, Selby VC, Niaura RS, Rossi JS. Self-efficacy and the stages of exercise behavior change. Res Q Exerc Sport 1992 Mar;63(1):60-66. [doi: 10.1080/02701367.1992.10607557] [Medline: 1574662]
13. Fukuoka Y, Vittinghoff E, Jong SS, Haskell W. Innovation to motivation--pilot study of a mobile phone intervention to increase physical activity among sedentary women. Prev Med 2010;51(3-4):287-289 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.006] [Medline: 20600263]
14. Sallis JF, Grossman RM, Pinski RB, Patterson TL, Nader PR. The development of scales to measure social support for diet and exercise behaviors. Prev Med 1987 Nov;16(6):825-836. [Medline: 3432232]
15. Centers for Disease Control and Prevention: National Diabetes Education Program. Atlanta, GA: Centers for Disease Control and Prevention; 1999. Barriers to being active quiz: What keeps you from being more active? URL: http://www.cdc.gov/diabetes/ndep/pdfs/8-road-to-health-barriers-quiz-508.pdf [accessed 2014-10-21] [WebCite Cache ID 6TUwM6Ucu]
16. Weissman MM, Sholomskas D, Pottenger M, Prusoff BA, Locke BZ. Assessing depressive symptoms in five psychiatric populations: a validation study. Am J Epidemiol 1977 Sep;106(3):203-214. [Medline: 900119]
17. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Amer Stat Assoc 1999;94(446):496-509.
18. Nagler RH, Ramanadhan S, Minsky S, Viswanath K. Recruitment and retention for community-based eHealth interventions with populations of low socioeconomic position: strategies and challenges. J Commun 2013 Feb 1;63(1):201-220 [FREE Full text] [doi: 10.1111/jcom.12008] [Medline: 23439871]
19. Brown BA, Long HL, Gould H, Weitz T, Milliken N. A conceptual model for the recruitment of diverse women into research studies. J Womens Health Gend Based Med 2000;9(6):625-632. [doi: 10.1089/15246090050118152] [Medline: 10957751]
20. Moroshko I, Brennan L, O'Brien P. Predictors of dropout in weight loss interventions: a systematic review of the literature. Obes Rev 2011 Nov;12(11):912-934. [doi: 10.1111/j.1467-789X.2011.00915.x] [Medline: 21815990]
21. Warner ET, Glasgow RE, Emmons KM, Bennett GG, Askew S, Rosner B, et al. Recruitment and retention of participants in a pragmatic randomized intervention trial at three community health clinics: results and lessons learned. BMC Public Health 2013;13:192 [FREE Full text] [doi: 10.1186/1471-2458-13-192] [Medline: 23496916]

Abbreviations

BMI: body mass index
CES-D: Center for Epidemiological Studies Depression Scale
HbA1c: hemoglobin A1c
J2ME: Java 2 Platform Micro Edition
mPED: mobile phone-based physical activity education
RCT: randomized controlled trial
SHR: subdistribution hazard ratio

Edited by G Eysenbach; submitted 07.10.14; peer-reviewed by M Duncan, R Freak-Poli; comments to author 15.12.14; revised version received 20.01.15; accepted 07.02.15; published 13.04.15.

Please cite as:
Fukuoka Y, Gay C, Haskell W, Arai S, Vittinghoff E
Identifying Factors Associated With Dropout During Prerandomization Run-in Period From an mHealth Physical Activity Education Study: The mPED Trial
JMIR mHealth uHealth 2015;3(2):e34
URL: http://mhealth.jmir.org/2015/2/e34/
doi:10.2196/mhealth.3928
PMID:25872754

Original Paper

Cartographic Analysis of Antennas and Towers: A Novel Approach to Improving the Implementation and Data Transmission of mHealth Tools on Mobile Networks

William Brown III[1,2], DrPH, MA; Mobolaji Ibitoye[1,3], MPH; Suzanne Bakken[2,4], RN, PhD; Rebecca Schnall[4], RN, PhD; Iván Balán[1,5], PhD; Timothy Frasca[1], MPH; Alex Carballo-Diéguez[1,5], PhD

[1]HIV Center for Clinical and Behavioral Studies, New York State Psychiatric Institute, Columbia University, New York, NY, United States

[2]College of Physicians and Surgeons, Department of Biomedical Informatics, Columbia University, New York, NY, United States

[3]Mailman School of Public Health, Sociomedical Sciences, Columbia University, New York, NY, United States

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

[5]Columbia University Medical Center, Department of Psychiatry, Columbia University, New York, NY, United States

Corresponding Author:
William Brown III, DrPH, MA
HIV Center for Clinical and Behavioral Studies
New York State Psychiatric Institute
Columbia University
HIV Center, Unit 15
1051 Riverside Drive
New York, NY,
United States
Phone: 1 646 774 6963
Fax: 1 212 543 6003
Email: wb2253@columbia.edu

Abstract

Background: Most mHealth tools such as short message service (SMS), mobile apps, wireless pill counters, and ingestible wireless monitors use mobile antennas to communicate. Limited signal availability, often due to poor antenna infrastructure, negatively impacts the implementation of mHealth tools and remote data collection. Assessing the antenna infrastructure prior to starting a study can help mitigate this problem. Currently, there are no studies that detail whether and how the antenna infrastructure of a study site or area is assessed.

Objective: To address this literature gap, we analyze and discuss the use of a cartographic analysis of antennas and towers (CAAT) for mobile communications for geographically assessing mobile antenna and tower infrastructure and identifying signal availability for mobile devices prior to the implementation of an SMS-based mHealth pilot study.

Methods: An alpha test of the SMS system was performed using 11 site staff. A CAAT for the study area's mobile network was performed after the alpha test and pre-implementation of the pilot study. The pilot study used a convenience sample of 11 high-risk men who have sex with men who were given human immunodeficiency virus test kits for testing nonmonogamous sexual partners before intercourse. Product use and sexual behavior were tracked through SMS. Message frequency analyses were performed on the SMS text messages, and SMS sent/received frequencies of 11 staff and 11 pilot study participants were compared.

Results: The CAAT helped us to successfully identify strengths and weaknesses in mobile service capacity within a 3-mile radius from the epicenters of four New York City boroughs. During the alpha test, before CAAT, 1176/1202 (97.84%) text messages were sent to staff, of which 26/1176 (2.21%) failed. After the CAAT, 2934 messages were sent to pilot study participants and none failed.

Conclusions: The CAAT effectively illustrated the research area's mobile infrastructure and signal availability, which allowed us to improve study setup and sent message success rates. The SMS messages were sent and received with a lower fail rate than those reported in previous studies.

(JMIR mHealth uHealth 2015;3(2):e63) doi:10.2196/mhealth.3941

KEYWORDS

cartographic analysis; mHealth; mobile health; antenna; short message service; text messaging; SMS; wireless; HIV

Introduction

Short message service (SMS) text messaging is one of the most ubiquitous digital forms of communication in the world—an average of 350 billion text messages a month are sent across the world's mobile networks [1]. A recent US survey reported that 84.06% (1914/2277, margin of error 3 percentage points) of adults own mobile phones and of those 73.00% (1212/1914) send and receive SMS text messages. Young adults age 18 to 24 lead this trend, exchanging an average of 109.5 messages daily [2]. Due to the ubiquity of mobile communication networks, mHealth—particularly the use of SMS—is growing as an innovative remote data collection and intervention method in biomedical research.

Searches performed on the National Library of Medicine's MEDLINE/PubMed database indicate that the biomedical literature on SMS nearly doubles each year. Title and abstract searches for the terms "short message service" and "text messaging" yielded 563 articles published between December 2000 and January 2013 (Table 1). Sorting these results by publication year revealed the yearly growth in the number of publications on these topics, indicating a substantial increase in the implementation of SMS in biomedical research. There was a 43.2% increase in the number of papers in 2010 compared to 2009, a 92.1% increase in 2011 compared to 2010, and a 94.2% increase in 2012 compared to 2011.

Researchers studying the use of remote data collection and transmission in biomedical research have demonstrated the potential for data collection efficiency and accuracy, real-time behavioral reporting, and assisting adherence to biomedical protocols [3-9]. However, multiple systematic literature reviews on SMS applications for disease prevention in developing countries reported several barriers to SMS system use [5,10]. One barrier is mobile network signal fluctuations, often due to tower positioning and obstructions. The second is short signal ranges. Both of these barriers affect feasibility, implementation, and acceptability [5,10-12]. As an example, one study included 9000 youth using SMS for seeking and reporting health information. Of the 2160/9000 (24%) dissatisfied users, 259/2160 (12.00%) did not receive a response to their SMS due to mobile network fluctuations, and 1080/2160 (50.00%) complained about related timing delays in receiving responses [5].

Identifying areas with a greater number of antenna towers and more optimal positioning of antenna towers can address signal fluctuations and short signal ranges. Currently, there are no studies in the literature that detail if and how the mobile infrastructure of a study area was assessed. We sought to fill this gap in the literature by conducting a cartographic analysis of the mobile infrastructure in our area prior to the implementation of a SMS pilot study assessing the utility of SMS text messaging to improve adherence and data collection to a biomedical protocol. Though the pilot study used SMS, the standardized communication protocols used by SMS are identical to those used to transmit data to and from most mHealth tools in biomedical research [8,13-15]. As such, the aim of this paper is to describe the use of a cartographic analysis of antennas and towers (CAAT) for mobile communications for geographically assessing mobile antenna and tower infrastructure and identifying signal availability for mobile devices. The findings from this study are widely applicable to other studies using mHealth tools.

Table 1. SMS articles published between December 2000 and January 2013.

Publication year	Number of PubMed publications	Annual increase from previous Year (%)
Dec 2000-Jan 2013	563	—
2012	235	94.2%
2011	121	92.1%
2010	63	43.2%
2009	44	—

Methods

SMS System Alpha Test

Before conducting the pilot study, a test team at the research site conducted simulated operational testing. Eleven staff members were asked to test the system and act as alpha test participants (N=11). Each of the eleven staff members/participants reported mock behavior via SMS once a day for one week and answered three to five branched questions per reporting session. Questions varied based on the answers that were given. After participants finished their reporting period, they were prompted for general feedback via email. Participants were asked, open-endedly, to discuss thoughts, ideas, suggestions, and concerns. Topics reported included system speeds, errors, system typos, reporting typos, question structure changes, SMS algorithm changes, reporting session length, and overall acceptability. Descriptive statistics were calculated on SMS text message sent and received data.

Analysis of SMS sent and received data from the alpha test identified several failed message deliveries that were due to weak mobile service signals. A failed message status, as

indicated by our open-source SMS system, means the message was unable to be sent to the mobile service provider for delivery. After confirming all hardware and software was functioning correctly, weak mobile service signal was the last possible reason for failed messages. Weak signal was confirmed using the signal strength gauge on a mobile phone with the same service provider. Thus, signal strength was a potential barrier to the successful implementation and collection of the SMS data. To better determine the feasibility of implementing the pilot study in New York City and to identify geographic areas of mobile service signal strength and weakness, we assessed the robustness of antenna tower networks within the research area.

Cartographic Methods

The research area that was alpha tested for the pilot study was New York City, which has five boroughs and a varied landscape of buildings, foliage, and altitudes. The site and location of our text messaging system was the HIV Center for Clinical and Behavioral studies at the Columbia University Medical Center (CUMC) in the northern region of the borough of Manhattan.

Data collection success depends on both sending messages as well as each participant receiving messages. Therefore, it was important to assess the research area where the participants live and conduct most of their daily activities. Thus, in addition to upper Manhattan, cartographic analyses were performed on lower Manhattan and the other three most populous of the five boroughs of New York City (Brooklyn, Queens, and the Bronx). Each borough was assessed in a 1 to 3 mile radius (depending on borough) of each outer borough's epicenter. Areas of mobile service signal strengths and weakness were identified, and our implementation strategy for the subsequent pilot study was modified accordingly.

Pilot Study

This pilot study was approved by the New York State Psychiatric Institute-Columbia University Department of Psychiatry Institutional Review Board. The pilot study drew its biomedical protocol and study population from a parent study "HIV Home Test and Decision-Making Among HIV-Negative Men" [16]. Eleven participants from the parent study were recontacted and enrolled in the pilot study. Study participants comprised an ethnically diverse sample of men who reported frequently engaging in unprotected anal intercourse with nonmonogamous male partners. All men were New York City residents at the time of enrollment, male, 18 years of age or older, fluent in English, human immunodeficiency virus (HIV)-negative (confirmed at enrollment), and aware that unprotected anal intercourse (UAI) with a partner of HIV-positive or unknown serostatus carries risk of HIV transmission.

Participants were then enrolled in the study and given 12 OraQuick Rapid HIV-1/2 test kits, with the option to use them to test their nonmonogamous sexual partners and/or themselves before intercourse over the course of the 10-week study period. Condoms were provided with each test kit.

Pilot Study SMS Procedures

Participants were trained to use an SMS system to report their HIV test kit use. Participants reported their sexual behavior and HIV test kit use daily via SMS. They were also asked to specify which time of the day they preferred to receive text messages reminding them to send their report into the system. After the first 3-week period of HIV test kit use and SMS reporting, participants were asked to stop reporting for one month so that researchers could do a preliminary system assessment. During this month participants were allowed to continue HIV test kit use. After the 1-month period, participants were asked to continue reporting to the system for another 3-week period. In case of a positive HIV test result or other problems, participants could contact study staff via an emergency hotline or SMS text message.

