JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 9 (2021), Issue 11 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Original Papers

Reviews

Original Paper

"A Question of Trust" and "a Leap of Faith"—Study Participants' Perspectives on Consent, Privacy, and Trust in Smart Home Research: Qualitative Study

Mari-Rose Kennedy[1], BSc, MSc, PhD; Richard Huxtable[1], LLB, MA, PhD; Giles Birchley[1], BSc, MSc, PhD; Jonathan Ives[1], BA, MPhil, PhD; Ian Craddock[2], BEng, PhD

[1]Centre for Ethics in Medicine, University of Bristol, Bristol, United Kingdom

[2]Department of Electrical & Electronic Engineering, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Mari-Rose Kennedy, BSc, MSc, PhD
Centre for Ethics in Medicine
University of Bristol
Bristol Medical School
39 Whatley Road
Bristol, BS8 2PS
United Kingdom
Phone: 44 117 331 4512
Email: mari-rose.kennedy@bristol.ac.uk

Abstract

Background: *Ubiquitous*, *smart* technology has the potential to assist humans in numerous ways, including with health and social care. COVID-19 has notably hastened the move to remotely delivering many health services. A variety of stakeholders are involved in the process of developing technology. Where stakeholders are research participants, this poses practical and ethical challenges, particularly if the research is conducted in people's homes. Researchers must observe prima facie ethical obligations linked to participants' interests in having their autonomy and privacy respected.

Objective: This study aims to explore the ethical considerations around consent, privacy, anonymization, and data sharing with participants involved in SPHERE (Sensor Platform for Healthcare in a Residential Environment), a project for developing smart technology for monitoring health behaviors at home. Participants' unique insights from being part of this unusual experiment offer valuable perspectives on how to properly approach informed consent for similar smart home research in the future.

Methods: Semistructured qualitative interviews were conducted with 7 households (16 individual participants) recruited from SPHERE. Purposive sampling was used to invite participants from a range of household types and ages. Interviews were conducted in participants' homes or on-site at the University of Bristol. Interviews were digitally recorded, transcribed verbatim, and analyzed using an inductive thematic approach.

Results: Four themes were identified—motivation for participating; transparency, understanding, and consent; privacy, anonymity, and data use; and trust in research. Motivations to participate in SPHERE stemmed from an altruistic desire to support research directed toward the public good. Participants were satisfied with the consent process despite reporting some difficulties—recalling and understanding the information received, the timing and amount of information provision, and sometimes finding the information to be abstract. Participants were satisfied that privacy was assured and judged that the goals of the research compensated for threats to privacy. Participants trusted SPHERE. The factors that were relevant to developing and maintaining this trust were the trustworthiness of the research team, the provision of necessary information, participants' control over their participation, and positive prior experiences of research involvement.

Conclusions: This study offers valuable insights into the perspectives of participants in smart home research on important ethical considerations around consent and privacy. The findings may have practical implications for future research regarding the types of information researchers should convey, the extent to which anonymity can be assured, and the long-term duty of care owed to the participants who place trust in researchers not only on the basis of this information but also because of their institutional affiliation. This study highlights important ethical implications. Although autonomy matters, trust appears to matter the most. Therefore, researchers should be alert to the need to foster and maintain trust, particularly as failing to do so might have deleterious effects on future research.

(JMIR Mhealth Uhealth 2021;9(11):e25227) doi:10.2196/25227

KEYWORDS

smart homes; assistive technology; research ethics; informed consent; privacy; anonymization; trust

Introduction

Background

Recent technological advances have made it possible to embed computer devices in everyday environments and objects [1,2]. *Ubiquitous*, *smart*, and assistive computing technology such as sensors, cameras, and interfaces, which can wirelessly connect and communicate, can aid humans in numerous ways [1,2]. For example, such technology has the potential to improve health and health care by monitoring medical conditions and providing in-home assistance [3-5]. The advent of COVID-19 has notably hastened the move to remote delivery of many health services, such as primary care [6].

Before technology can deliver on its promise, robust research is needed, which (in part) requires attention to the needs and perceptions of intended users. Research in this area rightly tends to be participatory, with potential users involved in study development and evaluation [5,7]. However, conducting research on ubiquitous technologies can pose both practical and ethical challenges, particularly if the study situates technology in real homes [8]. Participants should not expect to benefit directly from such interventions [9], and even if some benefit might accrue, the participants might be drawn from vulnerable populations and have complex needs [7].

Whoever the research participant is, they have interests in their autonomy and privacy being respected wherever and whenever they might contribute to the study. Therefore, there are prima facie ethical obligations to observe. First, participation must be *consensual*. Respect for autonomy requires the provision of consent, which is voluntarily given by a (mentally) competent individual who is sufficiently well informed. However, consent can be challenging, particularly in this context. Some participants, such as young children in the household or those with dementia, whom such research *might* come to benefit, may have absent or diminished competence to consent. Information about the study (such as what the technology does, the extent to which pseudonymization offers privacy, or the security of the technology) can also be difficult to grasp, even by those participants whose competence is unimpaired. Full disclosure of information to participants will not always be possible in any event, given potential unanticipated uses to which the data gathered might be put in the future; in this case, broad consent for the secondary uses of data is required. Researchers need to be attentive to such challenges and to the means of overcoming them, for example, by ensuring that consent is a process rather than an event and being open about the uncertain uses data gathered with broad consent may be put to in the future [10].

Second, some claim that researchers should also be mindful of the need to respect *privacy* [11,12]. Concerns about privacy might be especially acute when equipment is installed in and data gathered from people's homes [8,13]. However, there is debate about whether privacy is a concern in itself or whether privacy concerns can be met by attention to other protections, so responding to privacy concerns is not straightforward [14]. Furthermore, different people rate the importance of privacy differently for a myriad of reasons across different circumstances: what one person considers a problematic invasion of privacy, another might not [12] (eg, differing attitudes toward sharing data with social media platforms [15]). Difficulties in gaining consent and the contested nature of privacy make it clear that, when collecting potentially sensitive personal data, there is a need for researchers to take care in ensuring that arrangements to protect participants' interests are as effective and appropriate as possible.

Objectives

Informed by these background ethical considerations, we sought participants' views on consent and privacy in the context of smart home research. Our participants were drawn from households involved in the SPHERE (Sensor Platform for Healthcare in a Residential Environment) project. SPHERE is an ambitious project led by Professor Ian Craddock at the University of Bristol, involving collaborators from other universities; local third sector and community collaborators, such as Bristol City Council and Knowle West Media Centre; and international partners from the industry (IBM and Toshiba) [16]. The aim of SPHERE is "to develop a multi-purpose, multimodal sensor platform for monitoring people's health inside their homes" [9]. SPHERE, as described herein, is exploratory in its approach and not directed toward the specific needs of particular user groups. The participants encompass a variety of households, such as couples or families recruited with no particular focus on health conditions; however, subsequent SPHERE studies (outside the scope of the present contribution) specifically include patients with cardiovascular conditions, dementia, and Parkinson disease and those recovering from orthopedic surgery. The team has developed multiple technologies, including hardware, software, and machine learning, and more specifically an in-house Internet of Things platform, comprising environmental, video, and wearable sensors. The focus of this paper is to describe the SPHERE-CARED (Consent and Anonymization: A Review of Ethical Dimensions) study, exploring SPHERE participants' perspectives on the ethical aspects of informed consent, anonymity, privacy, and data sharing. For a list of the SPHERE sensor technologies, including wearables, ambient sensors, and cameras, the interested reader should refer to the literature [8,17]. SPHERE aims to test the technology, primarily by obtaining data about human movement and ambient measures that could in the future result in home health monitoring applications or the ability to address health research questions over long periods in the patient's own home, including measuring the characteristics of health-related activities of daily living, such as sleep, cooking, walking, moving between rooms, and transitions from sitting to standing. Prototypes are first

rigorously tested for robustness and stability before the sensors are deployed in participants' homes [9].

The SPHERE system was designed in accordance with the principles of privacy by design [18], including those of data minimization, end-to-end security, respect for user privacy, and the use of data protection impact assessments within the project. Many SPHERE design decisions were arrived at through workshops with members of the public during the system design process.

Before (and during) deployment, SPHERE has sought the views of the public at large via engagement events and involvement in its *Friends of SPHERE* group. Research participants, which include households comprising entire cohabiting families, must consent to participation and are assured that the data gathered are anonymized. Beyond contributing to the generation of knowledge that might prove beneficial in the future, participants do not receive any direct benefit from participation (however, they are recompensed for increased domestic electricity costs arising from their participation).

Participation in SPHERE does not intrude on participants' bodily integrity (meaning that participation does not require the use of, or intrusion into, participants' bodies), and careful attention has been paid to protecting the identities of participants. However, the project does involve monitoring, with not only environmental sensors but also cameras, in their homes. Moreover, many participants lived with this level of monitoring in their houses for months or even years, which may be regarded as a considerable invasion of their privacy. Therefore, SPHERE participants are uniquely well-positioned to contribute to research exploring ethical questions regarding surveillance, privacy and confidentiality, and the provision of informed consent in smart home research. The SPHERE-CARED project comprises an empirical study whose objectives are to qualitatively explore SPHERE participants' views about these ethical issues and use those data to inform reflections on the ethical dimensions of smart home research to benefit and enlighten similar future research. The study was guided by the following two research questions:

1. How do participants of SPHERE understand and think about the ethical issues arising from the SPHERE around informed consent, anonymity, privacy, and data sharing?
2. How can the insights from SPHERE participants about their consent experience help inform general thinking about how informed consent should be approached in future smart home research?

Methods

Sampling and Recruitment

SPHERE-CARED participants were recruited from the SPHERE cohort as whole households. Households that were currently participating or had previously participated in SPHERE were eligible for inclusion. SPHERE households recruited via the National Health Service and those who did not consent to follow-up research were excluded. As a result, of the total 50 (95 participants) SPHERE households, 26 (52%; participants: 53/95, 56%) were eligible for SPHERE-CARED. Eligible households were then purposively sampled to obtain the maximum variation [19] within the bounds of quite a limited subset of households. The final sample included participants from a range of age groups and from the following four household types: individuals, families, couples, and shared residences.

Recruitment was conducted in 3 stages. First, the SPHERE deployment officer identified eligible households and made initial contact with them via their preferred method of communication. Second, details of interested households were passed to the SPHERE-CARED researcher (MK), who arranged a telephone call to further explain the project. Finally, interviews were organized with households once it was confirmed that all members were willing to participate. In the first stage, 28% (14/50) of households (participants: 31/95, 33%) were contacted. Of these 14 households, 7 (50%) households (participants: 16/31, 52%) expressed interest in the research and were subsequently contacted by MK. All 7 households were recruited. All 4 household types were represented, and participants' ages ranged across 9 decades (Tables 1 and 2). The ages are presented as a range to preserve participant anonymity.

Table 1. Household demographics (households: n=7; participants: n=16).

Household ID	Household type	Total occupants, n (%)	Association with university	Attendance of >1 pre-event
H01	Family	4 (25)	Yes	Yes
H02	Family	4 (25)	Yes	Yes
H03	Family	2 (13)	Yes	No
H04	Couple	2 (13)	Yes	Yes
H05	Individual	1 (6)	No	Yes
H06	Individual	1 (6)	No	Yes
H07	Shared residence	2 (13)	Yes	No

Table 2. Participant demographics (households: n=7; participants: n=16).

Household ID and participant ID	Gender	Age (years), range
H01		
P01	Male	51-60
P02	Female	51-60
P03	Female	11-16
P04	Female	11-16
H02		
P05	Male	31-40
P06	Female	31-40
P07	Male	0-5
P08	Male	0-5
H03		
P09	Female	41-50
P10	Female	0-5
H04		
P11	Female	41-50
P12	Male	41-50
H05		
P13	Male	81-90
H06		
P14	Female	61-70
H07		
P15	Female	31-40
P16	Female	31-40

Ethics

SPHERE-CARED was reviewed and granted a favorable opinion by the University of Bristol Faculty of Social Science and Law research ethics committee as an amendment to the original SPHERE project, which was reviewed and granted a favorable opinion by the University of Bristol Faculty of Engineering research ethics committee (reference: FREC40403). In accordance with university policies, the study also required the completion of 2 risk assessments on working with children and lone-working, and a lone-worker protocol was implemented.

As these were household interviews, children were included as participants. Signed informed consent was sought from all participants aged ≥16 years; informed (age-appropriate) verbal assent was sought from younger children, with written informed consent for their inclusion also provided by parents. Adults and children were informed of their right to revoke consent or assent at any time during the interview. In interviews that included younger children, the interviewer (MK) was vigilant for any signs of distress, which would require a review of the assent with the household. All households were provided with a £20 (US $27.2) shopping voucher to thank them for their involvement in this research.

Data Collection

Qualitative interviews were conducted with 7 households; of the 7 interviews, 6 (86%) were conducted in participants' homes, and 1 (14%) was conducted on-site at the university in an adapted house equipped with the study smart home technology to aid participants' recollections of living with ubiquitous technologies. A topic guide was developed to explore participants' involvement in and experiences of the main SPHERE project. Questions were designed to focus on participants' views regarding information provision and informed consent, anonymity, and data use, including probing for areas where they may have experienced problems, such as misunderstandings, requirements for further information, or any disagreements among the household (Textbox 1). Interviews were conducted with the whole household in one instance and were semistructured, allowing the researcher (MK) flexibility to follow up on issues raised by participants during the interviews. To facilitate a more comfortable interview environment for younger children, age-appropriate materials were provided; for example, a soft toy was provided to children aged between 0 and 5 years [20]. Children were invited to contribute to the interviews as they wished, and, where appropriate, questions were rephrased to make them more accessible. Although we conducted whole household interviews,

for obvious reasons, we did not expect to gain any data from infants who had not yet developed speech and language comprehension. To account for potential power dynamics between parents and children in the interviews, the interviewer explained from the outset that they were interested in hearing from all members of the household and that there may be differing views and experiences within the group that were all relevant to the study. During the interviews, the interviewer was careful to provide space and encouragement to facilitate the children's contributions, such as asking follow-up questions and prompts to children directly. After each interview, field notes were written, recording such aspects as household and interview dynamics and the researcher's initial reflections.

Textbox 1. Overview of the topic guide.

Consent

- Motivations for participation and expectations of participating
- The consent process:
  - What do you remember?
  - Any household or individual concerns
  - Any household disagreements
  - General feelings about the process
  - Would you take part again?
- Information understanding and needs:
  - Did the household understand everything they needed to know to participate?
  - Areas of difficulty
  - Areas where more information is desired
  - What information should future participants be told?

Anonymization and data sharing

- Feelings about different information collected by SPHERE (Sensor Platform for Healthcare in a Residential Environment)
- Household use of SPHERE Genie
- Expectations about ongoing data use and sharing
- Importance of privacy
- Views about data sharing beyond project researchers

Feedback on the original SPHERE participant information sheet

- Views on content and clarity
- Any missing information
- Any extraneous information
- How should information about burdens and risks be presented?

Data and Analysis

All interviews were double audio-recorded using 256-bit encrypted Olympus digital recorders and an Olympus omnidirectional microphone. Audio recordings were transcribed using a university-approved transcription company. MK checked all transcripts against the recordings for accuracy and removed obvious (direct or indirect) identifying data as part of the process of anonymization. An inductive thematic analysis was undertaken to make sense of the data [21]. This was an iterative process focusing on identifying codes and subsequently developing themes from the data (transcripts) from the bottom-up rather than interrogating the data deductively using predefined codes. Transcripts of interviews were coded and recoded as data collection progressed. The initial coding captured diverse features of the interview texts, and the researcher (MK) developed a coding list with each successive transcript, facilitated by NVivo 10 software. Once all the transcripts were coded, the researcher (MK) used the codes to explore and develop themes to explain the data relevant to the research focus and aims [21]. During this process, members of the research team (MK, RH, JI, and GB) conducted multiple coding across 3 transcripts to check the researcher's coding interpretations and met to discuss and agree on the development of themes as the analysis progressed, whereby any disagreements were resolved by consensus. The themes were agreed upon by the whole research team and are outlined in the *Results* section.

Results

Overview

We derived the following four themes from our data: (1) motivation for participating; (2) transparency, understanding, and consent; (3) privacy, anonymity, and data use; and (4) trust in research. Drawn to participate by their early exposure to the project at a public event or through their existing affiliation with the university, adult participants primarily (and uniformly) described being motivated by an altruistic desire to support research directed toward the public good. Participants described satisfaction with the SPHERE consent process, although they pointed to difficulties in recalling and understanding the information received, the timing and amount of information provision, and the fact that the process seemed, at least initially, to be rather abstract. Participants also reported satisfaction that privacy was assured, taking comfort from the fact that the data were anonymous, not sensitive, and unobtrusively collected, and they felt that any small threat to privacy was compensated by the fact that the data were collected for a worthwhile reason. Participants' trust in the project and the team was evident, and among the factors relevant to developing and maintaining that trust was the perceived trustworthiness of the research team, the provision of necessary information, the control participants had over participation, and the participants' positive prior experiences of research involvement.

Motivations for Participating

Participants described how they came to participate in SPHERE in terms of causal factors, including their early exposure to the project and (for some) their links with the university and their motivations for doing so, which were primarily concerned with benefiting others (other-regarding).

In terms of *causal factors*, households primarily became involved in SPHERE after attending public engagement events or because of existing links to the university. Approximately 71% (5/7) of the households had at least one member who had attended an engagement event. These events were described as informative and "interesting" (P14 from H06). Some participants also noted how attendance had enabled them to contribute to the development of the project, for example, following a "lively debate" about the "level of intrusion we might be prepared to accept" (P01 from H01). Participation in such events was not mentioned by the other 29% (2/7) of households, although those households did refer to an existing affiliation with the university. This was also mentioned by 60% (3/5) of those who attended engagement events.

Participants had a range of *motivations* for becoming involved; however, all had in common an altruistic desire to benefit others, which found expression in participation as they expected SPHERE to generate future health-related benefits. Although the participants were aware that the research would not directly benefit them, some hoped for future personal benefits, as the research might eventually help to address familial circumstances or conditions. Moreover, many anticipated a "[k]ind of public good" (P06 from H02), as they expected the research to benefit others (including the health service at large). Some such benefits were indicated during engagement events:

> *The second event…it was, kind of like, "This is what it might look like…." In fact, that was one...where for me, the penny dropped about how it might impact on the NHS [National Health Service], given that resources are never going to match demand and here is an electronic solution to a number of monitoring issues.* [P01 from H01]

Regardless of whether they attended a prestudy event, many of the participants echoed the sentiment that "Most research is useful, eventually, somehow and somewhere" (P13 from H05). Many participants were also motivated by their interest in the research, particularly, as one put it, "the technology side" (P05 from H02). Another participant elaborated as follows:

> *I thought it sounded interesting, and I think I just thought, "Oh, I'll just go along and see what it's all about." ...I suppose that's sort of the future use of technology, and supporting people with health conditions and stuff like that seemed interesting and important.* [P11 from H04]

Other- and self-regarding motivations were dominant and framed as positive reasons for wanting to take part, whereas some participants described a more passive motivation. For example, one (older) child participant felt that the project was interesting and agreed to participate as the other members of the family were willing to do so:

> *I thought it was interesting, but I just, kind of, went along...because everyone else was.* [P03 from H01]

No household reported any disagreement about the household's decision to participate, although this participant implies that the majority view in a given household might prove influential in securing the consent or assent of those who are less motivated. This finding highlights a potential risk with household research that some members may not be actively consenting (or perhaps even assenting) but merely acquiescing and, at worst, doing so as they feel unable to decline to participate. How acceptable this is might depend on the age or autonomy of the individuals in question and the level of potential insult to privacy and autonomy. One way to address this concern is to ensure that all participants freely assent [22] or at least do not dissent when they have a clear opportunity to do so [23]. However, it is not necessarily clear how we should establish assent or dissent in a family environment: consider, for example, a preteen's mild complaints, a toddler's tantrums, or an adult with dementia who states that cameras make her feel nervous.

Transparency, Understanding, and Consent

Whatever their specific motivations for taking part, all households expressed satisfaction with the consent process, including the information received and the opportunity to have their questions answered. However, 5 areas of difficulty and tension also emerged, concerning recall of information, understanding of that information, information overload, timing of information provision, and the possibility that consent might be more theoretical than actual.

First, participants experienced some difficulty in *remembering the consent process*, including whether this was a single event or an ongoing process. One household recalled providing

consent initially at a pre-event and then again at home before the equipment was installed; however, other households seemed only to recall giving consent at home before or during installation.

Second, various participants revealed *gaps in their understanding*. Some were uncertain about the general nature and remit of the project, with one participant, who had not attended a pre-event, feeling "a bit less clued up...about what the project was" (P06 from H02). Others queried more specific aspects of the project, including what the different pieces of equipment monitored are and why, what the data would look like and how it would be used, and the processes associated with governance and data sharing. Some appeared confused about the practicalities of participation, such as the correct ways to use the equipment and the data that would then be captured, although they were sometimes able to gain reassurance from members of the SPHERE team:

> *[The technician] came out and explained all the equipment, and how it worked, and showed me some diagrams, and I was aware, at that point, that you were...a silhouette image – so it wouldn't be taking detailed pictures of you. I don't think I realised, at that time, that it didn't take a detailed picture of the background though. I was always a bit conscious that...Because sometimes...the washing up stacks up. I was always thinking "They'll be able to see all my washing up"...[the researcher] showed me, when she came over, and then I probably felt more comfortable after seeing that. I know [the technician] did show me on a piece of paper, but [the researcher] re-showed me.* [P09 from H03]

Third, despite these apparent problems with recall and understanding, many households commended the consent process as "thorough" (P02 from H01 and P06 from H02) and "rigorous" (P02 and P04 from H01); however, for at least one household, this also meant that the process was *burdensomely time-consuming*:

> *So that kind of thing [obtaining informed consent] was only difficult because of my particular pressures that I was dealing with.... So I was like, "Come on, like speed it up." But obviously I know you've got to go through everything because that's how you make it thorough, so that's not really a thing anyone can change.* [P06 from H02]

Finally, participants detected *temporal dimensions* in the consent process; not only did they find it difficult to recall the process many months later, but they also noted how their informational needs had changed over time. Several participants described how they were satisfied that pertinent information had been disclosed before enrollment; however, the length and complexity of the research meant that they had since forgotten that information:

> *I think, at the time, I felt confident...I felt like everything was covered. Well, they did go into...I'm sure they went into it all. I think it's just faded in my mind.* [P11 from H04]

An older participant (P14 from H06) described not only having difficulties recalling information but also preferring not to receive too much written information. Others indicated that their informational needs had changed over the course of the study, with one subsequently becoming interested in how the data generated might be shared with other researchers, observing that "I hadn't really thought that far down the line with it" (P16 from H07), when initially providing consent. No participants recounted having sought additional information at any point, strongly implying that participants will not necessarily seek further information, particularly about technical and practical matters, after the initial consent interaction.

However, equipped with their knowledge of what participation had involved, some offered suggestions about disclosure to future participants to *make consent more informed and practical*. Many expressed an interest in receiving feedback on the data generated from their household, as well as about how the researchers were using the data collected. However, the latter was more for interest—"There's no reason for me to know it anyway" (P13 from H05)—and some participants appreciated that the research team would not be able to predict all future uses of the data, as "[i]t's kind of exploratory research in that way" (P05 from H02). However, many households did indicate that they would have liked more explanation of the practical ramifications of and daily burdens associated with participation rather than those items typically included in information leaflets:

> *I think, definitely just, I know it's such a physical thing rather than anything else, but just how much the technology does take up of space around your house, and things...I know it's so silly, those little things, but I guess it's just making sure, because I think in your head you've got this idea that technology is, well, if you're doing research it's going to be really advanced, and it's all going to be wireless, and everything is going to be like a smart home, type idea, but I don't think that is what SPHERE was. So, maybe just that, a little bit.* [P15 from H07]

Practical examples would help participants to better appreciate how the equipment might occupy the home and the routines that would be associated with (for example) using wearable devices; the latter appeared to be the most burdensome (and confusing) aspect of participation. Participants similarly felt that consent might be more authentic and informed if participants could see, in advance, what the data might look like, as otherwise "it's almost like you're consenting in theory to something that you don't quite know what they're doing, in a way" (P11 from H04).

Privacy, Anonymization, and Data Use

Participants confirmed that privacy and confidentiality were important to them; however, they expressed few concerns about these, provided that 4 conditions were satisfied: the data collected (and shared) were anonymous, the data were nonsensitive, data collection was unobtrusive; and, consistent with their altruistic motivation, the research (including any follow-on studies) was directed toward the public good.

Privacy was recognized as "[d]efinitely" (P05 from H02) important to participants who were comforted first by SPHERE's assurance of *anonymity*. The fact that cameras only recorded silhouettes and the data were not being transmitted in real time were considered important privacy safeguards. Participants said they would have been reluctant to participate if more identifiable data were collected: "if we put our names, I don't think we'd have done it realistically..." (P05 from H02). The provided data were anonymous, and participants were content for it to be used, including by other research teams:

> *I just think we're here to, like, provide useful information for other people, but I don't really mind what they do with it. I'm not really sure what they're doing with it, but that doesn't bother me because it's anonymous.* [P03 from H01]

Second, participants took comfort from the fact that collected data were *nonsensitive*, even mundane. Ambient sensors and energy- and water-use monitors were viewed as unproblematic—"It's not like I'm going to be too upset about someone knowing how humid my bathroom is..." (P05 from H02)—and, even if such data were leaked (eg, web-based), anyone viewing said data would "just know...we just walk around the house" (P04 from H01). Some participants were similarly unconcerned about the data collected by the cameras:

> *I'm not bothered [about camera data] because it is anonymised and, anyway, if someone could find out who it was it, kind of, wouldn't be very interesting, it wouldn't matter, and I wouldn't mind.* [P02 from H01]

Therefore, the mundanity of the data reassured participants: "We're not that exciting or perverse..." (P12 from H04). However, the participants did anticipate feeling differently about data collection and sharing, if the data were more sensitive:

> *I just don't understand how that would have a negative effect on me. So, as long as it's not financial or sensitive information, I don't really mind who knows that about me.* [P16 from H07]

Third, participants seemed reassured by the *unobtrusive* nature of data collection, which meant that their privacy was not noticeably invaded. SPHERE participants have a mechanism to pause or delete data capture using the SPHERE Genie app; however, it was apparent that use of this was inconsistent across households and some problems with usability were reported. Some households reported that visitors expressed discomfort with the equipment running, and some participants reported being occasionally mindful of the presence of cameras in the home:

> *When [P10] was little, sometimes I'd be trying to get her in her coat in her pram. It probably looked like I was having a bit of a wrestling match with her. I sometimes thought, "I wonder what, if they see that, they'll think of that." Me trying to get her into the pushchair and things, I don't know what that would look like on there.* [P09 from H03]

However, many participants soon overcame their initial apprehension and forgot about the monitoring:

> *I thought that the technology might be more intrusive than it was, so I thought it might get in the way...but it sort of disappeared a lot in your consciousness, isn't it? Once it was in for a while.* [P06 from H02]

Awareness of the cameras, for example, "probably lasted for about a week" (P16 from H07), with many participants similarly commenting, "I just forget the equipment is there, and don't think about it" (P09 from H03) and "I don't really notice it" (P03 from H01). Indeed, some implied that they would be happy for more data to be collected, so as to provide a fuller picture of their lives beyond their activities in selected rooms. Some participants were also willing to have cameras placed in more private indoor spaces, such as bedrooms and bathrooms (P13 from H05), although the following 2 participants focused instead on outdoor behaviors:

> *It feels like, then, people should know that I do go out and I do exercise, or I do go out and do a course.* [P14 from H06]

> *I just felt like people would get a picture of my life and it's not how I had it in my head. Does that make any sense? I see myself as quite an active person, quite a social person, but I don't feel like that would be shown because what they're actually watching is just me sat down, doing nothing....* [P15 from H07]

Fourth, participants took comfort in this research being aimed at the *public good*. Despite their evident support for research, this support was limited to research with such aims, such as university-led, health-related research. Furthermore, they tended to feel their data could be shared, provided the recipients had a legitimate interest in the data, as (for example) they were health researchers as opposed to "the News of the World" (P14 from H06) or "organisations like Facebook and WhatsApp who...I think misuse information" (P13 from H05). One participant put it as follows:

> *I think if they said, "Okay. We're actually using your information because we want to look at something completely random that's nothing to do with healthcare," I don't think I would feel as comfortable about that, because I don't feel like, then, that's what I signed up for.* [P15 from H07]

Although some were willing to have the data shared beyond the originally envisaged purposes of the project (eg, P03 from H01, quoted earlier), not everyone could recall the possible uses of their data or the processes for handling data access requests.

Participants accepted that the university might create "commercial spin-offs" (P05 from H02) using their data; however, they drew the line at sharing data with commercial companies:

> *I don't mind them [the University] making money, but I wouldn't have felt comfortable with our data about how we live our lives being used to either manipulate us in terms of sales things or really make money for a private company.* [P05 from H02]

Some participants strongly and consistently expressed their resistance to the data being placed in "some evil company's hands" (P11 from H04) or being transferred to "some horrible

company" (P11 from H04). The corollary of this is that participants appeared to trust SPHERE, the university, and its data-sharing arrangements:

> *I suppose, potentially, it could go to some unscrupulous researcher who was using it to make vast profits at people's expense, I don't know. But, it's hard to imagine how that could happen, really.* [P02 from H01]

Ultimately, participants expressed strong support for research that might prove beneficial, which meant they were willing to accept invasions of their privacy:

> *The level of intrusion you're prepared to accept probably depends on the level of benefit you see.... If it was more intrusive I might need to be convinced that there was more to be gained...but where we are at the moment makes perfect sense.* [P01 from H01]

Trust in Research

Participants exhibited a great deal of trust in the project and the team, which underpinned their perceptions of the adequacy of the arrangements around informed consent, privacy, and data use. Four factors appeared relevant to developing and maintaining their trust: being a trustworthy project and team, transparently providing participants with necessary information, allowing participants control over their participation, and the participants having prior positive research participation experiences.

First, participants felt that the project and team were *trustworthy*. How, exactly, the determination of trustworthiness was made is unclear; however, it seemed connected to the research team meeting the participants' expectations, their personal experiences with the people undertaking (and overseeing) the research, and the open ethical standards in operation. As we have seen, all 7 households were altruistically motivated, and their trust in the research and the research team that SPHERE was designed to serve the public good (even in the absence of a precise understanding of what that good would or might be) outweighed any concerns they might have had about privacy:

> *The fact that it was public money that was funding it and all those sort of things was quite reassuring from that point of view, that they weren't going to be just selling the data on or that their priority was around public good rather than being around monetising the data.* [P05 from H02]

The involvement of a university rather than a commercial enterprise appeared to instill trust:

> *I know it sounds crazy that it's a university, because they are really businesses as well, but they have a set of ethics, they have a set of rules that they've got to go by. They've got to keep people's information secure. There has got to be someone that agrees to an industrial company coming in and using that information. So, for me, I guess that's why I'm not [concerned], because I have a bit more trust in that. It's research that has come from a research place, I guess.* [P15 from H07]

Some participants were affiliated with the university, which inclined them to take *a leap of faith* and trust in the project and its team:

> *I still feel a bit sort of vague about some aspects of it [the research]. But it felt like, I mean [I'm associated with the University], so I kind of feel like there's a bit of a trust.... I sort of felt like, nice, the sort of researchers at the [University]. So, they must be alright, which maybe explains why, yes, I've just, a bit like a leap of faith, I've just gone along with things. Whereas, maybe if it had been done by an external company, or a business...I would've probably asked more questions. I think that was quite an important aspect of it....* [P11 from H04]

Even those not affiliated with the (or a) university reported feeling *reassured* that the research was undertaken and overseen in an academic setting: "I felt safe with what was happening" (P16 from H07). SPHERE was judged "to be a very, very ethical operation all the way through" (P01 from H01) that merited trust, whereas, as we have seen, participants were more suspicious of commercial companies:

> *The SPHERE project, I think, the information collected wouldn't be misused in any way. But I don't feel I would have the same confidence about an organisation like Facebook.* [P13 from H05]

Prior knowledge of or experience with research processes (eg, around privacy, data protection, and data sharing) also inclined participants to trust in SPHERE:

> *I assume that somebody, someone in SPHERE maybe...has some control over who gets the data, but maybe not. I mean, there must be some mechanism for doing that. I would just assume it's like with Biobank, I know that the data goes to...people and researchers who have been approved in some way. So, I've, kind of, always just assumed that it [SPHERE] would be the same....* [P02 from H01]

Furthermore, participants recognized their vulnerability, as they were dependent on the project team; however, they felt that the team was mindful of this and deserved their trust:

> *I think it's just a question of trust and the objectives and the people who are running it. One gets an impression – impressions are sometimes wrong of course – but you do feel there's something worthwhile where they would be careful about data.* [P13 from H05]

> *You're assured that it's anonymous, and to some extent you're relying on the experts in terms of ethics people and all that to make sure that that's done properly....* [P05 from H02]

> *I just presumed it [data] would be in the right hands and that's where it would stay....* [P16 from H07]

Second, participants were inclined to trust as the project *transparently provided what they considered to be appropriate information*. Despite some difficulties with recall, all participants felt sufficiently informed to consent to the participation. The provision of detailed information and the availability of research

team members who could answer questions helped to instill trust:

> *[The researcher] went through loads of forms and talked me through everything and made sure I was very comfortable with everything to do with it [SPHERE].* [P06 from H02]

> *A point of contact is good, just so that you've got someone to help, isn't it, and to reassure you....* [P16 from H07]

Some participants had previous exposure to research or were well acquainted with the research process and research environment and appeared relaxed, even blasé, about providing their consent:

> *You skim over it [participant information sheet and consent]. You're so used to doing consent forms for everything, aren't you? It's like, "Yes, yes, yes."* [P09 from H03]

> *This is going to sound bad, but...how much do you read it and take it in? Again, it's that sort of a bit of a leap of faith thing, where like, "Oh, yes, na, na, na, na, na. It's all the usual stuff," to a certain extent. It's good that it's all there, but it also makes me think...I probably skim read it [participant information sheet] at the time. How much did you take in when you first read it?* [P11 from H04]

Indeed, some participants appeared to trust the project, seemingly irrespective of the detailed information provided:

> *I just didn't think into it too much, I don't think. I just took it for what it was, and what I was told, and that was okay with me.* [P16 from H07]

However, others recognized that not everyone would be so trusting:

> *My father didn't trust anybody as far as you could throw them, about anything. He wouldn't have been involved in anything. Like my friend's husband said, "No, I'm not having strangers coming in and finding out about what we do."* [P14 from H06]

Therefore, perceived trustworthiness appeared to play a role in deciding to participate, although participants suspected that less trusting individuals might still come to trust (and participate) in research, provided that they were given detailed information:

> *If you want people to trust you to be secure, then they need to know that the information is like dot, dot, dot, dot, and nobody could tell who they are...I think it's important for trust, but also for you to get people.... To recruit people for whatever, because people need to have some belief or trust....* [P14 from H06]

Third, participants trusted SPHERE as they felt they *had control over their participation*. Several of those who had attended SPHERE public engagement events before their participation described this as giving them a sense of control, as they had had an opportunity to contribute to the project during its development:

> *Part of that might be that we've, kind of, grown up with the process...so, actually, to think of that equipment and think, "Actually, I know the process that has got to having that there," was very, very useful. Not everybody could have that, but for us, what gets recorded is partly influenced by what we said.* [P01 from H01]

> *It doesn't bother me [future data use/sharing], because as I say, I was part of the development. I trust the system....* [P14 from H06]

Participants also valued having control over the data collected, for example, by being able to pause collection or delete data using the *SPHERE Genie* app provided to them on a tablet. One household (H02) was surprised to have this level of influence, as it might adversely affect the volume of data collected. Although not every household used these functions, their value was recognized:

> *Because I'm not thinking about it [data collection and privacy], I don't think to use it. It's nice to think the delete function is there. I do think that's nice, to think there is that option. I felt more comfortable about the study knowing that was there, but it's just never been of any relevance to use it. So I've not deleted.* [P09 from H03]

Ultimately, the participants appeared to trust the research and the team. Indeed, all except one of the participants (who felt participation was burdensome) signaled that they would participate again, and 2 households had already agreed to participate for another year:

> *At the end, I was asked, "Would you be willing to go forward for another year?" "Certainly, yes," no reason why not.* [P13 from H05]

Therefore, *positive experiences* of research participation appear to be the fourth factor in maintaining or reinforcing trust (and future participation) in research.

Discussion

Principal Findings

For many of these participants, participation was *a question of trust*, and they felt able to take the *leap of faith* and enroll as they trusted the project and its team. The presence of trust seemed to permeate many of the observations that participants offered about their motivations for participating, the consent process, and privacy. Before noting some of the practical implications of these findings, we will first reflect on their ethical dimensions and implications, focusing on the concepts of trust and respect for autonomy.

Trust is a relatively contested concept [24], which is also rather neglected in bioethics [25]. However, Baier [26] has helped to plug the gap, suggesting that trust involves "reliance on others' competence and willingness to look after, rather than harm, things one cares about." Reliance alone does not suffice since, as Fritz and Holton [27] note, "Trusting someone involves both a behaviour – a readiness to rely on them – and an attitude," that is, an assumption that the trusted party is trustworthy, for

example, as she is concerned for one's interests. Nickel [28] further draws out the individual (*internalist*) and institutional (*externalist*) elements of trust, first noting the following:

> *People's interest in trust is not merely to have trust, but to have it in the right circumstances and for the right reasons. Normally, this aspect of trust is backed by having a reliable grasp of the interests, functions, and norms that motivate and explain the trusted entity's behavior.*

Internalist accounts of this aspect of trust tend to focus on the *trusting* party: Manson and O'Neill [29], for example, advance an ideal of *intelligent trust*, which emphasizes the characteristics (or virtues) of the truster, who makes sound choices about whom to trust. Externalist accounts look instead to the individual's social or physical environment and to the *trusted* party, emphasized in such notions as *sound trust* [30].