One dollar in compensation was provided for each SMS report completed, plus an additional $25 to offset participant SMS sending and receiving costs. The maximum possible compensation per participant was $165.

Open Source SMS System

We used FrontlineSMS to send and receive text messages. FrontlineSMS was chosen based on its versatility and breadth of system features. The SMS system consisted of four components: (1) an open source computer-based application where researchers logged in and managed sending and receiving messages, contact information, user preferences, and passcodes; (2) a message database; (3) a module to set message reminder frequency; and (4) a global system for mobile communications (GSM) modem to send and receive text messages.

To ensure the privacy of study participants, messages were processed through a third-party SMS gateway carrier with message encryption capability. All sent and received messages were encrypted. In addition, the service contract included an agreement that the messages processed on their servers would be encrypted and protected and that privacy would be maintained. At the study site a record of all the messages was stored on a password-protected server in a secured location.

Social Network Environmental Scan of GSM Modems

The GSM modem provides the direct link from the computer to the GSM network. FrontlineSMS provides a detailed list of tested GSM modems (and phones that can be used as GSM modems) that are compatible with FrontlineSMS's open-source software [17]. An environmental scan was performed within FrontlineSMS's user social network community on GSM modem use, which helped to designate the most effective modem to use for the study. We systematically searched the social network community forum discussion threads with the following keywords: modem, United States, works, GSM, price, Falcom, Huawei, Wavecom, Sierra Compass, Onda, ZTE Incorporated, Samba, E Series, Fastrack, 885, MC8781, mini modem, and MF627. After aggregating search results and analyzing the success rates and functionalities described in the forum, we chose the Huawei E Series GSM modem because it had the best reported performance and most positive responses, and we purchased it from a non-FrontlineSMS vendor.

Cartographic Analysis of Antennas and Towers

We conducted our cartographic analysis to develop a geographic map of the study area's antennas and towers and to obtain detailed service provider information. We used AntennaSearch, which provides information on service providers, antennas, and both existing and future towers, to facilitate the analysis. Antennas are the actual emitters of radio signals. Antennas can be placed on towers (multiple) or can be installed to stand alone on top of existing buildings. Stand-alone antennas are usually small (under 200 ft). It is also possible to check multiple antennas to determine which cell phone carriers are located on a particular tower. Existing towers can be registered or nonregistered structures where antennas are placed. Towers may be used for various services including cellular, paging, and microwave. Future towers indicate newly filed (or pending) applications to construct new towers. Application information includes location coordinates and detailed ownership data [18].

Our cartographic analysis of antennas and towers consisted of sorting mobile carriers by geographic proximity, distance imputation, and general urban topology. These methods were applied to detect geographic mobile service changes, identify resources, analyze proximities, and assess the general robustness of the study area's mobile service infrastructure. This involved identifying an address location as the true epicenter of the area to be surveyed and imputing that address into a web-based mobile service tower search and mapping engine. The data collected included information on service ownership, antenna and tower locations, distance from the specified site, owner contact information for antennas and existing towers, and projections for the development of future towers.

SMS Message Frequency Analysis

We calculated descriptive statistics of the SMS messages. For the data analysis, we analyzed all messages in both the alpha test phase and pilot study. Alpha test and SMS pilot data were analyzed together, disaggregated, and categorically. Messages were sorted first by message type (pending, sent, received, failed), then by date and time, and finally by message content.

Results

Results From Cartographic Analysis of Antennas and Towers

One hundred and sixteen towers (3 registered, 113 not registered) were found within three miles of our study site; in addition, applications for three future towers were found. There were a total of 965 antennas found. Though the general area was robust and had a significant amount of towers within a mile of the study site, the building that the study was housed in was located in a service gap (Figures 1-3). For the alpha testing phase of implementation, messages were serviced by T-Mobile USA, Inc. The cartographic analysis showed service providers within a three-mile radius of the study site and the number of antennas they owned. At the time of our analysis, Nextel Communications, Inc. was the largest commercial service provider, followed by Clearwire Spectrum Holdings, Northrop Grumman Systems Corp, and ATT Corp (Table 2). There were no antenna towers privately owned and serviced by T-Mobile. Nextel had no text-message-only service plans and required a contract beyond the life of the study, which would not have been cost effective. Therefore, for the actual pilot study we switched to ATT Corp for its robustness and signal strength at the study site as determined by the cartographic analysis.

The images below demonstrate the breadth of service provided in the cartographic analysis. Zoomed images provide a magnified view of service gaps that may not be evident at lower magnifications and broader geographic views. Blue/single antennas indicate a small (below 100ft) stand-alone antenna on top of a structure, and a red/double antenna indicates multiple antennas sharing a high tower structure. Black circles indicate gaps in service—areas where mobile service antennas are either not present or distant.

Table 2. Service providers and antennas in the study area.

Service providers	Number of antennas
City Of New York	58
Nextel of New York Inc.	57
New York City Police Department	56
Horizon Communications	33
Clearwire Spectrum Holdings III LLC	31
Northrop Grumman Systems Corp	28
New York City Transit Authority	19
ATT Corp	13
FCI 900 Inc.	12
New York City of Manhattan	12

In Queens, there were 242 towers (4 registered, 238 not registered), 3 new tower applications, and 762 antennas detected (Figure 4). In Brooklyn, there were 121 towers (3 registered, 118 not registered), 3 new tower applications, and 648 antennas detected (Figure 5). The study site was located in upper Manhattan. Thus, we assessed the mobile service topology for upper Manhattan around the study site (Figures 1-3). In lower Manhattan, the radius of analysis was reduced to one mile to compensate for its narrow latitude. There were 183 tower structures (8 registered, 175 not registered), 7 new tower

applications, and 2618 antenna locations detected (Figure 6). In the Bronx there were 105 tower structures (5 registered, 100 not registered), 2 new tower applications, and 805 antenna locations detected (Figure 7). All four boroughs and sites qualified as high volume signal areas and were deemed robust enough to support the study needs.

Unlike the cartographic analysis of the study site, the cartographic analysis of the boroughs only demonstrates analyses and views of large geographic areas (Figures 4-7). This was done because there was no specific location from which participants were expected to always send and receive messages. Thus, a general assessment of the area was most appropriate and provided enough content and context to assess the availability of mobile signal to study participants.

Figure 1. Antenna sites: 1050 Riverside Dr, New York, NY 10032. [Magnification 1]. Blue/single antennas indicate a small (below 100ft) stand-alone antenna on top of a structure, and a red/double antenna indicates multiple antennas sharing a high tower structure.

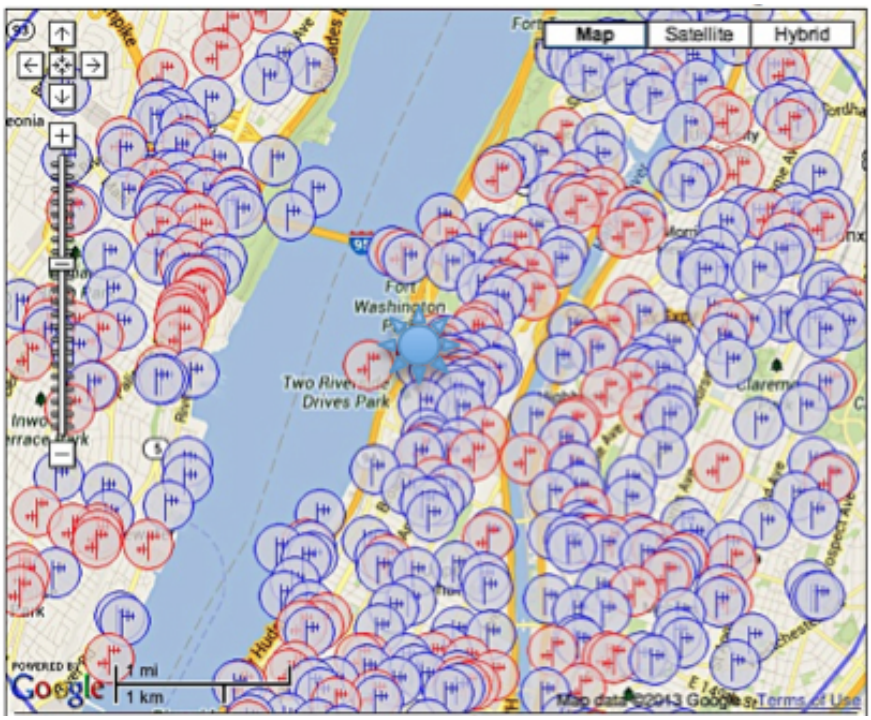


Figure 2. Antenna sites: 1050 Riverside Dr, New York, NY 10032. [Magnification 2].

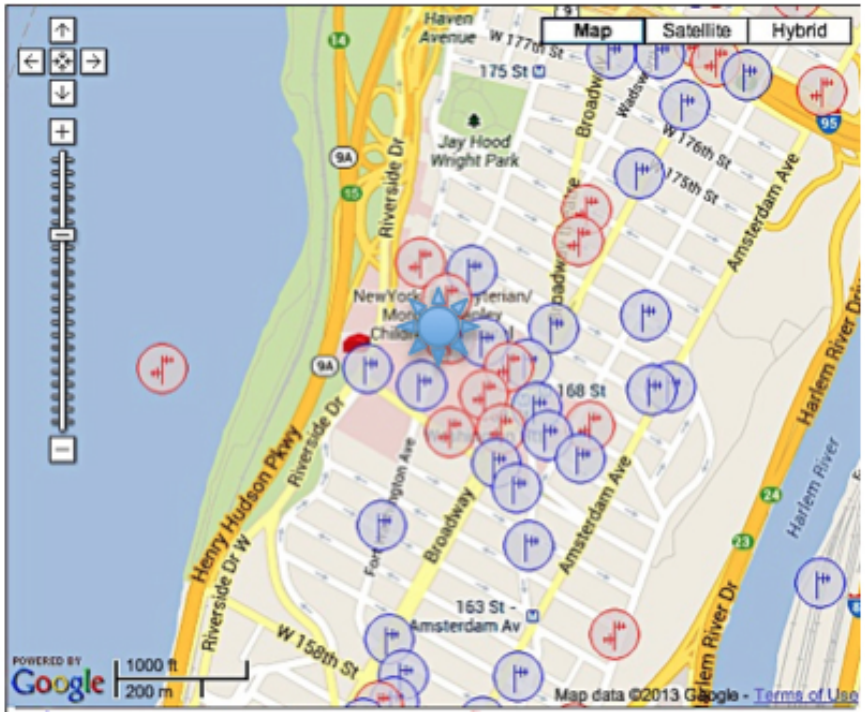


Figure 3. Antenna sites: 1050 Riverside Dr, New York, NY 10032. [Magnification 3].

Figure 4. Antenna sites: Queens Borough, New York. Black circles indicate gaps in service.

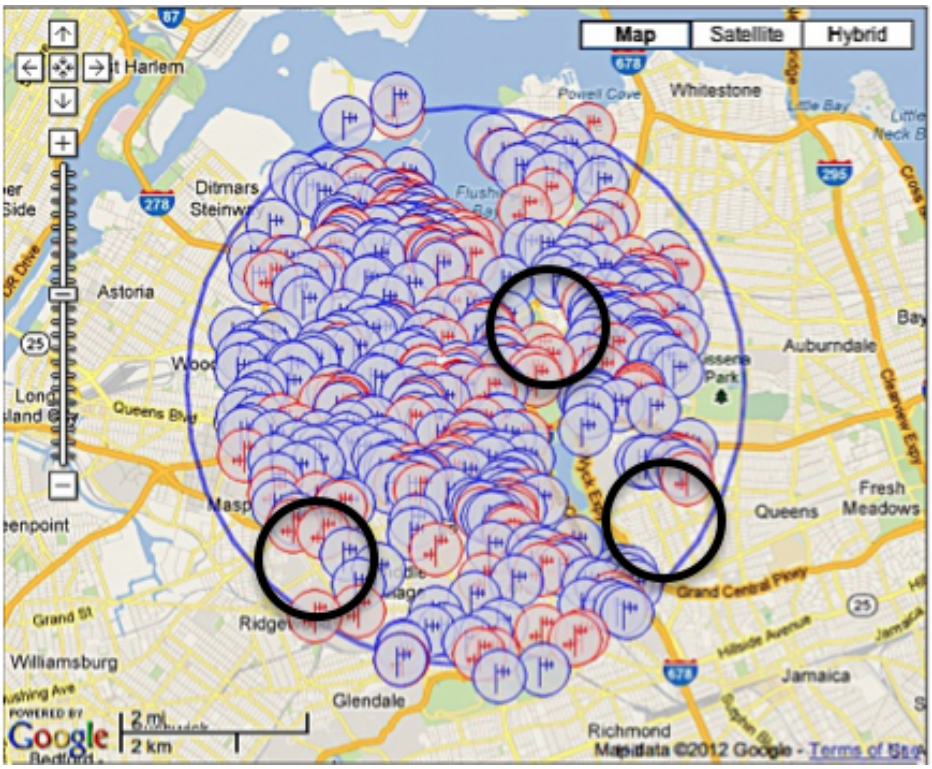


Figure 5. Antenna sites: Brooklyn Borough, New York. Black circles indicate gaps in service.

Figure 6. Antenna sites: Manhattan Borough, New York. Black circles indicate gaps in service.

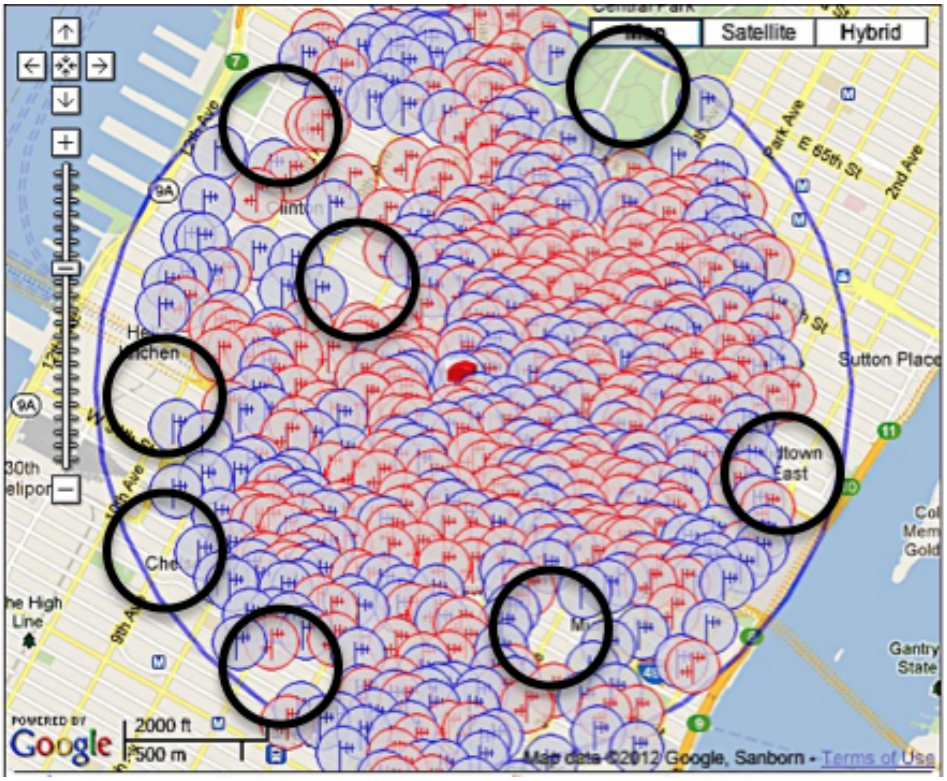
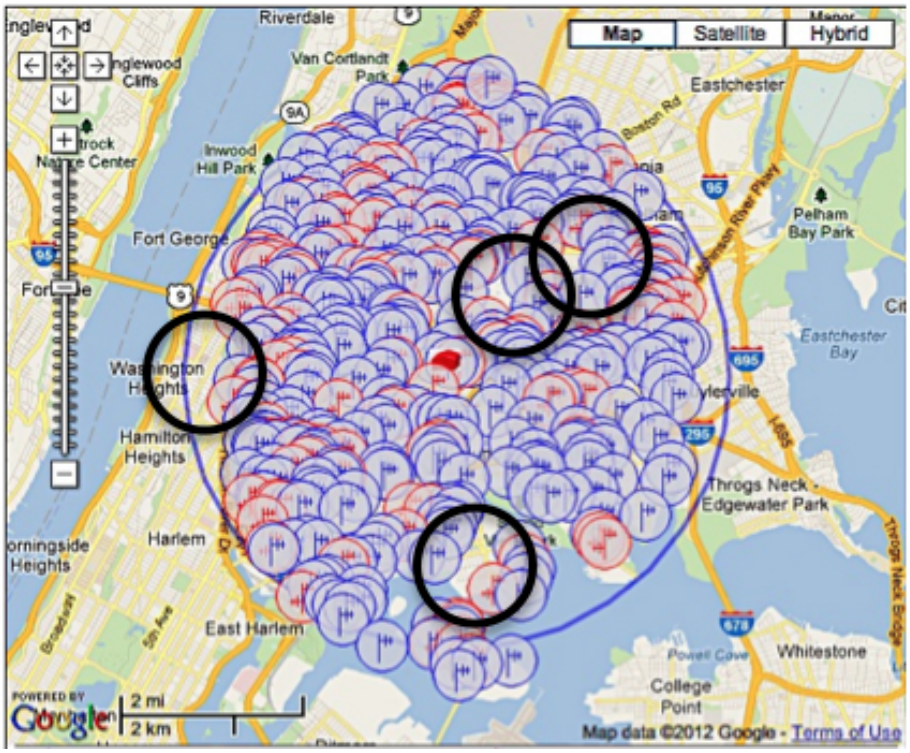


Figure 7. Antenna sites: Bronx Borough, New York. Black circles indicate gaps in service.



Results From SMS Message Frequency Analysis

We downloaded data from FrontlineSMS as a .csv file and organized the data in Microsoft Excel. For the SMS message frequency analysis we separated the data into two main categories, message type and message status, which work together to characterize a performed action. Message type describes the kind of message and is divided into two subcategories, sent and received. Message status describes the state or action of the message and is divided into four subcategories: sent, received, pending, and failed. Sent, pending, or failed message statuses are only applied to sent message types. The received message status is only applied to the received message type. A sent message status means the message was successfully transmitted to the mobile service provider for delivery to participants. Pending means the message is still being processed by the program's algorithm and/or the message has yet to be sent to the mobile service provider. A failed message status means the message was unable to be sent to the mobile service provider for delivery to the participant. A received message status means the system successfully received the participant's message from the mobile service provider.

Over the course of the alpha test and the pilot study there were a total of 6223 messages, excluding miscellaneous messages used for pinging or programming the system (see Table 3). Of the 6223 messages, 4136 (66.46%) were sent messages that originated from the FrontlineSMS program and 2087 (33.53%) were received messages that originated from alpha test or pilot study participants. Of the 4136 messages sent, 4109 (99.35%) were sent successfully, 26 (0.63%) failed to be sent, and 0 were pending or unprocessed. Of the 2087 messages received, all were received properly with no message processing errors.

Table 3. Comparison of the alpha test and pilot study SMS data.

SMS messages	Alpha test (%)	Pilot study (%)	Total (%)
All sent	1202/1817 (66.15)	2934/4406 (66.59)	4136/6223 (66.46)
All received	615/1817 (33.85)	1472/4406 (33.41)	2087/6223 (33.54)
Successful sent	1176/1202 (97.84)	2934/2934 (100.00)	4109/4136 (99.35)
Successful received	615/615 (100.00)	1472/1472 (100.00)	2087/2087 (100.00)
Failed sent	26/1202 (2.16)	0	26/4136 (0.63)
Failed received	0	0	0
Total	1817	4406	6223

Comparison of the alpha test and pilot study data showed system improvement in successful data transmission. Of the 1817 messages processed for the alpha test, there were 1202 (66.15%) messages sent to participants and 615 (33.85%) received from participants. Of the total sent messages, 1176 (97.84%) had a sent status and 26 (2.16%) had a failed status, with 0 pending. There were 615 received messages and all had a received status. In contrast, 4406 messages were processed in the pilot study.

Of those, 2934 (66.59%) were sent and 1472 (33.41%) were received. All sent messages had a sent message status, and all received messages had a received message status, which means all 4406 messages (100.00%) were processed successfully. Thus, after performing a cartographic analysis for antennas and towers we were able to choose the optimal service provider in the area, resulting in zero failed messages.

Discussion

Synthesis and Findings

The use of SMS text messaging in health research is growing exponentially, as is the number and diversity of mHealth tools. As mobile technology becomes a staple of everyday life, researchers, practitioners, and interventionists continue to find ways to leverage these technologies to improve the way we advance the health sciences. Since the state of health science and by proxy the state of public health depends on the information collected from those served, then speed, accuracy, and completeness become imperative.

The goal of the SMS pilot study was to understand the feasibility and acceptability of using SMS technology to improve data collection, participant behavior reporting, and adherence to biomedical protocols in research studies. Signal fluctuations and short signal ranges are major challenges to implementing SMS text messaging and other mHealth tools as research technologies. Though the literature on applying SMS technology to research is growing, no studies have scientifically and systematically addressed this challenge until now. We have demonstrated that signal fluctuations and short signal ranges can be mitigated by performing a cartographic analysis to identify areas with a greater number of antennas and towers where the mHealth technology will be used.

Assessing the mobile service infrastructure of the study site and research areas was fundamental to optimizing both data transmission and collection. After performing the cartographic analysis of antennas and towers, we modified our SMS pilot study protocol to take advantage of areas with stronger service and greater numbers of antennas and towers by locating our open-source SMS system in an area with greater signal strength and choosing the provider that had the most service antennas in our research areas. Since areas with greater numbers of antennas and towers provided greater signal strength, we saw an increase in the number of messages that were successfully sent and received. Other factors precipitated by the service provider or participant can have an impact on sent and received success rates. Thus, it was not possible to determine a causal or statistically significant relationship between adjusting the service provider according to the results of the cartographic analysis and the sent/received success rates. However, our analysis indicates a strong correlation between the two. These results elucidate the need and usefulness of performing a cartographic analysis prior to implementing mHealth technology in a research study.

Limitations in mobile service infrastructure could have prevented some messages from being sent and associated messages from being returned during our pilot study. This could also have prevented the behavioral reminders from going out on time, impacting participants' adherence to study protocol. The cartographic analysis not only informed us of whether our participants and study site would have enough signal availability to send and receive data, but also helped us to choose the optimal service providers in the area. Understanding the service availability of a geographic area is also important for international studies, particularly those in rural and/or less industrialized locations.

During the alpha test the SMS system's messages were serviced by T-Mobile USA Inc. The cartographic analysis showed that Nextel was the largest commercial service provider in the three-mile radius of the study site, followed by Clearwire Spectrum Holdings, Northrop Grumman Systems Corp, and ATT Corp, respectively. There were no antenna towers privately owned and serviced by T-Mobile. Thus, there was a greater chance for our T-Mobile service to have issues. Moreover, message send and receive data suggested there was a problem with service availability. Nextel had the most antennas in the area and is operated by Sprint Corp. However, at the time of the study, Sprint/Nextel had no text-message-only service plans and required a contract beyond the life of the study. Ultimately, using Sprint/Nextel would not have been cost effective. As a result, we switched to ATT Corp, which was determined by the cartographic analysis to have the second largest number of antennas in the area from a brand name carrier. It also had a cost effective, terminable text-message-only service plan.

Another important outcome of this study is the importance of alpha-testing mHealth tools before implementing them in research. Pilot testing can lead to the identification of bugs in the system that may negatively impact the feasibility and acceptability of a system and allow for troubleshooting prior to deployment. Our alpha test identified study-planning issues that would have a direct impact on our ability to get complete data and may have negatively impacted the experience of the participants. Issues included elements of sentence construction in the SMS message text and comprehension issues. Most importantly, the SMS sent and received data from the alpha test helped us discern that there were mobile service issues. Conducting an alpha test of the system not only helped to identify issues but helped us develop and test solutions to address those issues.

Limitations and Future Work

AntennaSearch, the information resource used for the cartographic analysis for antennas and towers, only reports data on US states and territories and thus is a limitation for translating these results to future international studies. For researchers wishing to use this technique internationally, cartographic analysis can also be performed using comparable digital resources or paper-based mobile service maps. Though our CAAT depicts antennas with assumed distances and assumed identical capabilities, other factors may affect signal strength, including power of the transmitter or the user cellular phone, antenna positioning, high use periods, and refraction and absorption by buildings and other structures. Thus, it is recommended that researchers be mindful of the information provided by cartographic resources and choose a cartographic

resource that best fits their studies' mobile system and geographic needs.

Moreover, no power analysis was conducted due to the small pool of potential candidates from the parent study (n=27 from "HIV Home Test and Decision-Making Among HIV-Negative Men"), as well as the budgetary constraints of our pilot funds, which only allowed for enrollment of a fraction of the parent study's participants. Thus, the pilot study was not powered for statistical significance, limiting the generalizability of the findings. Nevertheless, the cartographic analysis demonstrated in this study is unique in implementation and method of application. Future work should include a larger sample size and power calculation and identifying alternative resources for cartographic analysis outside of the US. This method should also be retested in research studies with other mHealth tools. Lastly, our CAAT was performed in New York City, a heavily urban environment. AntennaSearch is likely to yield different results in a rural setting due to differences in elevation and natural and manmade barriers.

Conclusions

As the literature in health research is beginning to demonstrate, SMS is quickly growing as a desired tool for real-time data collection and monitoring of biomedical adherence in research protocols. Since SMS works on a lowest common denominator mobile technology, it has potential to reach more people and to be user friendly when conducting research internationally, with marginalized populations, or across different languages. If properly implemented and scaled, SMS provides efficient data collection and reliable surrogate markers of adherence to assess viability, safety, acceptability, or efficacy of new biomedical and behavioral interventions.