Our findings chime with these theoretical reflections. The participants noted how some nonparticipating individuals, whether *intelligently*, were not ready to rely on others and, therefore, not prepared to take part in research; however, this was not the attitude of these participants. Rather, our participants evidently were ready *to* trust and, indeed, were trusting *of* the research and the teams involved. They appeared to judge the project and team as worthy of trust and implicitly justified that trust as sound in various ways—usually by appealing to the behaviors of the research team (reflecting an externalist account of trust). The provision of detailed information and assurances about anonymity and data security helped to foster trust; however, participants appeared particularly to base their trust on the fact that the research was, like them, altruistically motivated, a mark of which was that it was led by a university, as opposed to a more self-interested commercial entity.

The research institution in which trust is placed and how it conducts and governs its research appears to matter, and participants were evidently reassured by the ethical goals of the research and the visible ethical standards in operation, which suggests that transparency about ethical standards and ethical practice is important. Time also plays a role here: as Fritz and Holton [27] note, "A trusting relationship will typically be built up over time as we gain evidence that our trust is well placed." This helps explain why some participants were willing to trust immediately: it was not that they trusted quickly, easily and naively, but rather that they had already built trust in the institution through their existing links and previous experiences. This existing trust enabled them to be unconcerned about areas of uncertainty or, potentially, sanguine about follow-on studies.

The importance of trust in research, which has been noted in other studies [31,32], has ethical implications for future research in smart homes (and, most likely, elsewhere), two of which we note here. First, the institution leading the research should ensure that it is indeed trustworthy. This can be justified not only in ethical but also in prudential terms: our participants imply that any damage to trust will be likely to jeopardize recruitment to future research. Second, attention should be directed to recruitment in any event, as our participants seemed to have trusting attitudes; however, this raises the question of whether less trusting individuals are being (self-) excluded from research. If this is the case, then there is a risk that the research will be biased. As we note below, our participants were all White, (necessarily) self-selecting, and (as they themselves revealed) many were affiliated with the university or at least able to attend pre-events. We wonder, then, what might be missed by failing to engage with and foster trust among other potential participants from more diverse ethnic, cultural, and social backgrounds.

Although trust appeared to be a significant ethical dimension of the findings, participants were also keen to ensure that their autonomy was respected. Respect for autonomy—literally, self-rule—is a dominant concept in bioethics. Amenable to different readings [33], the concept—whether alone or in combination with others—nevertheless underpins or at least connects with such obligations as the need to obtain consent, respect privacy, and maintain confidentiality. Furthermore, it is arguable that trust only has the value and role it does as it is autonomously given in response to trustworthiness.

Turning first to consent, participants were keen to ensure that they had sufficient information before agreeing to participate. They reported being satisfied with the *thorough* information imparted before enrollment and also valued the opportunity to ask questions later to a team member. However, not every participant availed themselves of the opportunity to have their questions answered, and there was evidence of confusion, gaps in knowledge, and desires for further (more practical and less abstract) information to be provided. Problems with participant recall have been recognized in other studies [34]; however, what emerges particularly clearly here is how autonomy is sometimes traded off against or otherwise outweighed by considerations of trust. Autonomy matters, but trust seems to matter more. As such, these participants appeared unconcerned about their inability to recall or at the time fully comprehend study information *just because* they trusted the project and the team, an attitude perhaps best exemplified by the relatively relaxed attitude taken by some participants to study invitations, where participants reported skim reading the information before providing consent.

Some may consider this finding surprising, and it seems at first glance to suggest that participants are less concerned with being in control—with expressing their autonomy—than we might assume. Indeed, this appears to be at odds with the findings of an earlier study that explored the ethical perspectives of (early- and midcareer) smart home researchers [14]. Those participants indicated that they saw the provision of choice to smart home households as offering a solution to the ethical dilemmas, primarily those relating to privacy, which might arise. The current participants certainly valued choice, but they appeared most inclined to trust the researchers, which, contrary to what the researchers themselves indicated, implies that the research team rather than the households is ultimately in control. However, perhaps the respectful and careful attitude of the researchers in their dealings with the participants was central to the trusting relationship that developed.

Trust also appeared to take priority over autonomy when, second, the participants reflected on privacy and confidentiality. Autonomy was again valued in these contexts, with participants pleased to have the opportunity to control the data that were

(not) collected and reassured that any such data would be anonymous, mundane, and unobtrusively collected and would (they assumed) only be shared with trustworthy researchers undertaking research in the public interest. However, in practice, not all participants chose to pause data collection or delete the data that had been collected. Furthermore, although anonymity seemed to be important, at least some of the participants signaled that they would be willing to accept greater levels of intrusion and share more data about themselves. These participants appeared willing to trade off both their privacy and their autonomy to support beneficial research, which was linked to their (altruistic) motivations for participating and, ultimately, their trust in the research. The participants' willingness to have their data used in future health research (but not necessarily in other research) indicates that they may have in mind what Nissenbaum [35] has termed "norms of appropriateness," which "dictate what information about persons is appropriate, or fitting, to reveal in a particular context." Sharing beyond such boundaries might result in a privacy violation on Nissenbaum's account [12,35]. Provided that commercial companies would not receive their data for private gain, participants were also unconcerned that they did not (and perhaps could not) know the future uses to which their data might be put. Sheehan [10], for one, detects no problem with the *broad consent* these participants appear to have in view, as this is still consent and therefore respectful of autonomy. Indeed, according to O'Neill, "[n]either accountability nor informed consent is improved by aiming for high detail and specificity" [36]; rather, she suspects—like our participants—that "any regress of control mechanisms has eventually to end in a decision to place – or refuse – trust" [36].

Relinquishing control to this extent makes participants vulnerable to potential harm; however, participants appeared willing to make themselves vulnerable in this way as they had made the decision to trust. When that decision has been made intelligently, then that trust is an expression of autonomy, and we simply have an obligation to ensure that trust is not breached. Danger may occur when that trust is blind and insufficiently informed to be considered autonomous. However, the appropriate responses appear to be the same.

Therefore, these findings appear to have ethical implications, as far as they indicate that, despite there being a strong emphasis on informing participants and ensuring an autonomous decision in research participation, establishing and maintaining trust may be an essential part of smart home research and that allowing and respecting that trust may be an appropriate way to respect autonomy. If that trust is to be repaid and extended to future studies, researchers should be aware that participants' agreement to take part in research imposes on the researchers a long-term duty of care toward those participants.

Notwithstanding this general finding that it may be appropriate and acceptable, in future smart home research (and, surely, in many other fields), for respect for autonomy to be viewed and understood through this lens of trust, our findings do also have practical implications for fostering trust. First, the *types of information* imparted in the consent process might be usefully expanded. Beyond the usual information contained in patient information sheets, these participants appeared keen to learn more about the practical ramifications of living with technology. Second, consideration should be given to the *methods of communicating information*. Written information appears useful, but these participants particularly valued (ongoing) in-person communication, including communication before their recruitment during the public engagement events they attended. Third, the thorny question of *feeding findings back to participants* arose here. Some participants sought informational rewards for their participation. However, this might prove challenging for studies of this sort involving whole households, given the need to maintain anonymity and the likely difficulties in disentangling information about multiple participants in a household. This is something that needs to be discussed and, if necessary, negotiated during the household consenting process (with researchers alert to the unequal power dynamics that exist within households, especially those with older children). Fourth, the *promise of anonymity* itself raises questions. These participants appeared to assume that anonymity was guaranteed. We query the viability of an absolute guarantee, and, in the first instance, researchers need to be transparent about this from the outset, given the role that transparency appears to play in fostering trust. However, the presence of trust also implies a long-term duty of ensuring that participant data remain as anonymous as possible. Such a duty may imply that data custodians should work closely with researchers conducting secondary analyses to ensure that participants do not become more identifiable as a result of secondary research. Finally, thought should also be given to *commercial involvement* in the research. Universities appear to be trusted entities, so they may need to exercise caution when collaborating or sharing data with industries. At a minimum, participants will expect to know if or how this is happening; however, the presence of trust may imply longer-term obligations to ensure that the trust is repaid beyond the immediate terms for the consent that has been given.

Limitations

This study had some limitations. First, although the numbers are adequate for qualitative research with a nonrandomized sample [37], the generalizability or transferability of our findings is necessarily limited, partly because this is qualitative research, but moreover because the sample size was small and the participants were all drawn from a single smart home project, so their views might not be reflective of those involved in other projects. Nevertheless, we consider the exposure of SPHERE participants to prolonged and invasive ubiquitous monitoring to be unique, and thus, research soliciting their insights is justified on this basis. Second, interviewing and conducting inductive thematic analysis are necessarily subjective processes, so other researchers might have derived different themes. However, we achieved thematic saturation and involved the (multidisciplinary) research team in the analysis, so we believe that we have at least given a fair account of the data collected. Third, we could only capture the opinions of a self-selecting sample of participants. The sample only included White participants, and we acknowledge this lack of ethnic diversity as a limitation. To our knowledge, there are no other similar research studies that specifically report on ethnically diverse populations, and unfortunately this lack of diversity appears to be a common problem across other types of research involving

human participants [38,39]. Providing opportunities for involvement is a key part of increasing diversity in research [40]; at the time of writing, SPHERE is funding a program of outreach to ethnic minority groups in the city, with the explicit aim that future studies can access the views of more diverse individuals. For SPHERE-CARED, eligible individuals were those who had agreed to participate not only in the SPHERE project but also in further research linked to the project. This means that we could only directly access and represent the views of those willing to participate in the research; those who are unwilling to participate in or even distrustful of the research might well offer different insights. Nevertheless, we note that, in addition to some areas of agreement, our data also captured diverse opinions, including (indirectly) the views of those resistant to participation in research. Finally, although we endeavored to include young children in the interviews (aged 0-5 years), in practice, this proved difficult to achieve as they were too young to follow the interview in full, and when the interviewer posed simpler questions directly to the children, they were shy or unable to articulate themselves; therefore, we were only able to meaningfully use data from older children in the analysis. Furthermore, although every effort was made to create space and opportunity for older children to express themselves during group interviews, it remains possible that existing family power dynamics may have prevented some children (or, indeed, less confident adults) from expressing their opinions [41].

Conclusions

The SPHERE study offered a distinct opportunity to access the experiences and opinions of participants involved in a smart home research project. Our qualitative study invited willing households to reflect on the practical and ethical dimensions of consent to participation, privacy, anonymization, and data sharing. Although a small study, participants offered insights that might inform future research in this area (and, perhaps, beyond).

Drawn to the project via their existing links to the university or an introductory engagement event, the participants were mainly and uniformly motivated to participate by an altruistic desire to support (health-related) research directed toward the public good. Despite valuing the thorough consent process, the participants revealed certain difficulties with recalling and comprehending the information received, the timing and amount of the information provided, and the fact that the process seemed, at least initially, to be somewhat abstract. Participants also acknowledged the importance of privacy and confidentiality but were reassured by the anonymity and nonsensitive nature of the data collected, its unobtrusive collection, and their belief that they were supporting valuable research, consistent with their altruistic motivation. Notably, participants' perceptions of informed consent, privacy, and data use, all appeared to be informed by their trust in the project. Among the factors relevant to developing and maintaining their trust were the trustworthiness of the research team, the provision of necessary information, the control participants had over participation, and the participants' positive prior experiences of involvement in research.

The findings may have practical implications for future research, regarding not only (for example) the types of information researchers should convey and the extent to which anonymity can be assured but also the long-term duty of care owed to participants who had trusted them not only on the basis of this information but also because of their institutional affiliation. Moreover, the propensity to trust according to prior experiences with research or affiliation with research institutions raises an important concern regarding diversity in research participation, whereby researchers should be aware that individuals without these prerequisites may not be so forthcoming in trusting research and offering their participation. There also appear to be important ethical implications: although autonomy matters, trust appears to matter most to these participants. Therefore, researchers should be alert to the need to foster and maintain trust, particularly as failing to do so might have deleterious effects on future research.

Acknowledgments

The authors would like to thank the members of the SPHERE (Sensor Platform for Healthcare in a Residential Environment) team for facilitating this research, particularly Dannie Hale, who was involved in the sampling and recruitment of participants; Julia Carey and Lorraine McManus for providing administrative support; and Alison Burrows for providing advice on the design and initial setup of this work.

SPHERE-CARED (Consent and Anonymization: A Review of Ethical Dimensions; principal investigator: RH) was kindly funded by SPHERE (principal investigator: IC), so this work was performed under the SPHERE Interdisciplinary Research Collaboration funded by the UK Engineering and Physical Sciences Research Council (grant EP/K031910/1). RH and JI are partly funded by the National Institute for Health Research Biomedical Research Centre at University Hospitals Bristol and Weston National Health Service Foundation Trust and the University of Bristol. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

MK, RH, JI, GB, and IC were responsible for developing the research protocol for the qualitative research. MK recruited participants and conducted the interviews. MK led the qualitative analysis, to which RH, JI, and GB contributed. MK prepared most of the first draft of the manuscript; RH prepared a revised, complete second draft, to which all other authors then contributed. All authors approved this manuscript for publication.

Conflicts of Interest

None declared.

References

1. Lyytinen K, Yoo Y. Issues and challenges in ubiquitous computing. Communications of the ACM. 2002 Dec. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.135.3184&rep=rep1&type=pdf [accessed 2021-09-24]
2. Brey P. Freedom and privacy in ambient intelligence. Ethics Inf Technol 2005 Sep;7(3):157-166. [doi: 10.1007/s10676-006-0005-3]
3. Chan M, Estève D, Escriba C, Campo E. A review of smart homes - present state and future challenges. Comput Methods Programs Biomed 2008 Jul;91(1):55-81. [doi: 10.1016/j.cmpb.2008.02.001] [Medline: 18367286]
4. Hofmann B. Ethical challenges with welfare technology: a review of the literature. Sci Eng Ethics 2013 Jun;19(2):389-406. [doi: 10.1007/s11948-011-9348-1] [Medline: 22218998]
5. Amiribesheli M, Benmansour A, Bouchachia A. A review of smart homes in healthcare. J Ambient Intell Human Comput 2015 Mar 14;6(4):495-517. [doi: 10.1007/s12652-015-0270-2]
6. Majeed A, Maile EJ, Bindman AB. The primary care response to COVID-19 in England's National Health Service. J R Soc Med 2020 Jun;113(6):208-210 [FREE Full text] [doi: 10.1177/0141076820931452] [Medline: 32521196]
7. Jones S, Hara S, Augusto J. eFRIEND: an ethical framework for intelligent environments development. Ethics Inf Technol 2014 Dec 14;17(1):11-25. [doi: 10.1007/s10676-014-9358-1]
8. Woznowski P, Burrows A, Diethe T, Fafoutis X, Hall J, Hannuna S, et al. SPHERE: a sensor platform for healthcare in a residential environment. In: Angelakis V, Tragos E, Tragos HC, Kapovits A, Bassi A, editors. Designing, Developing, and Facilitating Smart Cities: Urban Design to IoT Solutions. Cham: Springer; 2017:315-333.
9. Elsts A, Oikonomou G, Fafoutis X, Piechocki R. Internet of things for smart homes: Lessons learned from the SPHERE case study. In: Proceedings of the Global Internet of Things Summit (GIoTS). 2017 Presented at: Global Internet of Things Summit (GIoTS); June 6-9, 2017; Geneva, Switzerland. [doi: 10.1109/giots.2017.8016226]
10. Sheehan M. Can broad consent be informed consent? Public Health Ethics 2011 Nov;4(3):226-235 [FREE Full text] [doi: 10.1093/phe/phr020] [Medline: 22102849]
11. Schoeman FD. Privacy and Social Freedom (Cambridge Studies in Philosophy and Public Policy). Cambridge: Cambridge University Press; 1992.
12. Nissenbaum H. Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford: Stanford University Press; 2009.
13. Burrows A, Coyle D, Gooberman-Hill R. Privacy, boundaries and smart homes for health: an ethnographic study. Health Place 2018 Mar;50:112-118. [doi: 10.1016/j.healthplace.2018.01.006] [Medline: 29428688]
14. Birchley G, Huxtable R, Murtagh M, Ter Meulen R, Flach P, Gooberman-Hill R. Smart homes, private homes? An empirical study of technology researchers' perceptions of ethical issues in developing smart-home health technologies. BMC Med Ethics 2017 Apr 04;18(1):23 [FREE Full text] [doi: 10.1186/s12910-017-0183-z] [Medline: 28376811]
15. Masur PK, Scharkow M. Disclosure management on social network sites: individual privacy perceptions and user-directed privacy strategies. Soc Media Society 2016 Feb 24;2(1):1-13. [doi: 10.1177/2056305116634368]
16. SPHERE - a sensor platform for healthcare in a residential environment. University of Bristol. 2018. URL: https://www.irc-sphere.ac.uk/partners-team [accessed 2021-06-01]
17. Holmes M, Nieto MP, Song H, Tonkin E, Grant S, Flach P. Modelling patient behaviour using iot sensor data: a case study to evaluate techniques for modelling domestic behaviour in recovery from total hip replacement surgery. J Healthc Inform Res 2020 May 03;4(3):238-260. [doi: 10.1007/s41666-020-00072-6]
18. Cavoukian A. Privacy by design: the 7 foundational principles. Information and Privacy Commissioner of Ontario. 2009. URL: https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf [accessed 2021-09-24]
19. Palys T. Purposive sampling. In: Given LM, editor. The SAGE Encyclopedia of Qualitative Research Methods. Thousand Oaks, California, United States: SAGE Publications; 2008:1-698.
20. Lund LK, Zimmerman TS, Haddock SA. The theory, structure, and techniques for the inclusion of children in family therapy: a literature review. J Marital Fam Ther 2002 Oct;28(4):445-454. [doi: 10.1111/j.1752-0606.2002.tb00369.x] [Medline: 12382553]
21. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
22. Tait AR, Geisser ME. Development of a consensus operational definition of child assent for research. BMC Med Ethics 2017 Jun 09;18(1):41 [FREE Full text] [doi: 10.1186/s12910-017-0199-4] [Medline: 28599638]
23. Bos W, Westra A, de Beaufort I, van de Vathorst S. To stop or not to stop: dissent and undue burden as reasons to stop participation in paediatric research. J Med Ethics 2017 Aug;43(8):519-523. [doi: 10.1136/medethics-2016-103788] [Medline: 27934770]
24. McLeod C, Zalta EN. "Trust", The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University. 2006. URL: https://plato.stanford.edu/entries/trust/ [accessed 2021-09-24]

25. Nie J, Li L, Gillett G, Tucker JD, Kleinman A. The crisis of patient-physician trust and bioethics: lessons and inspirations from China. Dev World Bioeth 2018 Mar;18(1):56-64. [doi: 10.1111/dewb.12169] [Medline: 28922581]
26. Baier A. Trust and antitrust. Ethics 1986 Jan;96(2):231-260. [doi: 10.1086/292745]
27. Fritz Z, Holton R. Too much medicine: not enough trust? J Med Ethics 2019 Jan;45(1):31-35 [FREE Full text] [doi: 10.1136/medethics-2018-104866] [Medline: 30367013]
28. Nickel PJ. The ethics of uncertainty for data subjects. In: Krutzinna J, Floridi L, editors. The Ethics of Medical Data Donation. Cham: Springer; Jan 16, 2019:55-74.
29. Manson NC, O'Neill O. Rethinking Informed Consent in Bioethics. Cambridge: Cambridge University Press; 2007.
30. Voerman SA, Nickel PJ. Sound trust and the ethics of telecare. J Med Philos 2017 Feb;42(1):33-49. [doi: 10.1093/jmp/jhw035] [Medline: 27932397]
31. Damschroder LJ, Pritts JL, Neblo MA, Kalarickal RJ, Creswell JW, Hayward RA. Patients, privacy and trust: patients' willingness to allow researchers to access their medical records. Soc Sci Med 2007 Jan;64(1):223-235. [doi: 10.1016/j.socscimed.2006.08.045] [Medline: 17045717]
32. Aitken M, Cunningham-Burley S, Pagliari C. Moving from trust to trustworthiness: experiences of public engagement in the Scottish Health Informatics Programme. Sci Public Policy 2016 Oct;43(5):713-723 [FREE Full text] [doi: 10.1093/scipol/scv075] [Medline: 28066123]
33. Christman JP. The Inner Citadel: Essays on Individual Autonomy. Oxford: Oxford University Press; 1989.
34. Robinson JO, Slashinski MJ, Wang T, Hilsenbeck SG, McGuire AL. Participants' recall and understanding of genomic research and large-scale data sharing. J Empir Res Hum Res Ethics 2013 Oct;8(4):42-52 [FREE Full text] [doi: 10.1525/jer.2013.8.4.42] [Medline: 24169421]
35. Nissenbaum H. Privacy as contextual integrity. Washington Law Review 2004;79:119-158 [FREE Full text]
36. O'Neill O. Accountability, trust and informed consent in medical practice and research. Clin Med (Lond) 2004;4(3):269-276 [FREE Full text] [doi: 10.7861/clinmedicine.4-3-269] [Medline: 15244364]
37. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]
38. Hussain-Gambles M, Atkin K, Leese B. Why ethnic minority groups are under-represented in clinical trials: a review of the literature. Health Soc Care Community 2004 Sep;12(5):382-388. [doi: 10.1111/j.1365-2524.2004.00507.x] [Medline: 15373816]
39. Guerrero S, López-Cortés A, Indacochea A, García-Cárdenas JM, Zambrano AK, Cabrera-Andrade A, et al. Analysis of racial/ethnic representation in select basic and applied cancer research studies. Sci Rep 2018 Sep 18;8(1):13978 [FREE Full text] [doi: 10.1038/s41598-018-32264-x] [Medline: 30228363]
40. UK standards for public involvement in research. National Institute for Health Research. 2019. URL: https://sites.google.com/nihr.ac.uk/pi-standards/home [accessed 2021-09-24]
41. Reczek C. Conducting a multi family member interview study. Fam Process 2014 Jun;53(2):318-335. [doi: 10.1111/famp.12060] [Medline: 24410452]

Abbreviations

CARED: Consent and Anonymization: A Review of Ethical Dimensions
NHS: National Health Service
SPHERE: Sensor Platform for Healthcare in a Residential Environment

Edited by R Kukafka, G Eysenbach; submitted 23.10.20; peer-reviewed by R Hendricks-Sturrup, A Alasmari, H Biermann; comments to author 20.11.20; revised version received 15.01.21; accepted 01.08.21; published 26.11.21.

Please cite as:
Kennedy MR, Huxtable R, Birchley G, Ives J, Craddock I
"A Question of Trust" and "a Leap of Faith"—Study Participants' Perspectives on Consent, Privacy, and Trust in Smart Home Research: Qualitative Study
JMIR Mhealth Uhealth 2021;9(11):e25227
URL: https://mhealth.jmir.org/2021/11/e25227
doi:10.2196/25227
PMID:34842551

Original Paper

Understanding the Predictors of Missing Location Data to Inform Smartphone Study Design: Observational Study

Anna L Beukenhorst[1,2], PhD; Jamie C Sergeant[1,3], PhD; David M Schultz[4,5], PhD; John McBeth[1], PhD; Belay B Yimer[1], PhD; Will G Dixon[1,6], PhD

[1]Centre for Epidemiology Versus Arthritis, Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom

[2]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Harvard University, Boston, MA, United States

[3]Centre for Biostatistics, University of Manchester, Manchester, United Kingdom

[4]Centre for Atmospheric Science, Department of Earth and Environmental Sciences, University of Manchester, Manchester, United Kingdom

[5]Centre for Crisis Studies and Mitigation, University of Manchester, Manchester, United Kingdom

[6]NIHR Greater Manchester Biomedical Research Centre, Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom

Corresponding Author:
Anna L Beukenhorst, PhD
Centre for Epidemiology Versus Arthritis
Manchester Academic Health Science Centre
University of Manchester
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 2622757506
Email: beuk@hsph.harvard.edu

Abstract

Background: Smartphone location data can be used for observational health studies (to determine participant exposure or behavior) or to deliver a location-based health intervention. However, missing location data are more common when using smartphones compared to when using research-grade location trackers. Missing location data can affect study validity and intervention safety.

Objective: The objective of this study was to investigate the distribution of missing location data and its predictors to inform design, analysis, and interpretation of future smartphone (observational and interventional) studies.

Methods: We analyzed hourly smartphone location data collected from 9665 research participants on 488,400 participant days in a national smartphone study investigating the association between weather conditions and chronic pain in the United Kingdom. We used a generalized mixed-effects linear model with logistic regression to identify whether a successfully recorded geolocation was associated with the time of day, participants' time in study, operating system, time since previous survey completion, participant age, sex, and weather sensitivity.

Results: For most participants, the app collected a median of 2 out of a maximum of 24 locations (1760/9665, 18.2% of participants), no location data (1664/9665, 17.2%), or complete location data (1575/9665, 16.3%). The median locations per day differed by the operating system: participants with an Android phone most often had complete data (a median of 24/24 locations) whereas iPhone users most often had a median of 2 out of 24 locations. The odds of a successfully recorded location for Android phones were 22.91 times higher than those for iPhones (95% CI 19.53-26.87). The odds of a successfully recorded location were lower during weekends (odds ratio [OR] 0.94, 95% CI 0.94-0.95) and nights (OR 0.37, 95% CI 0.37-0.38), if time in study was longer (OR 0.99 per additional day in study, 95% CI 0.99-1.00), and if a participant had not used the app recently (OR 0.96 per additional day since last survey entry, 95% CI 0.96-0.96). Participant age and sex did not predict missing location data.

Conclusions: The predictors of missing location data reported in our study could inform app settings and user instructions for future smartphone (observational and interventional) studies. These predictors have implications for analysis methods to deal with missing location data, such as imputation of missing values or case-only analysis. Health studies using smartphones for data collection should assess context-specific consequences of high missing data, especially among iPhone users, during the night and for disengaged participants.

(JMIR Mhealth Uhealth 2021;9(11):e28857) doi:10.2196/28857

KEYWORDS

geolocation; global positioning system; smartphones; mobile phone; mobile health; environmental exposures; data analysis; digital epidemiology; missing data; location data; mobile application

Introduction

Smartphones offer opportunities to collect sensor data frequently from people's daily lives and to determine their exposures or behaviors. Smartphone location data can be collected frequently (eg, daily, hourly, continuously) over sustained periods of time [1]. Studies have used these data to quantify exposure to weather [2,3], air pollution [4], vicinity to tobacco outlets [5], or to deliver context-aware messages when participants visited health facilities [6,7]. Smartphones can provide complete and accurate location data, especially when participants are provided with study smartphones, studies are short, and data are collected nearly continuously [8,9]. However, in large-scale epidemiological studies, location data are often collected for longer periods, less frequently, and from participants' own smartphones. In these cases, missing data are more common than when using research-grade location trackers [4,10,11]. In observational research studies, missing data can result in the loss of power, selection bias, and misclassification of participants' exposure or behavior [12]. In trials, it could hamper safe and effective delivery of context-aware interventions that rely on location data [13].

To anticipate the potential impact of missing location data on study findings, we need to better understand how often, when, and why location data are missing. Previous smartphone studies have reported the amount of missing location data [4,10,14,15]. However, they typically did not investigate differences in missing data over time [4,10,14,15], between participants [4,10,14,15], or between operating systems [4,14]. In addition, they have limitations of small sample sizes.

We therefore investigated the distribution of missing location data over time, predictors of missing location data, and between-participant differences. We used data from a longitudinal smartphone study with 9665 participants using Android phones or iPhones. We anticipate that understanding the predictors of missing location data could inform researchers who want to improve data completeness during study design and data collection.

Methods

Ethics Approval and Consent to Participate

The University of Manchester Research Ethics Committee (reference, ethics/15522) and the National Health Service Integrated Research Application System (reference 23/NW/0716) approved this study. Participants were required to provide electronic consent for study inclusion. Further details are available elsewhere [2,3].

Study Design

We performed a secondary analysis of the data from an observational smartphone study that analyzed the association between weather conditions and chronic pain in the United Kingdom (study name: *Cloudy with a Chance of Pain*) [3]. In this study, we collected self-reported pain levels from a large group of people with chronic pain such as osteoarthritis, rheumatoid arthritis, or migraine. The exposure of interest was daily average weather conditions (ie, temperature, relative humidity, wind speed, and air pressure). To determine what daily average weather conditions participants were exposed to, the app recorded participants' geolocation, which we could link to weather reports from local weather stations. The analysis of the weather and pain association and the details of data collection are described elsewhere [2,3].

Data Collection

People with chronic pain downloaded the app onto their Android phones or iPhones, provided informed consent, and reported baseline participant characteristics (eg, sex, year of birth, self-reported weather sensitivity). At local time of 6:24 PM each day, participants received a push notification to complete a survey, rating 10 aspects of symptoms, behavior, and well-being. To obtain weather data from the closest weather station, geolocation was required. The app was programmed to record geolocation each hour on the hour; thus, the app would ideally obtain 24 geolocations each day. The app used GPS (outdoors) and network signals (inside buildings) to determine the latitude and longitude. The app's ability to record geolocations depended on (1) the participant granting the app access to their geolocation and (2) the participant switching on the location services on their phone. Upon downloading the study app, the participants were requested access to their geolocation. Access to geolocation was voluntary; participants who provided the app with access to their geolocation could retract access at any time or switch off location services temporarily or permanently, in which case the app would not be able to record the participant's location. The app recorded the operating system of the smartphone, but this feature was introduced 1 week after the recruitment launch and was not collected for early enrollers.

Data Preparation and Eligible Participants

We investigated location-data completeness on calendar days that a participant completed the survey. Participants were eligible if they completed the survey at least once, excluding the day of enrollment. This exclusion ensured comparability of participant days, as recording 24 geolocations would be unlikely on the day of download. For each participant, we selected all days with survey data. For each full clock hour, we added indicators for (1) location data (1 if observed, 0 if missing), (2) number of days since the most recent survey completion (0 if less than 24 hours ago, 1 if 24-47 hours ago, etc), (3) time in study (days since first survey submission), and (4) time (weekday or weekend, part of the day where night was considered as midnight to 5:59 AM, morning as 6 AM to 11:59

AM, afternoon as noon to 5:59 PM, evening as 6 PM to 23:59 PM, and hour of the day). In addition, we added indicators for variables that did not change over time: (1) participant characteristics (eg, sex, age, self-reported weather sensitivity) and (2) operating system (eg, iPhone operating system, Android, or unknown).

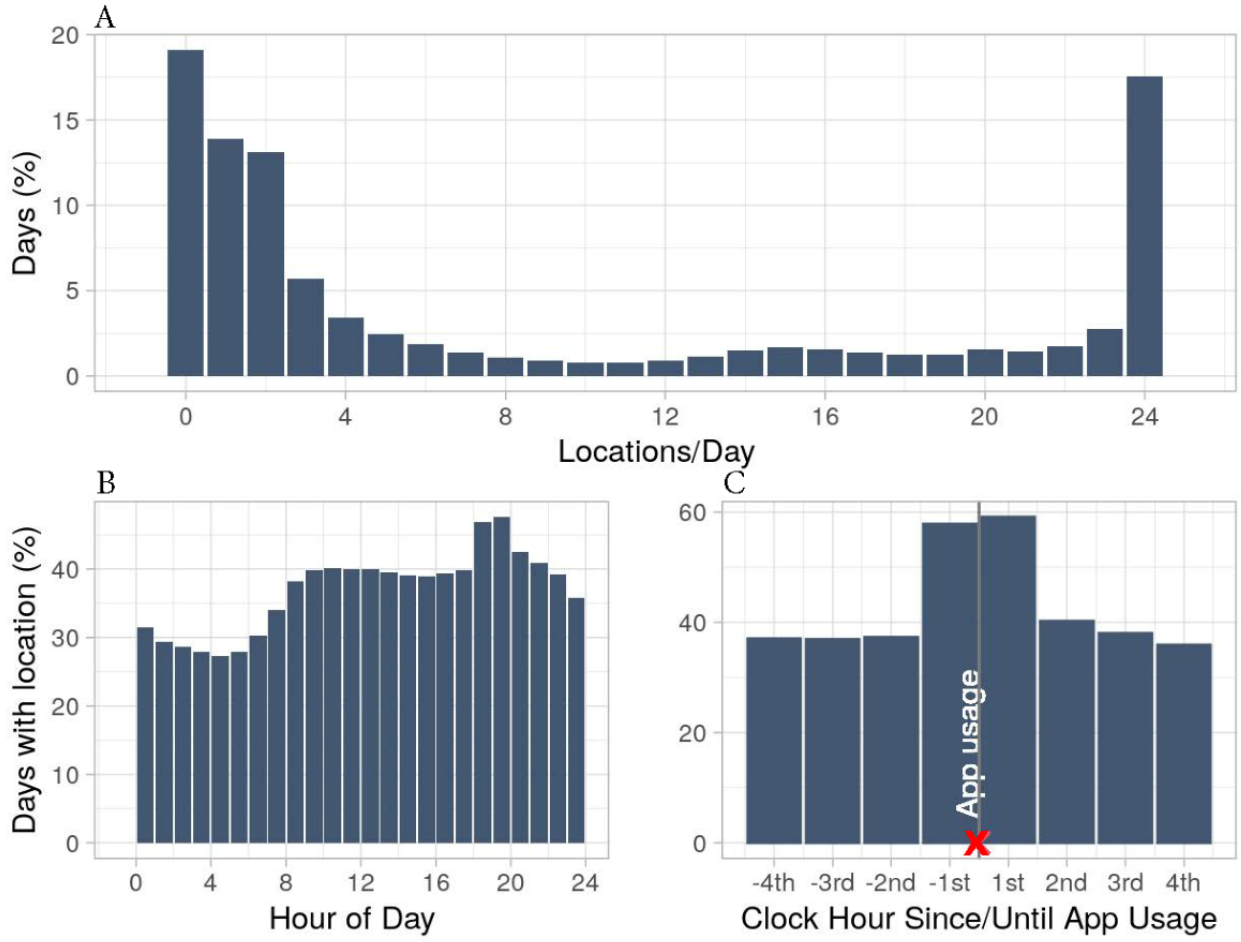
Data Analysis

We reported the number of eligible participants and their characteristics. We reported location-data completeness (1) per day (number of recorded locations during a day), (2) per hour for each clock hour (percentage of participant days with a recorded location data at that hour), (3) per hour for the 4 hours before and after survey completion, and (4) averages per participant (median number of recorded locations) for all participants and stratified by operating system. We investigated predictors of the outcome "presence of a location data point" (0 if missing, 1 if observed for a given full clock hour) with a logistic regression model with a participant-specific random intercept for within-participant correlation between repeated measurements [16-18]. A multivariable model identified whether the likelihood of the missing location data were associated with time indicators (ie, weekdays vs weekend days, part of the day), participant characteristics (ie, age, sex, self-reported weather sensitivity dichotomized around the median), operating system on their phone, survey compliance (ie, days since previous survey entry), or time in study (ie, days since first survey entry). Only participants with complete data for all covariates contributed information to the model. We estimated 95% CIs with 1000 simulations as recommended in [19]. Models were fitted in R (R Core Team) version 3.6 with the package lme4 [18]; odds ratios (ORs) and CIs were estimated using the merTools package [20].

Results

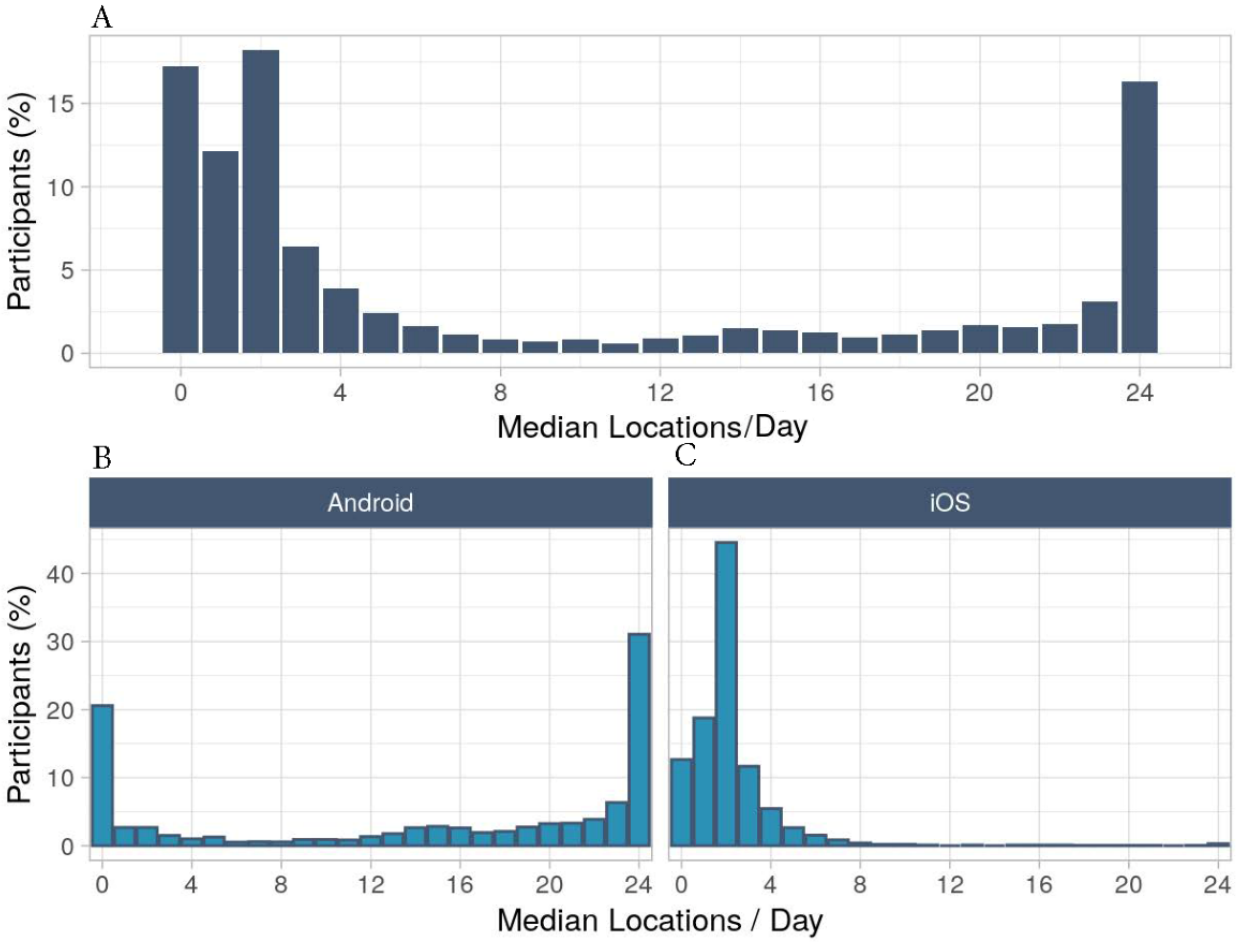
The app was downloaded by 13,207 participants, of which 9665 were eligible for inclusion (mean age 49 [SD 13] years; females, 7211/9665, 74.6%). These participants contributed to 488,400 participant days (median 14 eligible days/participant; IQR 4-60 days/participant). Of 9665 participants, 3109 (32.2%) used an Android phone, 1930 (19.9%) an iPhone, and the operating system was unknown for the remaining 4626 (47.6%) participants. We expected 11.72 million location data points or clock hours: 24 for each hour in the 488,400 participant days. Of 11.72 million hours, the app collected only 4.36 million clock hours (37.2%), resulting in missing data for the remaining 7.36 million clock hours. Data completeness per participant day varied from no location data (0/24) to fully complete data (24/24, Figure 1A, median 3, IQR 1-19). Location data were complete (24/24) for 17.5% (85,606/488,400) of participant days. Participant days with no location (93,255/488,400, 19.1%), 1 location (67,963/488,400, 13.9%), or 2 locations (64,207/488,400, 13.1%) were also common. Location was most often recorded at 7 PM (232,295/488,400, 47.5% of participant days; Figure 1B) just after the default notification of 6:24 PM. Locations were least often recorded between midnight and 6 AM. Location data were often recorded for the hour before survey completion (281,767/487,391, 57.8%) and the hour after survey completion (257,743/436,263, 59.1%; Figure 1C).