This study contributes to the growing literature on mHealth use in research and informs the development and improvement of mHealth data collection and adherence tools for biomedical research. The significance of this paper rests in its use of innovative analytics, specifically the use of a cartographic analysis to assess antennas and towers for mobile service availability at our study site and in our broader study area. Comparison of the alpha test and pilot study data show the utility of this type of cartographic analysis in improving SMS system success rates for sent messages. Due to the vast geographic reach of mHealth studies, as well as the large amount of data transmission events, it is imperative to assess mobile service infrastructure and robustness for mHealth studies. The results from this study suggest that cartographic analysis of antennas and towers for mobile service can be used to improve study planning and implementation for mHealth research studies using mobile tools.

Acknowledgments

William Brown III is supported by National Library of Medicine (NLM) T15-LM007079 (principal investigator Hripcsak). At the time of this research William Brown III was supported by a training grant from the National Institute of Mental Health (NIMH) (T32-MH19139, Behavioral Sciences Research in HIV Infection; principal investigator Theo Sandfort, PhD). This research was supported by a center grant from the NIMH to the HIV Center for Clinical and Behavioral Studies at New York State Psychiatric Institute and Columbia University [P30-MH43520; principal investigators Anke A. Ehrhardt (1987-2013) and Robert H Remien, PhD (2013-2018)]. The content is solely the responsibility of the authors and does not necessarily represent the official views of NLM, NIMH, or the NIH.

Authors' Contributions

Alex Carballo-Diéguez was the principal investigator for the overall study. William Brown III was a co-investigator for the overall study and devised and implemented the cartographic analysis of antennas and towers (CAAT). All authors significantly contributed to the research and/or writing of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

SMS text message programming script.

[PDF File (Adobe PDF File), 62KB - mhealth_v3i2e63_app1.pdf]

References

1. Mobile Marketing Association. 2014. Industry Overview URL: http://www.mmaglobal.com/about/industry-overview [accessed 2015-05-27] [WebCite Cache ID 6YqsjLJfG]
2. Smith A. Pew Internet & American Life Project. 2011 Sep 19. Americans and text messaging URL: http://www.pewinternet.org/files/old-media//Files/Reports/2011/Americans%20and%20Text%20Messaging.pdf [accessed 2015-06-01] [WebCite Cache ID 6YyTwUZCz]
3. Bourne C, Knight V, Guy R, Wand H, Lu H, McNulty A. Short message service reminder intervention doubles sexually transmitted infection/HIV re-testing rates among men who have sex with men. Sex Transm Infect 2011;87:229-231.

4. Coomes CM, Lewis MA, Uhrig JD, Furberg RD, Harris JL, Bann CM. Beyond reminders: a conceptual framework for using short message service to promote prevention and improve healthcare quality and clinical outcomes for people living with HIV. AIDS Care 2012;24(3):348-357. [doi: 10.1080/09540121.2011.608421] [Medline: 21933036]
5. Déglise C, Suggs LS, Odermatt P. Short message service (SMS) applications for disease prevention in developing countries. J Med Internet Res 2012;14(1):e3.
6. Bouri N, Ravi S. Going mobile: how mobile personal health records can improve health care during emergencies. JMIR Mhealth Uhealth 2014;2(1):e8.
7. Bock B, Heron K, Jennings E, Morrow K, Cobb V, Magee J. A text message delivered smoking cessation intervention: the initial trial of TXT-2-Quit: randomized controlled trial. JMIR Mhealth Uhealth 2013;1(2):e17.
8. Furberg RD, Uhrig JD, Bann CM, Lewis MA, Harris JL, Williams P. Technical implementation of a multi-component, text message-based intervention for persons living with HIV. JMIR Res Protoc 2012;1(2):e17.
9. Schoenberger YM, Phillips J, Mohiuddin MO, McNees P, Scarinci I. Acceptability of delivering and accessing health information through text messaging among community health advisors. JMIR Mhealth Uhealth 2013;1(2):e22.
10. Nhavoto JA, Gronlund A. Mobile technologies and geographic information systems to improve health care systems: a literature review. JMIR MHealth UHealth 2014;2(2):e21.
11. Ehrler F, Wipfli R, Teodoro D, Sarrey E, Walesa M, Lovis C. Challenges in the implementation of a mobile application in clinical practice: case study in the context of an application that manages the daily interventions of Nurses. JMIR Mhealth Uhealth 2013;1(1):e7.
12. Kruse CS, DeShazo J, Kim F, Fulton L. Factors associated with adoption of health information technology: a conceptual model based on a systematic review. JMIR Med Inform 2014;2(1):e9.
13. 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 2013;72(5):564-583.
14. GSM Association. 2006. Universal access: how mobile can bring communications to all URL: http://www.gsma.com/publicpolicy/wp-content/uploads/2012/03/universalaccessfullreport.pdf [accessed 2012-08-26] [WebCite Cache ID 6YLkwvRET]
15. Hansen MS, Dørup J. Wireless access to a pharmaceutical database: A demonstrator for data driven wireless application protocol applications in medical information processing. J Med Internet Res 2001;3(1):e4.
16. Carballo-Diéguez A, Frasca T, Balan I, Ibitoye M, Dolezal C. Use of a rapid HIV home test prevents HIV exposure in a high risk sample of men who have sex with men. AIDS and Behavior 2012;16(7):1753-1760.
17. FrontlineSMS overview. URL: http://www.frontlinesms.com/technologies/frontlinesms-overview/ [accessed 2015-05-07] [WebCite Cache ID 6YLyGgLD1]
18. AntennaSearch: search for cell towers, cell reception, hidden antennas and more.: General Data Resources Inc URL: http://antennasearch.com/ [accessed 2014-10-12] [WebCite Cache ID 6THFCd0OM]

Abbreviations

CAAT: cartographic analysis of antennas and towers
CUMC: Columbia University Medical Center
SMS: short message service
NLM: National Library of Medicine
HIV: human immunodeficiency virus
UAI: unprotected anal intercourse
GSM: global system for mobile communications

Edited by G Eysenbach; submitted 12.10.14; peer-reviewed by YM Schoenberger, C Kruse, L Fulton; comments to author 03.11.14; revised version received 05.01.15; accepted 04.02.15; published 04.06.15.

Please cite as:
Brown III W, Ibitoye M, Bakken S, Schnall R, Balán I, Frasca T, Carballo-Diéguez A
Cartographic Analysis of Antennas and Towers: A Novel Approach to Improving the Implementation and Data Transmission of mHealth Tools on Mobile Networks
JMIR mHealth uHealth 2015;3(2):e63
URL: http://mhealth.jmir.org/2015/2/e63/
doi:10.2196/mhealth.3941
PMID:26043766

Tutorial

A Platform to Build Mobile Health Apps: The Personal Health Intervention Toolkit (PHIT)

Randall Peter Eckhoff[1*], BS; Paul Nicholas Kizakevich[1*], MS; Vesselina Bakalov[1*], MS; Yuying Zhang[1*], MS; Stephanie Patrice Bryant[1*], MS; Maria Ann Hobbs[1*], BS

RTI International, Research Triangle Park, NC, United States

[*]all authors contributed equally

Corresponding Author:
Randall Peter Eckhoff, BS
RTI International
3040 E. Cornwallis Rd.
Research Triangle Park, NC,
United States
Phone: 1 919 541 7158
Fax: 1 919 541 6158
Email: reckhoff@rti.org

Abstract

Personal Health Intervention Toolkit (PHIT) is an advanced cross-platform software framework targeted at personal self-help research on mobile devices. Following the subjective and objective measurement, assessment, and plan methodology for health assessment and intervention recommendations, the PHIT platform lets researchers quickly build mobile health research Android and iOS apps. They can (1) create complex data-collection instruments using a simple extensible markup language (XML) schema; (2) use Bluetooth wireless sensors; (3) create targeted self-help interventions based on collected data via XML-coded logic; (4) facilitate cross-study reuse from the library of existing instruments and interventions such as stress, anxiety, sleep quality, and substance abuse; and (5) monitor longitudinal intervention studies via daily upload to a Web-based dashboard portal. For physiological data, Bluetooth sensors collect real-time data with on-device processing. For example, using the BinarHeartSensor, the PHIT platform processes the heart rate data into heart rate variability measures, and plots these data as time-series waveforms. Subjective data instruments are user data-entry screens, comprising a series of forms with validation and processing logic. The PHIT instrument library consists of over 70 reusable instruments for various domains including cognitive, environmental, psychiatric, psychosocial, and substance abuse. Many are standardized instruments, such as the Alcohol Use Disorder Identification Test, Patient Health Questionnaire-8, and Post-Traumatic Stress Disorder Checklist. Autonomous instruments such as battery and global positioning system location support continuous background data collection. All data are acquired using a schedule appropriate to the app's deployment. The PHIT intelligent virtual advisor (iVA) is an expert system logic layer, which analyzes the data in real time on the device. This data analysis results in a tailored app of interventions and other data-collection instruments. For example, if a user anxiety score exceeds a threshold, the iVA might add a meditation intervention to the task list in order to teach the user how to relax, and schedule a reassessment using the anxiety instrument 2 weeks later to re-evaluate. If the anxiety score exceeds a higher threshold, then an advisory to seek professional help would be displayed. Using the easy-to-use PHIT scripting language, the researcher can program new instruments, the iVA, and interventions to their domain-specific needs. The iVA, instruments, and interventions are defined via XML files, which facilities rapid app development and deployment. The PHIT Web-based dashboard portal provides the researcher access to all the uploaded data. After a secure login, the data can be filtered by criteria such as study, protocol, domain, and user. Data can also be exported into a comma-delimited file for further processing. The PHIT framework has proven to be an extensible, reconfigurable technology that facilitates mobile data collection and health intervention research. Additional plans include instrument development in other domains, additional health sensors, and a text messaging notification system.

(JMIR mHealth uHealth 2015;3(2):e46) doi:10.2196/mhealth.4202

KEYWORDS

intervention studies; mHealth; mobile apps; platform; software engineering; telemedicine; tool; toolkit

Introduction

With 968 million units sold worldwide, mobile phones accounted for 53.6% of the overall mobile phone sales in 2013 [1]. In the United States in 2013, 56% of adults own mobile phones [2] and 34% own tablets [3]. Because mobile devices have become more prevalent, mobile health care (mHealth) apps will play a growing role for those managing their health concerns [4]. Nielsen's Connected Life Report from November 2013 indicates that approximately 46 million users in the United States have accessed apps in the fitness and health category, an 18% increase over the previous year [5]. Unfortunately, many of the available mHealth apps are not evidence based. For example, a review of 98 smoking-cessation apps found most to have a low level of adherence to proven methods defined by the US Public Health Service's Clinical Practice Guidelines for Treating Tobacco Use and Dependence [6].Therefore, mHealth app development and evaluations should be conducted in collaboration with a health researcher who understands the science and can objectively transfer this knowledge to the app developer. Although researchers desire to build quality evidence-based apps to test the mHealth interventions, app development may impose costs of US $50,000-150,000 [7]. For grant-funded research, this can be a significant fraction of project funds, leaving fewer resources for validation studies and efficacy trials. Personal Health Intervention Toolkit (PHIT) helps address this problem by providing a common platform and reusable content for both development and evaluation.

The PHIT platform was conceived to support the PHIT for Duty research projects addressing secondary prevention of psychological and behavioral health problems in persons experiencing symptoms of post-traumatic stress that had not yet risen to the level of a psychological disorder, such as post-traumatic stress disorder (PTSD). In doing this work, we soon realized that a common process model could be derived for mHealth intervention research and implemented in a way to support cost-effective reuse in other health domains and research apps [8]. We therefore set out to design and develop our PHIT framework, with the following goals:

- Creating a common platform from which other mHealth intervention apps can be developed;
- Standardizing how data collection instruments and interventions are implemented, fostering reuse from a common cross-study library; and
- Masking the complexities of software development to reduce development time and enable researchers to focus on the research aims.

This paper describes the PHIT platform and illustrates our high-level programming tool set, which facilitates implementation of mHealth apps through reuse of existing software content and easy development of new content according to study requirements.