Figure 1. Data completeness. A: Distribution of participant days with a recorded location, stratified per hour of the day (N=488,400). B: Data completeness per hour of the day (N=24 x 488,400). C: Data completeness around the moment of survey completion (N=24 x 488,400). The red X marks app usage, and 1st is the first full clock hour after data entry.



For most participants, the app collected a median of 2 out of a maximum of 24 locations (1760/9665, 18.2% of participants), no location data (1664/9665, 17.2%), or complete location data (1575/9665, 16.3%; Figure 2A). Stratification by phone operating system and participant characteristics showed that 31.0% (965/3109) of Android users had 24 recorded locations on average versus less than 1% (6/1930) of iPhone users (Figure 2B). Android users usually had averages of 24 (out of 24) (965/3109, 31.0%) or 0 (out of 24) locations (640/3109, 20.6%). iPhone users usually had averages of 2 (out of 24) (859/1930, 44.5%) or 1 (out of 24) (362/1930, 18.8%) location, while only 0.03% (6/1930) had averages of 24 (out of 24) locations.

Figure 2. Median locations per day per participant. A: All participants (N=9665). B: Stratified by operating system for 3109 Android users and 1930 iPhone users. iOS: iPhone operating system.



The generalized linear mixed-effects model estimated whether time indicators, operating system, time since previous survey completion, or participant characteristics predicted the presence of a location data point (N=4435). The presence of a location data point was strongly predicted by the operating system and the part of the day (Table 1). The odds of a recorded location were the highest for Android phones (OR 21.91, 95% CI 19.53-26.87, referent: iPhone operating system) and during the afternoon (OR 1.18, 95% CI 1.18-1.20, referent: morning). The odds of a recorded location were lower in the weekends (OR 0.94, 95% CI 0.94-0.95, referent: weekdays) and if previous survey completion was longer ago (OR 0.95 per additional day, 95% CI 0.95-0.95) and marginally lower if a participant's time in study was longer (OR 0.998, 95% CI 0.9984-0.9985). Participant characteristics (eg, age, sex, self-reported weather sensitivity) did not predict the probability of location data.

Table 1. Results of the generalized linear mixed-effects model estimating the odds of having a recorded location (N=4435).

Variable	category	Odds ratio (95% CI)
Day of the week		
	Weekdays	Referent
	Weekend	0.94 (0.94-0.95)
Part of the day		
	Morning	Referent
	Afternoon	1.19 (1.18-1.20)
	Evening	1.11 (1.10-1.11)
	Night	0.37 (0.37-0.38)
Time in study	Per day	0.99 (0.99-1.00)
Operating system		
	iPhone operating system	Referent
	Android	22.91 (19.53-26.87)
Time since previous survey completion	Per day	0.96 (0.96-0.96)
Age	Per 10 years	1.00 (1.00-1.01)
Sex		
	Female	Referent
	Male	0.99 (0.80-1.22)
Weather sensitivity		
	Weak	Referent
	Strong	0.98 (0.84-1.15)

Discussion

Principal Findings

In our study, location data collected from participants' smartphones were missing for 63% of the intended hours (7.36 million/11.72 million). This percentage is higher than that reported in 5 other studies, reporting 26% [4], 28% [14], and 50% [10,15,21] of missing data. This difference may be due to the choices during the analysis: 3 studies excluded participants with the highest amounts of missing data and only investigated Android users, possibly resulting in an underestimation of the overall percentage of missing location data [4,14,21]. The other 2 studies sampled location continuously multiple times per hour for a few minutes, suggesting that our findings may not generalize to higher frequencies of location data collection [10,15].

Why Do Time Indicators and Operating System Predict Location-Data Completeness?

Missing data were predicted by part of the day, time since previous survey completion, and participants' operating system. Missing data at night might be caused by people being indoors where GPS signals are unavailable [11] or by their phones being switched off in airplane mode or out of battery [11,22]. Location data were most complete in the hour before and after survey completion, showing that apps are more likely to record the last known location upon restarting the app and the location on the clock hour after. In addition, we found a small but significant reduction in odds of a recorded location over time. Lower location-data completeness when people stay longer in a study is in line with the findings reported previously [22]. Less than 1% of iPhone users had complete location data. Other studies of smartphone data corroborate our finding of higher missing sensor data in iPhone users compared to Android users. iPhone's operating system refuses geolocation requests by apps more often compared to Android. Reasons for refusing geolocation requests are, for example, to reduce the phone's power consumption or to prioritize sensor data collection by other apps [10,15,23,24]. Of note, some studies have succeeded in obtaining higher coverage location data from iPhones compared to Android phones in spite of these operating system–specific differences [22,25]. This finding suggests that the research app used to collect data and the way this app interacts with the operating system may influence the amount of missing data. Experimental studies could further investigate this, as we cannot exclude the role of other differences between this study and our own study, such as the investigated population (eg, mean age 48 years in our study, but mean age 25 years in [22]) and sampling frequency (once an hour in our study; continuously for 1 minute every 10 minutes in [15,22]).

Implications: Consequences of Missing Data Are Context-Specific

Although missing location data reduce precision, they do not necessarily reduce a study's validity. For example, missing data

during the night may not be a problem for a study interested in identifying daytime behaviors from location data. In our study, we calculated daily average exposure to the weather based on the 24-hourly weather reports from participants' location [3]. For days with missing data, we imputed participant location. As UK weather stations are approximately 40 km apart, missing information on small relocations would not result in assigning participants to the wrong weather station. Furthermore, misclassification would only occur if the weather conditions at the "wrong" weather station were sufficiently different to change a participant's daily average exposure. Most previous studies investigating weather and pain measured participants' location only once and used daily weather reports, rather than hourly [26]. Compared to those studies, weather exposure in our study is less likely to be misclassified, even for participants with only 1 or 2 observed locations per day.

Participant age and sex did not predict missing location data, suggesting that data completeness is not associated with those 2 demographic factors. However, the difference in location-data completeness between iPhone and Android users could be a source of bias. Just-in-time interventions that depend on location data could be less safe and effective for iPhone users compared to Android users. On average, Android users have a lower socioeconomic status than iPhone users—a factor that is related to many health outcomes and may be associated with health disparities in underprivileged groups [27-29]. In observational studies, this difference could introduce selection bias. For example, exclusion of participants with incomplete data (complete case analysis) could lead to results that do not generalize to wealthier iPhone users.

Observational studies could impute missing location data based on participants' past behavior [30,31]. In that case, it is important to assess whether the imputation algorithm is also valid for iPhone users who may have fewer past data points available. If imputation is not feasible, researchers may want to consider using different devices to collect location data, such as a GPS tracker, which may be more suitable to answer certain research questions requiring complete location data for short periods of time [4,9]. Of note, although the imputation would mitigate some threats to internal validity due to selection bias, they do not address external validity. Study results may still not be generalizable to the wider population, especially not to underserved communities that tend to use health technologies less and may have fewer financial resources to purchase smartphones and pay for connection maintenance [29].

Improving Location-Data Completeness

At study design, researchers should optimize app settings and user instructions to improve location-data completeness. Our study showed that location was more often recorded around survey completion and around push notifications. Thus, encouraging participants to complete surveys and sending push notifications may improve location-data completeness as well as survey responses. As Android phone users have higher location-data completeness than iPhone users, restricting participation to Android users could improve location-data completeness. However, it could introduce important limitations to generalizability, given that many people have iPhones (market share 27% worldwide [32] and 54% in the United States [33]).

Conclusion

Missing hourly smartphone location data is common: in our study, 63% of hourly data points were missing. Missing data were more likely for iPhone users, during the night, on weekend days, and if participants had not recently used the app to complete a survey. Participant age and sex did not predict missing location data. Differences in location-data completeness between iPhone and Android users may impact the validity of observational or interventional studies. The predictors of missing data can help researchers at study design to optimize app settings and user instructions for higher location-data completeness. In addition, it may inform their assessment of context-specific consequences of missing location data.

Acknowledgments

We thank Sabine van der Veer for providing feedback on this manuscript. ALB is supported by a Medical Research Council Doctoral Training Partnership grant (MR/N013751/1). *Cloudy with a Chance of Pain* was supported by Versus Arthritis (grant reference 21225). This research was further supported by the Centre for Epidemiology Versus Arthritis (grant 21755) and the National Institute for Health Research Manchester Biomedical Research Centre.

Conflicts of Interest

WGD has received consultancy fees from AbbVie and Google, which is unrelated to this work.

References

1. Amor J, James C. Setting the scene: Mobile and wearable technology for managing healthcare and wellbeing. 2015 Presented at: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 25-29 Aug; Milan, Italy p. 7752-7755. [doi: 10.1109/EMBC.2015.7320189]
2. Schultz DM, Beukenhorst AL, Yimer BB, Cook L, Pisaniello H, House T, et al. Weather Patterns Associated With Pain In Chronic-Pain Sufferers. Bull Am Meteorol Soc 2020 2020 May 01;101(5):E555-E566. [doi: 10.1175/BAMS-D-19-0265.1]
3. Dixon WG, Beukenhorst AL, Yimer BB, Cook L, Gasparrini A, El-Hay T, et al. How the weather affects the pain of citizen scientists using a smartphone app. NPJ Digit Med 2019;2:105 [FREE Full text] [doi: 10.1038/s41746-019-0180-3] [Medline: 31667359]

4. Glasgow ML, Rudra CB, Yoo E, Demirbas M, Merriman J, Nayak P, et al. Using smartphones to collect time-activity data for long-term personal-level air pollution exposure assessment. J Expo Sci Environ Epidemiol 2016 Jun;26(4):356-364. [doi: 10.1038/jes.2014.78] [Medline: 25425137]
5. Lee EW, Bekalu MA, McCloud R, Vallone D, Arya M, Osgood N, et al. The Potential of Smartphone Apps in Informing Protobacco and Antitobacco Messaging Efforts Among Underserved Communities: Longitudinal Observational Study. J Med Internet Res 2020 Jul 07;22(7):e17451 [FREE Full text] [doi: 10.2196/17451] [Medline: 32673252]
6. Marsh A, Hirve S, Lele P, Chavan U, Bhattacharjee T, Nair H, et al. Validating a GPS-based approach to detect health facility visits against maternal response to prompted recall survey. J Glob Health 2020 Jun;10(1):010602 [FREE Full text] [doi: 10.7189/jogh.10.010602] [Medline: 32426124]
7. Dale LP, White L, Mitchell M, Faulkner G. Smartphone app uses loyalty point incentives and push notifications to encourage influenza vaccine uptake. Vaccine 2019 Jul 26;37(32):4594-4600. [doi: 10.1016/j.vaccine.2018.04.018] [Medline: 29699784]
8. Donaire-Gonzalez D, Valentín A, de Nazelle A, Ambros A, Carrasco-Turigas G, Seto E, et al. Benefits of Mobile Phone Technology for Personal Environmental Monitoring. JMIR Mhealth Uhealth 2016 Nov 10;4(4):e126 [FREE Full text] [doi: 10.2196/mhealth.5771] [Medline: 27833069]
9. Goodspeed R, Yan X, Hardy J, Vydiswaran VV, Berrocal VJ, Clarke P, et al. Comparing the Data Quality of Global Positioning System Devices and Mobile Phones for Assessing Relationships Between Place, Mobility, and Health: Field Study. JMIR Mhealth Uhealth 2018 Aug 13;6(8):e168 [FREE Full text] [doi: 10.2196/mhealth.9771] [Medline: 30104185]
10. Boonstra TW, Nicholas J, Wong QJ, Shaw F, Townsend S, Christensen H. Using Mobile Phone Sensor Technology for Mental Health Research: Integrated Analysis to Identify Hidden Challenges and Potential Solutions. J Med Internet Res 2018 Jul 30;20(7):e10131 [FREE Full text] [doi: 10.2196/10131] [Medline: 30061092]
11. Vhaduri S, Poellabauer C, Striegel A, Lizardo O, Hachen D. Discovering places of interest using sensor data from smartphones and wearables. In: IEEE Xplore.: IEEE; 2017 Presented at: SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI; June 28; San Francisco. [doi: 10.1109/UIC-ATC.2017.8397495]
12. Krenn PJ, Titze S, Oja P, Jones A, Ogilvie D. Use of global positioning systems to study physical activity and the environment: a systematic review. Am J Prev Med 2011 Nov;41(5):508-515 [FREE Full text] [doi: 10.1016/j.amepre.2011.06.046] [Medline: 22011423]
13. Marsh A, Hirve S, Lele P, Chavan U, Bhattacharjee T, Nair H, et al. Validating a GPS-based approach to detect health facility visits against maternal response to prompted recall survey. J Glob Health 2020 Jun;10(1):010602 [FREE Full text] [doi: 10.7189/jogh.10.010602] [Medline: 32426124]
14. Do TMT, Dousse O, Miettinen M, Gatica-Perez D. A probabilistic kernel method for human mobility prediction with smartphones. Pervasive and Mobile Computing 2015 Jul;20:13-28. [doi: 10.1016/j.pmcj.2014.09.001]
15. Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela JP. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. NPJ Digit Med 2018;1:15 [FREE Full text] [doi: 10.1038/s41746-018-0022-8] [Medline: 31304300]
16. Gelman A. Analysis of variance—why it is more important than ever. Ann. Statist 2005 Feb 1;33(1):1-53. [doi: 10.1214/009053604000001048]
17. Vittinghoff E, Glidden D, Shiboski SM. Repeated measures and longitudinal data analysis. In: Regression Methods in Biostatistics. Boston: Springer; 2012:261-308.
18. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, et al. Package 'lme4'. Linear mixed-effects models using S4 classes. R Package Version 2021. 2021. URL: https://cran.r-project.org/web/packages/lme4/index.html [accessed 2021-12-10]
19. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press; 2006.
20. Knowles J, Frederick C. merTools: tools for analyzing mixed effect regression models. R Package version 050 2019. URL: https://cran.r-project.org/web/packages/merTools/index.html [accessed 2021-10-12]
21. Bähr S, Haas G, Keusch F, Kreuter F, Trappmann M. Missing Data and Other Measurement Quality Issues in Mobile Geolocation Sensor Data. Social Science Computer Review 2020 Aug 06:089443932094411. [doi: 10.1177/0894439320944118]
22. Kiang MV, Chen J, Krieger N, Buckee C, Alexander M, Baker J, et al. Sociodemographic Characteristics of Missing Data in Digital Phenotyping. Scientific Reports 2021 Jul 29;11(1):15408 (forthcoming). [doi: 10.1101/2020.12.29.20249002]
23. Beukenhorst AL, Schultz DM, McBeth J, Lakshminarayana R, Sergeant JC, Dixon WG. Using Smartphones for Research Outside Clinical Settings: How Operating Systems, App Developers, and Users Determine Geolocation Data Quality in mHealth Studies. Stud Health Technol Inform 2017;245:10-14. [Medline: 29295042]
24. Torous J, Kiang MV, Lorme J, Onnela JP. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
25. Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela JP. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. NPJ Digit Med 2018;1:15 [FREE Full text] [doi: 10.1038/s41746-018-0022-8] [Medline: 31304300]

26. Beukenhorst AL, Schultz DM, McBeth J, Sergeant J, Dixon W. Are weather conditions associated with chronic musculoskeletal pain? Review of results and methodologies. Pain 2020 Apr;161(4):668-683. [doi: 10.1097/j.pain.0000000000001776] [Medline: 32195783]
27. Dorsey ER, Chan YF, McConnell MV, Shaw SY, Trister AD, Friend SH. The Use of Smartphones for Health Research. Acad Med 2017 Feb;92(2):157-160. [doi: 10.1097/ACM.0000000000001205] [Medline: 27119325]
28. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology 2021 Jan;46(1):45-54. [doi: 10.1038/s41386-020-0771-3] [Medline: 32679583]
29. Lee E, Viswanath K. Big Data in Context: Addressing the Twin Perils of Data Absenteeism and Chauvinism in the Context of Health Disparities Research. J Med Internet Res 2020 Jan 07;22(1):e16377 [FREE Full text] [doi: 10.2196/16377] [Medline: 31909724]
30. Barnett I, Onnela JP. Inferring mobility measures from GPS traces with missing data. Biostatistics 2020 Apr 01;21(2):e98-e112 [FREE Full text] [doi: 10.1093/biostatistics/kxy059] [Medline: 30371736]
31. Song C, Qu Z, Blumm N, Barabási AL. Limits of predictability in human mobility. Science 2010 Feb 19;327(5968):1018-1021. [doi: 10.1126/science.1177170] [Medline: 20167789]
32. O'Dea S. Mobile operating systems: market share worldwide from January 2012 to October 2021. Statista. 2012. URL: https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/ [accessed 2021-05-05]
33. O'Dea S. Market share of mobile operating systems in North America from January 2018 to June 2021. Statista. URL: https://www.statista.com/statistics/1045192/share-of-mobile-operating-systems-in-north-america-by-month/ [accessed 2021-05-05]

Abbreviations

OR: odds ratio

Edited by L Buis; submitted 16.03.21; peer-reviewed by I Barnett, E Lee, H Mehdizadeh; comments to author 22.05.21; revised version received 11.06.21; accepted 27.08.21; published 16.11.21.

Please cite as:
Beukenhorst AL, Sergeant JC, Schultz DM, McBeth J, Yimer BB, Dixon WG
Understanding the Predictors of Missing Location Data to Inform Smartphone Study Design: Observational Study
JMIR Mhealth Uhealth 2021;9(11):e28857
URL: https://mhealth.jmir.org/2021/11/e28857
doi:10.2196/28857
PMID:34783661

Original Paper

App Designs and Interactive Features to Increase mHealth Adoption: User Expectation Survey and Experiment

Allison J Lazard[1,2], PhD; J Scott Babwah Brennen[3], PhD; Stephanie P Belina[1], BS

[1]Hussman School of Journalism and Media, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[3]Center on Science and Technology Policy, Duke University, Durham, NC, United States

Corresponding Author:
Allison J Lazard, PhD
Hussman School of Journalism and Media
University of North Carolina at Chapel Hill
384 Carroll Hall
Chapel Hill, NC, 27599
United States
Phone: 1 3173317807
Email: lazard@unc.edu

Abstract

Background: Despite the ubiquity of smartphones, there is little guidance for how to design mobile health apps to increase use. Specifically, knowing what features users expect, grab their attention, encourage use (via predicted use or through positive app evaluations), and signal beneficial action possibilities can guide and focus app development efforts.

Objective: We investigated what features users expect and how the design (prototypicality) impacts app adoption.

Methods: In a web-based survey, we elicited expectations, including presence and placement, for 12 app features. Thereafter, participants (n=462) viewed 2 health apps (high prototypicality similar to top downloaded apps vs low prototypicality similar to research interventions) and reported willingness to download, attention, and predicted use of app features. Participants rated both apps (high and low) for aesthetics, ease of use, usefulness, perceived affordances, and intentions to use.

Results: Most participants (425/462, 92%) expected features for navigation or personal settings (eg, menu) in specific regions (eg, top corners). Features with summary graphs or statics were also expected by many (395-396 of 462, 86%), with a center placement expectation. A feature to “share with friends” was least expected among participants (203/462, 44%). Features fell into 4 unique categories based on attention and predicted use, including *essential features* with high (>50% or >231 of 462) predicted use and attention (eg, calorie trackers), *flashy features* with high attention but lower predicted use (eg, links to specific diets), *functional features* with modest attention and low use (eg, settings), and *mundane features* with low attention and use (eg, discover tabs). When given a choice, 347 of 462 (75%) participants would download the high-prototypicality app. High prototypicality apps (vs low) led to greater aesthetics, ease of use, usefulness, and intentions, (for all, $P<.001$). Participants thought that high prototypicality apps had more perceived affordances.

Conclusions: Intervention designs that fail to meet a threshold of mHealth expectations will be dismissed as less usable or beneficial. Individuals who download health apps have shared expectations for features that should be there, as well as where these features should appear. Meeting these expectations can improve app evaluations and encourage use. Our typology should guide presence and placement of expected app features to signal value and increase use to impact preventive health behaviors. Features that will likely be used and are attention-worthy—essential, flashy, and functional—should be prioritized during app development.

(JMIR Mhealth Uhealth 2021;9(11):e29815) doi:10.2196/29815

KEYWORDS

smartphone; interactive design; mobile apps; preventive health; mental models; prototypicality; attention; affordances

Introduction

Background

With the rapid increase in the use of mobile technologies and smartphones for health information [1,2], mobile apps present one possible solution for communicating preventive health information to the public [3-5]. Over the past decade, hundreds of health mobile apps have been produced—many designed by public health interventionists and researchers for cancer and other chronic disease prevention by encouraging healthy eating and physical activity [6-8]. While it remains unclear how successful these apps have been in reducing the incidence of cancer or improving health outcomes for other chronic diseases, there is a call for an increase in the accountability, reliability, and standardizations of evidence-based health apps developed by the research community [8-10].

Despite the potential of mobile health (mHealth) apps for communicating up-to-date, evidence-based prevention information and helping users maintain or implement healthy habits, there is very little guidance on how these intervention apps should be designed to ensure adoption [11]. Designing apps so they are appealing and used is a critical first step for apps to have an impact [12]. Visual and interactive design influences initial user evaluations, which are made within milliseconds, and serve as gateways for subsequent user engagement (eg, use) of apps as mHealth interventions [13-15]. Ignoring design can detrimentally impact the communication of evidence-based science to health consumers and undercut the effectiveness of mHealth interventions; yet, few mHealth interventions mirror the look and function of popular, industry-developed apps. Thus, our study objective was to explore app features expectations and examine how meeting expectations with high- (vs low-) prototypicality apps may influence predictors of app adoption.

How apps are designed (visual display) and the features they include (interactivity) can influence users' experience of and willingness to engage with apps. Individuals use salient cues that match their expectations, or mental models, to evaluate web-based information [16,17]. These expectations are met (or not) by the level of prototypicality or the degree to which an app resembles others in its comparative group [17,18]. Based on included design cues, in the form of interactive features, apps can range from having high prototypically (looks like others and meets expectations well) to low prototypicality (does not resemble others nor meet expectations) [19]. Users are often quicker and more willing to attend to apps that have high prototypicality—when designs align with one's mental models for how an app should look and function [19-21]. Indeed, users look for and pay attention to expected, salient features as guides to orient themselves to novel apps and platforms [21]. When these expected features are present, they increase familiarity and potential use of the app [19-21]; however, little is known on how attention for specific features translates into individual feature use versus overall app use.

The perceived affordances, or perceived action possibilities (eg, learn health tips), that users sense from app features also directly impact a user's experience and likelihood to engage with a design [22-24]. Specifically for mediated communication, including apps, design communicates what the viewer can do or gain from the use of an app, through interface symbols. Thus, not only must mHealth interventions have evidence-based content to drive use, but also apps must incorporate an evidence-based design to appeal to and engage audiences.

Design features influence the appeal or perceived aesthetics of the app and the likelihood for use [25,26]. To be effective, health apps must surely be used. It is necessary to understand how objective design features (the visible objects or designs in an app) influence subjective evaluations for initial appeal on the basis of theories of aesthetics [27-29] and antecedents for technology adoption in the Technology Acceptance Model (TAM); that is, perceived ease of use, perceived usefulness, and intentions to use [30,31]. Aesthetics, including facets for how information is organized and displayed, function as a precursor to perceptions for technology acceptance [28,31]. Accounting for users' expectations of features and placements within apps will shed light on how prototypicality impacts evaluations critical for future adoption.

Utility also drives evaluation of an app's usefulness and potential adoption, according to Nielsen et al's [32] well-established usability study. Utility refers to the inclusion of necessary features—whether an app provides the elements an individual needs or wants. When utility is paired with usability—when features are perceived as easy (perceived ease of use) and pleasant (aesthetics) to use—individuals are encouraged to engage or interact. In other words, interactivity is dependent on a user's willingness to engage with specific design features, if present (utility) and function properly (usability). In our work, we focus on the former—how app features that are needed (utility) or expected (prototypical) are the gateway to potential adoption.

Goal of This Study

In sum, engagement with and use of an app is driven by initial impressions and perceptions of what the app can do for the user. Top-rated industry-developed apps often incorporate a user-focused sleekness and are feature loaded; in comparison, pared-down mHealth interventions—despite the inclusion of theory-based content—may not appeal to audiences who need them [33]. When resources are not abundant, health researchers and interventionists need evidence-based guidance for design investments. Thus, we explored app expectations for the presence and placement of potential features, how these features garner attention and predict use, and how high-prototypicality apps (vs low-prototypicality apps) may influence app adoption through app choice and predictors of use. We asked the following research questions: What features do people expect and where do they expect these features to be placed (RQ1)? What specific features are associated with attention and predicted use of the app features (RQ2)? Last, we also examined whether high prototypicality, resembling that of top downloaded apps (vs low-prototypicality apps, resembling research intervention apps) would increase app choice (H1), aesthetics (H2), perceived ease of use (H3), perceived usefulness (H4), intentions to use the app (H5), and perceived affordances or action possibilities with the app (H6).

Methods

Overview

To explore app features expectations and examine how meeting expectations with high-prototypicality apps (vs low-prototypicality apps) may influence predictors of app adoption, we conducted a web-based survey with an embedded within-subjects experiment. Participants first responded to survey items about expectations for specific app features to answer RQ1-2 and an app choice (preview of apps with high vs low prototypicality) to address H1. Participants were then asked to rate their perceptions of the app overall, with the exposure order of condition (high vs low) randomized, to address H2-6.

Participant Recruitment

Using G*Power, our a priori power analysis indicated a required sample of at least 450 participants to detect a small-to-medium effect (Cohen *f*=0.14) for within-subjects comparison of the high and low prototypicality apps. Participants (n=462) were recruited from Amazon's Mechanical Turk (MTurk), a web-based crowdsourcing platform often used for social science research [34-36], through a link open to individuals over the age of 18 years. Participants were eligible if they were aged 18 years or older, resided in the United States, and had a task approval rate of 85% or higher on the MTurk platform, which indicates valid participation or completion of previous tasks. Participants received US $3 as compensation for their time (approximately 15 minutes). The institutional review board of University of North Carolina approved this study.

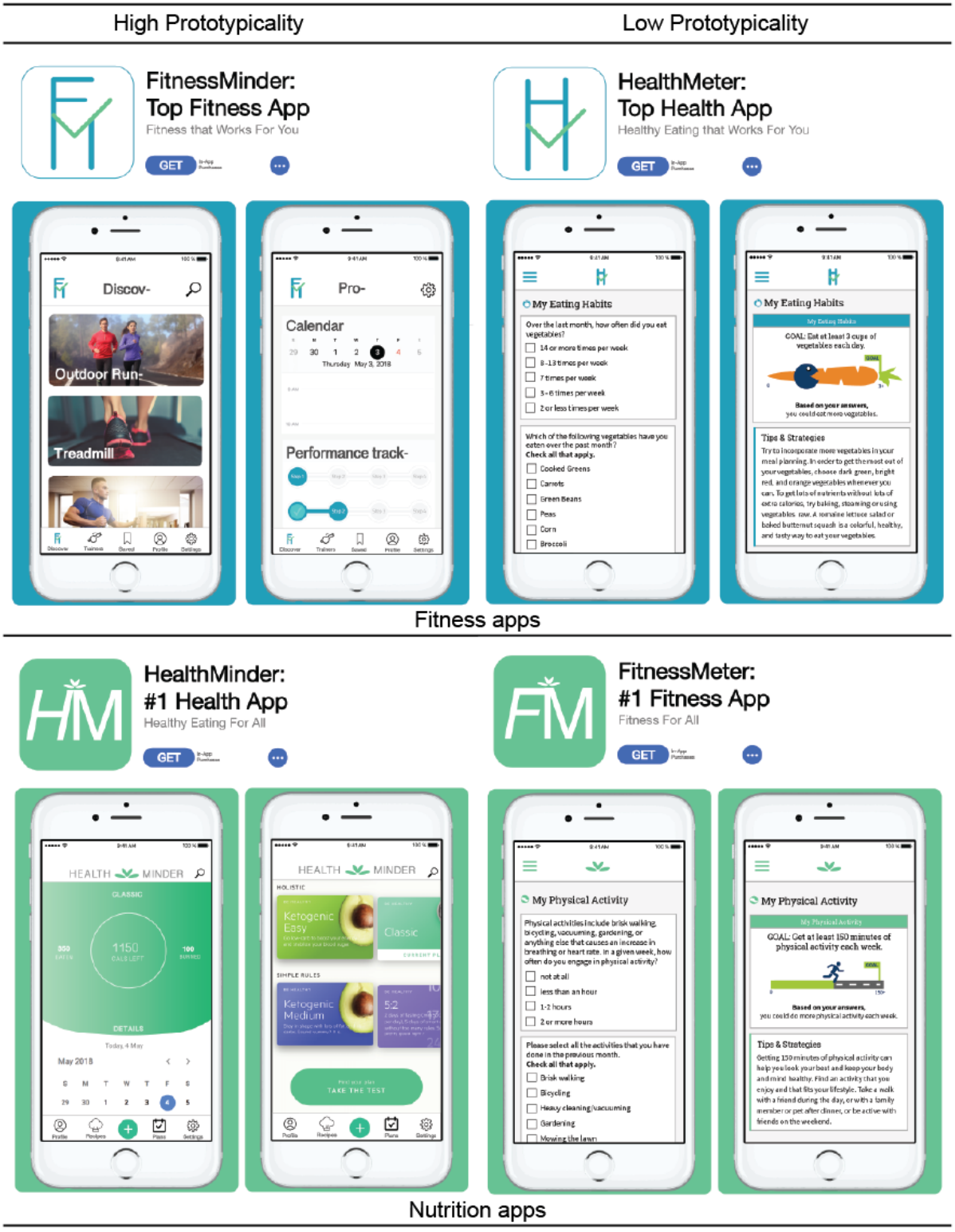
Procedure

Following consent, participants selected features (from a list) they would expect to find in a health app. For all expected features, participants were shown an outline of a smartphone and asked where that feature would be located in a typical health app. Participants were then randomly assigned to 1 of the 2 app types for the remainder of the study: fitness apps or nutrition apps. Participants selected the app they would most like to download from 2 previews (prototypical: high vs low). On subsequent pages, participants indicated what features grabbed their attention and what features they predicted they would use (predicted use) on their preferred app. Participants were shown the app previews again (one at a time, in a random order) and asked closed-ended items for perceived aesthetics, ease of use, usefulness, intentions to use the app in the future, and perceived affordances. Lastly, demographic, health, and health app information were collected from all participants. Closed-ended items and response options are described below (see Measures) and provided in Multimedia Appendix 1.

App Stimuli

To assess the impact of prototypicality on app perceptions, app previews were created for four fictitious brands: 2 fitness and 2 nutrition health apps (Figure 1). We designed previews for each app as they would appear if searched for in a mobile app store, including the app icon, brand name, and 2 preview screens of the app. High-prototypicality apps were developed on the basis of structure and content from top rated apps (Aaptiv, Lifesum) in the Health & Fitness section of the App Store. Low-prototypicality apps were designed to mirror the mobile interface of an interactive intervention (Carolina Health Assessment and Research Tool) for data collection and tailored feedback for preventive health behaviors [37].

Figure 1. App preview stimuli.



Measures

Feature Selection and Placement

Participants selected features from a list they "would expect to find in a health app." The list was generated from structured interviews about fitness tracker apps [38] and included 12 features: menu, search option, settings option, logo, log/input data option, share with friend option, summary statistics, summary graph/chart, calendar, page title, login, and user profile. For each expected (ie, selected) feature, respondents were shown a smartphone screen divided into a grid of 60 distinct clickable hot spot regions. Respondents selected as many regions of each screen as necessary for expected placement.

App Choice

Participants were instructed to "select the app you would most likely download." The 2 response options were the low

prototypical app and the high prototypical app, for their randomly assigned app type (physical activity or nutrition).

Feature Attention and Predicted Use

To identify features that attracted participants' attention and predicted use, participants were shown the app preview they selected during app choice. Participants were asked, "What elements in the app caught your attention?" and instructed to "select all elements that grabbed your attention within the app preview." On the following page of the questionnaire the app preview was shown again; participants were asked, "What elements in the app do you think you would use?" and selected the elements in the preview. As performed in previous studies [39,40], a priori hot spots were constructed around each app feature (Multimedia Appendix 1). Hot spots were not visible until participants selected the feature and then the feature was highlighted.

Perceived Aesthetics

The validated Visual Aesthetics of Website Inventory (VisAWI) assessed 4 facets of aesthetics with 18 items for simplicity, "The layout appears well structured"; diversity, "The layout appears dynamic"; colorfulness, "The colors are appealing"; and craftsmanship, "The app is designed with care" [28]. Response options ranged from "strongly disagree" (coded as 1) to "strongly agree" (5). Responses were averaged for each facet (α=.76-.90).

Perceived Ease of Use

Participants' perceived ease of use, or belief that using the technology would not be difficult, were assessed with 3 adapted Likert-type items [30]: "The app was clear and understandable," "Getting the app to function does not require much mental effort," and "I find the app to be easy to use." Response options ranged from "strongly disagree" (coded as 1) to "strongly agree" (5). Responses were averaged (α=.84-.87).

Perceived Usefulness

The degree to which one believes that the technology will enhance their life was assessed with 3 adapted Likert-type items [30]: "Using the app would improve my health," "Using the app would make me more likely to meet my health goals," and "I would find the app useful for achieving my health goals." Response options ranged from "strongly disagree" (coded as 1) to "strongly agree" (5). Responses were averaged (=.85-.88).

Intentions to Use

Intentions or plans to use the app "if the app were available" were assessed with 2 Likert-type items [30]. Participants rated their agreement to statements that they "intend" and "predict" they would use the app next month with response options that ranged from "strongly disagree" (coded as 1) to "strongly agree" (5). Responses were averaged (r=0.88-0.93).

Perceived Affordances

Participants reported perceived action possibilities from the app with the item, "This app would allow me to…" Response options included a list of 13 dichotomous items generated from evidence-based behavior change techniques and reasons for eHealth adoption, such as "set health goals," "track my progress," "earn rewards," and "share my health data with friends" [41,42].

Participant Characteristics

Demographic items assessed age, gender, race, ethnicity, and education. Additionally, we asked about one's health and mental health status with the item: "in general, would you say your [mental] health is…" Response options ranged from "very poor" (coded as 1) to "very good" (5). We also asked whether participants "use a health app" with a "yes"/"no" response option.

Data Analyses

We used n (%) values to describe app feature expectations, placement, app choices, attention, predicted use, and perceived affordances. Frequencies for attention vs predicted use and for perceived affordances of the high- vs low-prototypicality apps were compared with McNemar chi-square tests. Prior to this analysis for direct effects of prototypicality, a multivariate analysis of variance (MANOVA) was used to determine if there are any significant differences in perceptions among the app types (fitness and nutrition) across aesthetics and TAM outcomes. No differences were observed for high prototypicality (aesthetics outcomes: Wilks λ=0.98; $F_{4,454}$=1.08; P=.10; TAM outcomes: Wilks λ=0.99; $F_{3,454}$=1.08; P=.36) or low prototypicality (aesthetics outcomes: Wilks λ=0.99; $F_{3,452}$=0.68; P=.61; TAM outcomes: Wilks λ=1.00; $F_{3,454}$=0.10; P=.96), so data within conditions (high vs low prototypicality) were combined for analyses. Two repeated measure (RM) MANOVAs and analyses of variance (ANOVAs) were then conducted with high vs low prototypicality as the predictor; 1 for aesthetic outcomes (simplicity, diversity, colorfulness, and craftsmanship) and 1 for technology acceptance outcomes (perceived ease of use, usefulness, and intentions to use).