Methods

PHIT Model

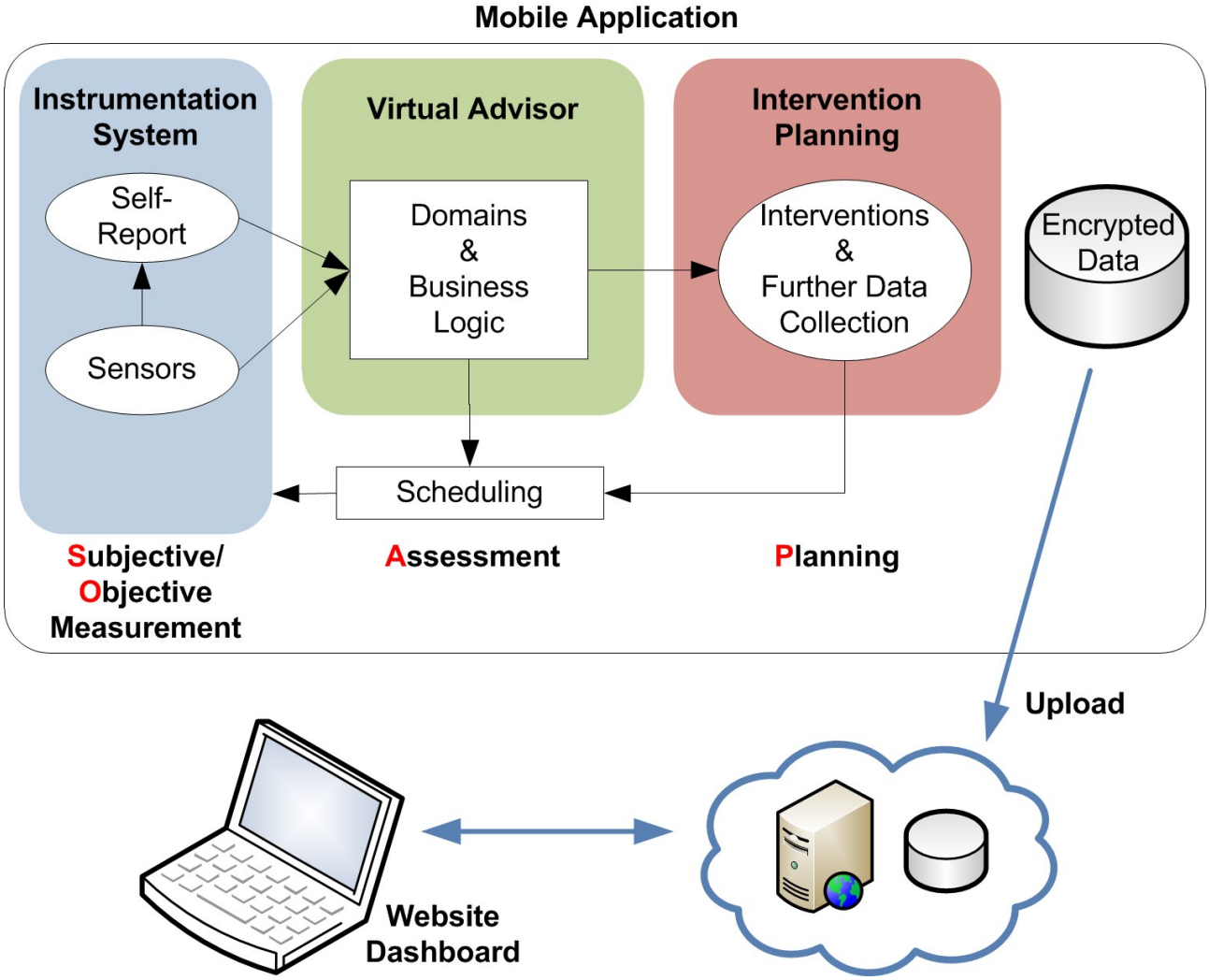
Overview

The PHIT architectural model is based on the subjective and objective measurement, assessment, and plan note methodology [9] for health status analysis, intervention recommendation, self-help activities, and data presentation, creating a feedback loop of personalized health (Figure 1). All data collection, analysis, and planning are performed locally on the mobile device, rather than via Internet services, with secure local data storage. PHIT has the following primary features:

- Integrates self-reported and physiological sensor instruments;
- Analyzes the data in real time on the device via an intelligent virtual advisor (iVA);
- Presents a suite of custom self-help activities and interventions;
- Collects data and adjusts the activities and interventions over time, tailoring the app to the individual needs of each user; and
- Transparently transfers data to a centralized database for conducting data analysis.

Research studies are supported using data objects that tag data with information on the study, protocol, participant identification, and other related information to facilitate analysis. Data access is facilitated using a website dashboard allowing the researcher to monitor the state of longitudinal studies and download study data into comma-separated files for easy analysis. Furthermore, 90% of the app configuration is done via extensible markup language (XML) making it easy to change the behavior of the app.

Figure 1. Personal Health Intervention Toolkit model utilizing the subjective, objective, assessment, and plan note methodology.



Instrumentation System: Data Collection

The PHIT architecture allows the researcher to reuse data-collection instruments or create their own instruments. Using XML, each instrument has a series of forms (screens) composed of data-collection entities, which represent a data point the researcher wants to collect. For example, a user history instrument may ask for age, weight, and gender information on the first screen and vital signs such as blood pressure and body temperature on the second. These two screens are considered forms in PHIT, and age, weight, gender, systolic and diastolic measures, and temperature are entities. Entities represent the data the researcher is collecting and have two facets to them, namely, (1) the internal, logic side and (2) the graphical user interface. If the app is configured for data storage, the user's responses are automatically saved. If desired, a researcher may configure the PHIT app to periodically upload data from the local secure database to a backend server, thereby allowing the researcher to examine the collected data and monitor study progress.

The majority of instruments are data-entry instruments, using familiar entities such as text fields, checkboxes, and selection lists. Many of these implement well-established *subjective* health screeners and assessment instruments previously administered through paper forms. Others had been developed ad hoc as needed for a particular study, such as the military deployment history instrument for the PHIT for Duty project. These also can be reused, or adapted as needed, to support new studies. PHIT has a library of over 50 such standardized instruments, including the examples in Table 1.

Table 1. Some of the standardized instruments the Personal Health Intervention Toolkit platform has implemented.

Category	Self-report data instrument
Alcohol use	Alcohol Use Disorder Identification Test [10]
Anger	Clinical Anger Scale [11]
Anxiety	General Anxiety Disorder-7 [12]
Combat exposure	Combat Exposure Scale [13]
Coping	Brief Coping Scale [14]
Concussion	Rivermead Post-Concussion Questionnaire [15]
Depression	Patient Health Questionnaire-8 [16-18]
Emotion regulation	Difficulties in Emotion Regulation Scale [19]
Mindfulness	Five Facet Mindfulness Questionnaire [20,21]
Pain	Brief Pain Inventory [22]
Post-traumatic stress disorder	Post-Traumatic Stress Disorder Checklist-Military Version [23]
Resilience	Connor-Davidson Resilience Scale [24]
Sleep quality	Pittsburgh Sleep Quality Index [25,26]
Stress	Perceived Stress Scale-10 [27,28]
Stressors	Impact of Event Scale [29,30]

In addition to the standard data entry fields, radio buttons, and checkboxes, other entities provided by PHIT include audio playback, charts, date picker, image display, Likert scale, and specialty entities, such as game-like cognitive tests (eg, reaction time). These can be combined in various ways for building both data-collection and intervention instruments. For example, following several alcohol reduction-related screens, a series of slides can be combined with audio narration for constructing an integrated multimedia alcohol education instrument.

Instrument Coding

By coding the entire instrument in XML, the question definitions, codebook responses, data validation, skip logic, and overall flow of the instrument are defined in one place, making it easy to adjust as necessary. The XML definition in Textbox 1 drives the first form from the PTSD checklist instrument used in our “Flight Attendant Wellness” app (Figure 2).

Using simple descriptive text, the form is made up of a single entity, named Q1, comprising a question with a series of radio buttons for the responses. The codebook values are defined in the code attribute of the item element and both the user text and the code are stored in the database. When the user selects one of the five radio buttons, a variable is created as instrumentName_entityName (ie, PCL_Q1), and the variable, a string type, is set to the selected code attribute.

Forms are not restricted in the number of entities they can contain. For purposes of this paper, the form example was kept simple. Textbox 2 shows a more complex example showing 4 entities on a form and the use of the vertical and horizontal elements to control user interface layout (Figure 3).

Textbox 1. XML definition.

```
<form name="F1">
  <entity name="Q1" type="radio" required="true">
    <text>
      To what extent are you bothered by repeated, disturbing
      memories, thoughts, or images of a stressful experience
      from the past?
    </text>
    <item code="1">Not at all</item>
    <item code="2">A little bit</item>
    <item code="3">Moderately</item>
    <item code="4">Quite a bit</item>
    <item code="5">Extremely</item>
  </entity>
  <logic event="onFormExit">
    <![CDATA[
      if ("{PCL_Q1}">="3") then
        set {PCL_Bcount} = "{PCL_Bcount}" + "1";
    ]]>
  </logic>
</form>
```

Textbox 2. A more complex example showing 4 entities on a form and the use of the vertical and horizontal elements to control user interface layout.

```
<form name="f2" title="Multiple Entities”>
  <vertical styleName="layoutVGroup1">
    <text name="sometextname" save="true" styleName="layoutTextLabel">
      Tell us about your family</text>
    <text styleName="layoutTextLabel2">First the adults</text>
    <horizontal styleName="layoutHGroup100Percent">
      <entity name="mom" layout="itemHorizontal" type="radio"
          styleName="radioWidget" required="true">
        <text>Mom</text>
        <item code="yes">Yes</item>
        <item code="no">No</item>
      </entity>
      <entity name="dad" type="radio">
        <text>Dad</text>
        <item code="yes">Yes (showing word wrap)</item>
        <item code="no">No</item>
      </entity>
    </horizontal>
    <text styleName="layoutTextLabel2">And about those siblings</text>
    <entity name="sib" layout="fullHorizontal" type="radio">
      <text>Do you have any?</text>
      <item code="Y">Yes</item>
      <item code="N">No</item>
    </entity>
    <entity name="numSibs" layout="itemHorizontal" type="text">
      <text>How many?</text>
    </entity>
  </vertical>
</form>
```

In addition to the instrument and form definitions, business logic is included. There are a number of different events that are triggered during the course of an instrument (Figure 4). Consider again the first form of the Post-Traumatic Stress Disorder Checklist (PCL) instrument. When the user presses the next arrow, the form's onFormExit code is executed before the next form is displayed. In this case, when the user responds with "Moderately" or higher, the variable PCL_Bcount is incremented by 1.

In Figure 4, the second form (F2) is expanded to highlight the different events. Of particular note are the onValueChanged and onValidate events. The event is first sent to the entity where the data change occurred to provide specific entity-level processing, and then passed up to the form itself where the form can look at all the entities collectively. All scripted logic is written with simple commands (Textbox 3).

Logic within the instrument XML file allows the researcher to decide exactly what validation and skip logic to have, to set initial conditions for instrument variables, and to execute code when the instrument terminates, such as calculating an overall score or saving data to the local secure database. With the instrument completely defined in the XML file, it becomes a reusable object to be shared across apps with the same expected behavior. Other than content used by the instrument, no other dependencies are required.

Textbox 3. Commands for scripted logic.

```
set {userHx_ageMonths} = "-1"; // initialize user age


if ("{userHx_userID}" == "" && "{userHx_userID_YN}" == "not asked") then
  call message("Are you sure you don't wish to personalize the
       application instrument? If so, hit Next again.");


if ("{userHx_ageMonths}" < "0") then
  begin;
    call message("User's birthday must be before today.");
    goto F1_Birthday;
  end;
```

Figure 2. Definition of the first form of the Post-Traumatic Stress Disorder Checklist (PCL) instrument from the Flight Attendant Wellness app.

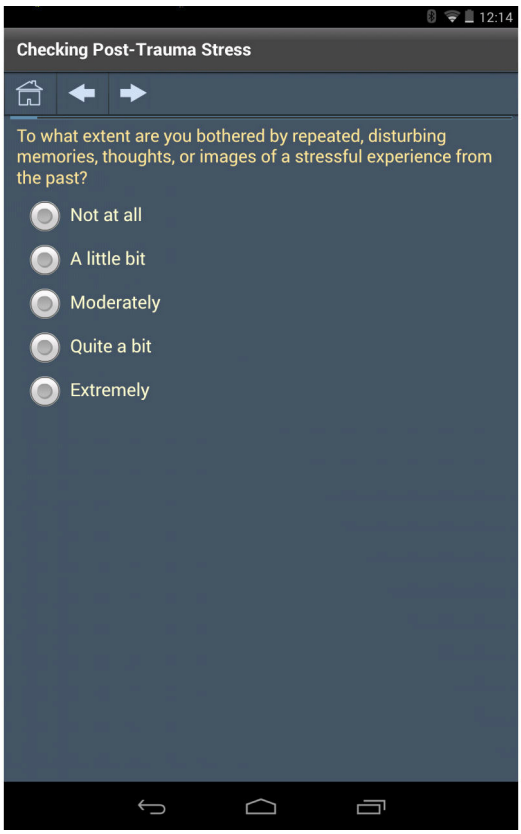


Figure 3. A form containing multiple entities.