Results

Participants

Participants (n=462) were aged 18 to 70 years (mean age 35.03 years, SD 10.02 years) and half of them were female (50%, 232/462). Participants identified as White (78%, 358/462), African American (13%, 58/462), Asian (8%, 35/462), or multiracial/other; additionally, 48 of 462 participants (10%) reported their ethnicity as Hispanic. Education levels included high school to some college (33%, 153/462), associate degree (13%, 60/462), bachelor's degree (43%, 197/462), master's degree (10%, 45/462), and doctoral or professional degree (2%, 7/462). Most participants reported their health as good (48%, 220/462) or very good (17%, 78/462), although some did report that their health was fair (30%, 138/462), poor (4%, 20/462), or very poor (1%, 4/462). Over half of the participants (53%, 248/462) reported currently using health apps.

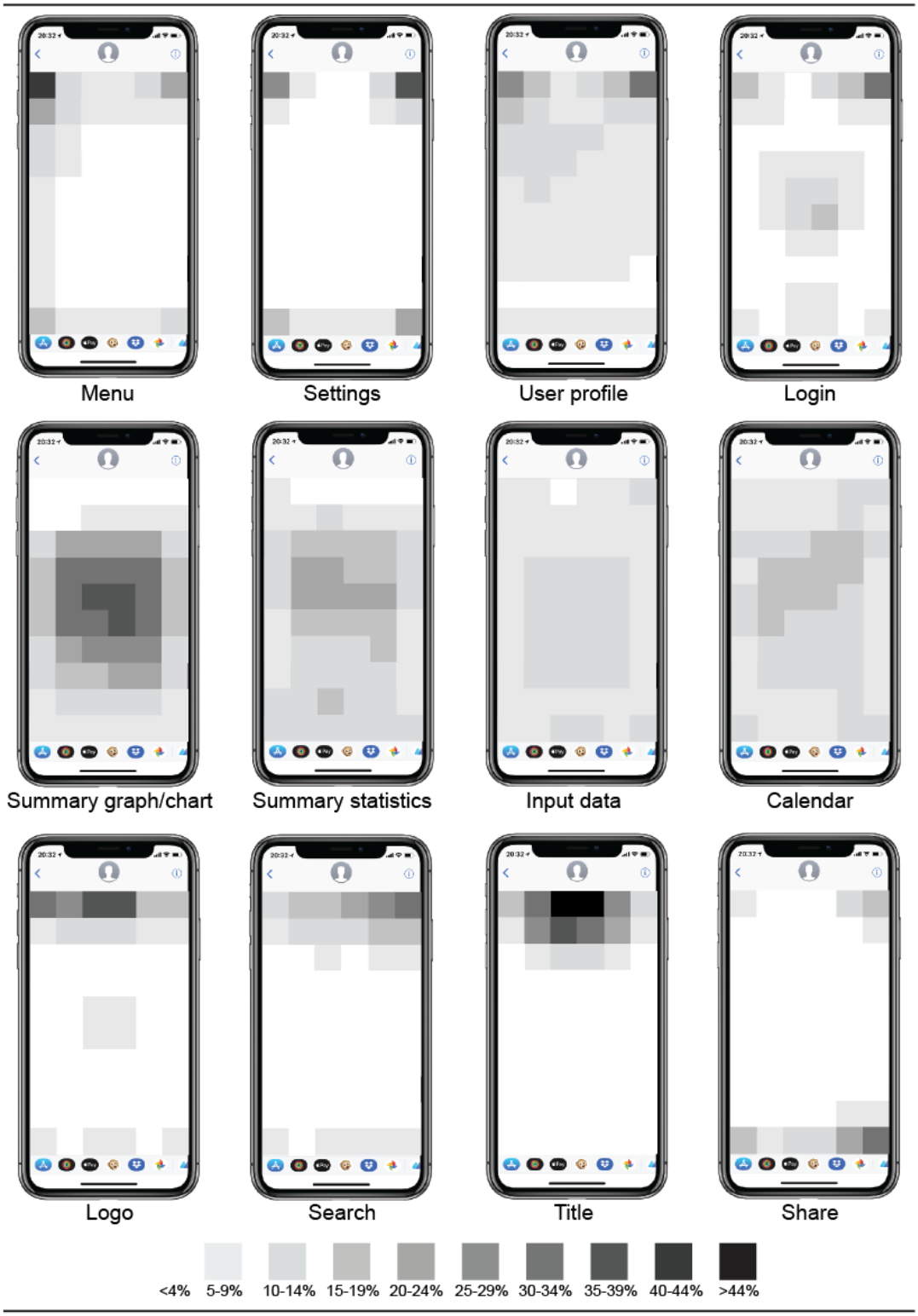
App Feature Selection and Placement

Each of the 12 features was selected by at least 44% (203/462) of participants (RQ1). The majority of participants (92%, 425/462) selected a menu, settings options, and user profile; notably, these features (ie, menu, settings option, and user profile) were selected an equal number of times but not by the

same respondents. Additional features were expected, including the following: login (88%, 406/462), summary graph/chart (86%, 396/462), summary statistics (86%, 395/462), input data feature (80%, 368/462), calendar (77%, 354/462), logo, (77%, 357/462), search (69%, 321/462), page title (62%, 286/462), and an option to "share with friends" (44%, 203/462).

Most features were expected in similar locations (Figure 2) among participants who had expected features (n=425). Menus were consistently expected to be in the top-left, while search and login options are placed in the top-right corner. Other features—title, logo, profile, and settings—were expected along the top, in the center, or either side. Sharing capability was expected to appear in the bottom-right of the app, although expectations of where to log input data were more diffuse. Users expect summary statistics, graphs, and calendars to be shown across the center of the app.

Figure 2. Expected app feature location (n=425).



Attention and Predicted Use of App Features

Respondents selected features of their preferred app, which caught their attention and they would use (Table 1 and Multimedia Appendix 1). Attention and predicted use patterns of the high-prototypicality apps indicate 4 distinct categories of mHealth app features. *Mundane features* are those that have similar low attention and predicted use values. In the fitness app, the footer menu options "Discover" and "Saved" represent mundane features. *Functional features* have higher predicted use than attention, but predicted use remains low (<50%, <231/462) among participants, such as the settings icon in both apps. *Flashy features* are elements identified as attention-capturing by most participants (>50%, >231/462), and attention is significantly higher than the predicted use. In the nutrition app, large photo-based links for the "Ketogenic Easy" and "Ketogenic Medium" diets represent flashy features. *Essential features* are elements that most participants (>50%, >231/462) thought they would use, and where predicted use is higher than or similar to attention, as with the "Calorie Tracker" in the nutrition app. Not included in these 4 categories are elements that have higher attention than predicted use, but the attention remains low (<50%, <231/462); the only features with these characteristics were logos and app titles, as well as 2 features partially obscured in the design.

Table 1. Reported attention and predicted use of app features (n=462).

App	Feature	Attention, n (%)	Predicted use, n (%)	Chi-square (*df*)	*P* value
Mundane					
Fitness	Footer menu option "Discover"	48 (29)	52 (32)	0.20 (1)	.66
Fitness	Footer menu option "Saved"	41 (25)	42 (26)	0.00 (1)	>.99
Nutrition	Footer menu option "Plus"	23 (13)	28 (15)	0.46 (1)	.50
Functional					
Fitness	Footer menu option "Settings"	49 (30)	69 (42)	7.22 (1)	.007
Nutrition	Search Icon	21 (12)	38 (21)	5.95 (1)	.02
Nutrition	Footer menu option "Profile"	25 (14)	58 (32)	19.32 (1)	<.001
Flashy					
Fitness	Activity 1 "Outdoor Running"	109 (66)	63 (38)	32.66 (1)	<.001
Fitness	Acitivty 2 "Treadmill"	104 (63)	53 (32)	38.46 (1)	<.001
Nutrition	Ketogenic Easy feature	109 (60)	71 (39)	20.74 (1)	<.001
Essential					
Fitness	Performance Tracker feature	142 (86)	157 (95)	N/A[a]	.003
Nutrition	Calorie Tracker feature	156 (86)	160 (88)	N/A	.54
Nutrition	Calendar feature	88 (48)	114 (63)	11.57 (1)	.001

[a]N/A: chi-square values are not applicable if fewer than 25 discordant pairs; binominal distributions are used for exact 2-tailed significance in these comparisons.

Effects of Prototypicality on App Choice, Aesthetics, and Technology Acceptance

When asked to choose between the high-prototypicality app and one designed to look more like a typical health intervention (low prototypicality), 347 of 462 (75%) participants indicated they would download the high-prototypicality app (H1).

Prototypicality had a significant main effect on all facets of aesthetics and technology acceptance outcomes (Table 2). High-prototypicality apps (vs low-prototypicality apps) had significantly higher ratings of aesthetics for simplicity ($F_{1,455}$=291; P<.001), diversity ($F_{1,455}$=578; P<.001), colorfulness ($F_{1,455}$=295; P<.001), and craftsmanship ($F_{1,455}$=462; P<.001). Similarly, the high-prototypicality app was rated higher than the low prototypicality app for perceived ease of use ($F_{1,455}$=84; P<.001), usefulness ($F_{1,455}$=116, P<.001), and intentions to use the app ($F_{1,455}$=170; P<.001). H2-5 were supported.

Table 2. Main effects of prototypicality on aesthetics and technology acceptance (n=456).

Attributes	High prototypicality, mean (SD)	Low prototypicality, mean (SD)	*F* test (*df*)	*P* value
Simplicity	4.26 (0.74)	3.19 (1.00)	291 (*1,455*)	<.001
Diversity	4.10 (0.74)	2.48 (1.09)	578 (*1,455*)	<.001
Colorfulness	4.38 (0.74)	3.41 (0.94)	295 (*1,455*)	<.001
Craftsmanship	4.25 (0.75)	2.83 (1.07)	462 (*1,455*)	<.001
Perceived ease of use	4.26 (0.75)	3.74 (0.97)	84 (*1,455*)	<.001
Perceived usefulness	4.08 (0.74)	3.58 (0.91)	116 (*1,455*)	<.001
Intentions to use	3.83 (1.00)	2.95 (1.28)	170 (*1,455*)	<.001

Impact of Prototypicality on Perceived Affordances

Participants reported that the app would allow them to carry out various actions in both the high- and low-prototypicality design (Table 3). Almost all perceived affordances had significantly higher endorsement for the high-prototypicality (vs low-prototypicality) apps (P<.01), partially supporting H6; to "learn health tips" was the only affordance endorsed similarly in both conditions. The most highly endorsed affordances (>60% across conditions or >277/462) were the following: "track my progress" (high: 93%, 430/462; low: 70%, 325/462), "set health goals" (high: 88%, 405/462; low: 73%, 339/462), "improve my health" (high: 74%, 342/462; low: 63%, 293/462), "learn health tips" (high: 73%, 336/462; low: 76%, 353/462), and "give me more information about my health" (high: 70%, 325/462; low: 63%, 292/462).

Table 3. Frequencies and McNemar chi-square differences for perceived affordances (n=462).

Affordances	High prototypicality, n (%)	Low prototypicality, n (%)	Chi-square (*df*)	*P* value
Track my progress	430 (93.1)	325 (70.3)	79.70 (*1*)	<.001
Set health goals	405 (87.7)	339 (73.4)	30.75 (*1*)	<.001
Improve my health	342 (74.0)	293 (63.4)	20.15 (*1*)	<.001
Learn health tips	336 (72.7)	353 (76.4)	2.59 (*1*)	.11
Give me more information about my health	325 (70.3)	292 (63.2)	6.86 (*1*)	.009
Create new health habits	310 (67.1)	265 (57.4)	11.06 (*1*)	.001
Increase my control over my health	323 (69.9)	239 (51.7)	46.86 (*1*)	<.001
Make meeting my health goals easier	292 (63.2)	195 (42.2)	51.28 (*1*)	<.001
Have fun with technology	256 (55.4)	135 (29.2)	79.57 (*1*)	<.001
Interact with others	120 (26.0)	47 (10.2)	56.63 (*1*)	<.001
Share my health data with friends	100 (21.6)	47 (10.2)	35.12 (*1*)	<.001
Share my health data with a healthcare provider	74 (16.0)	50 (10.8)	11.50 (*1*)	.001
Earn rewards	57 (12.3)	34 (7.4)	10.30 (*1*)	.001

Discussion

Principal Findings

For mHealth to have an impact on reducing risk for chronic disease, intervention apps must be designed to effectively reach wide audiences to promote preventive health behaviors. Identifying the impact of prototypicality—the extent to which apps meet expectations—on app reception and adoption is a critical step in mHealth intervention research. Designs that match users' perceptions of organization and content evoke prototypicality and can influence intentions to use web-based tools, including health resources [21,31,43]. Our study on prototypicality serves as an antecedent to positive app reception and technology acceptance in preventive health apps. We also found designs that contradict what users typically expect from apps (eg, low prototypicality), leading to a suboptimal first impression and diminishing users' expectations [19].

It is likely that the actual use of multiple apps influences preventive behavior [44]; thus, identifying key features, or classes of features, to increase orientation and facilitate ease of use and usefulness are needed to guide intervention development. Our findings for user attention and predicted use of features point to 4 distinct types of mHealth features that should be considered when developing mHealth. Of these, 3 categories serve as useful features of mHealth: driving attention, perceived use, or both.

Functional features have higher predicted use than attention, and a majority "would expect to find" these sorts of features in a health app. To meet expectations, salient functional features such as search options, settings, and menus should be included,

in their expected corner placement. Even if these features do not draw attention as much as others, users still expect to see them in mobile apps, and meeting baseline expectations can reduce time and cognitive demand for initial orientation and web-based information processing [21]. Arguably, these functional features constitute a sort of prototypical milieu or background environment for mHealth apps to likely help users orient themselves within new and unfamiliar apps.

Flashy features garner significantly more attention from users; these attention-capturing features may be most influential for positive initial impressions. Flashy features often incorporated photographs or novel design elements, which have been shown to increase attention and appeal [43,45]. Beyond meeting expectations, flashy features represent the unique category that should be treated *differently* in designs: using visuals to highlight salient benefits and perceived affordances.

Essential features—including those selected by most users as features that they predict to use and garner their attention—are also important components of mHealth designs. It is important to note, however, that the essential features seen in this study are all familiar: calendar, calorie counter, and performance tracker. Even though some designers may assume that features as basic as a calendar are not worth the time and effort to include, respondents strongly indicated that these features remain important components of mHealth apps.

Our findings also highlight a distinct category that can be skipped or given little attention in development: mundane features. Mundane features, such as app title and tabs for discovering or saving, elicited little attention and predicted use and are a good indication not to waste precious resources on these elements.

Potential mHealth users had consistent expectations for some features by region (eg, middle or top corner), but not necessarily a specific location. Essential features, such as a calendar, were expected to be shown across the center of the app. Other features, such as function features including *search* and *settings*, had more narrow placement expectations. Understanding these location expectations is critical to ensure that feature placement matches individual models [21].

Higher prototypicality led to higher ratings for aesthetics, perceived ease of use, usefulness, and intentions to use apps. Individuals also expect greater function, possibilities, and valuable outcomes from apps with higher prototypicality. Low prototypicality led to lower rankings for aesthetics, perceived ease of use, and perceived usefulness. Additionally, low prototypicality runs the risk of users initially dismissing the app. Negative product evaluations—where expectations are not met—can also lower satisfaction with product interaction [46].

Limitations

This study is limited to the specific health apps manipulated herein; these apps do not represent all available mHealth strategies. Although we evaluated placement, attention, and predicted use, we could have reviewed more features within apps. Our findings are also limited to a convenience sample of participants of a web-based panel. It is possible that our participants have more digital literacy or skills than the general population or diverse subgroups.

Future Work

Future studies should consider assessing actual use after download, instead of solely predicted use. Replication with more diverse audiences, varied app designs, and expanded methodological approaches are needed to generalize our findings. Notably, future research should account for additional personal characteristics, such as health literacy or the ability to obtain, process, and understand health information [47], to examine how these skills affect both first impressions for app adoption and actual use to determine the effectiveness of health apps.

Conclusions

Mobile apps can communicate critical health information for preventive health behaviors through readily available and consumer-friendly tools. Apps that are thoughtfully designed to match potential users' expectations, with increased prototypicality, will support app use. Conversely, designs that do not include a threshold of expected features will be dismissed, thus undermining the potential of app-based interventions. Designing mHealth apps to account for user expectations will increase the likelihood of adoption and impact from actual use. Prototypicality is positively related to favorable reception and expectations for future use of health apps. These findings provide guidance for user expectations of feature presence and location.

Acknowledgments

This work was supported by an award from the University of North Carolina Lineberger Comprehensive Cancer Center. The funders had no role in the study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
App hot spots and survey items.
[PDF File (Adobe PDF File), 680 KB - mhealth_v9i11e29815_app1.pdf]

References

1. Smith A. U.S. Smartphone Use in 2015. Pew Research Center. 2015 Apr 01. URL: https://www.pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015/ [accessed 2021-10-01]
2. Fox S, Duggan M. Mobile Health 2012. Pew Research Center. 2012 Nov 08. URL: https://www.pewresearch.org/internet/2012/11/08/mobile-health-2012/ [accessed 2021-10-01]
3. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
4. Ribeiro N, Moreira L, Almeida AMP, Santos-Silva F. Pilot study of a smartphone-based intervention to promote cancer prevention behaviours. Int J Med Inform 2017 Dec;108:125-133. [doi: 10.1016/j.ijmedinf.2017.10.013] [Medline: 29132617]
5. Chen J, Gemming L, Hanning R, Allman-Farinelli M. Smartphone apps and the nutrition care process: Current perspectives and future considerations. Patient Educ Couns 2018 Apr;101(4):750-757. [doi: 10.1016/j.pec.2017.11.011] [Medline: 29169863]
6. Bender JL, Yue RYK, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec 23;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
7. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
8. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019;2:54 [FREE Full text] [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]
9. Pandey A, Hasan S, Dubey D, Sarangi S. Smartphone apps as a source of cancer information: changing trends in health information-seeking behavior. J Cancer Educ 2013 Mar;28(1):138-142. [doi: 10.1007/s13187-012-0446-9] [Medline: 23275239]
10. Riley WT, Oh A, Aklin WM, Wolff-Hughes DL. National Institutes of Health Support of Digital Health Behavior Research. Health Educ Behav 2019 Dec;46(2_suppl):12-19. [doi: 10.1177/1090198119866644] [Medline: 31742453]
11. Covolo L, Ceretti E, Moneda M, Castaldi S, Gelatti U. Does evidence support the use of mobile phone apps as a driver for promoting healthy lifestyles from a public health perspective? A systematic review of Randomized Control Trials. Patient Educ Couns 2017 Dec;100(12):2231-2243. [doi: 10.1016/j.pec.2017.07.032] [Medline: 28855063]
12. Iten G, Troendle A, Opwis K. Aesthetics in context? The role of aesthetics and usage mode for a website's success. Interacting with Computers 2018;30(2):133-149. [doi: 10.1093/iwc/iwy002]
13. Lindgaard G, Fernandes G, Dudek C, Brown J. Attention web designers: You have 50 milliseconds to make a good first impression!. Behav Inf Technol 2006 Mar;25(2):115-126. [doi: 10.1080/01449290500330448]
14. Lazard A, Mackert M. User evaluations of design complexity: the impact of visual perceptions for effective online health communication. Int J Med Inform 2014 Oct;83(10):726-735. [doi: 10.1016/j.ijmedinf.2014.06.010] [Medline: 25034279]
15. Lazard AJ, Brennen JS, Troutman Adams E, Love B. Cues for Increasing Social Presence for Mobile Health App Adoption. J Health Commun 2020 Feb 01;25(2):136-149. [doi: 10.1080/10810730.2020.1719241] [Medline: 32026759]
16. Roth SP, Schmutz P, Pauwels SL, Bargas-Avila JA, Opwis K. Mental models for web objects: Where do users expect to find the most frequent objects in online shops, news portals, and company web pages? Interact Comput 2010 Mar;22(2):140-152. [doi: 10.1016/j.intcom.2009.10.004]
17. Leder H, Belke B, Oeberst A, Augustin D. A model of aesthetic appreciation and aesthetic judgments. Br J Psychol 2004 Nov;95(Pt 4):489-508. [doi: 10.1348/0007126042369811] [Medline: 15527534]
18. Lazard AJ, Mackert MS. e-health first impressions and visual evaluations. Commun Des Q Rev 2015 Sep 17;3(4):25-34. [doi: 10.1145/2826972.2826975]
19. Tuch AN, Presslaber EE, Stöcklin M, Opwis K, Bargas-Avila JA. The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. Int J Hum Comput Stud 2012 Nov;70(11):794-811. [doi: 10.1016/j.ijhcs.2012.06.003]
20. Oulasvirta A, Karkkainen L, Laarni J. Expectations and memory in link search. Comput Hum Behav 2005 Sep;21(5):773-789. [doi: 10.1016/s0747-5632(04)00035-4]
21. Roth SP, Tuch AN, Mekler ED, Bargas-Avila JA, Opwis K. Location matters, especially for non-salient features–An eye-tracking study on the effects of web object placement on different types of websites. Int J Hum Comput Stud 2013 Mar;71(3):228-235. [doi: 10.1016/j.ijhcs.2012.09.001]
22. Gibson JJ. The Ecological Approach to Visual Perception. East Sussex: Psychology Press; 1986.
23. Norman D. The Design of Everyday Things. New York, NY: Basic Books; 2002.
24. Norman DA. Affordance, conventions, and design. interactions 1999 May;6(3):38-43. [doi: 10.1145/301153.301168]
25. Seet B, Goh T. Exploring the affordance and acceptance of an e‐reader device as a collaborative learning system. The Electronic Library 2012 Aug 03;30(4):516-542. [doi: 10.1108/02640471211252229]
26. Withagen R, de Poel HJ, Araújo D, Pepping G. Affordances can invite behavior: Reconsidering the relationship between affordances and agency. New Ideas Psychol 2012 Aug;30(2):250-258. [doi: 10.1016/j.newideapsych.2011.12.003]
27. Moshagen M, Musch J, Göritz AS. A blessing, not a curse: experimental evidence for beneficial effects of visual aesthetics on performance. Ergonomics 2009 Oct;52(10):1311-1320. [doi: 10.1080/00140130903061717] [Medline: 19787509]

28. Moshagen M, Thielsch MT. Facets of visual aesthetics. Int J Hum Comput Stud 2010 Oct;68(10):689-709. [doi: 10.1016/j.ijhcs.2010.05.006]
29. Lavie T, Tractinsky N. Assessing dimensions of perceived visual aesthetics of web sites. Int J Hum Comput Stud 2004 Mar;60(3):269-298. [doi: 10.1016/j.ijhcs.2003.09.002]
30. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
31. Lazard AJ, Watkins I, Mackert MS, Xie B, Stephens KK, Shalev H. Design simplicity influences patient portal use: the role of aesthetic evaluations for technology acceptance. J Am Med Inform Assoc 2016 Apr;23(e1):e157-e161 [FREE Full text] [doi: 10.1093/jamia/ocv174] [Medline: 26635314]
32. Nielsen J. Usability Engineering. Burlington, MA: Morgan Kaufmann; 1994.
33. Hingle M, Patrick H, Sacher PM, Sweet CC. The Intersection of Behavioral Science and Digital Health: The Case for Academic-Industry Partnerships. Health Educ Behav 2019 Feb;46(1):5-9. [doi: 10.1177/1090198118788600] [Medline: 30041556]
34. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? Perspect Psychol Sci 2011 Jan;6(1):3-5. [doi: 10.1177/1745691610393980] [Medline: 26162106]
35. Sheehan KB. Crowdsourcing research: Data collection with Amazon’s Mechanical Turk. Commun Monogr 2017 Jul 04;85(1):140-156. [doi: 10.1080/03637751.2017.1342043]
36. Kees J, Berry C, Burton S, Sheehan K. An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk. J Advert 2017 Jan 23;46(1):141-155. [doi: 10.1080/00913367.2016.1269304]
37. What is Chart? Learn how this assessment tool adds value to UNC research. UNC Lineberger. URL: https://chart.unc.edu/ [accessed 2017-03-01]
38. Brennen JS, Lazard AJ, Adams ET. Multimodal mental models: Understanding users' design expectations for mHealth apps. Health Informatics J 2020 Sep;26(3):1493-1506 [FREE Full text] [doi: 10.1177/1460458219882271] [Medline: 31674881]
39. Horrell L, Knafl GJ, Brady T, Lazard A, Linnan L, Kneipp S. Communication Cues and Engagement Behavior: Identifying Advertisement Strategies to Attract Middle-Aged Adults to a Study of the Chronic Disease Self-Management Program. Prev Chronic Dis 2020 Jun 25;17:E48 [FREE Full text] [doi: 10.5888/pcd17.190413] [Medline: 32584754]
40. Mays D, Villanti A, Niaura RS, Lindblom EN, Strasser AA. The Effects of Varying Electronic Cigarette Warning Label Design Features On Attention, Recall, and Product Perceptions Among Young Adults. Health Commun 2019 Mar;34(3):317-324 [FREE Full text] [doi: 10.1080/10410236.2017.1372050] [Medline: 29236529]
41. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
42. Kunst A. Major reasons for adoption of e-health applications and devices by U.S. adults as of 2017. Statista. 2019. URL: https://www.statista.com/statistics/328661/reasons-for-patient-use-of-mhealth-apps-and-services/ [accessed 2021-10-01]
43. Lazard AJ, King AJ. Objective Design to Subjective Evaluations: Connecting Visual Complexity to Aesthetic and Usability Assessments of eHealth. Int J Hum-Comput Int 2019 Apr 24;36(1):95-104. [doi: 10.1080/10447318.2019.1606976]
44. Kim K, Lee C, Hornik RC. Exploring the Effect of Health App Use on Fruit and Vegetable Consumption. J Health Commun 2020 Apr 02;25(4):283-290. [doi: 10.1080/10810730.2020.1745962] [Medline: 32286924]
45. Pieters R, Wedel M, Batra R. The Stopping Power of Advertising: Measures and Effects of Visual Complexity. Journal of Marketing 2010 Sep 01;74(5):48-60. [doi: 10.1509/jmkg.74.5.048]
46. Raita E, Oulasvirta A. Too good to be bad: Favorable product expectations boost subjective usability ratings. Interact Comput 2011 Jul;23(4):363-371. [doi: 10.1016/j.intcom.2011.04.002]
47. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med 2005 Feb;20(2):175-184 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.40245.x] [Medline: 15836552]

Abbreviations

ANOVA: analysis of variance
MANOVA: multivariate analysis of variance
mHealth: mobile health
MTurk: Mechanical Turk
RM: repeated measures
TAM: technology acceptance model

Edited by L Buis; submitted 21.04.21; peer-reviewed by S Bhattacharjya, E Brainin; comments to author 30.06.21; revised version received 23.08.21; accepted 23.09.21; published 04.11.21.

Please cite as:
Lazard AJ, Babwah Brennen JS, Belina SP
App Designs and Interactive Features to Increase mHealth Adoption: User Expectation Survey and Experiment
JMIR Mhealth Uhealth 2021;9(11):e29815
URL: https://mhealth.jmir.org/2021/11/e29815
doi:10.2196/29815
PMID:34734829

Original Paper

Acceptance of an Informational Antituberculosis Chatbot Among Korean Adults: Mixed Methods Research

Agnes Jihae Kim[1,2], PhD; Jisun Yang[3], MSc; Yihyun Jang[4], MSc; Joon Sang Baek[5], PhD

[1]National Rehabilitation Center, Seoul, Republic of Korea

[2]Department of Industrial Design, Yonsei University, Wonju, Republic of Korea

[3]Department of Psychology, Yonsei University, Seoul, Republic of Korea

[4]Department of Cognitive Science, Yonsei University, Seoul, Republic of Korea

[5]Department of Human Environment and Design, Yonsei University, Seoul, Republic of Korea

Corresponding Author:
Joon Sang Baek, PhD
Department of Human Environment and Design
Yonsei University
Samsung Hall, 7th Fl.
Yonsei-ro 50, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2123 3137
Email: joonsbaek@yonsei.ac.kr

Abstract

Background: Tuberculosis (TB) is a highly infectious disease. Negative perceptions and insufficient knowledge have made its eradication difficult. Recently, mobile health care interventions, such as an anti-TB chatbot developed by the research team, have emerged in support of TB eradication programs. However, before the anti-TB chatbot is deployed, it is important to understand the factors that predict its acceptance by the population.

Objective: This study aims to explore the acceptance of an anti-TB chatbot that provides information about the disease and its treatment to people vulnerable to TB in South Korea. Thus, we are investigating the factors that predict technology acceptance through qualitative research based on the interviews of patients with TB and homeless facility personnel. We are then verifying the extended Technology Acceptance Model (TAM) and predicting the factors associated with the acceptance of the chatbot.

Methods: In study 1, we conducted interviews with potential chatbot users to extract the factors that predict user acceptance and constructed a conceptual framework based on the TAM. In total, 16 interviews with patients with TB and one focus group interview with 10 experts on TB were conducted. In study 2, we conducted surveys of potential chatbot users to validate the extended TAM. Survey participants were recruited among late-stage patients in TB facilities and members of web-based communities sharing TB information. A total of 123 responses were collected.

Results: The results indicate that perceived ease of use and social influence were significantly predictive of perceived usefulness (P=.04 and P<.001, respectively). Perceived usefulness was predictive of the attitude toward the chatbot (P<.001), whereas perceived ease of use (P=.88) was not. Behavioral intention was positively predicted by attitude toward the chatbot and facilitating conditions (P<.001 and P=.03, respectively). The research model explained 55.4% of the variance in the use of anti-TB chatbots. The moderating effect of TB history was found in the relationship between attitude toward the chatbot and behavioral intention (P=.01) and between facilitating conditions and behavioral intention (P=.02).

Conclusions: This study can be used to inform future design of anti-TB chatbots and highlight the importance of services and the environment that empower people to use the technology.

(JMIR Mhealth Uhealth 2021;9(11):e26424) doi:10.2196/26424

KEYWORDS

tuberculosis; chatbot; technology acceptance model; mobile phone

Introduction

Background

Tuberculosis (TB) is a highly infectious disease and one of the top 10 causes of death worldwide, claiming approximately 4000 lives a day [1]. Each year, millions of people continue to fall ill with TB, a preventable and curable disease [2]. Among the member countries of the Organization for Economic Cooperation and Development, South Korea has the highest incidence of and mortality rates due to TB [2]. It remains a debilitating disease in the South Korean context in that the treatment is already generalized, but its prevalence and mortality rates are unevenly distributed among social classes [1,3]. Its eradication has been difficult owing to both stigmatization and insufficient understanding of the disease that cause delays in diagnosis and treatment [4]. Approximately one-quarter of the world's population is estimated to be infected with TB, and approximately 5%-10% of those at risk of infection develop active TB in their lifetime [1,5].

Mobile Health Interventions for TB Control

In recent years, mobile health (mHealth) has rapidly emerged as a vehicle for delivering better health services at a lower cost, regardless of time and place [6]. It is used to treat a wide range of infectious diseases, including TB. An extensive investigation on the use of digital technologies for TB control reports various mobile technologies applied for treatment adherence, program management, and e-learning related to TB [7]. These technologies include video-observed treatment (VOT), SMS text messages, mobile apps, voice calls, and mobile phone 3D-printed induration. mHealth apps assist medical staff with patient adherence monitoring (eg, apps for direct observed treatment [DOT] and VOT), dosage adjustment based on patient conditions, and provision of information about diagnosis and management of TB [7,8]. They inform patients and people vulnerable to TB about the disease and its therapy, provide diagnostics based on data input, and evaluate treatment costs. They are also used to trace people who have been exposed to the disease, monitor and track patients, and create laboratory reports [7]. The number of mHealth apps has more than doubled since 2016, evidencing the increasing demand for a new approach to TB control. It is also noteworthy that 39 out of the total 55 apps (71%) are only provided in English, thereby limiting access to non–English-speaking countries, where the highest prevalence of TB cases is observed [7].

Chatbots

Chatbots are a conversational agent, a software program that interacts with natural language, and have emerged as a new form of mHealth service [9,10]. Chatbots are useful for providing information to users with low literacy: users interact with them through dialog, a universal form of interaction. Furthermore, they can provide information in formats that are accessible to people with low literacy, such as images, sounds, and videos [9]. Thus, they are relatively easy to learn and are also age-friendly. From the perspective of health care providers, chatbots can save time and labor [11], in addition to providing continuous treatment management plans, motivation for patients with chronic diseases, and access to real-time information [8,12].

However, despite the expansion of mHealth solutions for TB control and the potential of chatbots, little research has been conducted on applying these tools to the management of TB. To the best of our knowledge, few studies have attempted to develop chatbots and virtual agents to support information accessibility for patients with TB [8]. It is essential to understand the exact factors that predict the acceptance of chatbots by potential users before we introduce them to a TB eradication program, which indeed underscores that the success of digital interventions in health care will depend on how well users accept the technology [13]. Furthermore, understanding the factors that increase the use of chatbots would accelerate the acceptance of this technology among the people most at risk of contracting TB. For this reason, we developed an anti-TB chatbot to bridge the gap between technology and people and studied its acceptance based on the factors that predict potential users, using the Technology Acceptance Model (TAM).

Context of Study: Anti-TB Chatbot

In 2019, we developed an anti-TB chatbot that provides information about the disease, its treatment, and TB hospitals and facilities. It targets people vulnerable to TB, as well as those affected by it. Textbox 1 presents the features of the chatbot.

Textbox 1. Antituberculosis chatbot feature summary.

Feature summary

1. Providing information on tuberculosis
   - Functions
     - Overview of tuberculosis
     - Diagnosis of tuberculosis
     - Tuberculosis treatment
     - Information on drugs to treat tuberculosis
     - Side effects of tuberculosis drugs
     - Screening for contact and latent tuberculosis infection
2. Providing information on hospitals and facilities
   - Functions
     - Institutions for tuberculosis screening and treatment
     - Tuberculosis treatment support project
     - Tuberculosis treatment support facility
     - Information on welfare and administration related to tuberculosis
     - Information on welfare facilities related to tuberculosis

The chatbot was built on an open-source platform and operates within an instant messenger app called Kakao Talk. An advantage of using this platform is that the medium through which users interact with the chatbot, that is, the messenger app, is widely used in South Korea, with over 72% of the total population or roughly 36.6 million people using it [14]. This makes the chatbot highly accessible as most people already have experience in using the app. The open-source platform builder uses machine learning to respond and adapt to diverse conversation patterns. This allows for the accuracy and relevance of the chatbot responses to improve as more user data accumulate.

The knowledge base was obtained from the information provided by the Korea Disease Control and Prevention Agency. We acquired the content with permission and then reorganized it in a dialog format. In addition to the text information, multimedia content was actively adopted, considering the tendency of low health literacy level of the poor and older people [15], who are characterized by a higher-than-average incidence of TB [16]. The curated content was examined by medical staff at a Seoul municipal hospital before publication. Gamification elements, including quizzes and prizes, were also adopted to motivate learners to engage with the chatbot [17].

We gave the chatbot the personality of a doctor. A chatbot with identity cues, such as a name, profile, and language style, is perceived as more empathetic, friendly, and personal [18,19]. Dr Colochman, the personality of the chatbot, is a retired doctor with a long record of treating patients with TB at a municipal hospital and is now working voluntarily for TB hospitals and support facilities. Its identity is conveyed through portraits, names, and intonation. Users encounter Dr Colochman for the first time during the tutorial that provides information on the chatbot and the instructions on how to use it in (what is supposedly) Dr Colochman's voice.

The chatbot provides graphic and text information on the disease, its treatment, and neighboring TB facilities. Users navigate the content by scrolling the page vertically and horizontally. They communicate with the chatbot by selecting menus at the bottom of the screen, pushing buttons, or typing texts (Figure 1).

Figure 1. Antituberculosis chatbot user interface.

TAM and Chatbots

Davis et al [20] developed the TAM to investigate users' intent to accept various technologies, including chatbots, and the factors that predict their decisions [20,21]. The key determinants used to study the acceptance of new technologies with the TAM are perceived usefulness, perceived ease of use, attitude, and behavioral intention. In this study, perceived usefulness is defined as the extent to which users think using anti-TB chatbots is helpful for TB management, and perceived ease of use is the extent to which users think using anti-TB chatbots is convenient and low-effort. According to the TAM, the adoption of a particular technology is governed by individual perceptions of usefulness and ease of use [21].

The TAM is widely used in technology acceptance research; however, it can predict only approximately 40% of the overall explanatory power [22]. A number of extended TAMs have been proposed to overcome the limitations of the original model. Venkatesh and Davis [23] developed the TAM2, adding social influence processes such as subjective norms, voluntariness, and image as external constructs of perceived usefulness. Social influence in our context is the extent to which users think that important others believe in using the anti-TB chatbot. Venkatesh and Davis [24] developed the Unified Theory of Acceptance and Use of Technology, which is the latest derivative of the TAM, adding facilitating conditions as a determinant of behavioral intention. Facilitating conditions here refer to the extent to which users think organizational and technical infrastructure exists to support the use of anti-TB chatbots.

Among the studies that have validated the TAM, some extended the model to address different contexts and populations, including the acceptance and continuous use of chatbots. For example, Huang and Chueh [25] reported that perceived accuracy and ease of use increased pet owners' satisfaction with veterinary consultation chatbots. Ashfaq et al [26] found that perceived enjoyment, usefulness, and ease of use are significant predictors of the continuance intention of chatbot-based customer service. In a study that investigated the acceptance of the health chatbot, Softic´ et al [27] identified the lack of users' trust and qualified medical opinion as barriers; data confidentiality, speed of access to information, information security, and ease of use as facilitators; and reduced time spent on visiting doctors, increased access and care of patients, and enhanced protection of patient data as motivators for using a chatbot.

In the absence of studies that explain the acceptance of chatbots in the context of TB control, we aim to explore the benefits and concerns regarding accepting an anti-TB chatbot as perceived by potential users, to provide an extended TAM that can better predict the acceptance of anti-TB chatbots. Thus, we present studies 1 and 2. Study 1 aims to identify the factors that predict the acceptance of anti-TB chatbots through interviews with patients with TB and homeless facility personnel. On the basis of the interview results, we derived an operational definition of the questionnaire items and identified the factors for the extended TAM. Study 2 aims to verify the proposed theoretical model and identify the factors predicting the acceptance of an anti-TB chatbot.

Methods

Study 1

Data Collection

To collect data for study 1, we conducted interviews with potential users of our anti-TB chatbot. Interviewees were recruited by posting a notice at a municipal TB hospital. The participants were selected using convenience sampling among people who have or had TB. People who could neither understand nor respond to the questionnaire provided in Korean were excluded. In total, 16 patients with TB received a gift worth US $50. We also conducted a focus group interview with 10 experts on TB from the academia, hospitals, shelters, support facilities, and housing providers for homeless people who have worked for patients with TB and thus have sufficient knowledge about them and are willing to use the chatbot or introduce it to them. Participant information is presented in Table 1.

Table 1. Participant information of study 1 (N=26; site: Seoul; year: 2020).

Demographics	Values, n (%)
Patients with TB[a]	
Gender	
Male	16 (100)
Female	0 (0)
Age (years)	
30s	2 (13)
40s	3 (19)
50s	3 (19)
60s	6 (38)
70s	2 (13)
Experience of smartphone use	
Yes	9 (56)
No	7 (44)
Experience of chatbot use	
Yes	0 (0)
No	16 (100)
Experts in treating TB	
Academia	1 (10)
Hospital	1 (10)
Shelters	2 (20)
Support facilities	5 (50)
Housing provider	1 (10)

[a]TB: tuberculosis.

Procedure

Data collection followed the protocols of the American Psychological Association (APA) ethical principles and code of conduct [28]. However, institutional review board approval was not sought. The interviewees were presented with the aim of the study, its procedure and duration, anticipated benefits, and data protection policy. Written informed consent was obtained from those who agreed to participate in the research for recording texts, images, and voices. All the data were transcribed and pseudoanonymized. The interviews were conducted in the following order: (1) introduction to the research, (2) explanation of the data protection policy and collection of informed consent, (3) introduction to the chatbot and instructions on how to use it, (4) installation of the messenger app (if not already installed) and trial of the chatbot, and (5) interview session. All interviews were conducted in the Korean language.

Data Analysis

We analyzed the collected data using thematic analysis in ATLAS.ti, a qualitative data analysis and research tool. Three researchers designed the coding frame to analyze the interviewees' attitude or intent to accept the anti-TB chatbot. We classified the results into a set of subthemes, which were clustered into the main themes. These main themes were assigned as TAM factors.

Study 2

Hypotheses

Study 2 aimed to evaluate factors that predict the acceptance of the anti-TB chatbot. According to Davis et al [20], perceived usefulness and perceived ease of use were the primary factors that predicted the attitude toward a new technology under the TAM. Moreover, perceived ease of use was associated with perceived usefulness. Finally, the attitude toward the technology determined the behavioral intention [23]. Thus, we proposed the following hypotheses:

Hypothesis 1: Attitude toward the chatbot would be positively predicted by perceived usefulness.

Hypothesis 2: Attitude toward the chatbot would be positively predicted by perceived ease of use.

Hypothesis 3: Perceived usefulness would be positively predicted by perceived ease of use.

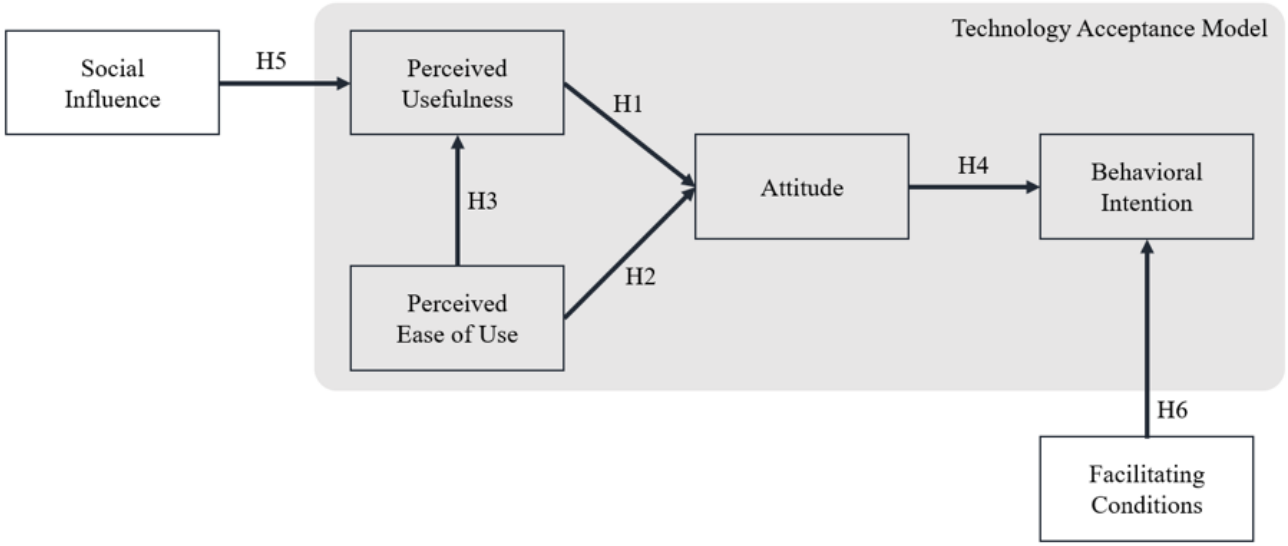
Hypothesis 4: Behavioral intention would be positively predicted by attitude toward the chatbot.

Study 1 demonstrated that social influence and facilitating conditions were relevant to the acceptance of the chatbot by patients with TB. In previous studies that modeled technology acceptance, social influence such as subjective norm, voluntariness, and image is known as an external construct of perceived usefulness [23], while facilitating conditions such as internal and external resources are determinants of behavioral intention [24]. We thus built additional hypotheses as follows and conducted a study that considered them as variables in the research model (Figure 2):

Hypothesis 5: Perceived usefulness would be positively predicted by social influence.

Hypothesis 6: Behavioral intention would be positively predicted by facilitating conditions.

Figure 2. Hypothetical model of study 2.



Data Collection

We conducted both offline and web-based surveys, considering that older adults and other vulnerable groups have limited access to the internet. In the offline survey, we recruited participants at TB facilities that were mainly used by patients in the late stage of TB treatment, who can take medication on their own after discharge from the hospital. The research team visited the facility and instructed and provided assistance to those who expressed their willingness to participate. In the web-based survey, participants were recruited from web-based communities that share information on TB. The web-based survey was distributed among the potential users of the anti-TB chatbot, and their responses were collected via Google Forms. All participants received a monetary reward worth US $5.

Procedure

As in study 1, the data collection process in study 2 was guided by the protocols of the APA ethical principles and code of conduct [28]. All survey participants were asked to read (or were told, if they could not read) the introduction page of the survey describing the purpose of anti-TB chatbot use, its procedure and duration, anticipated benefits, and the data protection policy. Written informed consent was obtained from those who agreed to participate in the research for collecting texts. We then introduced the main screen and dialogs of the app, which informed the participants of the character and functionality of the chatbot. All the data were pseudoanonymized.

Questionnaire Development

The questionnaire was developed based on the theoretical framework of the TAM and the findings from study 1. It consisted of 32 items inquiring about demographic and attitudinal data—participants were asked general questions on demography and experience with chatbots and specific questions regarding their attitude toward the anti-TB chatbot. The attitudinal components were measured using a 7-point Likert-type scale, where the choice of answers ranged from *strongly disagree* (score=1) to *strongly agree* (score=7). The language used in the questionnaire was revised to consider the context of TB and reflect the digital literacy of potential users, as inferred from the results of study 1. The details of the questionnaire items for each construct are presented in Multimedia Appendix 1.

Data Analysis

A total of 127 cases were collected in March 2020. After the screening, 4 cases were excluded: there were missing values in 3 cases, and a straight line was found in 1 case. We used the partial least squares structural equation modeling (PLS-SEM) approach to statistically analyze and process the collected data using SmartPLS 3.0, a dedicated structural equation program with a strong verification power, even for small sample sizes. First, we used the PLS algorithm to evaluate the measurement model. This was followed by bootstrapping and blindfolding techniques for evaluation and hypothesis testing of the structural models.

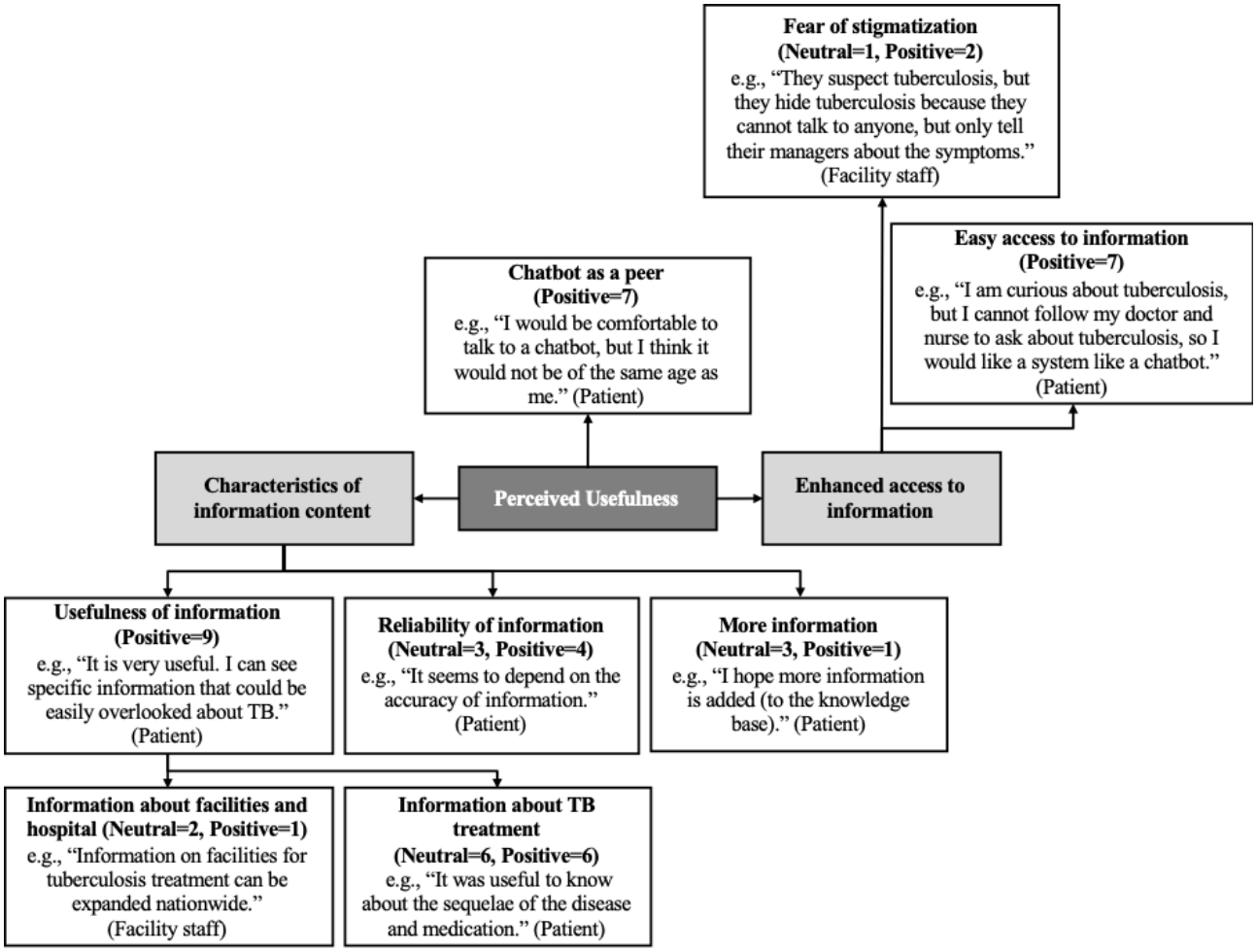
Results

Study 1

Perceived Usefulness

The interviewees noted that the usefulness of the chatbot was associated with the characteristics of the information content, the chatbot's ability to communicate in a similar manner as a peer, and enhanced access to information (Figure 3). In terms of information content, they expected not only useful and reliable information about TB and its treatment but also more content. For example, they considered the fact that the chatbot currently provides information on facilities for TB treatment in Seoul only, which is a limitation. They found information on TB treatment, including hospitals and support facilities for patients with TB, to be most useful. Finally, they anticipated that the chatbot could help reduce the workload of medical staff while increasing patients' access to the necessary information and reducing the risk of stigmatization.

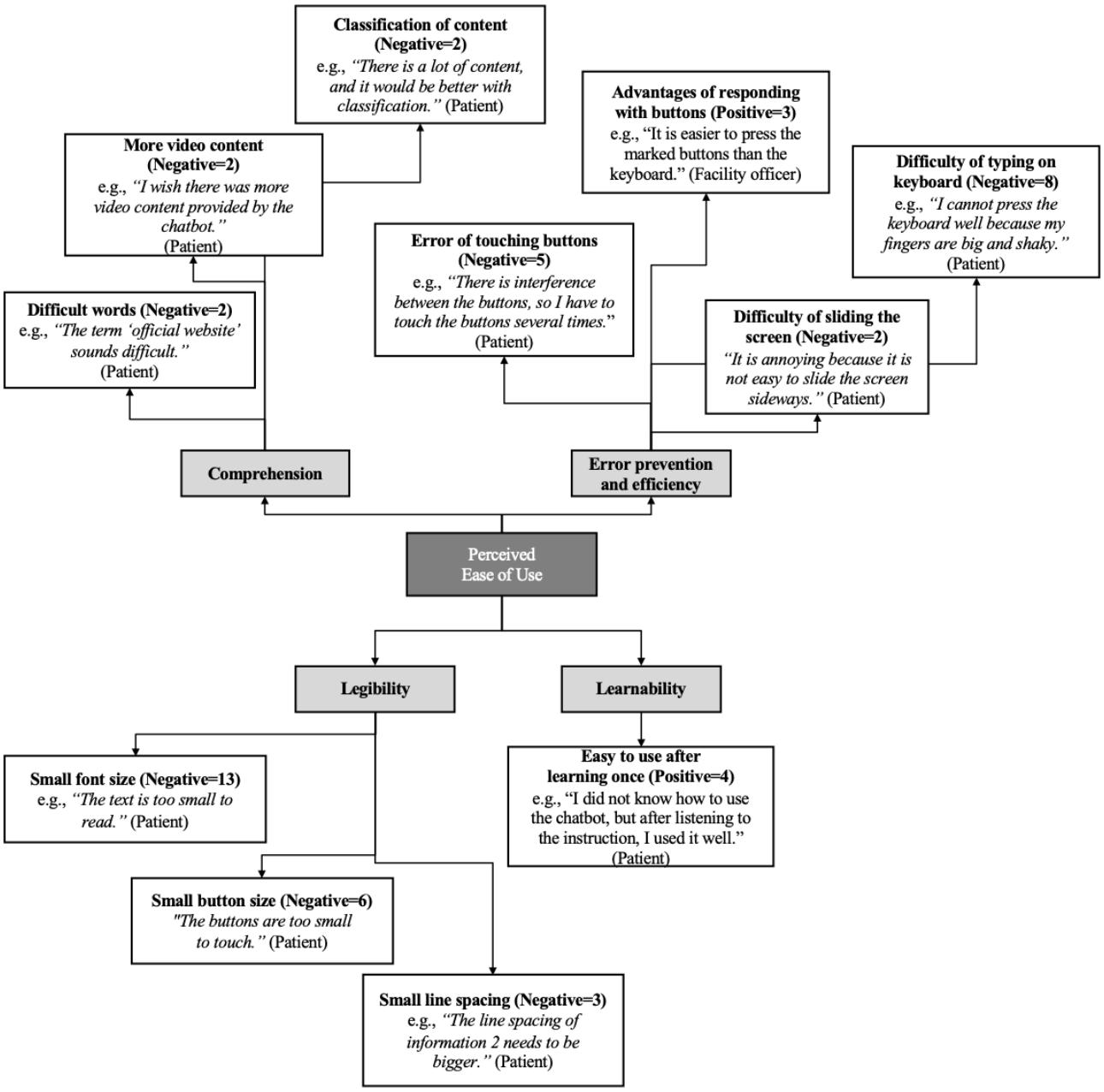
Figure 3. Perceived usefulness of the antituberculosis chatbot among potential users (n=the number of times a theme was mentioned by the interviewees of study 1).



Perceived Ease of Use

Regarding the perceived ease of use of the anti-TB chatbot, the interviewees mentioned the following themes: legibility, comprehension, error prevention and efficiency, and learnability (Figure 4). Legibility is defined as the ability "to see, distinguish, and recognize the characters and words in a text" and is influenced by visual design [29]. In the anti-TB chatbot, legibility issues included inadequate font and button sizes and narrow line spacings. Comprehension *measures whether a user can understand the intended meaning of a text and can draw the correct conclusions from the text* [29]. Related issues included difficult wording, audio-visual information, and unorganized information. Error prevention and efficiency were often related to usability functions supported by the chatbot development platform. The open-builder platform we used provided a simple but functionally constrained environment to develop the chatbot. For example, users can respond to a question from the chatbot either by touching a button or by typing on a virtual keyboard. Interviewees found it difficult to type answers due to the small button size and distance between them, which however could not be adjusted on the platform. The interviewees mentioned that horizontally navigating the information by sliding the screen sideways was troublesome. In terms of learnability, the interviewees quickly learned how to navigate the chatbot after they were given proper instructions.

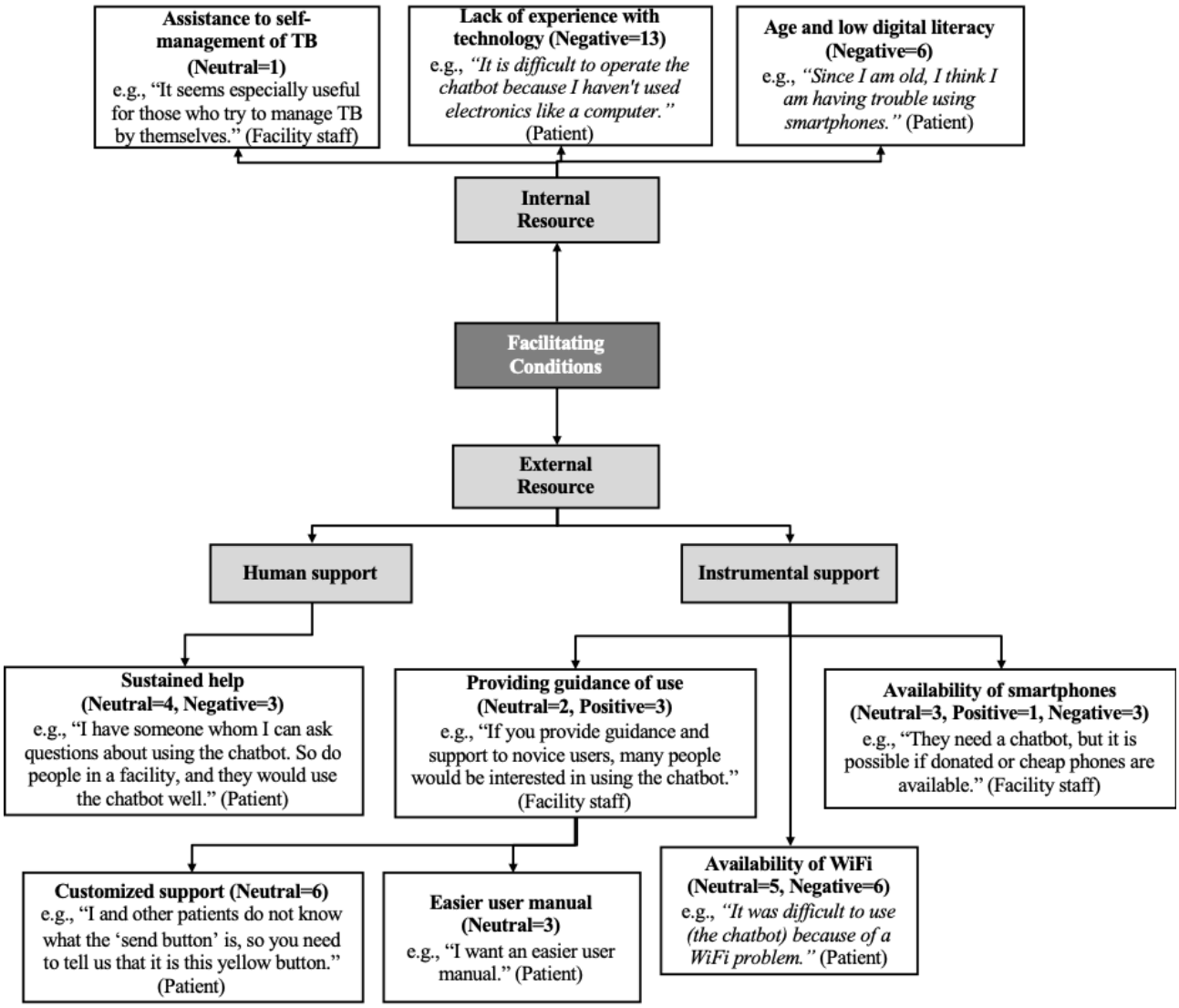
Figure 4. Perceived ease of use of the antituberculosis chatbot among potential users (n=the number of times a theme was mentioned by the interviewees of study 1; positive and neutral comments in normal and negative comments in italics).



Facilitating Conditions

Facilitating conditions associated with the interviewees' acceptance of the anti-TB chatbot were classified as internal and external resources (Figure 5). The former included the user's will to self-manage the disease, their experience of using smartphones and computers, and their age. For example, older interviewees who had no experience of using a smartphone hesitated to use the chatbot. External resource was further subdivided into instrumental and human supports. The former included the availability of instructions and guidance on how to use the chatbot, availability of a smartphone or computer, and access to the internet. Several interviewees lacked basic digital literacy skills and required explanations for simple tasks such as touching the *send message* button. This further suggests the need for an easy-to-understand user manual. The interviewees' attitude toward the acceptance of the chatbot differed depending on the availability of human support, that is, someone who could help them use the chatbot effectively.

Figure 5. Facilitating conditions of the antituberculosis chatbot (n=the number of times a theme was mentioned by the interviewees of study 1; positive and neutral comments in normal and negative comments in italics).



Social Influence

The social influence on the use of the anti-TB chatbot was governed by recommendations from professionals treating TB and the context of use. Interviewees mentioned that recommendations from hospitals would facilitate their adoption of the chatbot (Figure 6). The fact that their peers used the chatbot would also motivate them to accept the new technology.

Figure 6. Social influence of the antituberculosis chatbot in potential users. It should be noted that n=number of times a theme was mentioned by the interviewees of study 1.

Study 2

Demographics

Participants' ages ranged from 22 to 85 years, with almost equal participation by men and women. Most respondents did not have any history of TB, and approximately half had no experience using chatbots. Out of 123 participants, 120 (97.5%) participants had already used the messenger app (Table 2).

Table 2. Participant demographics of study 2 (N=123; site: Seoul; year: 2020).

Demographics	Values, n (%)
Gender	
Female	60 (48.7)
Male	63 (51.2)
Age (years)	
22 to 29	26 (21.1)
30 to 39	33 (26.8)
40 to 49	34 (27.6)
50 to 59	9 (7.3)
60 to 85	21 (17.1)
History of tuberculosis	
Yes	16 (13)
No	107 (86.9)
Experience of using the messenger app	
Yes	120 (97.5)
No	3 (2.5)
Chatbot experience	
Yes	61 (49.5)
No	62 (50.4)

Evaluation of the Measurement Model

The measurement models of study 2 using PLS-SEM were evaluated for internal reliability, convergent validity, and discriminant validity. The internal reliability was assessed using Cronbach α and composite reliability, in which a value greater than .70 for each indicates acceptable internal consistency [30]. To assess the convergent validity, the average variance extracted (AVE) was used, with a recommended value of 0.50 [31]. The results are presented in Table 3. Cronbach α ranged from .798 to .932, and the composite reliability ranged from 0.868 to 0.951, indicating strong internal reliability. Table 3 also presents the estimated construct loading for the study, which ranged from 0.801 to 0.941, and the AVE, which ranged from 0.625 to 0.831, which are greater than the corresponding recommended levels. Therefore, the conditions for convergent validity were satisfied in this study.

Table 3. Reliability and convergent validity of the measurement model in study 2.

Construct and items	Factor loadings	Cronbach α	Composite reliability coefficient	Average variance extracted
PU[a]		.854	0.902	0.697
PU 1	0.893			
PU 2	0.859			
PU 3	0.860			
PU 4	0.783			
PEOU[b]		.927	0.948	0.822
PEOU 1	0.870			
PEOU 2	0.892			
PEOU 3	0.941			
PEOU 4	0.920			
SI[c]		.858	0.904	0.702
SI 1	0.801			
SI 2	0.852			
SI 3	0.887			
SI 4	0.808			
FC[d]		.798	0.868	0.625
PR 1	0.860			
PR 2	0.852			
PR 3	0.635			
PR 4	0.847			
ATC[e]		.899	0.930	0.768
ATC 1	0.834			
ATC 2	0.893			
ATC 3	0.899			
ATC 4	0.878			
BI[f]		.932	0.951	0.831
BI 1	0.851			
BI 2	0.931			
BI 3	0.924			
BI 4	0.937			

[a]PU: perceived usefulness.

[b]PEOU: perceived ease of use.

[c]SI: social influence.

[d]PR: facilitating conditions.

[e]ATC: attitude to chatbot.

[f]BI: behavioral intention.

Discriminant validity was assessed using the square root of the AVE in the cross-loading matrix. To establish a satisfactory discriminant validity of the model, the square root of the AVE for a given construct should be greater than its correlation with other constructs [31]. This, in turn, implies that the diagonal elements must be larger than the entries in the corresponding columns and rows of the matrix. The results shown in Table 4 reveal that all the constructs in this study confirm the discriminant validity of the data.

Table 4. Discriminant validity of the measurement model in study 2.

Constructs	Perceived usefulness	Perceived ease of use	Social influence	Facilitating conditions	Attitude to chatbot	Behavioral intention
Perceived usefulness	0.835	0.512	0.81	0.519	0.714	0.588
Perceived ease of use	0.512	0.906	0.422	0.707	0.325	0.410
Social influence	0.81	0.422	0.838	0.508	0.743	0.664
Facilitating conditions	0.519	0.707	0.508	0.791	0.421	0.494
Attitude to chatbot	0.714	0.325	0.743	0.421	0.876	0.713
Behavioral intention	0.588	0.410	0.664	0.494	0.713	0.911

The Structural Model

The results of the structural model for the TAM are shown in Figure 7 and Table 5. The significance of the path coefficients was assessed using bootstrapping with 5000 samples. The results indicate that attitude toward the chatbot was positively predicted by perceived usefulness (hypothesis 1 supported; $P<.001$) but was not significantly predicted by perceived ease of use (hypothesis 2 not supported; $P=.88$). Perceived usefulness was positively predicted by perceived ease of use (hypothesis 3 supported; $P<.001$) and social influence (hypothesis 5 supported; $P<.001$). Social influence and perceived ease of use explained 67.5% of the variance in perceived usefulness. Finally, behavioral intention was positively predicted by attitude toward the chatbot (hypothesis 4 supported; $P<.001$) and facilitating conditions (hypothesis 6 supported; $P=.03$). Overall, attitude toward the chatbot and facilitating conditions explained 55.4% of the variance in behavioral intention.

Figure 7. Path analysis results for study 2.

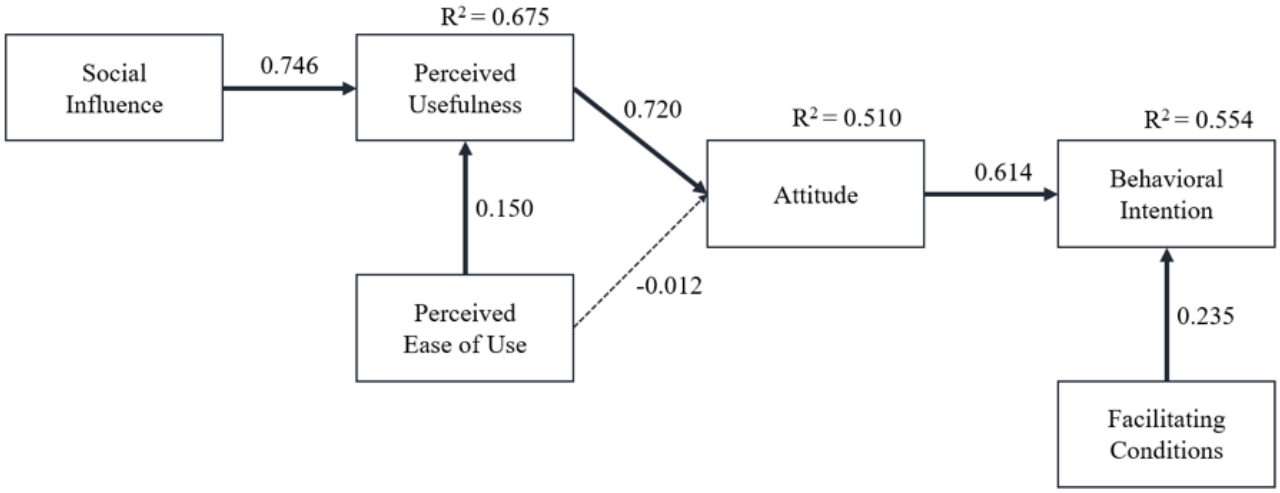


Table 5. Results of the structural model in study 2.

Endogenous variable and exogenous variable	β	*t* value	*P* value
Perceived usefulness			
Perceived ease of use	.15	2.062	.04
Social influence	.746	12.023	<.001
Attitude to chatbot			
Perceived usefulness	.720	11.314	<.001
Perceived ease of use	−.012	0.151	.88
Behavioral intention			
Facilitating conditions	.235	2.242	.03
Attitude to chatbot	.614	7.438	<.001

Multigroup Analysis

We also performed a PLS multigroup analysis (PLS-MGA) by dividing the participants into 2 groups based on their history of TB. There were 107 participants with a history of TB and 16 with no history of TB. The results indicated that perceived usefulness was positively predicted by social influence in both groups (Table 6). Facilitating conditions were predictive of behavioral intention in the TB history group, whereas the attitude toward the chatbot was predictive of behavioral intention in the non–TB history group.

Table 6. Results of the multigroup analysis.

Path	TB[a] history group	Non–TB history group	Difference	TB history group	Non–TB history group	*P* value
	β	β	β	*P* value	*P* value	
PU[b]→ATC[c]	.662	.733	−.071	.002	<.001	.72
PEOU[d]→ATC	.186	−.046	.233	.41	.60	.34
PEOU→PU	−.118	.194	−.313	.56	.008	.13
ATC→BI[e]	.113	.66	−.547	.66	<.001	.01
SI[f]→PU	.906	.726	.18	<.001	<.001	.34
FC[g]→BI	.826	.175	.651	.002	.07	.02

[a]TB: tuberculosis.

[b]PU: perceived usefulness.

[c]ATC: attitude to chatbot.

[d]PEOU: perceived ease of use.

[e]BI: behavioral intention.

[f]SI: social influence.

[g]FC: facilitating conditions.

Discussion

Principal Findings

This study aimed to propose a chatbot that provides information for the prevention and treatment of TB and identify factors that predict the acceptance of the chatbot. We conducted interviews with 16 patients with TB and 10 experts in TB and identified the factors that predict the acceptance of the anti-TB chatbot in study 1. From the results, we found social influence and facilitating conditions as additional factors in the extended TAM model. In study 2, we proposed an extended TAM model capable of predicting the acceptance of the anti-TB chatbot and evaluated it. We found that social influence was a strong predictor of perceived usefulness, regardless of history of TB. Study 1 suggests that social influence can arise from both health care experts and peers. Regarding users' behavioral intention, the predictive factor varied in the participants' history of TB. Overall, our findings were consistent with those of other researchers [18,19,21], indicating that (1) perceived usefulness was predicted by social influence, (2) attitude was predicted by perceived usefulness, and (3) attitude toward the system and facilitating conditions predicted behavioral intention.

Perceived Usefulness

Our study confirmed that people needed information about the disease, as well as TB hospitals and support facilities. It also suggested that the reliability of the information provided by the chatbot is crucial to perceived usefulness and eventually the acceptance of the chatbot. Although this may sound rather obvious, existing mHealth apps that provide information on TB have been found to contain errors such as spelling and grammatical mistakes, outdated information, and wrong and potentially harmful content, according to a recent study that investigated 29 e-learning and information apps on TB [7]. The reliability of the information can be achieved by using trusted sources, having the content examined by experts before publication, and keeping it up to date through continuous maintenance.

Perceived usefulness was significantly predictive of people's attitude toward the anti-TB chatbot if they have experienced TB. When people seek information about TB, stigmatization and its consequences (eg, social isolation and reduced economic opportunities) can be barriers to active information seeking and timely access to necessary services [32]. We expect that the anti-TB chatbot can contribute to lowering this barrier by facilitating access to information and reducing the risk of stigmatization (see Figure 4 for a glimpse of the chatbot experience). For patients with TB, the primary channel through which they receive information related to the disease is the medical staff. However, due to limited time at hand, medical staff provide selective information. The anti-TB chatbot can reduce staff workload while providing patients with the necessary information when needed. In other words, it bridges the distance between patients and medical staff by acting as a virtual assistant [33]. It also mitigates the information asymmetry between the 2 parties by empowering patients with the ability to access the information they need.

Perceived Ease of Use

Study 2 confirmed that perceived ease of use was predictive of perceived usefulness but not predictive of the attitude toward technology. The latter result has been observed in studies where

participants were proficient in using the technology (eg, responses of experienced mobile phone users to a mobile app or a chatbot) [34,35]. The same trend was observed in study 2, where there was a roughly even distribution in age of the survey participants (22 to 85 years). Does this imply that participants of different ages, and possibly varied levels of digital literacy, were proficient in using the chatbot that they were introduced for the first time? If so, what aspects of the chatbot are associated with proficiency? We speculate that this may be due to the popularity of the platform on which the anti-TB chatbot runs, that is, the messenger app widely used by people. The familiar user interface of the chatbot may have been transferred to the perceived proficiency in the use of the chatbot and a positive opinion of its utility. Thus, we conclude that the perceived usefulness of a chatbot can increase when its user interface is familiar to the target users.

Social Influence

In study 1, we observed social influence acting on the interviewees when a staff member in the hospital or TB treatment facility recommended the use of the anti-TB chatbot or when a peer introduced them. Thus, social influence can have a positive impact on the perceived usefulness of the chatbot. In a study that investigated the acceptance of conversational agents for disease diagnosis, social influence was identified as a factor influencing users' intention to adopt or use a chatbot [9]. It has also been reported that users' trust in providers and chatbots predicts performance expectancy. Performance expectancy refers to the degree to which using a chatbot will provide benefits to users in improving their health conditions [24].

Social influence can be derived from the authority and credibility of the service provider (ie, hospital) and those who have (expert or user) knowledge about the disease and technology. Among the different types of social influence was the peer pressure from other people who use a smartphone and a chatbot. For example, a patient with TB whom we met in study 1 was among several patients who did not have a smartphone and were eager to learn to use the smartphone and the anti-TB chatbot (Figure 6). Considering that there is still a large population who cannot access mHealth solutions, our findings reiterate a barrier to these technologies and simultaneously a strong demand for them that remains to be met. It is beyond the scope of this study to discuss how to meet this demand, but we introduce some of the existing efforts and emphasize the need for facilitating conditions in the Facilitation Condition section.

Facilitating Condition

Facilitating condition is strongly associated with the acceptance of chatbots by patients with TB and thus should be considered when designing an anti-TB chatbot. TB occurs more commonly in older adults and low-income groups. It is also these groups who find it most challenging to access and use mHealth solutions. They are often reluctant to accept new solutions, such as chatbots, due to a lack of internal resources (eg, information on and capabilities to use mHealth solutions) according to study 1. However, this lack can be compensated by the provision of external support, such as a peer who can help them learn how to use a chatbot or smartphone tutorials. Existing chatbot-related studies tend to focus on the efficiency and usefulness of these technologies [11,36-38]. However, our findings suggest that it is equally important to design facilitating conditions from the perspective of users to encourage and accelerate their acceptance of chatbots. In other words, both services and an environment that empowers people to use the chatbots should be designed to eradicate TB.

Trans-sectoral efforts have been made to disseminate smartphones among homeless people as a strategy to reinforce self-sufficiency and mitigate poverty. Organizations, such as the Community Technology Alliance, Seoul Municipality, and Underheard in New York, have implemented smartphone giveaway projects in which donated smartphones were delivered to homeless people and used to find accommodation, economic opportunities, and fulfill other basic needs [39,40]. These examples demonstrate the possibility of making mHealth solutions accessible to the bottom of the pyramid, although they do not report any integration with mHealth. On the basis of our empirical study, we cautiously argue that there is sufficient demand for mHealth solutions, including the anti-TB chatbot, among poor and older people. The question is how to deliver them in a scaled-up and sustained manner. As we witness the rapid growth of the mHealth industry and anticipate a variety of solutions for TB control, including chatbots, the facilitating conditions are all the more important for the democratization of these technologies, that is, the development of technologies for people who are most affected by TB should be concurrent with sustained efforts to empower them.

Limitations

This study has several limitations. First, our hypotheses were evaluated using correlation methods; therefore, the derived model did not explain causal relationships among the identified constructs. Second, this study was conducted using a convenience sample, which limits the generalizability of our findings. Thus, future studies should conduct a more comprehensive inspection of how these individual differences are associated with the acceptance of technology using representative and larger samples [41]. With larger samples, we may also be able to identify additional external factors that are predictive of the acceptance of the anti-TB chatbot. Potential candidates include social support and stigma, which have been identified as relevant for treating TB [42-44]. Third, our sample with a history of TB is relatively small and homogeneous due to the invisibility caused by the fear of stigmatization. Although this study showed that the predictor of anti-TB chatbot acceptance depends on history of TB, the number of patients was not sufficient to obtain results with greater statistical and conceptual strengths. Finally, the impact of this study remains limited in the current environment, where technological advances are not accessible to many homeless people who can benefit from them. At the same time, we are reminded that technology alone cannot solve complex societal problems. We also need to invest in scaling up the ongoing efforts to empower these people (eg, digital literacy education) and build the necessary infrastructure (eg, provide mobile devices and services that they can afford and expand public Wi-Fi zones in low-income residential areas).