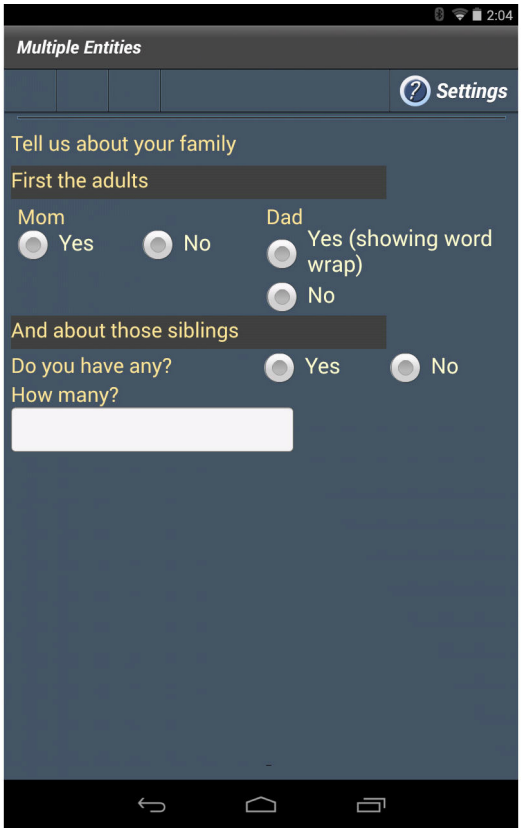
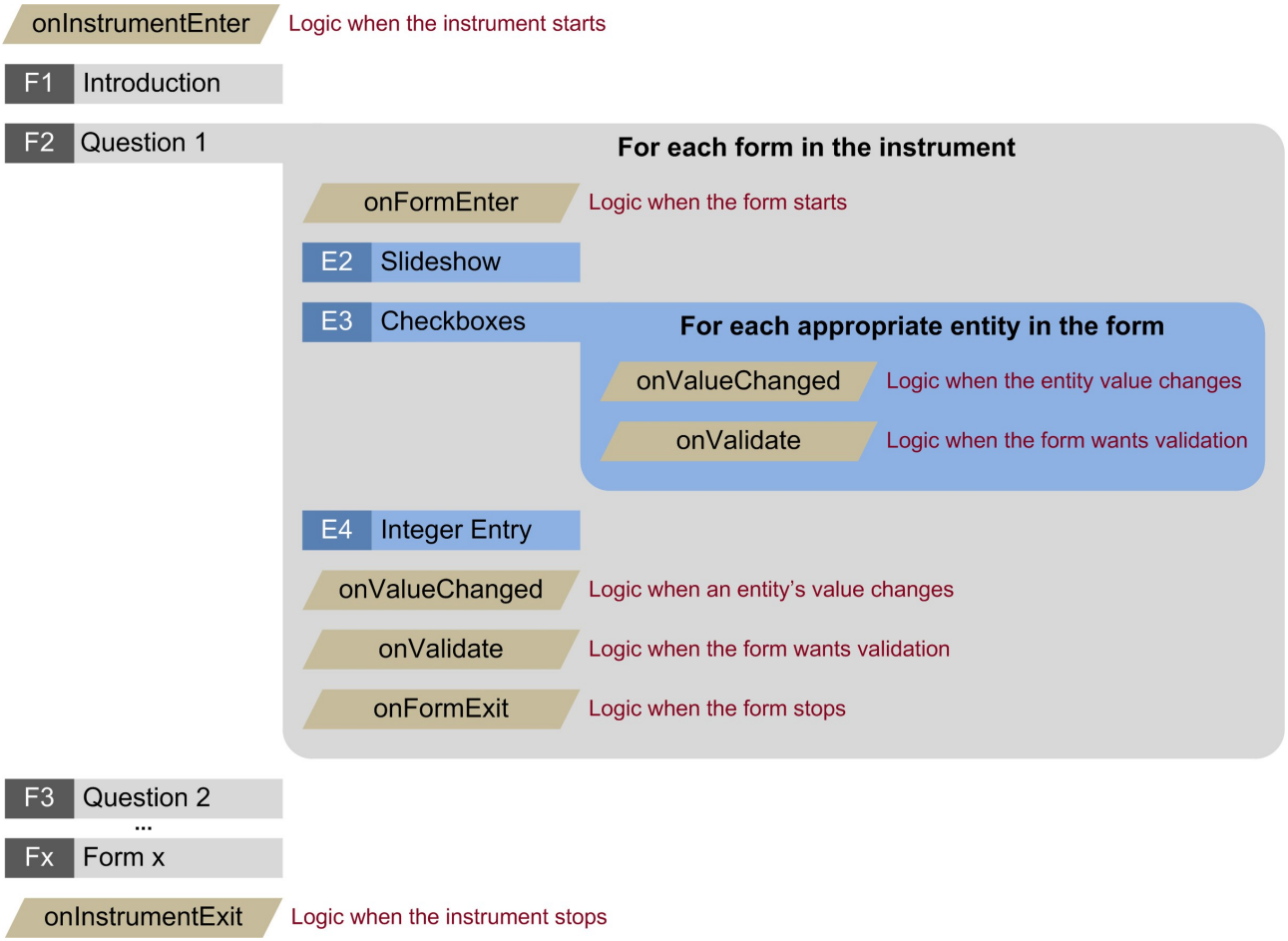


Figure 4. The various events that are triggered during the course of an instrument's lifetime.



Sensors

In addition to form-based data entry, the PHIT platform can also collect *objective* data from internal device sensors (eg, global positioning system coordinates) and external Bluetooth sensors (eg, heart rate monitor or fitness accelerometer). In the PHIT for Duty study, where individuals with post-traumatic stress are taught mindfulness exercises for stress reduction, the mobile app uses a heart rate monitor during the mindfulness meditation to calculate heart rate variability (HRV) and graphically show whether the user is achieving a more calm state. This is illustrated in the middle-line chart of Figure 5. Notice the rise in the middle graph after the onset of meditation as the user goes from a stressful state to a calm state.

Whenever sensor data are acquired (eg, the heart pulse rate) and processed to produce a derived measurement (eg, the HRV index), the PHIT software allows for saving interim data at each stage of data processing. Such storage facilitates verification of data-processing algorithms and supports both reanalysis and alternative analysis of raw data at a later date without repeating the data-collection activities. This facilitates exploratory analyses of mHealth data for determining optimal processing methodologies without the expense and effort of repeated field studies, saving a considerable amount of both cost and time.

Figure 5. During mindfulness training, an external heart rate monitor captures heart rate data to objectively determine if the user is relaxing or not.

Background Tasks

In addition to user-facing instruments, PHIT provides a means for performing background tasks using instruments without a user interface. Some examples are (1) querying the battery state, (2) uploading data, and (3) retrieving the current global positioning system location. These script instruments, because they run in the background, do not have forms or entities but they do have the onInstrumentEnter and onInstrumentExit events for which custom logic can be written. Such tasks can be developed to execute singularly on demand, or execute repeatedly at a specified interval, like every 5 minutes.

Home Screen

Tying it all together, the PHIT platform interprets the instrument definitions and creates the user experience, displaying the appropriate instruments in a task list on the “Home” screen. Attributes of the instrument such as title, description, icon filename, and menu index determine exactly what the user sees and in what order. To reflect a change in state, an XML logic can be written to modify these attributes to highlight changing conditions, alert the user to perform a critical task, or simply change from day to day according to a protocol.

To avoid overwhelming the user with many instruments, instruments can be scheduled. This minimizes burden on the user by cleaning up the user interface and only displaying what is currently relevant. Using the built-in scheduler and the hide and show commands callable from the XML logic, only those tasks appropriate at a certain point in time will be displayed. The PHIT scheduler extends RFC5545, the Internet Calendaring and Scheduling (iCalendar) specification [31], specifically the DTSTART, DTEND, DURATION, and RRULE properties of the VEVENT calendar component. For example, the PHIT scheduler can set an instrument to be scheduled to be on the task list each Friday at 8 am.

Virtual Advisor: Tailoring the App for the Users Based on Their Input

PHIT's iVA is an expert system logic layer where data are *analyzed* and *plans* are created in real time on the mobile device. The iVA tailors its analysis so that the help the user receives is personal and timely, with reanalysis occurring as frequently as the researcher wants it to happen (ie, daily, weekly, or monthly) using the PHIT scheduling function. The iVA program modules are used to stratify health assessments (eg, normal, moderate dysfunction) and to prescribe and schedule self-help activities (eg, exercise, meditation, alcohol reduction) according to the evidence-based criteria provided by the mHealth app researcher.

Consider this example in which a person with a sleep disorder is being evaluated using the Pittsburgh Sleep Quality Index (PSQI) instrument. Upon completing the PSQI, the sleep improvement protocol may recommend the following activities whenever the score exceeds a value of 16:

- Display a slide show on improving the sleeping environment;
- Provide a narrated meditation exercise at bedtime for stress relaxation; and
- Schedule the PSQI instrument to run every 3rd day at 8 am for reassessment until a downward trend is established in the PSQI score.

Just like the instrument logic, iVA logic is defined in XML that is organized by domain for easy reuse. Continuing the aforementioned example, the sleep *assessment* portion of the iVA script is shown in Textbox 4.

Textbox 4. Sleep assessment iVA script.

```
<assess domain="sleep">
  <logic>
  <![CDATA[
    // Assess the sleep risk
    if ("{PSQI_complete}" == "true") then begin;
      set {iVA_sleepIntervene} = "false";
      if ("{PSQI_score}"<="7") then set {iVA_sleepRisk} = "1";
      if ("{PSQI_score}">="8" && "{PSQI_score}"<="15") then
        set {iVA_sleepRisk} = "2";
      if ("{PSQI_score}">="16" && "{PSQI_score}"<="19") then
        set {iVA_sleepRisk} = "3";
      if ("{PSQI_score}">="20") then set {iVA_sleepRisk} = "4";
      if ("{iVA_sleepRisk}">"1") then set {iVA_sleepIntervene} = "true";
      set {iVA_scheduleSleep} = "true";
    end;
    if ("{iVA_sleepRisk }" >= "3") then begin;
      // Plan the sleep self-help intervention
      // See the Planning section below for details

    end;
  ]]>
  </logic>
</assess>
```

Planning is also handled within the iVA logic. Continuing the previous sleep example where the PSQI score is above 16, the iVA sets the iVA_sleepRisk variable based on the PSQI_score and tests that condition as defined by the protocol in Textbox 5.

Textbox 5. Protocol for testing the PSQI score.

```
<assess domain="sleep">
  <logic>
  <![CDATA[
    // Assess the sleep risk
    if ("{PSQI_complete}" == "true") then begin;

      // See the Virtual Advisor section above for details

    end;

    if ("{iVA_sleepRisk }" >= "3") then begin;

      // Plan the sleep self-help intervention

      // List the sleep environment and mindfulness meditation tasks
      call scheduleTask("sleep%environment");
      call scheduleTask("mindfulness%bodyScan");

      // Reschedule the PSQI assessment according to protocol
      call scheduleTask("PSQI", "P1DTA8H", null, "FREQ=DAILY,INTERVAL=3");
    end;
  ]]>
  </logic>
</assess>
```

The PHIT interventions and activities are implemented using the same scripting and script-processing methods as data-collection instruments; the XML constructs are identical. An intervention might display a series of slides, collect heart rate data during a relaxation exercise, or support behavior modification. For example, a series of images can be combined with audio narration to create an alcohol education module that can be reused across different PHIT apps. Interventions are study and protocol specific and with PHIT's library of reusable interventions (Table 2), many can be easily tweaked to be content specific for specific intervention needs.

Table 2. A partial list of Personal Health Intervention Toolkit (PHIT) interventions used in the PHIT for Duty [8] mobile app.

Category	Self-report data instrument
Stress relaxation	Relaxation breathing
Stress relaxation	Body scan meditation
Stress relaxation	Sitting meditation
Stress relaxation	Walking meditation
Stress relaxation	Loving kindness meditation
Stress relaxation	Heart rate variability biofeedback
Sleep quality	Improving your sleep
Sleep quality	Preparing for sleep
Sleep quality	Personal and environmental factors
Sleep quality	Reclaiming your bedroom
Sleep quality	Sleep smarter skills
Sleep quality	Nightmares
Risk alerts	Post-traumatic stress (Post-Traumatic Stress Disorder Checklist-Military version score > 50)
Risk alerts	Sleep quality (Pittsburgh Sleep Quality Index score > 22)
Risk alerts	Alcohol (Alcohol Use Disorder Identification Test score > 20)
Risk alerts	Anxiety (General Anxiety Disorder-7 score> 15)
Risk alerts	Depression (Patient Health Questionnaire-8 > 10)
Stress management	Arousal control
Stress management	Attention absorption
Alcohol use	Alcohol and stress
Alcohol use	Calories, costs, and consequences
Alcohol use	Drink smarter skills
Alcohol use	Blood alcohol level simulator
Alcohol use	Drinking reduction goals and action plan

Logic Processing

The logic written for either instruments or advisor coding is processed by the same logic processor. It supports common programming constructs in a simplified language format. Each logic statement reads like a sentence that ends with a semicolon.

Variables are used to keep track of information in the system, can be accessed globally across XML scripts, and persist until the app terminates. The PHIT naming convention for a variable is <objectName>_<variableName>, which provides a somewhat object-oriented variable naming construct. The PHIT variables are wrapped in "{}" (curly braces) to simplify runtime parsing of the logic code as in Textbox 6.

Textbox 6. PHIT variables in curly braces to simplify runtime parsing.

```
if ("{ptHx_status}"=="new") then
  begin;
    set {ptHx_status} = "started";
    set {ptHx_done} = "false";
  end;
```

When processing this example code statement, PHIT looks up the value of ptHx_status, compares it with the string *new* and if they are equal, sets ptHx_status to the string *started*. The variable ptHx_status refers to the instrument named ptHx (patient history), and the status entity within the ptHx instrument. As no other instrument, or object, may have the same name, the identity of the global variable is assured.

When a variable is evaluated, an attempt is made to determine whether it is a number, and if so, evaluates it as a number. Otherwise, it is treated as a string. If ptHx_age is set to 1, then *if ("{ptHx_age}" <= "2")* becomes *if (1 <= 2)*. The PHIT logic processor is not rigid; it will automatically convert a quoted number to a real number object when the evaluation is performed, allowing the nonprogrammer who is writing this

logic to have more flexibility. Boolean variables remain represented as strings, with “true/false” being the default, designated Boolean values.

A variety of programming statements are supported (Table 3), which are sufficient to meet most programming requirements. These statements are supplemented by the PHIT application programming interface (API), a set of global functions to provide scripted access to frequently used processes such as data conversion, database storage, and retrieval, playing sounds for effects and notifications, displaying pop-up messages, and formatting number variables as strings.

Table 3. A sample of Personal Health Intervention Toolkit logic statements and application programming interface function calls.

Statement	Example
Set a variable	set {ptHx_age} = “15”; set {ptHX_gender} = “M”;

Call a function	set {age_m} = call calculateAgeInMonths("{ptHx_birthday}");
	call Message("Patient birthday must be before today.");
	set {iVA_height}=call formatNumber({iVA_height}, 0);

Conditional logic	if ("{ptHx_ageMonths}">="0" && "{ptHx_ageMonths}"<="48" && "{iVA_isExposed}")then set {iVA_tobaccoCode}="Under4Exposure";
Skip logic	if ("{ptHx_ptID}"=="") then goto F0_PtID;
Exit the instrument	exit;
Move to next form in form stack	if (("{ptHx_today}"-"{ptHx_birthday}")<"{ptHx_one_week}") then nextForm;

Nested ifs and while	set {randytest_collectId1} = call generateCollectionId();
	if ("{randytest_collectId1}">="0") then begin;
	set {iVA_alcoholInstr} = "CAGE";
	set {iVA_alcoholInstrWording} = "past two weeks";
	if ("{randytest_collectId1}" == "0") then begin;
	set {iVA_another} = "foobar";
	end;
	set {iVA_alcoholOnQueue} = "foobar";
	set {iVA_alcoholStatus} = "monitoringScheduled";
	set {iVA_i} = 0;
	while ({iVA_i} < 3) begin;
	set {iVA_i} = {iVA_i} + 1;
	if ("{iVA_i}" == "1") then
	set {randytest_zzzz} = "zzzz";
	end;
	end;
	set {randytest_collectId2} = call generateCollectionId();
	var {randytest_abar} = "";

Statement blocks	begin;
	set {ptHx_status} = "started";
	set {ptHx_done} = "false";
	end;

To analyze historical trends, the API provides query functions to the local database as the following shows:

- “findLatest” retrieves the last saved value
- “findLatestN” retrieves the last N saved values
- “findByDate” retrieves values based on data range (eg, everything between June and September)
- “findByPlacement” retrieves values based on placement range (eg, everything between the fourth and eighth save)

Should a particularly complex XML script cause performance problems due to runtime compilation, you may recode the XML script in native code as an API function as “isPSQITrendingDownward.”

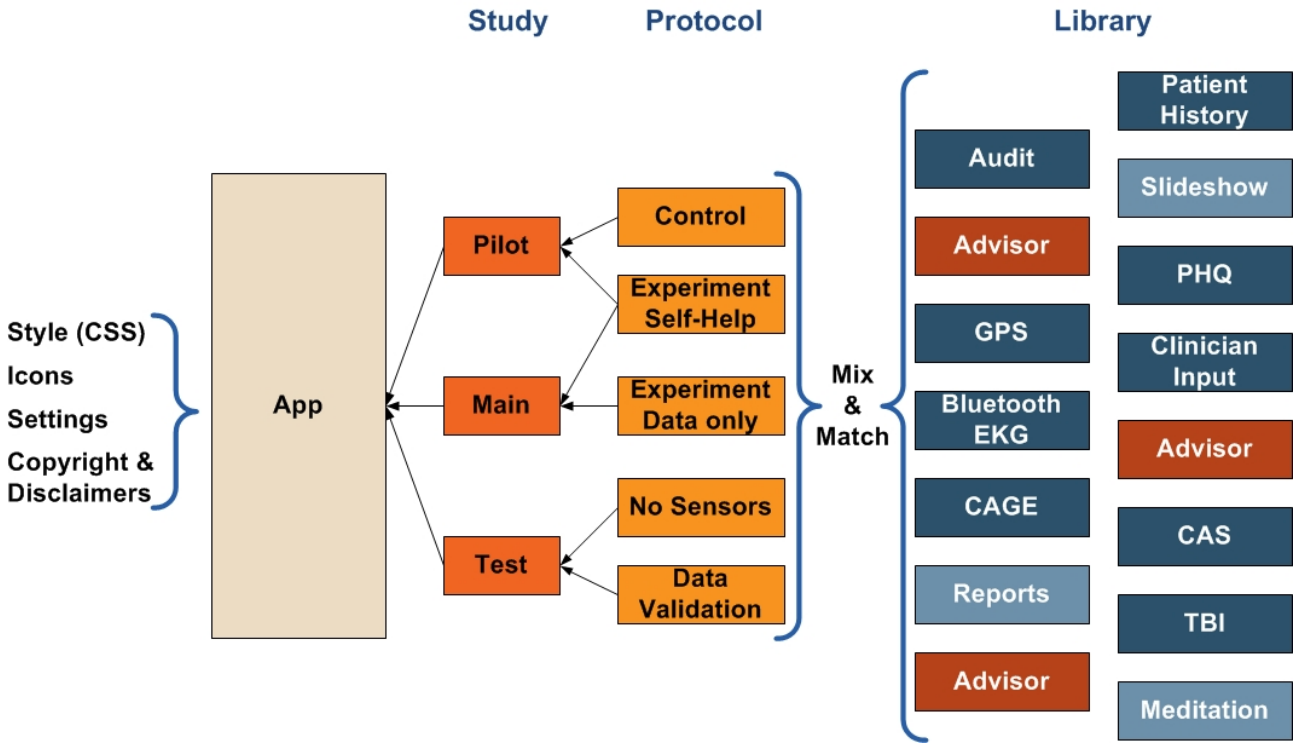
Although not common, any native code you write is automatically compiled into your PHIT app when the app is built.

Creating Your Own PHIT App: How the Pieces Fit Together

Unlike most mHealth apps, PHIT apps can be configured to support specific research requirements including multiple-treatment research studies. Using XML configuration files, you further define the app into studies and protocols. Each PHIT study contains one or more protocols and each protocol contains the instruments, virtual advisor, and interventions to be used for that protocol (Figure 6). In this way, different treatment groups are automatically embodied in the app.

With customization at the protocol level, each protocol in a study can have very different instruments, interventions, and virtual advisor; thus, creating a very different app. One example is a study in which Protocol A collects data and runs the virtual advisor to display appropriate interventions based on assessment scores, whereas Protocol B may merely collect data and have neither a virtual advisor nor any interventions.

Figure 6. Personal Health Intervention Toolkit app configuration.



Data Storage

Instrument-defined entity data are optionally saved in a local database with each record tagged with project id, study id, protocol id, case id, observation id, and a date-timestamp. Although not required, the PHIT platform supports, and strongly recommends that all data be stored using encryption to ensure data privacy. If the database is encrypted, PHIT automatically enforces the use of a password to access the app.

Although apps usually require a single database, the PHIT framework allows for multiple databases, including both create, update, and delete and read-only databases. A read-only database might be a lookup table to support app requirements, such as percentile growth chart for a pediatric wellness check [32].

Data Upload

In addition to the local data store, PHIT provides a data upload capability to a backend database. The following two upload options are available: (1) a user-initiated upload function, which initiates transfer and provides status feedback via an upload progress bar, and (2) a utility instrument for uploading to the server in the background. Because instruments can be scheduled, the mobile app can be configured to initiate a background data transfer on a prescheduled time, such as once a day at 12 am, thereby providing little disruption to the user. Of course, the app must be running on the device for this to occur. For privacy, data are uploaded over hypertext transfer protocol secure (HTTPS) with the option to encrypt the data before being uploaded over HTTPS, providing a doubly encrypted upload. Uploading data to a central server is an optional feature, which is most useful for research studies; however, use of a central data store is not a required element of a PHIT app.

To accommodate the typical software development process, three-dimensional different upload URLs can be specified to mimic the different phases of software development:

- Development;
- Test/staging; and
- Production.

This has the benefit of not polluting the production database with test data.

When merged into one dataset on a backend data server, data can be visualized, studied, and extracted. Access to the project data is controlled via an access control list to ensure data privacy. By default, PHIT uploads no personally identifiable information (PII) ensuring that all data are deidentified but allowing for data to be reported up to the case id level. However,

it is up to the app development team creating the mobile app to ensure that PII data are not misidentified, leading to accidental upload.

User Interface Customization

Mobile app developers need to address differences in screen size and density, and differences across mobile device software platforms. Each platform has a distinct user experience and human interface guidelines. The technology PHIT is built on uses a neutral look and feel across these platforms. The appearance of the mobile app can be changed through the use of cascading style sheets (CSS), custom skins, and icons, giving each app its own unique look and feel (Figures 7-10).

Figure 7. Flight Attendant Wellness app home screen.

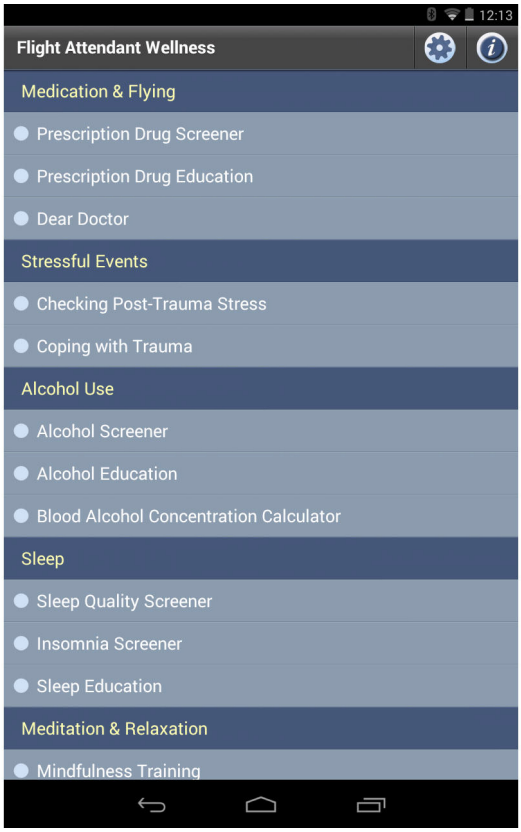


Figure 8. Flight Attendant Wellness app blood alcohol concentration calculator.

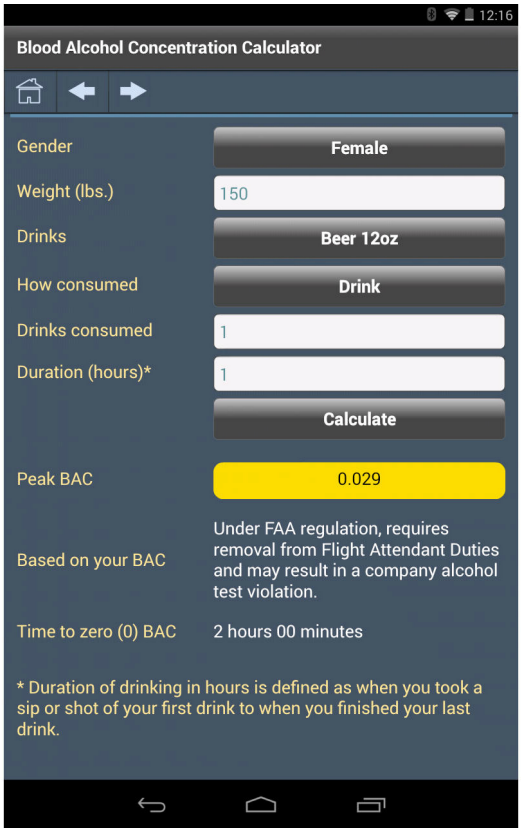


Figure 9. Clinical decision support tool for Pediatric Cardiovascular Risk Reduction app home screen.