Conclusions

Despite the expansion of mHealth solutions for TB control and the potential of chatbots to save costs and reduce the risk of stigma associated with the diagnosis and treatment of TB, few studies have sought to investigate the determinants of their adoption. In this context, we conducted 2 studies to develop an extended TAM that incorporates additional variables obtained from an empirical study with patients with TB and explain the intention to use a chatbot for TB control. The results showed that the intention to use the anti-TB chatbot was predicted by attitude toward the chatbot and facilitating conditions. Attitude toward the chatbot was positively predicted by its perceived usefulness but was not significantly predicted by perceived ease of use. The results also suggested that the perceived usefulness of the anti-TB chatbot was positively predicted by perceived ease of use and social influence. The importance of this study is to identify the underlying factors associated with the intention to use an anti-TB chatbot. These findings can be used to inform future design of anti-TB chatbots. For future work, it will be necessary to integrate the proposed model with other theories and factors that can help explain greater acceptance.

Acknowledgments

This work was supported by the Ministry of Science and Information and Communication Technology on the “Digital Social Innovation Project” of National Information Society Agency and Yonsei University Research Grant of 2021. The authors thank the Misohealing Centre and Seoul Seobuk Hospital for their cooperation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definition for the constructs used in study 2.
[DOCX File , 15 KB - mhealth_v9i11e26424_app1.docx]

References

1. World Health Organization. Global Tuberculosis Report 2013. Geneva, Switzerland: World Health Organization; 2013.
2. Global Tuberculosis Report 2020. World Health Organization. 2020. URL: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2020 [accessed 2021-09-19]
3. Christof C, Nußbaumer-Streit B, Gartlehner G. WHO guidelines on tuberculosis infection prevention and control. Gesundheitswesen 2020;82(11):885-889. [doi: 10.1055/a-1241-4321] [Medline: 32977345]
4. Gebremariam MK, Bjune GA, Frich JC. Barriers and facilitators of adherence to TB treatment in patients on concomitant TB and HIV treatment: a qualitative study. BMC Public Health 2010 Oct 28;10:651 [FREE Full text] [doi: 10.1186/1471-2458-10-651] [Medline: 21029405]
5. World Health Organization. Latent Tuberculosis Infection: Updated and Consolidated Guidelines for Programmatic Management. Geneva: World Health Organization; 2018.
6. Samhan B, Joshi K. Resistance of healthcare information technologies; literature review, analysis, and gaps. In: Proceedings of the 48th Hawaii International Conference on System Sciences.: IEEE; 2015 Jan Presented at: 48th Hawaii International Conference on System Sciences; January 5-8, 2015; Kauai, HI, USA. [doi: 10.1109/hicss.2015.362]
7. Keutzer L, Wicha SG, Simonsson US. Mobile health apps for improvement of tuberculosis treatment: descriptive review. JMIR Mhealth Uhealth 2020 Apr 21;8(4):e17246 [FREE Full text] [doi: 10.2196/17246] [Medline: 32314977]
8. Iribarren SJ, Wallingford J, Schnall R, Demiris G. Converting and expanding mobile support tools for tuberculosis treatment support: Design recommendations from domain and design experts. J Biomed Inform 2020 Mar;5:100066 [FREE Full text] [doi: 10.1016/j.yjbinx.2019.100066] [Medline: 34327316]
9. Laumer S, Maier C, Gubler F. Chatbot acceptance in healthcare: explaining user adoption of conversational agents for disease diagnosis. In: Proceedings of the 2019 European Conference on Information Systems. 2019 Presented at: 27th European Conference on Information Systems (ECIS); June 2019; Stockholm & Uppsala URL: https://cris.fau.de/converis/portal/publication/215960736?lang=en_GB
10. Laranjo L, Dunn A, Tong H, Kocaballi A, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
11. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017 Jul 21;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]
12. Chaix B, Bibault J, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. JMIR Cancer 2019 May 02;5(1):e12856 [FREE Full text] [doi: 10.2196/12856] [Medline: 31045505]

13. Ngwatu BK, Nsengiyumva NP, Oxlade O, Mappin-Kasirer B, Nguyen NL, Jaramillo E, Collaborative Group on the Impact of Digital Technologies on TB. The impact of digital health technologies on tuberculosis treatment: a systematic review. Eur Respir J 2018 Jan 11;51(1):1701596 [FREE Full text] [doi: 10.1183/13993003.01596-2017] [Medline: 29326332]
14. Piao M, Ryu H, Lee H, Kim J. Use of the healthy lifestyle coaching chatbot app to promote stair-climbing habits among office workers: exploratory randomized controlled trial. JMIR Mhealth Uhealth 2020 May 19;8(5):e15085 [FREE Full text] [doi: 10.2196/15085] [Medline: 32427114]
15. Sudore R, Mehta K, Simonsick E, Harris T, Newman A, Satterfield S, et al. Limited literacy in older people and disparities in health and healthcare access. J Am Geriatr Soc 2006 May;54(5):770-776. [doi: 10.1111/j.1532-5415.2006.00691.x] [Medline: 16696742]
16. Annual Report on the Notified Tuberculosis in Korea. Korea Centers for Disease Control and Prevention. 2020 Mar. URL: https://kdca.go.kr/board/board.es?mid=a20602010000&bid=0034&act=view&list_no=712904 [accessed 2021-10-06]
17. Glover I. Play as you learn: gamification as a technique for motivating learners. Sheffield Hallam University Research Archive (SHURA). 2013. URL: http://shura.shu.ac.uk/7172/1/glover_-_play_as_you_learn_-_proceeding_112246.pdf [accessed 2021-09-14]
18. Araujo T. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. Comput Hum Behav 2018 Aug;85:183-189. [doi: 10.1016/j.chb.2018.03.051]
19. Feine J, Gnewuch U, Morana S, Maedche A. A taxonomy of social cues for conversational agents. Int J Hum Comput Stud 2019 Dec;132:138-161. [doi: 10.1016/j.ijhcs.2019.07.009]
20. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
21. Venkatesh V. Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. Inform Sys Res 2000 Dec;11(4):342-365. [doi: 10.1287/isre.11.4.342.11872]
22. Legris P, Ingham J, Collerette P. Why do people use information technology? A critical review of the technology acceptance model. Inform Manag 2003 Jan;40(3):191-204. [doi: 10.1016/s0378-7206(01)00143-4]
23. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
24. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
25. Huang D, Chueh H. An analysis of use intention of pet disease consultation chatbot. In: Proceedings of the 4th International Conference on E-Society, E-Education and E-Technology. 2020 Presented at: ICSET'20: 2020 The 4th International Conference on E-Society, E-Education and E-Technology; August 15 - 17, 2020; Taipei Taiwan p. 1-5. [doi: 10.1145/3421682.3421693]
26. Ashfaq M, Yun J, Yu S, Loureiro SM. I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. Telemat Informat 2020 Nov;54:101473. [doi: 10.1016/j.tele.2020.101473]
27. Softić A, Husić J, Softić A, Baraković S. Health chatbot: design, implementation, acceptance and usage motivation. In: Proceedings of the 20th International Symposium INFOTEH-JAHORINA (INFOTEH).: IEEE; 2021 Presented at: 20th International Symposium INFOTEH-JAHORINA (INFOTEH); March 17-19, 2021; East Sarajevo, Bosnia and Herzegovina. [doi: 10.1109/infoteh51037.2021.9400693]
28. Psychological Association. Ethical principles of psychologists and code of conduct. Am Psychol 2002;57(12):1060-1073. [doi: 10.1037/0003-066x.57.12.1060] [Medline: 12613157]
29. Legibility, readability, and comprehension: making users read your words. Nielsen Norman Group. URL: https://www.nngroup.com/articles/legibility-readability-comprehension/ [accessed 2021-09-14]
30. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. London: Pearson; Feb 23, 2009.
31. Hair JJ, Hult G, Ringle C, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). California: SAGE Publications; 2016.
32. Courtwright A, Turner AN. Tuberculosis and stigmatization: pathways and interventions. Public Health Rep 2010;125 Suppl 4:34-42 [FREE Full text] [doi: 10.1177/00333549101250S407] [Medline: 20626191]
33. Fadhil A. arXiv.org. 2018. URL: https://arxiv.org/ftp/arxiv/papers/1803/1803.06000.pdf [accessed 2021-09-14]
34. Yuan S, Ma W, Kanthawala S, Peng W. Keep using my health apps: discover users' perception of health and fitness apps with the UTAUT2 Model. Telemed J E Health 2015 Sep;21(9):735-741. [doi: 10.1089/tmj.2014.0148] [Medline: 25919238]
35. Zarouali B, Van den Broeck E, Walrave M, Poels K. Predicting consumer responses to a chatbot on Facebook. Cyberpsychol Behav Soc Netw 2018 Aug;21(8):491-497. [doi: 10.1089/cyber.2017.0518] [Medline: 30036074]
36. Kim S, Eun J, Oh C, Suh B, Lee J. Bot in the bunch: facilitating group chat discussion by improving efficiency and participation with a chatbot. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April 25 - 30, 2020; Honolulu HI USA p. 1-13. [doi: 10.1145/3313831.3376785]
37. Kong B, Wang E, Li Z, Lu W. Study on the feature of electromagnetic radiation under coal oxidation and temperature rise based on multifractal theory. Fractals 2019 May 30;27(03):1950038. [doi: 10.1142/s0218348x19500385]

38. Zamora J. I'm sorry, dave, I'm afraid I can't do that: chatbot perception and expectations. In: Proceedings of the 5th International Conference on Human Agent Interaction. 2017 Presented at: HAI '17: The Fifth International Conference on Human-Agent Interaction; October 17 - 20, 2017; Bielefeld Germany p. 253-260. [doi: 10.1145/3125739.3125766]
39. Hear the voices from the underheard. Underheard in New York. 2009. URL: http://underheardinnewyork.com/ [accessed 2021-05-18]
40. Community Technology Alliance. 2019. URL: https://ctagroup.org/ [accessed 2021-05-18]
41. Lamela D, Figueiredo B, Morais A, Matos P, Jongenelen I. Are measures of marital satisfaction valid for women with depressive symptoms? The examination of factor structure and measurement invariance of the Couple Satisfaction Index-4 across depression levels in Portuguese women. Clin Psychol Psychother 2020 Mar;27(2):214-219. [doi: 10.1002/cpp.2420] [Medline: 31881102]
42. Milligan H, Iribarren SJ, Chirico C, Telles H, Schnall R. Insights from participant engagement with the tuberculosis treatment support tools intervention: Thematic analysis of interactive messages to guide refinement to better meet end user needs. Int J Med Inform 2021 May;149:104421. [doi: 10.1016/j.ijmedinf.2021.104421] [Medline: 33706032]
43. Paz-Soldán VA, Alban RE, Jones CD, Oberhelman RA. The provision of and need for social support among adult and pediatric patients with tuberculosis in Lima, Peru: a qualitative study. BMC Health Serv Res 2013 Jul 31;13:290 [FREE Full text] [doi: 10.1186/1472-6963-13-290] [Medline: 23899353]
44. Suliman Q, Lim P, Said S, Tan K, Zulkefli N. Survival analysis of time to early tb treatment interruption: a longitudinal study on psycho-social risk factors among newly diagnosed tb patients in a state of Malaysia. Res Square 2021 Feb:1-14. [doi: 10.21203/rs.3.rs-226208/v1]

Abbreviations

AVE: average variance extracted
DOT: direct observed treatment
mHealth: mobile health
PLS-MGA: partial least squares multigroup analysis
PLS-SEM: partial least squares structural equation modeling
TAM: Technology Acceptance Model
TB: tuberculosis
VOT: video-observed treatment

Edited by L Buis; submitted 08.01.21; peer-reviewed by M Drehlich, B Thies, A Kibballi Madhukeshwar; comments to author 19.03.21; revised version received 20.05.21; accepted 03.09.21; published 09.11.21.

Please cite as:
Kim AJ, Yang J, Jang Y, Baek JS
Acceptance of an Informational Antituberculosis Chatbot Among Korean Adults: Mixed Methods Research
JMIR Mhealth Uhealth 2021;9(11):e26424
URL: https://mhealth.jmir.org/2021/11/e26424
doi:10.2196/26424
PMID:34751667

Original Paper

To Use or Not to Use a COVID-19 Contact Tracing App: Mixed Methods Survey in Wales

Kerina Jones[1], BSc, PhD; Rachel Thompson[1], BA, BSc

Swansea University, Swansea, United Kingdom

Corresponding Author:
Kerina Jones, BSc, PhD
Swansea University
Population Data Science, Swansea University Medical School
Singleton Park
Swansea, SA28PP
United Kingdom
Phone: 44 01792602764
Email: k.h.jones@swansea.ac.uk

Abstract

Background: Many countries remain in the grip of the COVID-19 global pandemic, with a considerable journey still ahead toward normalcy and free mobility. Contact tracing smartphone apps are among a raft of measures introduced to reduce spread of the virus, but their uptake depends on public choice.

Objective: The objective of this study was to ascertain the views of citizens in Wales on their intended use of a COVID-19 contact tracing smartphone app, including self-proposed reasons for or against use and what could lead to a change of decision.

Methods: We distributed an anonymous survey among 4000 HealthWise Wales participants in May 2020. We adopted a mixed methods approach: responses to closed questions were analyzed using descriptive and inferential statistics; open question responses were analyzed and grouped into categories.

Results: A total of 976 (24.4%) people completed the survey. Smartphone usage was 91.5% overall, but this varied among age groups. In total, 97.1% were aware of contact tracing apps, but only 67.2% felt sufficiently informed. Furthermore, 55.7% intended to use an app, 23.3% refused, and 21.0% were unsure. The top reasons for app use were as follows: controlling the spread of the virus, mitigating risks for others and for oneself, and increasing freedoms. The top reasons against app use were as follows: mistrusting the government, concerns about data security and privacy, and doubts about efficacy. The top response for changing one's mind about app use from being willing to being unwilling was that nothing would; that is, they felt that nothing would cause them to become unwilling to use a contact tracing app. This was also the top response for changing one's mind from being unwilling to being willing to use contact tracing apps. Among those who were unsure of using contact tracing apps, the top response was the need for more information.

Conclusions: Respondents demonstrated a keenness to help themselves, others, society, and the government to avoid contracting the virus and to control its spread. However, digital inclusion varied among age groups, precluding participation for some people. Nonetheless, unwillingness was significant, and considering the nature of the concerns raised and the perceived lack of information, policy and decision-makers need to do more to act openly, increase communication, and demonstrate trustworthiness if members of the public are to be confident in using an app.

(JMIR Mhealth Uhealth 2021;9(11):e29181) doi:10.2196/29181

KEYWORDS

COVID-19; survey; Wales; contact tracing; app; mHealth; mobile apps; digital health; public health

Introduction

In common with many countries worldwide, Wales remains in the grip of the COVID-19 global pandemic, with a considerable journey still ahead toward normalcy and free mobility. With a population of 3.1 million individuals, Wales accounts a relatively small fraction of the 66.8 million individuals in the United Kingdom [1]. Nonetheless, at the time of writing, Wales has had a cumulative total of over 200,000 cases and over 5400 deaths due to COVID-19 with a confirmed positive test [2]. The Senedd—the Welsh Government—is committed to the use of

contact tracing smartphone app technology as part of a raft of measures to control and reduce spread of the virus. Along with the direct measures of individual testing and vaccinations, these also include society-wide travel restrictions; closure of schools, nonessential shops, and businesses; a track-and-trace system; and working from home wherever possible, such that the entire country has been on the highest level of lockdown used by the Senedd (alert level 4) [3]. As such, this is a novel situation in general and specifically concerning the introduction of a contact tracing app.

There has been considerable debate in the media about the use of contact tracing smartphone apps during the pandemic. Some concerns that have been raised are the perceived risk to individual privacy, data security, and the ethics of automated data collection with a person's own private device [4]. Although contact tracing is a longstanding part of public health surveillance for infectious diseases, the standard methods are seen as too slow in a pandemic situation when used alone. The use of smartphone apps is proposed to capture information more quickly and expedite rapid information sharing to enable action to prevent further spread [4]. In the spirit of openness and because of the concerns raised, there is considerable interest in citizens' views on the use of these smartphone apps. Since majority uptake is needed to achieve maximum effectiveness, it is important to gauge and understand citizens' views on the acceptability of contact tracing apps for smartphones.

The published literature shows that there have been large-scale quantitative studies with the public and small-scale qualitative studies to gain insight into citizens' perceptions about the use of an app in various countries. These have included a variety of study designs: surveys providing a menu of options for respondents to indicate their reasons for or against the use of an app [5-7]; discrete choice experiments where respondents are asked to select options based on trade-offs [8]; and focus groups where individuals were able to present their own reasons for or against an app [9]. The focus groups study conducted by Williams et al [9] included participants from all 4 UK countries but naturally had a small sample size. The multi-country survey carried out by Altmann et al [5] included people from the United Kingdom but did not specify the breakdown by country. Since public opinion on self-determination, personal choice, and responsibility can vary by culture and government administration across countries, we wanted to address the knowledge gap by finding out more about Welsh citizens' views. Our research question was as follows: What are the views and intentions of members of the Welsh public in relation to using a contact tracing app? To address this question, we designed a mixed methods study that included a survey composed of a combination of closed and open questions. In particular, to generate rich information, we allowed the participants to provide open responses for or against the use of an app and on what would lead them to change their mind regarding app use.

The objective of this study was to ascertain the views of citizens in Wales on their intended use of a COVID-19 contact tracing smartphone app, including self-proposed reasons for or against it, and what could lead to a change of decision.

Methods

We designed and distributed an anonymous survey among 4000 HealthWise Wales (HWW) participants. HWW is a cohort of approximately 40,000 people who have signed up to help shape the health and well-being of future generations in Wales. Compared with the wider population, there was a higher percentage of HWW participants older than 45 years. The percentage of women was higher than that in the general population (72% vs 51%). The percentage of participants in ethnic groups other than White ethnicity (2%) matched that in the Welsh population. Around half of participants are in higher managerial or professional occupations, which is significantly greater than the general population; however, each quintile of the Welsh Index of Multiple Deprivation was represented [10,11]. The survey was released on the internet on the TypeForm site and piloted to check for consistency before being released. To yield a rapid response, we asked HWW to select a random sample from their full cohort, correctively weighted for responses among men, and ethnic minorities of all genders as these groups were known to be underrepresented in the HWW cohort. Ethical approval was not required for the study, as no identifiable information was sought.

The survey was released from May 22 to 28, 2020, and was closed when the response rate had tailed off. At this time, the United Kingdom was testing the NHSX app which was intended to operate on a centralized data collection model. This model was abandoned shortly after the survey was conducted, and the National Health Service (NHS) COVID-19 contact tracing app (operating on a decentralized model) was rolled out in England and Wales on September 24, 2020. Both models collect data via Bluetooth technology, but in a decentralized model, data processing is performed on the smartphone rather than being transferred to and stored in a central database.

All the survey questions were in a closed or structured format apart from those about reasons for being willing or unwilling to use a contact tracing app, and reasons for a change of mind regarding app use. Participants were invited to provide up to 3 reasons for or against app use and were not asked to rank them. They were asked for one reason for why they might change their mind. These responses were open in the free-text format and were analyzed and grouped manually by a consensus with 2 researchers. The quantitative data were analyzed using SPSS (version 26), using descriptive (n [%]) and inferential (chi-square) statistics. A list of survey questions is provided in Multimedia Appendix 1.

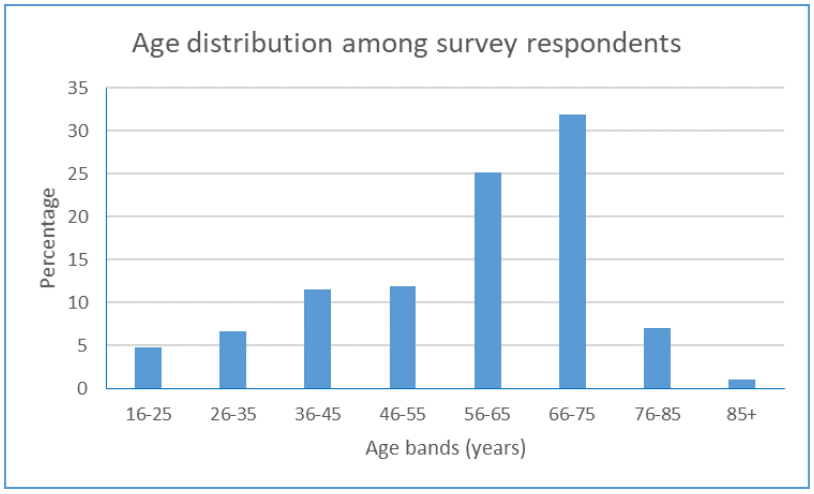
Results

Basic Demographics

A total of 976 (24.4%) full survey responses were received from across all main postcode areas of Wales (Cardiff, Llandrindod, Wrexham, Newport, Swansea, and Shrewsbury). Only the first part of the postcode was requested to preserve anonymity. Among 968 respondents, 504 (52.1%) responded as being male, 461 (47.6%) as being female, and 3 (0.3%) as being nonbinary. In terms of ethnicity (N=965), 923 (95.5%) identified as White, 7 (0.7%) as being Asian or Asian British, 4 (0.4%) as being

Black, African, Black British, or Caribbean, 26 (2.7%) as being mixed or of multiple ethnic groups, and 5 (0.5%) as being of other ethnicities. The age distribution of respondents (N=974) is shown in Figure 1.

Figure 1. Age distribution among the survey respondents. In total, 974 people (of 976 in total) provided their ages in 10-year bands.



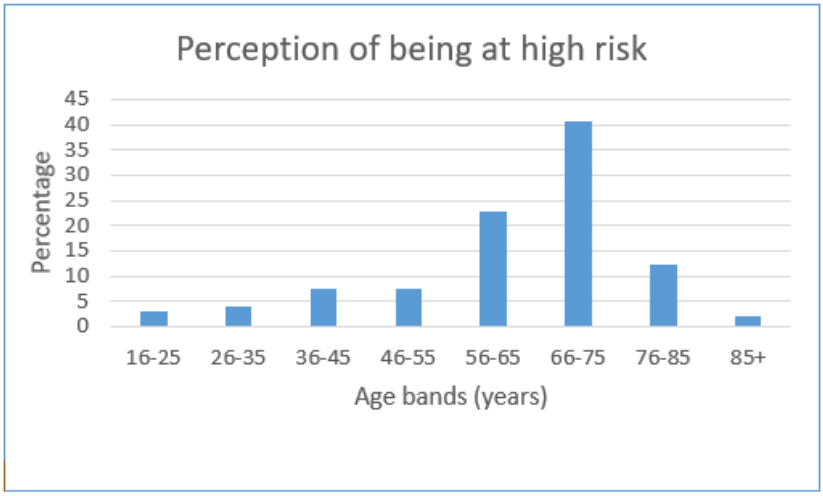
Smartphone Usage

Smartphone usage was reported as 91.5% overall, but this varied by age groups (P<.001). It was over 95% in age groups of up to 55 years, approximately 90% among individuals aged 56-75 years, 78% among the individuals aged 76-85 years, and only 40% among those aged over 85 years.

COVID-19 Risk and Experience

Of 971 respondents, 291 (29.8%) considered themselves at high risk of infection and 680 (69.7%) did not believe that they were at high risk. There was no significant difference in the responses among all ethnic groups. Minority ethnicities were considered a single group in this instance as the numbers among each group were small. The distribution of these responses showing differences in the perception of personal risk among these age bands, with older people being at higher risk (P<.001), is shown in Figure 2. Among 971 respondents, 34 (3.5%) indicated that they had had COVID-19, and 35.7% indicated that they knew someone who had been infected. Both of these questions addressed whether COVID-19 was diagnosed or self-reported by these individuals.

Figure 2. Respondents' perceptions of being high risk of COVID-19 infection. In total, 971 respondents (of 976 in total) provided their perception of their personal risk of COVID-19.



Knowledge and Use of Apps

Participants were asked about their knowledge and use of symptom tracking apps before proceeding to focus on contact tracing apps. A common symptom tracking app used in the United Kingdom is operated by the COVID Symptom Study, commonly referred to as "the Zoe app" [12]. In total, 974 (94%) respondents indicated that they were familiar with symptom tracking apps and 910 (37.7%) respondents used the app. Questions proceeded to ask about the awareness of plans for introducing a contact tracing app and whether people felt they had sufficient knowledge of potential benefits and risks. A total of 974 (97.1%) respondents were aware, and 973 (67.2%) respondents felt that they had sufficient knowledge. This was followed by a question about participants' plans to use a contact tracing app if one were to be introduced. The number of responses to this question was only 652, of whom almost three-fourth (73.9%) indicated that they would do so, 12.7%

refused, and 13.3% were unsure. Subsequent questions asked respondents about the reasons for their decision, and considering them, a yes/no/unsure response was inferred to the nonresponders from their reasons given for or against an app, yielding a total to 970 responses. The remaining 6 could not be inferred, as the respondents had not provided reasons. The inclusion of the inferred decisions resulted in the proportion for "yes" being reduced to just over half (55.7%), with the proportion for "no" increasing to 23.3% and that of "unsure" to 21.0% (Table 1).

There were no significant differences in the willingness to use a contact tracing app based on ethnicity or main postcode area. However, intentions varied by sex and by age. Females were more likely to be willing to use a contact tracing app than males (*P*=.045). Younger age groups tended to be less willing to use a contact tracing app than older age groups (*P*=.01). However, when assessed by sex, age was only a significant factor for females (*P*=.01).

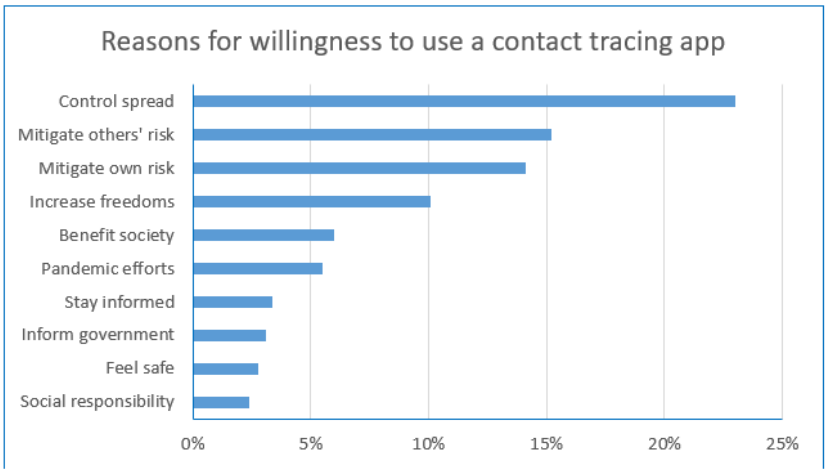
Table 1. Respondents' intentions on the use of a COVID-19 contact tracing app (N=970). This table shows the numbers and percentages of people intending to use, not intending to use, and unsure about using a COVID-19 contact tracing app.

Response type	Responses, n (%)
Direct responses (n=652)	
Yes	482 (73.9)
No	83 (12.7)
Unsure	87 (13.3)
Inferred from reasons (n=318)	
Yes	58 (18.2)
No	143 (45.0)
Unsure	117 (36.8)
Total responses (N=970)	
Yes	540 (55.7)
No	226 (23.3)
Unsure	204 (21.0)

Reasons for Being Willing to Use a Contact Tracing App

The top 10 reasons among people willing to use a contact tracing app are shown in Figure 3. Since the reasons were not requested to be ranked, we treated all reasons equally. As shown, the top reason was to control spread of the virus, followed by mitigating risk for others and for oneself, and a desire to increase freedoms.

Figure 3. Respondents' reasons for their willingness to use a COVID-19 contact tracing app. Participants provided their own, open responses on their reasons for being willing to use a contact tracing app. The top 10 reasons are shown here as percentages of those willing (N=540).



Quoted Responses From Willing Respondents

Some quoted responses from willing participants are given here as illustrations of their viewpoints.

Controlling Spread

> *I feel it will be an essential part of combating the spread of the virus, and gives me an element of control and decision making.*

We need to emulate those countries that have managed to control their pandemic by use of this type of technology.

To help stop the spread of COVID-19 and to help inform decision to ease lockdowns.

Mitigating Risk

I wish to be able to move safely in my residential area; wish to know if I have been in contact with anyone diagnosed with the virus; wish to keep up to date with latest developments.

For my own peace of mind.

For my own safety and that of others & so that scientists have good data.

Increasing Freedoms

I would use one as I am keen to get the country going again.

Want lockdown to end, want pubs back open.

To help to overcome Covid, get back to work and enable a more normal life.

Other Reasons

We collectively owe it to our country to participate in track and trace to improve our chances of getting on top of Covid19.

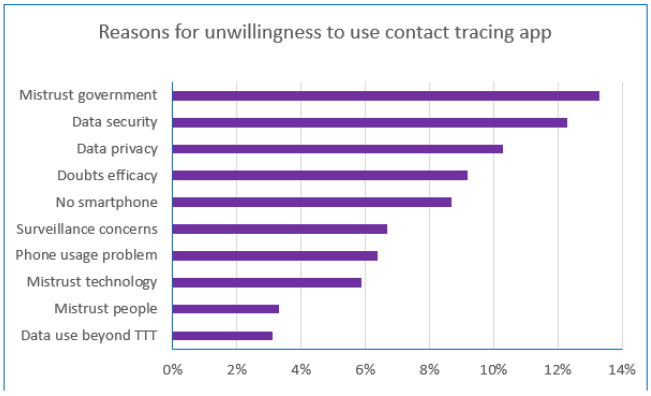
Effective contact tracing, alongside widespread virus testing, is the best answer we have to managing the ongoing Covid 19 pandemic.

The sooner everyone takes responsibility for learning as much as possible about COVID 19 the sooner we'll control it.

Reasons for Being Unwilling to Use a Contact Tracing App

The top 10 reasons among those unwilling to use a contact tracing app are shown in Figure 4. Again, all reasons were treated equally. As shown, the top reason was mistrust in the government, followed by concerns about data security, data privacy, and app efficacy.

Figure 4. Respondents' reasons for their unwillingness to use a COVID-19 contact tracing app. Participants provided their own, open responses on their reasons for being unwilling to use a contact tracing app. The top 10 reasons are shown here as percentages of those unwilling (N=226). TTT: test, track, and trace.



Quoted Responses From Unwilling Respondents

Some quotations from unwilling participants are given here as illustrations of their viewpoints.

Mistrusting Government

I do not like the idea of the Government storing my data on a centralised system.

Creepy, 1984 stuff. Given how incompetent and chaotic Westminster's response to the virus has been so far, I wouldn't bet on information remaining confidential.

I have zero faith that the Westminster administration would not use the data for purposes other than tracing the virus.

Data Security

An app forcibly enabling Bluetooth which is inherently insecure is not something I will let happen on my phone.

Not confident my information won't be hacked or misused.

Data Privacy

Insufficient evidence of preservation of privacy and lack of adverse effects on device performance / battery life / security.

I worry about personal privacy.

Doubts About Efficacy

I am concerned that I may get into a cycle of being informed to self isolate multiple times because I might have been near somebody who may have the virus as I'm shopping for my wife and several neighbours once a week at a supermarket.

Info is confusing. Not willing to self isolate until person I've been in close contact with has confirmed COVID-19.

Other Reasons

> *I am not paying for a smartphone because of the stupid [expletive] in government won't make an app that can be used on all mobile phones.*
>
> *Concerns about data held centrally.*
>
> *Mobile coverage in my area is patchy, this could be detrimental to the effectiveness of the app if you can't get a signal.*
>
> *Risk of false warnings by malicious persons.*

Change of Decision

When asked what would change their mind from being willing to being unwilling to use a contact tracing app, participants gave a range of reasons, with the top 10 reasons shown in Figure 5. As shown, the most frequent response to this question was nothing would; that is, they felt that nothing would cause them to become unwilling to use a contact tracing app. This was followed by a security breach, if the app proved ineffective, and if data were misused.

When asked what would change their mind from being unwilling to being willing to use a contact tracing app, there were a variety of reasons, with the top 10 reasons shown in Figure 6. Again, the most frequent response to this question was that nothing would; that is, they felt that nothing would cause them to become willing to use a contact tracing app. This was followed by a preference for a decentralized app, the provision of suitable tech (ie, a smartphone, a network connection, or both), and assurances of data safety.

Figure 5. Participants' responses to what would change their mind on app use from willing to unwilling. Participants who expressed willingness to use an app (N=540) provided their own, open responses on what would make them change to being unwilling to use a contact tracing app. The top 10 reasons are shown here. TTT: test, track, and trace.

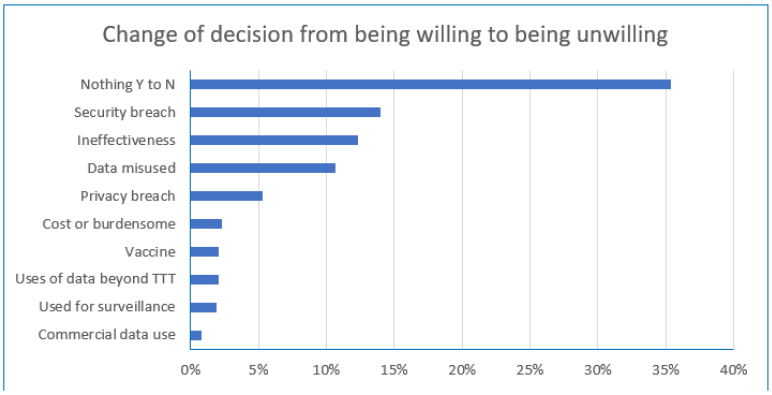
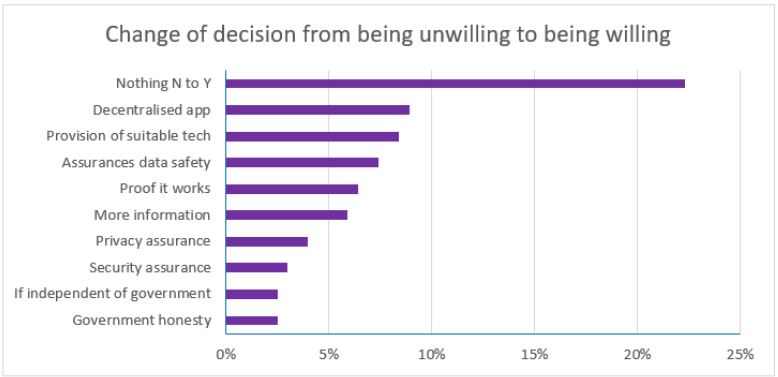


Figure 6. Participants' responses to what would change their mind on app use from being unwilling to being willing. Participants who expressed unwillingness to use an app (N=226) provided their own open responses on what would make them change their mind to being willing to use a contact tracing app. The top 10 reasons are shown here.



Unsure Respondents

As would be expected, respondents who were unsure (n=204) about using an app gave mixed reasons for and against app use. In response to the question of what would lead to a change of mind (or for them, decision-making), the most frequent response by far was the need for more information (27.5%). This was followed by a preference for a decentralized app (6.4%), being unsure what could change their mind (5.9%) and proof that the app is functional (5.4%). Some quotes from unsure participants are given here as illustrations of their viewpoints. As shown, respondents needed further information in a variety of areas including how the app would function, the data to be collected and its use, the risks to privacy, data security, and the impact on the phone battery, location tracking, and data usage.

More Information

> *Much more detailed understanding about how it works, and credibility of the organisation launching/running the app.*
>
> *Info about how it works and why it matters.*

More information on impact upon phone battery life and privacy (location tracking).

Very detailed, clear, public explanation of the App's findings.

More knowledge and a better understanding of the way it works.

Full disclosure about what is and isn't tracked and stored, and confidence in the people evaluating this and reporting it.

Complete honesty as to what happens to all the personal data collected.

I would happily use it if I understood more about what the potential risk to my personal information is and how that is mitigated.

Decentralized App

Decentralised data handling and storage.

Legally binding commitments on how data will be used and how long it will be stored. Fully anonymised decentralised system.

I'll use one once the tech has had a chance to bed in. I'd much prefer to use one which is coordinated with those in use in other countries so as to facilitate travel.

Being Unsure

Not sure but if there was some way to ensure the data would be safe (don't trust gov to do as they say necessarily).

Not sure cos all authorities lie.

I'm not sure, a lot of reassurance that it's secure.

Proof it Works

Good evidence that it works and is safe.

Independent confirmation of adequate security and usefulness of the app.

If it was widely used and therefore accurate.

Other Reasons

A guarantee that the information would not be used for anything else and it was secured.

Assurance in law that my data would be solely used for contact tracing and that no private companies would have the right to hold or use my data.

Some very clear advice on how to install, use, etc. Support with what kind of phone is needed.

The price of a smart phone. It might be just easier for me to wear a mask and maintain physical distancing.

Discussion

What This Study Adds

To our knowledge, this is the only study on the use of a COVID-19 contact tracing app to use a mixed methods approach and combine qualitative and quantitative data collection and analysis at this scale. This is also the only known study focusing on Welsh citizens, thus adding Wales to the countries studied. In particular, allowing participants to provide open responses on their reasoning, and what would cause them to change their mind, demonstrated the value of asking more than closed questions and provided rich information at scale to augment the simple categorical answers. This information is important if policy and other decision-makers are to address and respond to concerns and to support use of a contact tracing app.