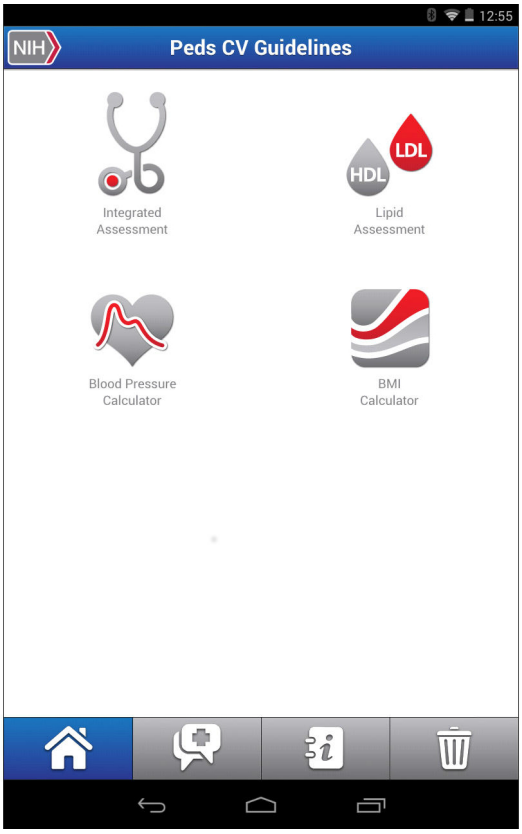
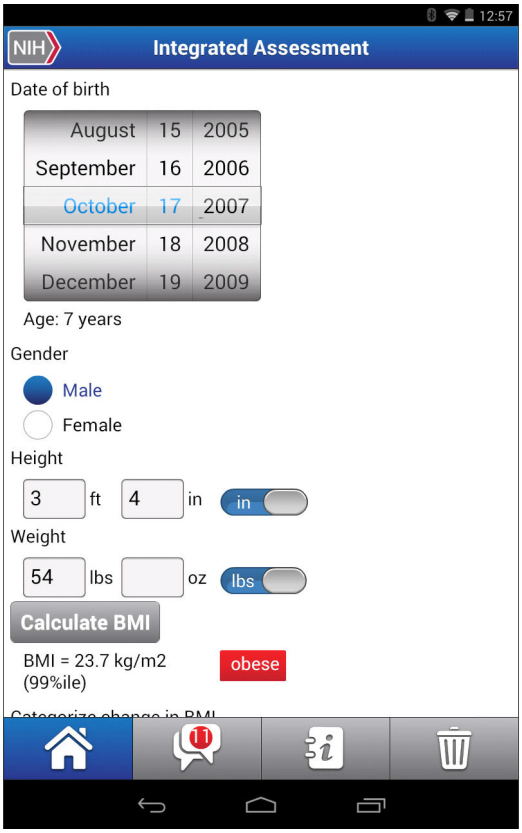


Figure 10. Clinical decision support tool for Pediatric Cardiovascular Risk Reduction app integrated assessment screen.



What Skills Do You Need?

With minimal customization of the app, only knowledge of XML and workflow logic is required to implement most mHealth apps. Custom apps require various software engineering skills. Some of these optional skills are as follows:

- Mobile app development familiarity;
- Graphics, video, and audio content design;
- CSS;
- Adobe Integrated Runtime, Apache Flex, and ActionScript;
- Java for Android native code;
- Objective-C for iOS native code;
- Encryption; and
- HTTP programming

Results

Overview

To date, we have completed 4 mobile apps and 1 desktop app using the PHIT platform. Several additional apps are in development and a half-dozen are in the concept phase. We have found that the time from concept to completion and the cost of implementation were substantially reduced in the later projects, compared with the initial work attesting to the flexibility of the PHIT platform and the reusability of developed components. Examples of our completed apps are as follows:

PHIT for Duty, a Personal Health Intervention Tool for Psychological Health and Traumatic Brain Injury

PHIT for Duty, deployed on Android devices, includes over 30 psychometric, personal/medical history, trauma exposure, and other data-collection instruments and evaluations [8]. Self-help interventions have been developed for stress, sleep problems, and alcohol abuse, including multimedia health information modules, stress relaxation exercises, and cognitive behavior therapies for sleep and alcohol.

Clinical Decision Support for Cardiovascular Health and Risk Reduction in Children and Adolescents

This app implements data collection, risk assessment, and intervention recommendation requirements of a subset of the Guidelines on Pediatric Cardiovascular Health and Risk Reduction [33]. The mobile app, intended for use by pediatricians aids to facilitate their use of the guidelines in daily clinical practice, is easy to use, and available for both Android and iOS devices.

Pre-Deployment Stress Inoculation Training (PRESIT)

This is a desktop app for training in stress reduction techniques as a preventative measure for reducing incidence of post-traumatic stress in service men and women.

ActiSleep

An app used for collecting research data in a study of sleep habits, sleep quality, and substance use in teenagers. The app includes daily diaries for prebedtime activities, substance use, and sleep quality. It also provides step-by-step multimedia instructions to enable participants to carry out biosample collection (ie, saliva) and facilitate use of a sleep activity monitor, thereby maintaining data quality in these ancillary data-collection processes.

Flight Attendant Wellness

An app providing screeners and education to support the prevention of prescription drug abuse, the federal model drug-free workplace, and the Workplace Prevention Research initiative.

Use of the PHIT framework for your apps does not limit you in how you distribute your apps nor do they require any oversight or verification. You are in complete control in distributing your PHIT app, whether it is made available via a public app store or a private distribution. The PHIT for Duty and ActiSleep apps are both for private research studies with a private distribution. The Flight Attendant Wellness app and the Clinical Decision Support for Cardiovascular Health and Risk Reduction in Children and Adolescents app are in the process of being made publically available and should be in the app stores soon.

Discussion

The PHIT framework has proven to be an extensible, reusable, and reconfigurable technology that facilitates mobile data collection and health intervention research. In addition to specific project requirements to enhance the platform, plans are to grow the library of instruments and interventions, add simple texting service prompting and notification, provide distributed advisor processing on the backend, and improve the Bluetooth layer for access to sensors, including wearable sensors.

Acknowledgments

This work was supported by the US Army Medical Research and Materiel Command, Office of Naval Research, Substance Abuse and Mental Health Services Administration, National Heart, Lung, and Blood Institute, National Institute on Drug Abuse, and RTI Internal Research and Development

Conflicts of Interest

All authors are salaried employees of RTI International with no stock options. The only personal benefit of widespread use of PHIT is scientific stature for all authors.

References

1. Gupta A, Cozza R, Lu C. Market Share Analysis: Mobile Phones, Worldwide, 4Q13 and 2013.: Gartner; 2014 Feb 12. URL: https://www.gartner.com/doc/2665319/market-share-analysis-mobile-phones [accessed 2015-05-23] [WebCite Cache ID 6Ygsx79IQ]
2. Smith A. Smartphone Ownership. Washington, DC: PewResearchCenter; 2013 Jun 05. URL: http://www.pewinternet.org/2013/06/05/smartphone-ownership-2013/ [accessed 2014-10-09] [WebCite Cache ID 6TCWlHLDd]
3. Zickuhr K. Tablet Ownership 2013. Washington, DC: PewResearchCenter; 2013. URL: http://www.pewinternet.org/2013/06/10/tablet-ownership-2013/ [accessed 2014-10-09] [WebCite Cache ID 6TCYgr2v0]
4. Levy D. Emerging mHealth: Paths for Growth.: PricewaterhouseCoopers; 2012. URL: http://www.pwc.com/en_GX/gx/healthcare/mhealth/assets/pwc-emerging-mhealth-full.pdf [accessed 2015-05-23] [WebCite Cache ID 6V1CbbSQb]
5. Nielsen. Hacking Health: How Consumers Use Smartphones and Wearable Tech to Track Their Health.: Nielsen; 2014 Apr 16. 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 2015-05-23] [WebCite Cache ID 6V1EEoD2Y]
6. Abroms LC, Westmass JL, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736. [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
7. Novak R. How Much Does It Cost to Build a Mobile App?. Arlington, VA: Segue Technologies; 2013 Jan 23. URL: http://www.seguetech.com/blog/2013/01/23/cost-build-mobile-app [accessed 2014-12-22] [WebCite Cache ID 6V1EmnUab]
8. Kizakevich PN, Hubal RC, Brown JM, Lyden JT, Spira JL, Eckhoff RP, et al. PHIT for duty, a mobile approach for psychological health intervention. Stud Health Technol Inform 2012;181:268-272. [doi: 10.3233/978-1-61499-121-2-268] [Medline: 22954869]
9. Cameron SS, Turtle-Song I. Learning to write case notes using the SOAP format. J Couns Dev 2002;80:286-292. [doi: 10.1002/j.1556-6678.2002.tb00193.x]
10. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. The alcohol use disorders identification test. In: Guidelines for Use in Primary Care 2nd Edition. Geneva, Switzerland: World Health Organization; 2001.
11. Snell WE, Gum S, Shuck RL, Mosley JA, Hite TL. The Clinical Anger Scale: Preliminary reliability and validity. J Clin Psychol 1995 Mar;51(2):215-226. [Medline: 7797645]
12. 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]
13. Keane TM, Fairbank JA, Caddell JM, Zimering RT, Taylor KL, Mora CA. Clinical evaluation of a measure to assess combat exposure. Psychol Assess 1989;1:53-55.
14. Carver CS. You want to measure coping but your protocol's too long: Consider the brief COPE. Int J Behav Med 1997;4(1):92-100. [doi: 10.1207/s15327558ijbm0401_6] [Medline: 16250744]

15. 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]
16. 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]
17. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
18. Dhingra SS, Kroenke K, Zack MM, Strine TW, Balluz LS. PHQ-8 Days: A measurement option for DSM-5 major depressive disorder (MDD) severity. Popul Health Metr 2011;9:11 [FREE Full text] [doi: 10.1186/1478-7954-9-11] [Medline: 21527015]
19. Gratz KL, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. J Psychopathol Behav Assess 2004;26:41-54. [doi: 10.1023/B:JOBA.0000007455.08539.94]
20. Baer RA, Smith GT, Hopkins J, Krietemeyer J, Toney L. Using self-report assessment methods to explore facets of mindfulness. Assessment 2006 Mar;13(1):27-45. [doi: 10.1177/1073191105283504] [Medline: 16443717]
21. Baer RA, Smith GT, Lykins E, Button D, Krietemeyer J, Sauer S, et al. Construct validity of the five facet mindfulness questionnaire in meditating and nonmeditating samples. Assessment 2008 Sep;15(3):329-342. [doi: 10.1177/1073191107313003] [Medline: 18310597]
22. Cleeland CS, Ryan KM. Pain assessment: Global use of the Brief Pain Inventory. Ann Acad Med Singapore 1994 Mar;23(2):129-138. [Medline: 8080219]
23. Weathers FW, Huska JA, Keane TM. PCL-M for DSM-IV. Boston, MA: National Center for PTSD; 1991.
24. Connor KM, Davidson JR. Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003;18(2):76-82. [doi: 10.1002/da.10113] [Medline: 12964174]
25. Germain A, Hall M, Krakow B, Shear KM, Buysse DJ. A brief sleep scale for posttraumatic stress disorder: Pittsburgh Sleep Quality Index Addendum for PTSD. J Anxiety Disord 2005;19(2):233-244. [doi: 10.1016/j.janxdis.2004.02.001] [Medline: 15533706]
26. Buysse DJ, Reynolds III CF, Monk TH, Berman SR, Kupfer DJ. Pittsburgh Sleep Quality Index (PSQI). 2000. URL: http://docslide.us/documents/psqidoc.html [accessed 2015-05-23] [WebCite Cache ID 6Yk7YjhM2]
27. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24(4):385-396. [Medline: 6668417]
28. Cohen S, Williamson G. Perceived stress in a probability sample of the United States. Thousand Oaks, CA: Sage; 1988 Presented at: The Social Psychology of Health: Claremont Symposium on Applied Social Psychology; 1988; Newbury Park, CA.
29. Horowitz M, Wilner N, Alvarez W. Impact of event scale: A measure of subjective stress. Psychosom Med 1979 May;41(3):209-218. [Medline: 472086]
30. Briere J. Psychological Assessment of Adult Posttraumatic States. Washington, DC: American Psychological Association; 1997.
31. Desruisseaux B. Internet Calendaring and Scheduling Core Object Specification (iCalendar). Fremont, CA: Internet Engineering Task Force; 2009. URL: http://tools.ietf.org/html/rfc5545 [accessed 2014-12-22] [WebCite Cache ID 6V1FoASn2]
32. Centers for Disease Control Prevention. Percentile Data Files with LMS Values. Atlanta, GA: Centers for Disease Control Prevention; 2009. URL: http://www.cdc.gov/growthcharts/percentile_data_files.htm [accessed 2014-12-22] [WebCite Cache ID 6V1G4Muu2]
33. Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents; National Heart, Lung, and Blood Institute. Expert panel on integrated guidelines for cardiovascular health and risk reduction in children and adolescents: Summary report. Pediatrics 2011 Dec;128(Suppl 5):S213-S256. [doi: 10.1542/peds.2009-2107C] [Medline: 22084329]

Abbreviations

API: application programming interface
CSS: cascading style sheets
HRV: heart rate variability
HTTPS: hypertext transfer protocol secure
iVA: intelligent virtual advisor
PCL: Post-Traumatic Stress Disorder Checklist
PHIT: Personal Health Intervention Toolkit
PII: personally identifiable information
PRESIT: Pre-Deployment Stress Inoculation Training
PSQI: Pittsburgh Sleep Quality Index

PTSD: post-traumatic stress disorder
XML: extensible markup language

Edited by G Eysenbach; submitted 02.02.15; peer-reviewed by G Lewis, W Greenleaf; comments to author 25.03.15; revised version received 26.03.15; accepted 03.04.15; published 01.06.15.

Please cite as:
Eckhoff RP, Kizakevich PN, Bakalov V, Zhang Y, Bryant SP, Hobbs MA
A Platform to Build Mobile Health Apps: The Personal Health Intervention Toolkit (PHIT)
JMIR mHealth uHealth 2015;3(2):e46
URL: http://mhealth.jmir.org/2015/2/e46/
doi:10.2196/mhealth.4202
PMID:26033047

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org