Principal Findings

Our mixed methods survey among citizens of Wales found that over half (55.7%) of the respondents explicitly stated they planned to use a contact tracing app; a further 21% were unsure and just less than a quarter (23.3%) stated that they were unwilling. These values are based on actual responses plus inferred choices for those participants who did not answer the question. In contrast with the other closed questions in our survey, which were completed by over 95% of people, the response rate for this question was only 66%. This is interesting as it suggests some reluctance to respond to this question; nonetheless, almost all nonresponders gave reasons for being for, against, or unsure of using a contact tracing app in the free-text responses. The reasons for this are unknown, but it might indicate forms of response bias, such as acquiescence or social desirability, since after inferring from reasons, the proportions shifted toward unwillingness and unsureness. By comparison, almost three-fourth (74.8%) of participants in a multi-country survey using Likert scales stated they would probably or definitely download a contact tracing app [5]; over 67.5% of US citizens and 84% of Irish citizens indicated that they would probably or definitely download an app [6,7]. The US survey was also conducted in the United Kingdom, Germany, Italy, and France, with definite and probable intention to use rates at least as high as or higher than those in the United States [13].

In February 2021, the NHS Test and Trace program released the first detailed data about app use since it was rolled out in England and Wales in September 2020, and it reported that 21.7 million people had downloaded the app [14]. With the population of England and Wales being 59.4 million individuals [15], this indicates that 36.5% of the population downloaded the app. We refer to the combined figures for England and Wales because separate figures for Wales alone were not reported. The actual download figures are considerably lower than those found in our survey or in other surveys conducted in the United Kingdom [5,13]. It is widely recognized that a majority uptake is needed for optimum app efficacy; however, thus far, the figures fall far short. The reasons for this are not known, but they might be partly owing to varying representativeness, a tendency among respondents to provide the survey response seen as desirable, changes in viewpoints over time, and intention not being borne out by action for any reason.

In our survey, the top reasons in favor of app use were controlling spread of the virus, mitigating risks for others and for oneself, and increasing freedoms to enable society to open up. By comparison, the top reasons among some other surveys, which were based on predefined choices were as follows. In the US survey [6], protecting family and friends, knowing about

the risk of infection, stopping the epidemic, and staying healthy were the top reasons. In the multi-country survey conducted by Altmann et al [5], the top reasons were protecting family and friends, stopping the epidemic, social responsibility, and knowing about the risk of infection. In the survey in Ireland [7], the top reasons were protecting family and friends, social responsibility, knowing about the risk of infection, and protecting oneself [7]. The top reasons for being against app use in our survey were mistrusting the government, concerns about data security and privacy, and doubts about app efficacy. Among other surveys, the top reasons were as follows. In the US survey conducted by Abeler et al [6]: concerns about government surveillance post pandemic, that no one else would use the app, phone being hacked, and increased anxiety. In the study by Altmann et al [5], the top 2 reasons were the same as those in the US survey, followed by increased anxiety and it being a major inconvenience to install the app. In O'Callaghan and colleagues' study [7], the top reasons were surveillance by technology companies after the pandemic, a response that none of the options apply, government surveillance post pandemic, and the phone being hacked. Although category names differ, it can be seen that the top reasons for or against app use are similar across our and other surveys.

We included a question in our survey on what could lead people to change their mind on app use as we expected it to yield interesting results, given that this was a first-of-its-kind app and global context. As observed, 24% of people who downloaded the NHS app in England and Wales are not using it, which indicated a change of mind from being willing to being unwilling [16]. In the government data release showing 21.7 million people had downloaded the app, it was also revealed only about 16.5 million people (27.8%) were currently actively using it. It has been proposed that this discrepancy is due to a combination of people turning off the contact tracing capability, uninstalling the app or never actually activating it [14,16]; but the reasons for the change are not known. Of those in our survey willing to use an app, the top response for changing their mind from being willing to being unwilling was that nothing would cause them to do so, indicating their firm intentions. This was followed by assessing whether there was a breach in data security, if the app proved to be ineffective, and if their personal data were misused. These stated reasons may shed some light on the loss of app users as fear of breaches and misuse, as well as low uptake figures being seen as poor efficacy. The latter point is somewhat ironic and could create a self-fulfilling prophecy, one person at a time [9]. The same response (ie, that nothing would cause them to change their mind) was also the most frequent for those unwilling to change their mind to be willing. This was followed by the use of a decentralized app, being provided with suitable tech (ie, a smartphone, a network connection, or both), and data safety assurance. Since at the time our survey was conducted the intention was to introduce a centralized app in England and Wales, the subsequent change to and roll-out of a decentralized app might at least partly reduce unwillingness. However, the other difficulties and concerns remain to be addressed. Among the unsure, it was the need for more and clearer information, which can be seen as positive as it suggests it can be remedied through better communication. The actual reasons for the 24% loss in app users are not known, nor what proportion of people had previously been against using the app but had changed their mind to be in favor. Further studies are required to obtain information on these questions.

Since contact tracing apps require the use of a smartphone and a suitable network connection, digital inclusion and exclusion and underlying links with socioeconomic status, are important factors. In our survey, although 91.5% of overall respondents reported being smartphone users, this was only 78% among individuals aged 76-85 years and only 40% among those aged over 85 years. Not having, not wanting to, or being unable to buy, a smartphone, difficulties in using their smartphone, lack of knowledge on how to download and use an app, and lack of a reliable network connection were among the free-text reasons given for not using the app. At least one of these reasons was given by almost 10% (n=93) of our respondents. In April 2020, the Ada Lovelace Institute [17] published a rapid evidence review on the technical considerations and societal implications of using technology to transition from the COVID-19 crisis [18]. This included consideration of various issues relating to the use of contact tracing apps, among which were the potential exclusion of vulnerable groups and exacerbation of pre-existing health inequalities. The report highlighted that the effectiveness of digital contact tracing needed to be established, that effectiveness relies on a high level of accuracy and ubiquity, and is dependent on public trust and confidence. It further warned about societal and financial implications for individuals required to self-isolate and the possibility of fake contact warnings and other scams. The report concluded that there was (at the time of publication) insufficient evidence to support the use of digital contact tracing as an effective technology to support the pandemic response. It recommended clear government commitment to the following: privacy by design in app development and function, robust regulation and oversight, time limitation on contact tracing, purpose limitation in data use, clear guidance on the enforcement and use of digital contact tracing, and transparency to enable public scrutiny [18]. These concerns accord with many of those raised by our survey respondents. The data release on app use showed that 1.7 million people in England and Wales had been told to self-isolate as a result of using the NHS app, which health ministers estimate has prevented about 600,000 cases of the disease [16]. This is certainly good news, as is the change from a centralized to decentralized app model with regard to the preferences of our respondents. However, little is known about government achievements on other recommendations and on public involvement. Accepting that COVID-19 is having widespread and unequal serious impacts on individuals and societies, there are still ethical issues, such as the relationship between liberty and privacy to be addressed, and it has been shown that moral reasoning plays an important part in decision-making on app use [4,9]. Considering that government mistrust was the most frequent reason given by our respondents unwilling to use an app, and it was high among the reasons in other surveys, policy makers and other decision-makers need to increase efforts to engage with citizens, provide clearer information and act transparently if societies are to get the best from smartphone-based contact tracing apps. These issues will only become more important if added functionalities are introduced, such as vaccine status which is under discussion in the United

Kingdom [19], in addition to the need for ongoing monitoring, since we have long progressed beyond common early thinking that the COVID-19 pandemic would end quickly.

Limitations

As is common in one-time surveys, our work was based on a defined period and, as such, presents a snapshot of citizen views at that time. Further, the timing of the survey may have had an impact as it was early in the "first wave" of the pandemic and at a time when the NHSX model was the only model being tested at scale in the United Kingdom. As the data were collected anonymously, we cannot repeat the survey with the same respondents to compare their intentions with their actual decisions on the use of a contact tracing app. We acknowledge that our respondents are not fully representative of the people of Wales in terms of age profile, digital literacy, and ethnic heterogeneity. Other survey models were piloted with the aim of hearing from underrepresented groups, with some success, but the results are not reported here owing to adaptations in method reducing the viability of comparison with the HWW cohort.

Recommendations

The following are some recommendations arising from our survey to inform decisions on enhancing the use of a contact tracing app to promote its effectiveness and build public trust. Although these arose from a survey with people in Wales, they are more widely applicable as in accordance with survey findings from other countries.

Concurrent with the transparency in a democratic society, there should be more engagement with the public to gain viewpoints, listen to concerns, and provide more information. This would also benefit decision-makers in developing transparent policy plans with social license.

There is an issue with digital inclusion among some groups, such as older people, being less likely to use a smartphone. In some cases, it is the lack of a smartphone or stable network connection, but for others it is a lack of knowledge on app use. For the latter, this could be at least partly addressed by an education program with straightforward information and a step-by-step guide to download and use the app.

The reasons people gave for being willing to use a contact tracing app demonstrate a keenness to help themselves, others, society, and the government to avoid the virus and control its spread. However, the reasons they might change their mind, notably, the need to safeguard against security breaches and data misuse, and to be able to demonstrate app effectiveness are critical to trust and success. Regularly updated reliable information is crucial to this.

The reasons people gave for their unwillingness to use an app were topped by mistrust in the government, followed by concerns about data security and privacy and the efficacy of the app. Policy and decision makers must address these issues and demonstrate trustworthiness if members of the public are to be confident their data are safe and that using an app is worthwhile.

In summary, we recommend greater public involvement in the development and implementation of policy and technologies from the outset and on an ongoing basis.

Future Work

As a separate question alongside the survey, we asked respondents to indicate which topics interested them for an in-depth discussion and to email us outside the survey if they would like to take part. These topics were as follows: (1) what counts as acceptable use of digital technologies including apps, (2) the development and implementation of Wales-specific policy responses to COVID-19, (3) the potential benefits and challenges of using personal data gathered in the COVID-19 response for research purposes beyond the pandemic, (4) public engagement with proposed government strategies prior to implementation (and ongoing), (5) the impact of digital technologies introduced in response to the COVID-19 crisis on disadvantaged groups, and (6) the ethical challenges of designing, developing, and implementing technologies that support the exit strategy. The most frequently chosen was topic 3. Accordingly, we have embarked on deliberative public involvement [20] to ascertain public views on this topic to present to decision-makers in due course. The adapted survey formats (mentioned above) and their findings will be reported in a separate study.

Conclusions

This is the only known citizen survey on the use of contact tracing apps to use a mixed methods approach, combining qualitative and quantitative data collection and allowing respondents to suggest their own reasons for and against app use, plus what would cause them to change their decision. Our findings show that citizens are intent on helping themselves, others, society, and government to avoid the virus and control its spread. The fact that contact tracing apps are necessarily smartphone-based raises issues of digital inclusion, such that participation is precluded for individuals who do not have a smartphone, have difficulty using one, or lack a stable network connection. However, the most prominent concerns raised about app use, namely, mistrusting the government, concerns about data security and privacy, and doubts about efficacy, could be addressed by greater efforts by policy and decision-makers to act openly, provide clearer information, and demonstrate trustworthiness. These actions are essential if the potential of contact tracing apps in contributing to controlling the pandemic are to be realized and may be useful in the ethical development and roll out of other health apps.

Acknowledgments

We would like to express our grateful thanks to the citizens of Wales who took part in the survey. We gratefully acknowledge the assistance of HealthWise Wales in administering the survey and disseminating findings. The work conducted in this study

was supported by Administrative Data Research (Wales) funded by the Economic and Social Research Council (ref: ES/S007393/1), and the SAIL Databank funded by Health & Care Research Wales (this funder does not use reference numbers).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions.
[DOCX File , 16 KB - mhealth_v9i11e29181_app1.docx]

References

1. Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2019. Office for National Statistics. 2020 Jun 24. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2019estimates [accessed 2021-11-10]
2. Rapid COVID-19 virology - Public. Public Health Wales Health Protection. URL: https://public.tableau.com/profile/public.health.wales.health.protection#!/vizhome/RapidCOVID-19virology-Public/Headlinesummary [accessed 2021-11-10]
3. Senedd Cymru Welsh Parliament. URL: https://senedd.wales/ [accessed 2021-11-10]
4. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020 Jul;46(7):427-431 [FREE Full text] [doi: 10.1136/medethics-2020-106314] [Medline: 32366705]
5. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of App-Based Contact Tracing for COVID-19: Cross-Country Survey Study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
6. Abeler J, Altmann S, Milsom L, Nosenzo D, Toussaert S, Zillessen H. Support in the US for app-based contact tracing of Covid-19. URL: https://files.osf.io/v1/resources/7vgq9/providers/osfstorage/5e94831fd69735038abdbb27?action=download&direct&version=1 [accessed 2021-11-10]
7. O'Callaghan ME, Buckley J, Fitzgerald B, Johnson K, Laffey J, McNicholas B, et al. A national survey of attitudes to COVID-19 digital contact tracing in the Republic of Ireland. Ir J Med Sci 2021 Aug;190(3):863-887 [FREE Full text] [doi: 10.1007/s11845-020-02389-y] [Medline: 33063226]
8. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 Contact Tracing Apps: Predicted Uptake in the Netherlands Based on a Discrete Choice Experiment. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e20741 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]
9. Williams S, Armitage C, Tampe T, Dienes K. Public attitudes towards COVID-19 contact tracing apps: A UK-based focus group study. Health Expect 2021 Apr;24(2):377-385 [FREE Full text] [doi: 10.1111/hex.13179] [Medline: 33434404]
10. HealthWise Wales Research outcomes. HealthWise Wales. URL: https://www.healthwisewales.gov.wales/ [accessed 2021-11-10]
11. Hurt L, Ashfield-Watt P, Townson J, Heslop L, Copeland L, Atkinson MD, et al. Cohort profile: HealthWise Wales. A research register and population health data platform with linkage to National Health Service data sets in Wales. BMJ Open 2019 Dec 02;9(12):e031705 [FREE Full text] [doi: 10.1136/bmjopen-2019-031705] [Medline: 31796481]
12. New daily cases across the UK. ZOE COVID Study. URL: https://covid.joinzoe.com/ [accessed 2021-11-10]
13. Milsom L, Abeler J, Altmann S, Toussaert S, Zillessen H, Blasone R. Survey of acceptability of app-based contact tracing in the UK, US, France, Germany and Italy. Center for Open Science. URL: https://osf.io/7vgq9/ [accessed 2021-11-10]
14. NHS Test and Trace (England) statistics: 21 January to 27 January 2021. Government of the United Kingdom. 2021 Feb 04. URL: https://www.gov.uk/government/publications/nhs-test-and-trace-england-statistics-21-january-to-27-january-2021 [accessed 2021-11-10]
15. Population of the United Kingdom in 2020, by region. Statista. URL: https://www.statista.com/statistics/294729/population-united-kingdom-uk-by-country/ [accessed 2021-11-10]
16. Covid-19: NHS app has told 1.7 million to self-isolate. BBC News. URL: https://www.bbc.co.uk/news/technology-55985562 [accessed 2021-11-10]
17. Ada Lovelace Institute. URL: https://www.adalovelaceinstitute.org/about/ [accessed 2021-11-10]
18. COVID-19 rapid evidence review. Ada Lovelace Institute. URL: https://www.adalovelaceinstitute.org/evidence-review/covid-19-rapid-evidence-review-exit-through-the-app-store/ [accessed 2021-11-10]
19. Elgot J. NHS app could show Covid vaccine status or latest test result. The Guardian. 2021 Feb 23. URL: https://www.theguardian.com/society/2021/feb/23/nhs-app-could-show-covid-vaccine-status-or-latest-test-result [accessed 2021-11-10]
20. Elstub S, Thompson R, Escobar O, Hollinghurst J, Grimes D, Aitken M, et al. The Resilience of Pandemic Digital Deliberation: An Analysis of Online Synchronous Forums. Javnost - The Public 2021 Oct 04;28(3):237-255. [doi: 10.1080/13183222.2021.1969616]

Abbreviations

HWW: HealthWise Wales
NHS: National Health Service

Edited by L Buis; submitted 29.03.21; peer-reviewed by J Ziegenfuss, T Ntalindwa, S Rostam Niakan Kalhori, A Teles; comments to author 17.08.21; revised version received 13.09.21; accepted 14.10.21; published 22.11.21.

Please cite as:
Jones K, Thompson R
To Use or Not to Use a COVID-19 Contact Tracing App: Mixed Methods Survey in Wales
JMIR Mhealth Uhealth 2021;9(11):e29181
URL: https://mhealth.jmir.org/2021/11/e29181
doi:10.2196/29181
PMID:34698645

Original Paper

Physicians' Attitudes Toward Prescribable mHealth Apps and Implications for Adoption in Germany: Mixed Methods Study

Florian Dahlhausen[1], MSc; Maximillian Zinner[1], MSc; Linn Bieske[2], MSc; Jan P Ehlers[1], DVM; Philip Boehme[2], MD; Leonard Fehring[2], MD

[1]Didactics and Educational Research in Health Care, Faculty of Health, School of Medicine, Witten/Herdecke University, Witten, Germany
[2]Faculty of Health, School of Medicine, Witten/Herdecke University, Witten, Germany

Corresponding Author:
Leonard Fehring, MD
Faculty of Health
School of Medicine
Witten/Herdecke University
Alfred-Herrhausen-Strasse 50
Witten, 58448
Germany
Phone: 49 2302 926 78608
Email: leonard.fehring@uni-wh.de

Abstract

Background: In October 2020, Germany became the first country, worldwide, to approve certain mobile health (mHealth) apps, referred to as DiGA (Digitale Gesundheitsanwendungen, in German, meaning digital health applications), for prescription with costs covered by standard statutory health insurance. Yet, this option has only been used to a limited extent so far.

Objective: The aim of this study was to investigate physicians' and psychotherapists' current attitudes toward mHealth apps, barriers to adoption, and potential remedies.

Methods: We conducted a two-stage sequential mixed methods study. In phase one, semistructured interviews were conducted with physicians and psychotherapists for questionnaire design. In phase two, an online survey was conducted among general practitioners, physicians, and psychotherapists.

Results: A total of 1308 survey responses by mostly outpatient-care general practitioners, physicians, and psychotherapists from across Germany who could prescribe DiGA were recorded, making this the largest study on mHealth prescriptions to date. A total of 62.1% (807/1299) of respondents supported the opportunity to prescribe DiGA. Improved adherence (997/1294, 77.0%), health literacy (842/1294, 65.1%), and disease management (783/1294, 60.5%) were most frequently seen as benefits of DiGA. However, only 30.3% (393/1299) of respondents planned to prescribe DiGA, varying greatly by medical specialty. Professionals are still facing substantial barriers, such as insufficient information (1135/1295, 87.6%), reimbursement for DiGA-related medical services (716/1299, 55.1%), medical evidence (712/1298, 54.9%), legal uncertainties (680/1299, 52.3%), and technological uncertainties (658/1299, 50.7%). To support professionals who are unsure of prescribing DiGA, extended information campaigns (1104/1297, 85.1%) as well as recommendations from medical associations (1041/1297, 80.3%) and medical colleagues (1024/1297, 79.0%) were seen as the most impactful remedies.

Conclusions: To realize the benefits from DiGA through increased adoption, additional information sharing about DiGA from trusted bodies, reimbursement for DiGA-related medical services, and further medical evidence are recommended.

(JMIR Mhealth Uhealth 2021;9(11):e33012) doi:10.2196/33012

KEYWORDS

mobile health; mHealth; digital health; apps; physicians; general practitioners; technology acceptance; adoption

Introduction

Health care systems worldwide are struggling with rising costs [1]. Great hopes are being pinned on digital health, such as mobile health (mHealth) apps, to address the root causes of these burdens [2]. mHealth apps are said to have great potential for improving health outcomes in numerous ways [3] (eg, increased health competence [4], better symptom management

[5], and improved adherence to chronic disease management [6]). Despite these benefits, several factors are hindering widespread adoption of mHealth solutions, including technological, social, and organizational factors [7], limited reimbursement [8,9], and further need for empirical research on the effectiveness of mHealth [10].

To address some of these challenges, in October 2020, Germany became the first country, worldwide, to grant statutorily insured individuals an entitlement to use certain mHealth apps at the expense of health insurers [11]. These apps are referred to as DiGA (Digitale Gesundheitsanwendungen, in German, meaning digital health applications), a subset of the over 280,000 health, fitness, and medical apps available worldwide at the end of 2020 [12,13]. DiGA are medical devices primarily based on digital technologies that support the detection, monitoring, treatment, mitigation, or compensation of disease, injury, or disability. Additionally, they must have successfully cleared an assessment of positive care effects and product qualities—most importantly, safety and suitability for use, data protection and security, and interoperability—by BfArM (Bundesamt für Arzneimittel und Medizinprodukte, in German, meaning the German Federal Institute for Drugs and Medical Devices) [14]. All such apps would then be included in the official DiGA directory of prescribable, reimbursable apps.

As app reimbursement is only possible when prescribed by a physician or psychotherapist or when approval had been directed by the health insurer, health care professionals—especially in the outpatient care sector—play an important role in the implementation process [15]. Five months after their introduction, only 3700 DiGA had been prescribed and reimbursed, increasing to 17,000 DiGA by 10 months after their introduction [16,17].

Vast research has investigated the technological, structural, and human factors that may influence technology adoption by health care professionals [18], most prominently through innovation adoption and diffusion theories by Rogers [19], the technology acceptance model [20], and the unified technology acceptance and use of technology theory [21]. What followed was empirical work introducing various country-specific surveys on health care professionals' mHealth adoption [22-24] as well as studies focused on specific medical disciplines and technologies, ranging from telemedicine and remote monitoring [25,26] to medical app use [27].

To our knowledge, no study has systematically examined adoption of mHealth apps by physicians and psychotherapists in the outpatient care sector—referred to as health care professionals in the following sections—in the context of institutionalized programs with reimbursement of government-certified, prescribable apps, as is the case with DiGA in Germany. This study aims to fill this gap by analyzing health care professionals' attitudes and prescription intentions toward DiGA, as well as barriers to adoption and potential remedies. It includes findings from the largest survey on mHealth adoption by health care professionals in Germany. Given Germany's unique and leading approach to mHealth app adoption, the findings can be applied to other countries looking to expand access to mHealth apps.

Methods

We used a mixed methods approach consisting of semistructured interviews followed by an online survey, which was developed based on the findings of the initial qualitative interviews.

Exploratory Interviews for Survey Questionnaire Design

We first conducted a structured literature review of both existing technology adoption literature and global case studies. Drawing on these bodies of literature, we developed a semistructured interview guide for interviews with physicians and psychotherapists about their views toward and experiences with DiGA (Multimedia Appendix 1). To ensure that a vast variety of profiles and views on DiGA were represented, we used a purposive sampling approach to identify heterogeneous interviewees across various age groups, medical specializations, attitudes toward digitization, and geographic locations in Germany.

Interviews were conducted one-on-one by three independent researchers via video conference, telephone, or face-to-face. Interviews were conducted until all researchers agreed that further interviews were unlikely to surface major new viewpoints or topics. In total, 18 interviews with physicians and psychotherapists were conducted. These lasted between 25 and 60 minutes and covered four question categories: (1) attitudes toward DiGA, (2) prescription behavior and intentions, (3) barriers to DiGA prescription, and (4) potential remedies.

During each interview, interviewers wrote extensive notes. These were subsequently aggregated and reviewed by an expert panel consisting of five members with multi-professional backgrounds in medicine, natural sciences, and business and used for survey questionnaire design. In the first round of iteration, 38 survey questions were generated. These were prioritized in the second round of iteration, resulting in 25 questions. Next, answer options were developed based on the results from the qualitative interviews. Questions were also rephrased as Likert-scale items, most often with responses ranging from 1 (strongly disagree) to 5 (strongly agree).

Online Survey

We next conducted a cross-sectional survey investigating health care professionals' interactions with DiGA along four key categories discussed in the qualitative interviews. To establish a similar understanding of DiGA compared to general health and wellness apps among all survey respondents, an introductory information page about DiGA was displayed. We pretested the survey questionnaire with five colleagues and additional health care professionals to ensure survey comprehensibility and clarity. Question wording, survey functionality, and/or the introductory information page about DiGA were adjusted after each pretest, where necessary. The final questionnaire (Multimedia Appendix 2) was administered using Qualtrics, a web-based survey tool [28].

The survey was conducted over a 6-week period between December 2020 and January 2021 in accordance with the Checklist for Reporting Results of Internet E-Surveys

(CHERRIES) guidelines [29]. A total of 19,196 German general practitioners, physicians, and psychotherapists were invited to participate in the anonymous online survey via their publicly accessible professional email addresses. To preserve privacy, respondents were not asked to provide any identifiable information. Additionally, we did not track which invited participants had started or completed the survey, limiting our ability to use reminders. To motivate participation, respondents could download a comprehensive, custom-made information package about DiGA for health care professionals after survey completion, addressing the various uncertainties and questions about DiGA that surfaced during our qualitative interviews and pretests. No financial incentive was offered.

In addition to insights from our qualitative interviews, we report findings from 17 out of 25 online survey questions asked. Besides descriptive analyses, dependencies between health care professional characteristics and attitudes toward DiGA as well as the likelihood of prescribing were tested in RStudio (version 1.3.1056) using chi-square tests or, when conditions for using chi-square tests were not met, Fisher exact tests with Monte Carlo approximation and 2000 replicates [30,31]. If respondents did not answer a particular question, they were excluded from the total number of respondents of this question in the analysis.

Ethics Approval

This study was approved by the ethics committee of Witten/Herdecke University (reference No. 278/2020).

Results

Qualitative Interviews

Most respondents viewed DiGA positively. More flexible access to care independent of a practice's opening hours and availability of therapy location, patient empowerment through increased sense of responsibility and self-efficacy, and improved adherence emerged as key potential benefits. While respondents had some experience with general mHealth apps, no respondent had prescribed DiGA so far. Some were generally open to doing so in the future. Yet, all respondents saw substantial barriers associated with prescribing DiGA, most importantly, lack of information, uncertainties regarding therapeutic benefits and medical evidence, and technical concerns. For some respondents, the low number of available DiGA relevant to their practice posed an additional barrier. All interviewees highlighted the desire to be informed more broadly. Some interviewees also called for stronger medical evidence and better compensation of services related to DiGA. These findings were further tested in the subsequent online survey.

Online Survey

Demographics

A total of 1308 health care professionals completed the questionnaire, with minor nonresponse to individual questions, corresponding to a response rate of 7%, in line with previous research [23,32,33], making this the largest study on health care professionals' mHealth adoption in Germany so far.

As shown in Table 1, the median age of respondents was 46 to 55 years, with 52.7% (682/1295) male and 47.2% (611/1295) female respondents, both representative of the overall German medical profession [34]. Most respondents hailed from urban areas (76.8%), predominately medium-sized cities between 20,000 and 100,000 inhabitants (406/1298, 31.3%), large cities between 100,000 and 500,000 inhabitants (304/1298, 23.4%), followed by small cities under 20,000 inhabitants (287/1298, 22.1%). A vast majority of respondents (1260/1296, 97.2%) were active in outpatient settings. About half of the respondents were active in single practices without physician and psychotherapeutic colleagues (613/1268, 48.3%), while the other half (655/1268, 51.7%) worked jointly with at least one colleague, a fact in line with doctors and psychotherapists in Germany overall [34]. Nearly all responding health care professionals participated in the German statutory health insurance scheme, although 93.6% (1171/1251) also accepted privately insured patients.

Table 1. Distribution of sociodemographic characteristics of all general practitioners, physicians, and psychotherapists (N=1308) who participated in the survey.

Characteristic	Respondents, n (%)
Age in years (n=1295)	
<26	1 (0.1)
26-35	49 (3.8)
36-45	233 (18.0)
46-55	415 (32.0)
56-65	477 (36.8)
>65	120 (9.3)
Gender (n=1295)	
Male	682 (52.7)
Female	611 (47.2)
Diverse	2 (0.2)
Practice location size: inhabitants (n=1298)	
<5000	85 (6.5)
5001-20,000	287 (22.1)
20,001-100,000	406 (31.3)
100,001-500,000	304 (3.4)
>500,000	216 (16.6)
Practice type (n=1296)	
Hospital	28 (2.2)
Single practice	613 (47.3)
Joint practice	647 (49.9)
Other occupation	8 (0.6)
Practice size: practicing physicians or psychotherapists (n=1268)	
1	613 (48.3)
2	270 (21.3)
3	139 (11.0)
4	101 (8.0)
5	41 (3.2)
6-10	64 (5.0)
>10	40 (3.2)
Patient population (n=1251)	
Statutory health insurance only	70 (5.6)
Private health insurance only	10 (0.8)
Both statutory and private health insurance	1171 (93.6)
Medical specialty (n=1260)	
Anesthesiology	24 (1.9)
Child and adolescent psychiatry and psychotherapy	61 (4.8)
Dermatology	22 (1.7)
Ear, nose, and throat medicine	38 (3.0)
General medicine	284 (22.5)
Gynecology	65 (5.2)

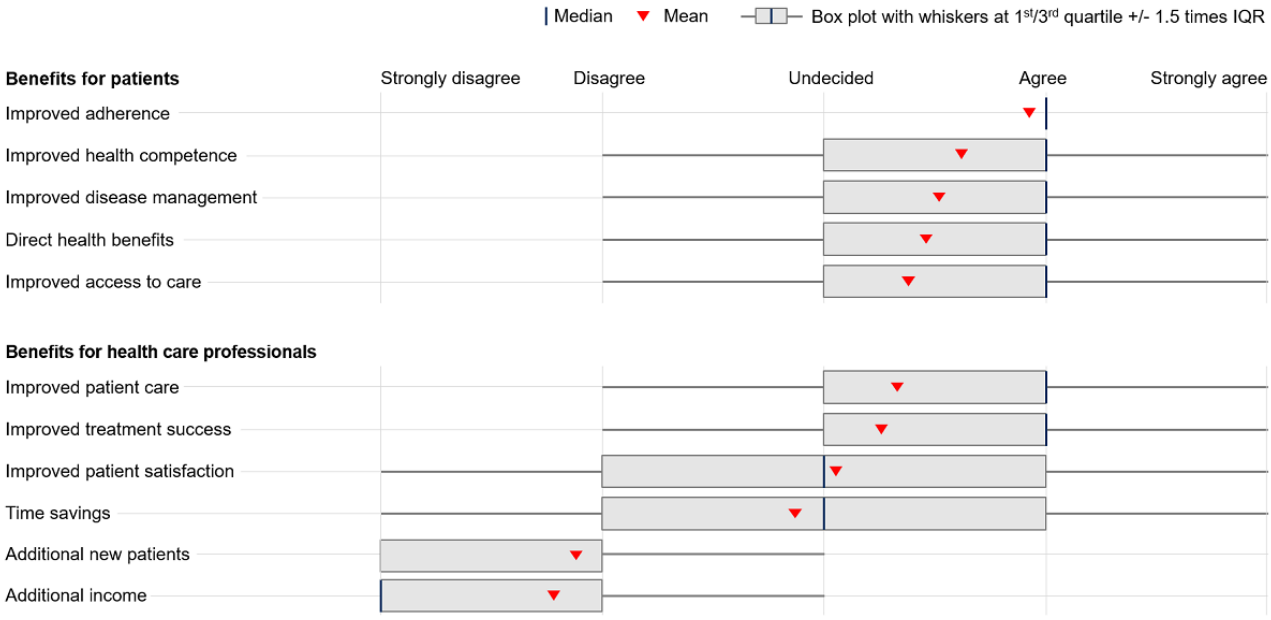
Characteristic	Respondents, n (%)
Internal medicine	130 (10.3)
Neurology	19 (1.5)
Ophthalmology	18 (1.4)
Orthopedics and trauma surgery	44 (3.5)
Pediatrics	50 (4.0)
Psychiatry and psychotherapy	65 (5.2)
Psychological psychotherapy	264 (21.0)
Psychosomatic medicine and psychotherapy	93 (7.4)
Surgery	19 (1.5)
Urology	19 (1.5)
Other specialties	45 (3.6)

Perceived Benefits From and Attitudes Toward DiGA

A total of 62.1% (807/1299) of health care professionals viewed the fact that physicians can prescribe DiGA as positive or very positive. Only 22.6% (293/1299) viewed this recent development as negative or very negative in addition to 15.3% (199/1299) who viewed it neutrally. While health care professionals who had higher digital affinity (χ^2_{36}=126.7, *P*<.001) or were female (Fisher exact *P*=.01) held significantly more positive attitudes, the strength of the association between digital affinity, measured as self-rating for job-related digital competency or gender on the one hand and attitude towards DiGA on the other hand was rather weak (Cramer V=0.16 and 0.09, respectively). Medical specialty significantly influenced attitudes toward DiGA (Fisher exact *P*=.001; Cramer V=0.14). Other professional characteristics, such as age, practice type, size, and location and patient population, did not show significant effects on attitude.

Positive attitudes toward DiGA may be explained by the various benefits that health care professionals expect from DiGA for both patients and physicians: health care professionals who perceived greater benefits from DiGA held significantly more positive attitudes toward them (χ^2_{16}=116.5-785.3, *P*<.001; Fisher exact *P*<.001; Cramer V=0.12-0.42, depending on the individual benefit; see Figure 1 for respective benefits). On average, benefits for patients were considered to be larger than those for physicians, as shown in Figure 1. With 77.0% of respondents (997/1294), improved therapy adherence was identified as a benefit for patients most often, followed by increased health competence (842/1294, 65.0%), improved disease management (783/1294, 60.5%), direct health benefits from using DiGA (733/1295, 56.7%), and improved access to care (705/1294, 54.4%). These benefits were seen to accrue primarily among younger patients. A total of 40.7% (527/1295) of health care professionals would prescribe DiGA primarily to younger patients.

Figure 1. Perceived benefits from DiGA for patients and health care professionals. Respondents were asked to indicate the extent to which they see various benefits from DiGA on 5-point Likert scales. DiGA: Digitale Gesundheitsanwendungen (digital health applications).



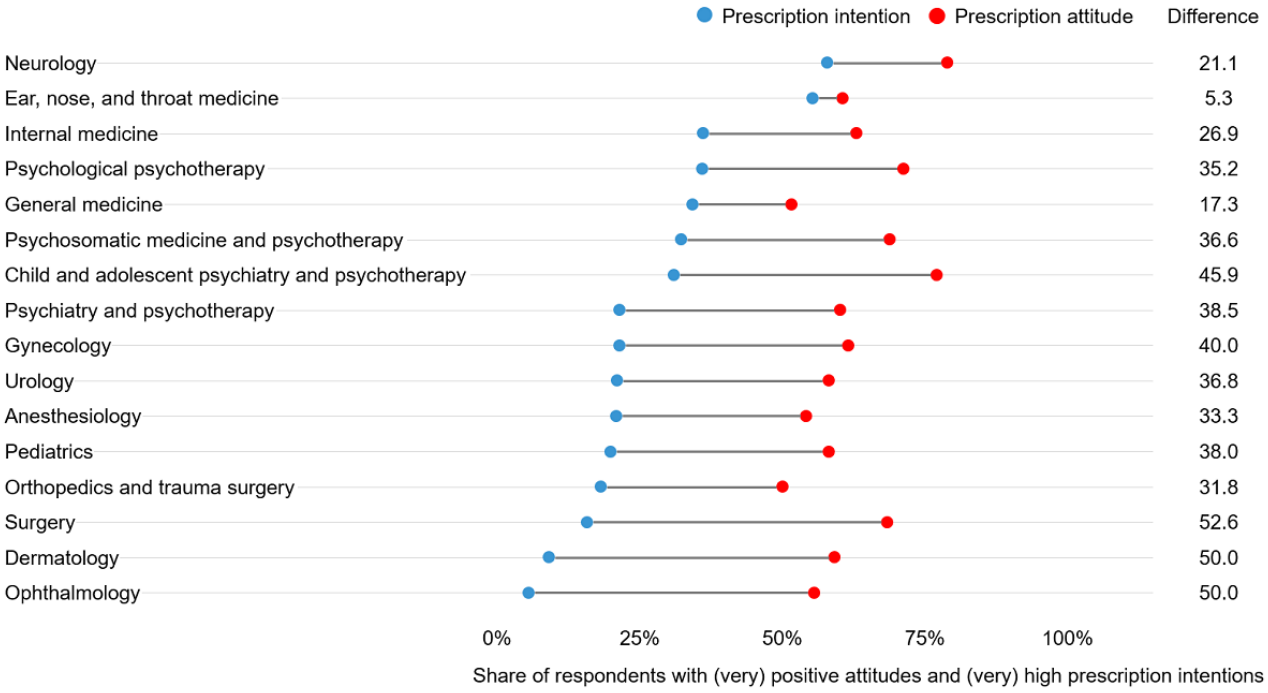
About 1 in 2 health care professionals saw improved patient care (727/1287, 56.5%) as a benefit of DiGA for physicians. Increased patient satisfaction was seen as a benefit by 43.2% of respondents (556/1287), followed by time savings (410/1287, 31.9%). Acquiring new patients (82/128, 76.3%) and receiving additional income through reimbursement for medical services related to DiGA (26/128, 72.0%) were rarely seen as benefits. At the same time, one-fifth of health care professionals (234/1287, 18.2%) indicated that they were unable to assess whether DiGA would lead to attractive reimbursement, significantly more than for other potential physician benefits.

Prescription Intentions

A large majority of health care professionals have not prescribed DiGA and did not intend to do so in the next year: less than 10% (103/1299) of health care professionals indicated that they had prescribed DiGA. Only 30.3% (393/1299) of health care professionals planned to prescribe DiGA in the next 12 months. A total of 19.9% (259/1299) were uncertain as to whether they would prescribe DiGA and 49.8% (647/1299) did not plan to do so. Those who held more positive attitudes toward DiGA (χ^2_{16}=570.3, *P*<.001; Cramer V=0.33) or saw larger benefits from DiGA (χ^2_{16}=215.4-409.0, *P*<.001; Fisher exact *P*<.001; Cramer V=0.11-0.30, depending on the individual benefit) were believed to be significantly more likely to prescribe. Apart from digital affinity (χ^2_{36}=79.0, *P*<.001; Cramer V=0.12), health care professionals' demographics were not significantly associated with prescription intentions.

Prescription intentions varied largely by medical specialty (Figure 2). Across all specialties, 30.3% (393/1299) of respondents indicated that they would be likely or very likely to prescribe DiGA in the coming year. Neurologists (11/19, 58%) and ear, nose, and throat doctors (21/38, 55%) held the highest prescription intentions. At the lower bound, only 6% (1/18) of professionals from ophthalmology intended to prescribe.

Figure 2. Prescription attitude and intention by medical specialty. Prescription attitude represents the share of respondents who expressed positive or very positive attitudes toward prescribing DiGA. Prescription intention represents the share of respondents who indicated that they would be likely or very likely to prescribe DiGA during the coming year. The difference shows the gap between prescription attitude and intention by medical specialty. See Table 1 for respective sample sizes per medical specialty. DiGA: Digitale Gesundheitsanwendungen (digital health applications).



Similar but smaller variations across specialties were found for attitudes toward DiGA prescription. Across all specialties, 62.1% (807/1299) of respondents held positive or very positive attitudes. Neurologists held the most positive attitudes toward DiGA (15/19, 79%). At the other end of the spectrum, only 50% (22/44) of orthopedists and trauma surgeons did so.

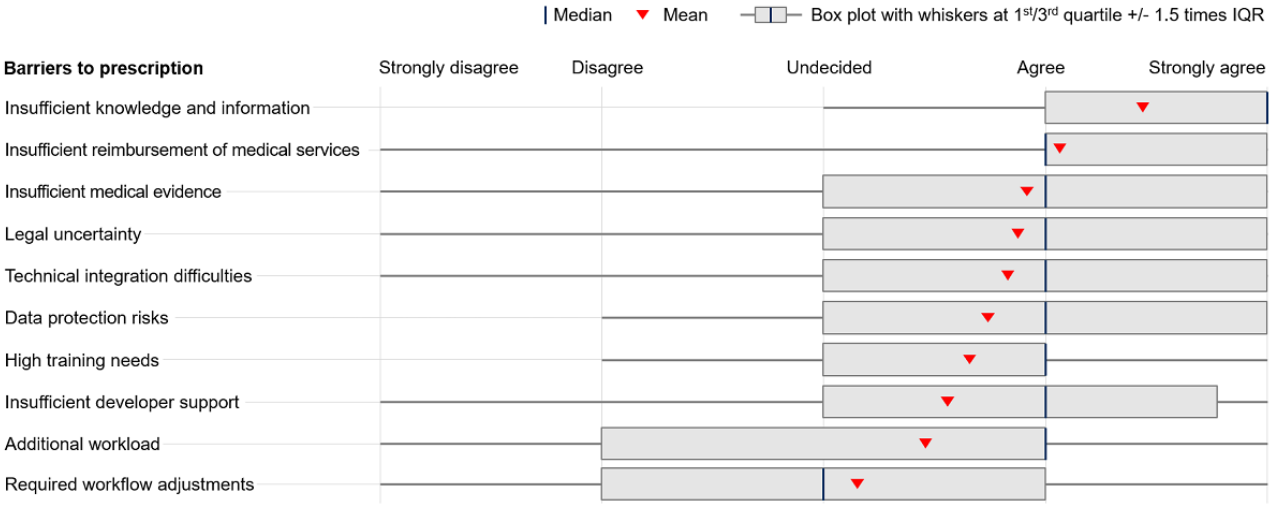
On average, prescription intentions were more than 30 percentage points lower than prescription attitudes. This gap was smallest for ear, nose, and throat doctors (5.3 percentage points): 61% (23/38) of responding ear, nose, and throat professionals displayed high prescription attitudes and 55% (21/38) displayed an intention to prescribe. The gap was largest for surgeons: 68% (13/19) held positive prescription attitudes, yet only 16% (3/19) reported prescription intentions, with a gap of 52.6 percentage points. Despite this general trend, some of the results for prescription intentions, attitudes, and their relative gap may also be influenced by the comparatively small sample size in some medical specialties.

Perceived Barriers to Prescription

As Figure 3 displays, health care professionals saw significant barriers to prescribing DiGA across several dimensions. Above all, 87.4% (1135/1299) of health care professionals viewed insufficient information as an obstacle to DiGA prescriptions. This translates into low perceived competence in dealing with DiGA: about 7 out of 10 health care professionals felt

insufficiently knowledgeable to differentiate bad from good DiGA (915/1298, 70.5%) and to advise patients regarding their application (905/1308, 69.2%). However, 92.4% (1208/1308) of health care professionals wanted to receive information about DiGA, thereby showing openness to address the key barrier to adoption of DiGA.

Figure 3. Perceived barriers to prescribing DiGA by health care professionals. Respondents were asked to indicate to what extent they believed various barriers prevented health care professionals from prescribing DiGA on 5-point Likert scales. DiGA: Digitale Gesundheitsanwendungen (digital health applications).

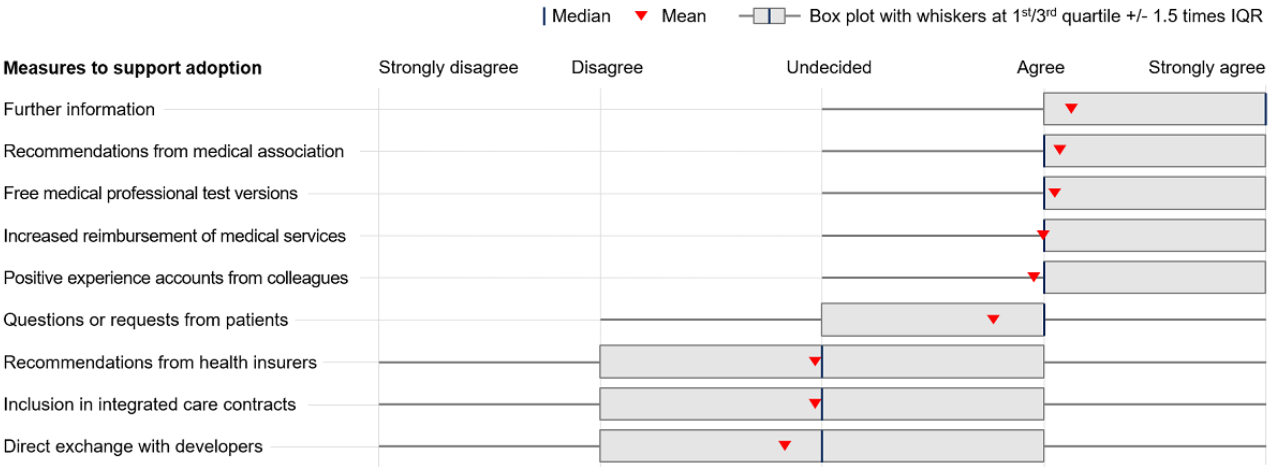


Additionally, a majority of health care professionals saw insufficient reimbursement of medical services related to DiGA (716/1299, 55.1%), insufficient evidence (712/1298, 54.9%), legal insecurities about potential liabilities for mistreatment (680/1299, 52.3%), and worries about data protection and security (658/1299, 50.7%) as clear barriers. Slightly less than half of the respondents believed that training needs for the respondent and potential staff (632/1299, 48.7%), perceptions of increased workload (584/1299, 45.0%), and technical integration issues (560/1299, 43.1%) were preventing health care professionals from adopting DiGA more broadly. Only about one-third of health care professionals saw workflow adjustment needs (431/1299, 33.2%) and missing support for health care professionals from DiGA providers (eg, for technical issues in daily operations; 372/1298, 28.7%) as obstacles to prescribing.

Measures to Support Adoption

Six measures were viewed positively by health care professionals to increase willingness to prescribe DiGA (Figure 4). Additional information about DiGA (1104/1297, 85.1%), recommendations by medical associations (1041/1297, 80.3%), positive experience reports about DiGA from medical colleagues (1024/1297, 79.0%), opportunities to test apps (1010/1297, 77.9%), and increased reimbursement for medical services related to DiGA (932/1297, 71.9%) have the potential to support health care professionals in the adoption of DiGA. When approached by patients, health care professionals also believed they would be more likely to engage with the topic and, thereafter, potentially prescribe DiGA (821/1297, 63.3%).

Figure 4. Measures to support health care professionals' adoption of DiGA. Respondents were asked to indicate to what extent they believed various measures could help health care professionals to adopt DiGA on 5-point Likert scales. DiGA: Digitale Gesundheitsanwendungen (digital health applications).



As displayed in Figure 4, other measures were viewed as neutral or ineffective. Recommendations to health care professionals by health insurers (529/1297, 40.8%), integrated care contracts (464/1296, 35.8%), and direct exchanges between health care professionals and developers (361/1297, 27.8%) were believed to have a weaker effect.

Discussion

Despite the high potential of mHealth to improve medical care at lower costs [35,36], broad adoption has been challenging in the past. To overcome these challenges, Germany embarked on a new path by being the first country, worldwide, to introduce DiGA as prescribable mHealth apps into regular care in October 2020. However, DiGA adoption has been relatively slow, even at a time when large numbers of health care professionals have adopted telemedicine due to the COVID-19 pandemic [37]. To our knowledge, this study was the first to systematically examine the dynamics underlying the adoption of prescribable mHealth apps.

Our findings show that a majority of health care professionals support the introduction of DiGA into standard care, as they see significant medical benefits for patients, most importantly, improved patient adherence, health literacy, disease management, access to care, and direct health benefits. Although further research on the evidence of mHealth apps is needed in general [10], patient benefits have already been confirmed for various DiGA in randomized controlled trials [38-41].

Countless studies have found the expectation of benefits, positive attitudes, or perceived usefulness of mHealth technologies to be core predictors of adoption [20,21]. Accordingly, health care professionals are more likely to use a technology when they believe it to be beneficial to their patients' care or themselves [9] and refrain from doing so when skeptical of its benefits for their practice [7].

While our findings confirmed a positive relationship between perceived usefulness and intention to use, the effect seems to be somewhat limited. Despite the multitude of benefits of DiGA seen by our respondents, only about one-third of health care professionals planned to prescribe DiGA in the future. Although this finding is in line with mHealth adoption rates in other countries [32], the share of health care professionals who have already prescribed DiGA is drastically smaller in Germany, seconding the need for further investigations of relevant factors.

While some studies consider gender and age as sociodemographic factors influencing technology adoption [21], others find this effect to be limited to attitude, not intention to prescribe [7]. The latter is true for our survey results. Only digital affinity had a significant and positive effect on both attitude and prescription intention. This may be due to the fact that health care professionals with greater digital affinity and information and communications technology experience anticipate greater ease of use when integrating DiGA into their work, a factor that has been found to be a strong predictor of technology adoption [9,42].

In addition to the potential effects of sociodemographics, two other factors may explain the low prescription intentions of DiGA. First, the availability of relevant DiGA is limited for some specialties, which may, therefore, result in these health care professionals not planning to prescribe DiGA, a factor also highlighted by our qualitative interviews. Looking at the 20 apps that have been approved so far, 10 of them are related to psychotherapy (eg, depression, phobias, and insomnia), 4 are related to neurology (eg, stroke, multiple sclerosis, and migraine), and 1 is related to nutrition (ie, obesity) [43]. These are largely irrelevant for professionals from specialties with the largest gap between positive attitudes and prescription intentions (ie, ophthalmology, dermatology, surgery, and other specialties). However, medically beneficial mHealth apps targeting diseases in these currently underrepresented specialties (eg, smartphone-based early detection of skin cancer [44] or treatment of ophthalmologic conditions, such as amblyopia and glaucoma [45]) are starting to emerge or are already under review by the German Federal Institute for Drugs and Medical Devices [46] and may increase DiGA prescriptions in the future.

Second, barriers to adoption may explain low prescription intentions. Barriers identified in our study include lack of information and medical evidence; insufficient reimbursement of medical services; concerns about medico-legal issues, such as liability and data protection risks; as well as workflow-related issues, including required workflow adjustments, training needs, and increased workloads. Most of these barriers are consistent with those identified by other studies from various countries and settings. A recent systematic review by Jacob et al [7] identified workflow-related factors; privacy, security, and medico-legal concerns; and monetary issues related to reimbursement and fees to be among the most studied and important social and organizational factors that influence technology adoption by health care professionals. Interestingly, lack of information—with over 87% of responses reporting this as the largest barrier for adoption in this study—has been studied significantly less [7]. This may be because past research has frequently studied conceptually more established and mature concepts, such as electronic health records [15], contrary to Germany's DiGA, which had only been available for under 3 months at the time of this study. For such novel technology, information may be an anteceding barrier that needs to be addressed first before health care professionals become fully aware of more frequently studied barriers to adoption.

To address these barriers and support adoption of DiGA, five concrete measures should be implemented. First, increasing health care professionals' level of information and trust in DiGA through recommendations from reliable bodies, such as medical associations, scientific societies, opinion leaders, and peers [7,24,47], and enabling health care professionals to experience DiGA themselves through free test versions may foster adoption. Here, it is critical to address the barriers perceived, such as medico-legal concerns around liability for mistreatment and data risks, as well as benefits from using DiGA for both patients and health care professionals. Second, introducing DiGA-related medical services into the remuneration system for statutory health insurance–accredited health care professionals may offer stronger financial incentives for adoption. Past research from Germany suggests that such measures may influence up to 85% of health care professionals in adopting a new technology [25].

Third, scientists should further investigate medical evidence of DiGA using robust study designs (eg, randomized controlled trials and meta-reviews according to Cochrane standards) and make findings freely available more than is currently the case. Moreover, given the widespread lack of awareness, previous results should be disseminated more effectively, starting with the national DiGA directory operated by the German Federal Institute for Drugs and Medical Devices. A more transparent, standardized, and, thus, more accessible presentation of evidence, with a clear indication of medical and structural effects for patients and study design conditions complied with, may promote trust in DiGA [48]. Fourth, training offerings related to DiGA should be expanded to help physicians make decisions about DiGA implementation within their own work at an extensive and intensive margin. Providing incentives for trainings, for instance, continuing medical education certification, may further aid this effort. Fifth, ensuring compatibility of DiGA with existing clinical practices, workflows, and infrastructure will be critical to remove barriers to adoption [15].

This study extends our understanding of the dynamics underlying the adoption of prescribable mHealth apps by health care professionals. Given that an online survey was used, our results may be subject to some self-selection bias and, therefore, bounded representativeness. Further research may, therefore, wish to validate these findings with an even larger, more representative sample.

In conclusion, three strands of research resulted from this study. First, given the criticality of greater information for prescription among medical professionals, future studies should investigate which channels appear to be most appropriate for delivering DiGA information and which types of content are most critical for health care professionals. Second, to reduce reliance on health care professionals who might remain reluctant to prescribe DiGA, other paths to support the adoption of medically beneficial DiGA should be explored. Third, further research should investigate whether health care professionals are reluctant to prescribe DiGA to some patient groups (eg, those lacking language or digital skills) and how such potential digital divides can be addressed to realize mHealth's full potential for patients and in the German health care system at large.

Acknowledgments

The authors thank Rebecca Janßen for her support and valuable comments on the manuscript. The authors also thank Martin Huber for sharing his perspective on the global health care industry.

Conflicts of Interest

This study was conducted as part of doctoral and habilitation research projects at the Witten/Herdecke University. All authors, except JPE, participated in these projects while on academic leave from (FD and MZ) or employed by (LB, PB, and LF) a large international consulting firm. The study was executed during the authors' personal time. The consulting firm was at no point involved in the research. There was also no funding, pay, or other commercial interest provided by the consulting firm.

Multimedia Appendix 1
Translated interview guide.
[PDF File (Adobe PDF File), 59 KB - mhealth_v9i11e33012_app1.pdf]

Multimedia Appendix 2
Translated survey questionnaire.
[PDF File (Adobe PDF File), 172 KB - mhealth_v9i11e33012_app2.pdf]

References

1. Keehan SP, Cuckler GA, Sisko AM, Madison AJ, Smith SD, Stone DA, et al. National health expenditure projections, 2014-24: Spending growth faster than recent trends. Health Aff (Millwood) 2015 Aug;34(8):1407-1417. [doi: 10.1377/hlthaff.2015.0600] [Medline: 26220668]
2. Idrish S, Rifat A, Iqbal M, Nisha N. Mobile health technology evaluation: Innovativeness and efficacy vs cost effectiveness. Int J Technol Hum Interact 2017;13(2):1-21. [doi: 10.4018/ijthi.2017040101]
3. Chib A, Lin SH. Theoretical advancements in mHealth: A systematic review of mobile apps. J Health Commun 2018;23(10-11):909-955. [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]
4. Simons LPA, Foerster F, Bruck PA, Motiwalla L, Jonker CM. Microlearning mApp raises health competence: Hybrid service design. Health Technol (Berl) 2015;5(1):35-43 [FREE Full text] [doi: 10.1007/s12553-015-0095-1] [Medline: 26097799]
5. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: A systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]

6. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: A systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
7. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: Systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
8. Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: Lessons and opportunities for other countries. NPJ Digit Med 2020;3:94 [FREE Full text] [doi: 10.1038/s41746-020-0306-7] [Medline: 32685700]
9. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. mHealth adoption by healthcare professionals: A systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220 [FREE Full text] [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
10. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: Systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
11. Bundesgesundheitsminister Jens Spahn im Bundestag zum Digitale-Versorgung-Gesetz (DVG). Bundesministerium für Gesundheit. 2020. URL: https://www.bundesgesundheitsministerium.de/presse/reden/dvg-23-lesung.html#c16594 [accessed 2020-12-15]
12. Anzahl der im Apple App-Store verfügbaren mHealth-Apps vom 1. Quartal 2015 bis zum 3. Quartal 2020. Statista. 2020. URL: https://de.statista.com/statistik/daten/studie/1191205/umfrage/anzahl-der-bei-apple-verfuegbaren-mhealth-apps/ [accessed 2021-02-25]
13. Most popular Google Play categories. AppBrain. 2020. URL: https://www.appbrain.com/stats/android-market-app-categories [accessed 2020-09-20]
14. Digital Health Applications (DiGA). Bundesamt für Arzneimittel und Medizinprodukte (BfArM). 2020. URL: https://www.bfarm.de/EN/MedicalDevices/DiGA/_node.html [accessed 2021-02-15]
15. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
16. Beeger B, Heeg T. Holpriger start für die "Apps auf Rezept". Frankfurter Allgemeine Zeitung. 2021 Jul 29. URL: https://www.faz.net/aktuell/wirtschaft/unternehmen/gesundheit-holpriger-start-fuer-die-apps-auf-rezept-17458459.html [accessed 2021-10-25]
17. Klöckner J. Umfrage: Kassen haben bislang rund 3700 "Apps auf Rezept" erstattet. Handelsblatt. 2021. URL: https://tinyurl.com/acs9cypz [accessed 2021-09-20]
18. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: A qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]
19. Rogers EM. Diffusion of Innovations. New York, NY: Free Press of Glencoe; 1962.
20. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
21. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. MIS Q 2003;27(3):425-784. [doi: 10.2307/30036540]
22. Ariens LF, Schussler-Raymakers FM, Frima C, Flinterman A, Hamminga E, Arents BW, et al. Barriers and facilitators to eHealth use in daily practice: Perspectives of patients and professionals in dermatology. J Med Internet Res 2017 Sep 05;19(9):e300 [FREE Full text] [doi: 10.2196/jmir.7512] [Medline: 28874336]
23. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. The use of smartphone health apps and other mobile health (mHealth) technologies in dietetic practice: A three country study. J Hum Nutr Diet 2017 Aug;30(4):439-452. [doi: 10.1111/jhn.12446] [Medline: 28116773]
24. El Amrani L, Oude Engberink A, Ninot G, Hayot M, Carbonnel F. Connected health devices for health care in French general medicine practice: Cross-sectional study. JMIR Mhealth Uhealth 2017 Dec 21;5(12):e193 [FREE Full text] [doi: 10.2196/mhealth.7427] [Medline: 29269336]
25. Leppert F, Dockweiler C, Eggers N, Webel K, Hornberg C, Greiner W. Economic aspects as influencing factors for acceptance of remote monitoring by healthcare professionals in Germany. J Int Soc Telemed eHealth 2015;3:1-9 [FREE Full text]
26. Fairbrother P, Ure J, Hanley J, McCloughan L, Denvir M, Sheikh A, Telescot Programme Team. Telemonitoring for chronic heart failure: The views of patients and healthcare professionals - A qualitative study. J Clin Nurs 2014 Jan;23(1-2):132-144. [doi: 10.1111/jocn.12137] [Medline: 23451899]
27. Payne KFB, Wharrad H, Watts K. Smartphone and medical related app use among medical students and junior doctors in the United Kingdom (UK): A regional survey. BMC Med Inform Decis Mak 2012 Oct 30;12:121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]
28. Survey tools. Qualtrics. 2020. URL: https://www.qualtrics.com/support/survey-platform/survey-module/survey-tools/survey-tools-overview/ [accessed 2021-03-02]

29. Eysenbach G. Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
30. North B, Curtis D, Sham P. A note on the calculation of empirical P values from Monte Carlo procedures. Am J Hum Genet 2002 Aug;71(2):439-441 [FREE Full text] [doi: 10.1086/341527] [Medline: 12111669]
31. Illiger K, Hupka M, von Jan U, Wichelhaus D, Albrecht UV. Mobile technologies: Expectancy, usage, and acceptance of clinical staff and patients at a university medical center. JMIR Mhealth Uhealth 2014 Oct 21;2(4):e42 [FREE Full text] [doi: 10.2196/mhealth.3799] [Medline: 25338094]
32. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among Australian general practitioners: Survey study. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: 10.2196/13199] [Medline: 31199343]
33. Leppert F, Dockweiler C, Eggers N, Webel K, Hornberg C, Greiner W. Acceptance of telemonitoring by health care professionals in Germany: A question of financial conditions. Value Health 2014 Nov;17(7):A422-A423 [FREE Full text] [doi: 10.1016/j.jval.2014.08.1045] [Medline: 27201078]
34. Statistische Informationen aus dem Bundesarztregister. Berlin, Germany: Kassenärztliche Bundesvereinigung; 2020. URL: https://www.kbv.de/media/sp/2020-12-31_BAR_Statistik.pdf [accessed 2021-11-05]
35. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
36. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: Evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
37. Matthies H. Here to stay: Digital health in times of COVID-19 - A German deep dive. Health Innovation Hub. 2020. URL: https://hih-2025.de/here-to-stay-digital-health-in-times-of-covid-19-a-german-deep-dive/#:~:text=Here [accessed 2021-02-25]
38. Twomey C, O'Reilly G, Bültmann O, Meyer B. Effectiveness of a tailored, integrative internet intervention (deprexis) for depression: Updated meta-analysis. PLoS One 2020;15(1):e0228100 [FREE Full text] [doi: 10.1371/journal.pone.0228100] [Medline: 31999743]
39. Pöttgen J, Moss-Morris R, Wendebourg JM, Feddersen L, Lau S, Köpke S, et al. Randomised controlled trial of a self-guided online fatigue intervention in multiple sclerosis. J Neurol Neurosurg Psychiatry 2018 Sep;89(9):970-976. [doi: 10.1136/jnnp-2017-317463] [Medline: 29549193]
40. Lorenz N, Heim E, Roetger A, Birrer E, Maercker A. Randomized controlled trial to test the efficacy of an unguided online intervention with automated feedback for the treatment of insomnia. Behav Cogn Psychother 2019 May;47(3):287-302. [doi: 10.1017/S1352465818000486] [Medline: 30185239]
41. Berger T, Urech A, Krieger T, Stolz T, Schulz A, Vincent A, et al. Effects of a transdiagnostic unguided internet intervention ('velibra') for anxiety disorders in primary care: Results of a randomized controlled trial. Psychol Med 2017 Jan;47(1):67-80. [doi: 10.1017/S0033291716002270] [Medline: 27655039]
42. Duplaga M. Searching for a role of nursing personnel in developing landscape of eHealth: Factors determining attitudes toward key patient empowering applications. PLoS One 2016;11(4):e0153173 [FREE Full text] [doi: 10.1371/journal.pone.0153173] [Medline: 27049525]
43. DiGA Verzeichnis. Bundesamt für Arzneimittel und Medizinprodukte (BfArM). 2021. URL: https://diga.bfarm.de/de/verzeichnis [accessed 2021-10-01]
44. Choi J, Cho Y, Woo H. mHealth approaches in managing skin cancer: Systematic review of evidence-based research using integrative mapping. JMIR Mhealth Uhealth 2018 Aug 02;6(8):e164 [FREE Full text] [doi: 10.2196/mhealth.8554] [Medline: 30072362]
45. Hogarty DT, Hogarty JP, Hewitt AW. Smartphone use in ophthalmology: What is their place in clinical practice? Surv Ophthalmol 2020;65(2):250-262. [doi: 10.1016/j.survophthal.2019.09.001] [Medline: 31541618]
46. Der DiGA-Fast-Track des BfArM: Die reise von beratung zum verzeichnis. YouTube. Bonn, Germany: Bundesamt für Arzneimittel und Medizinprodukte (BfArM); 2021. URL: https://www.youtube.com/watch?v=P_Jdk4dXRlU [accessed 2021-03-01]
47. Deng Z, Hong Z, Ren C, Zhang W, Xiang F. What predicts patients' adoption intention toward mHealth services in China: Empirical study. JMIR Mhealth Uhealth 2018 Aug 29;6(8):e172 [FREE Full text] [doi: 10.2196/mhealth.9316] [Medline: 30158101]
48. Busse R. Wege zu einer besseren Implementierung von digitalen Gesundheitsanwendungen in die Gesundheitsversorgung der GKV. Abschlussveranstaltung "Fast-Track 2.0: Ein Blick in die Zukunft der digitalen Patientenversorgung". Berlin, Germany: Technische Universität Berlin; 2020. URL: https://www.mig.tu-berlin.de/fileadmin/a38331600/I.DiGA_Abschlussveranstaltung_final.pdf [accessed 2021-02-07]

Abbreviations

BfArM: Bundesamt für Arzneimittel und Medizinprodukte (German Federal Institute for Drugs and Medical Devices)
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
DiGA: Digitale Gesundheitsanwendungen (digital health applications)
mHealth: mobile health

Edited by G Eysenbach; submitted 23.08.21; peer-reviewed by U Bork, C Jacob, B Arents, C Gibson; comments to author 14.09.21; revised version received 28.09.21; accepted 02.10.21; published 23.11.21.

Please cite as:
Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L
Physicians' Attitudes Toward Prescribable mHealth Apps and Implications for Adoption in Germany: Mixed Methods Study
JMIR Mhealth Uhealth 2021;9(11):e33012
URL: https://mhealth.jmir.org/2021/11/e33012
doi:10.2196/33012
PMID:34817385

Original Paper

How Identification With the Social Environment and With the Government Guide the Use of the Official COVID-19 Contact Tracing App: Three Quantitative Survey Studies

Annika Scholl[1], PD, PhD; Kai Sassenberg[1], PhD, Prof Dr

Social Processes Lab, Leibniz-Institut fuer Wissensmedien, Tuebingen, Germany

Corresponding Author:
Annika Scholl, PD, PhD
Social Processes Lab
Leibniz-Institut fuer Wissensmedien
Schleichstr 6
Tuebingen, 72076
Germany
Phone: 49 7071979 ext 257
Email: a.scholl@iwm-tuebingen.de

Abstract

Background: Official contact tracing apps have been implemented and recommended for use across nations to track and contain the spread of COVID-19. Such apps can be effective if people are *willing* to use them. Accordingly, many attempts are being made to motivate citizens to make use of the officially recommended apps.

Objective: The aim of this research was to contribute to an understanding of the preconditions under which people are willing to use a COVID-19 contact tracing app (ie, their use intentions and use). To go beyond personal motives in favor of app use, it is important to take people's social relationships into account, under the hypothesis that the more people identify with the *beneficiaries* of app use (ie, people living close by in their social environment) and with the *source* recommending the app (ie, members of the government), the more likely they will be to accept the officially recommended contact tracing app.

Methods: Before, right after, and 5 months after the official contact tracing app was launched in Germany, a total of 1044 people participated in three separate surveys. Structural equation modeling was used to test the hypotheses, examining the same model in all studies at these critical points in time.

Results: Across the three surveys, both identification with the beneficiaries (people living in their social environment) and with the source recommending the app (members of the government) predicted greater intention to use and use (installation) of the official contact tracing app. Trust in the source (members of the government) served as a mediator. Other types of identification (with people in Germany or people around the world) did not explain the observed results. The findings were highly consistent across the three surveys.

Conclusions: Attempts to motivate people to use new health technology (or potentially new measures more generally) not only for their personal benefit but also for collective benefits should take the social context into account (ie, the social groups people belong to and identify with). The more important the beneficiaries and the sources of such measures are to people's sense of the self, the more willing they will likely be to adhere to and support such measures.

(JMIR Mhealth Uhealth 2021;9(11):e28146) doi:10.2196/28146

KEYWORDS

COVID-19; SARS-CoV-2; contact tracing app; social identification; technology acceptance; pandemic; outbreak; health technology

Introduction

Background

The outbreak of COVID-19 in 2020 impacted individual people worldwide but also their communities, governments, and whole nations, with (often) unknown challenges and numerous new measures to be rapidly accepted and implemented. One important measure implemented in many countries to collectively contain the spread of the virus has been the use of new technological means, namely, an *official contact tracing app* [1]. Such apps aim at retracing chains of infection and

warning people in case they have had contact with a potentially ill person. In Germany, the official app was commissioned by the German government—launched on June 16, 2020—and has since been recommended by the government for (voluntary) use. In February 2021, almost 8 months after its launch, 25.4 million people (out of roughly 83.7 million living in Germany) were using the app, with 59% positive test results being entered [2]. Therefore, there is still substantial room to gain more users. This situation raises the question: What *motivates* people to support and make use of such new technology?

In an "era of massive technological advancement" [3] and this pandemic, studying the acceptance of such health-related apps (especially regarding contact tracing) is important both from theoretical and practical points of view to better understand and potentially foster people's willingness to use them. Going beyond prior work on personal motives for app use, this study examined the role of social relationships. Specifically, we targeted the question if the extent to which people identify with specific groups, namely (a) with the *beneficiaries* of app use (ie, people in their social environment) and (b) with the *source* recommending the app (ie, members of the government), predicts a greater willingness to use the contact tracing app. This relies on the idea that tracing apps gain their impact at the collective level (ie, when many people use them, these apps are beneficial for the community but not necessarily for the individual user).

We used three separate surveys to test this idea at crucial points during 2020: right before the official app was launched, right after its launch, and 5 months later when substantial extensions to track more user data were added. With this approach, we sought to contribute to a better understanding of the preconditions of people's willingness to use such new health-related technology from a motivational perspective, highlighting the importance of the social groups (collectives) that people belong to (rather than their personal benefits or pitfalls) as driving forces.

Prior Work on App Acceptance and the Role of Individual Motives as a Predictor

This study focused on the official contact tracing app in Germany. This specific app traces contact of the user with people who have been diagnosed with COVID-19, using Bluetooth over smartphones. If a "positive" contact is recorded during the day, the app notifies the user with a warning the next day, which did not reveal the potential contact's identity or the specific point in time of contact when the study was performed.

Prior work on how people respond to such apps focused on three main aspects. First, recent reviews compared the features of tracing apps and other types of apps developed during the COVID-19 pandemic (eg, for training, information sharing, or diagnosis) [4-6]. In these reviews, the focus was on the descriptions of the technology rather than the users. Second, researchers have performed surveys on general *acceptance rates* of COVID-19 tracing apps among different populations. Although some findings suggest relatively high support for the app across countries [7], other studies did point out the problem of low usage rates (eg, in France among health care students) [8]. In Germany, at the time of the launch of the official contact tracing app, 81% of an adult sample between 18-77 years old did possess the (technological and ability-related) requirements to use the app, but only 35% reported being willing to do so [9]. This points to the necessity to better understand the *motivational* preconditions of app acceptance.

The low acceptance rates stress the relevance of the third research question on what motivates people to make use of this app and disclose their personal data. In this regard, prior work has adopted a clear focus on *personal* benefits and detriments. From a personal perspective, using these technological means does carry benefits (eg, being informed about one's own positive contacts) but also potential barriers for each individual user (eg, providing personal data). Personal motives are known to influence use intentions. Recent evidence in this domain has shown that privacy concerns along with uncertainty about the app's effectiveness constitute two main personal *barriers* to accept such an app [5,7,8,10-12]; thus, balancing these two aspects poses a major challenge for app developers [3,13]. At the same time, personal conditions that *support* acceptance are related to already having adopted one's lifestyle during the pandemic [14], trusting in data security or authorities [7,10,11], perceiving high personal vulnerability to a health threat [12], and experiencing high personal self-efficacy [12]. Finally, providing users with an app that seems easy to use [15] or giving transparent information about the app contributes to greater app acceptance [11].

Overall, motivational factors on the level of the *individual* play a role in acceptance: people are more willing to use technological means if they expect it to benefit (rather than cost) them personally [16]. However, it is unclear whether such personal costs and benefits are the *only* motivational drivers, or if people might also be motivated to use the app (as a measure designed for a collective) not for personal reasons but because they care about *others*. Going beyond the prior work outlined above, we here present a novel perspective on the motivational drivers behind app acceptance. We reason that people may also be willing to use the app because they *identify* with those who benefit from the app and/or those who recommend using it.

The Role of Identification With Others (Beneficiaries) in Fostering App Use

One important aspect of technological means such as a contact tracing app during a pandemic is that its use does not only benefit oneself personally but it also benefits *other people* in one's social environment (ranging from friends or family living close by to unknown people who happen to buy their groceries in the same supermarket) by warning them in case one receives a positive test result. Accordingly, people may be more motivated to use the app the more *important* the welfare of these potential beneficiaries is to them.

Others' importance to oneself is a topic addressed Social Identity Theory [17,18]. People define themselves not only in terms of what makes them unique individuals ("I"; their personal identity) but also in terms of the social groups they belong to ("we"; their social identity). The more a person identifies with a social group they belong to, the more relevant this (in)group becomes to their definition of the "self." As a result, people start thinking more in terms of "we" (rather than "I") and they care more about the

interests of the group. Consequently, the more people identify with a social group, the more willing they are to engage in behaviors that benefit the group's members (eg, use technology to contribute knowledge for others at work) [19-21], potentially even at personal costs (eg, sacrificing privacy concerns to benefit the safety of the group; see the example of CCTV cameras in the United Kingdom [22]).

Applied to the present case, this implies that the more a person identifies with people living in their social environment (ie, potential beneficiaries from their app use), the more willing this person will likely be to contribute to these others' welfare and thus to use the app. This resulted in the following hypothesis:

H1: The more people identify with people living in their social environment, the more willing they are to use the contact tracing app.

The Role of Identification With the Source in Fostering App Use

A second important motivational predictor of people's willingness to use such new technology could be the level of identification with the *source* ("authority") recommending the app. In this sense, the more people trust in and identify with members of the government—as the source who commissioned the production of this app and now persistently recommends its use—the more willing they may be to use the app.

Members of the government represent a small group of (elected) people who act as representatives of a nation (ie, the broader ingroup of the people of their country that they may identify with). In the present case, we investigated people's identification with the *members of the German government*. During the early months of the COVID-19 pandemic, European and international news often stated that the German government was dealing relatively effectively with the challenges; during these times, members of the government have been meeting and communicating the results of such meetings repeatedly (eg, every other week) to the public to address citizens' concerns and needs (eg, building upon regular opinion surveys and including public addresses). As a result, citizens may have had the impression of having substantial (virtual) contact with members of the government, and such (positive) contact is known to create a feeling of closeness and identification to others [23,24]. In short, citizens may have had opportunities to *identify* with members of the government (even if citizens would typically not consider them to be members of their ingroup) just as people can generally identify with their leaders at work (eg, [25-28]).

The more people identify with others, the more positively they view these others [29-32]. This also means that when taking a "leap of faith" and being potentially vulnerable (as is likely the case during a pandemic), identification presumably makes people more willing to *trust* in and follow their ingroup members [33,34] and, potentially, even to trust more in and follow those people (authorities) who make decisions on behalf of their ingroup (eg, their community [35] or the people of their country). Supporting this idea, trust in the government was an important predictor of accepting measures during the Ebola outbreak [36,37] (for a similar argument, see [38]), and has also been considered important by people themselves during the COVID-19 pandemic [10]. This resulted in the following hypothesis:

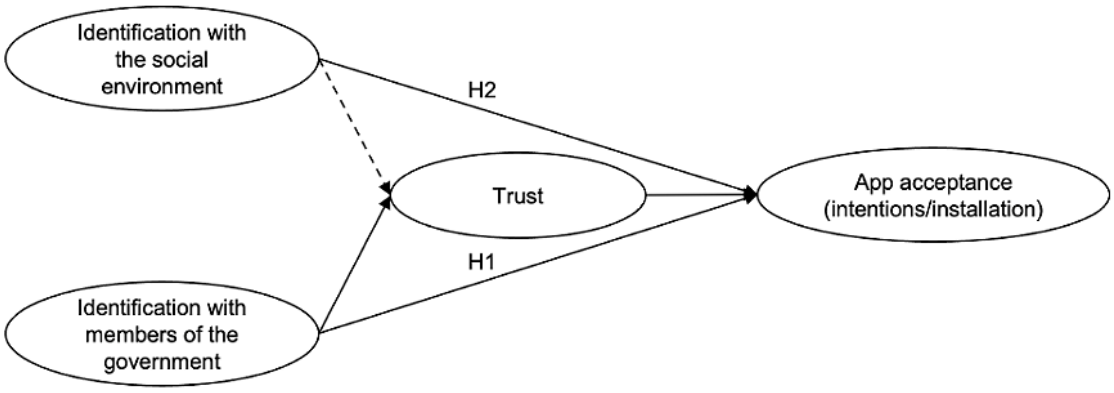
H2: The more people identify with members of the government, the more they trust in these members and, accordingly, the more they are willing to use the contact tracing app.

Study Objectives and Design

These two predictions were tested across three surveys at different important points in time throughout 2020 to perform a more comprehensive test of the hypothesized model. As an indicator of people's responses toward the contact tracing app, all surveys focused on the outcome willingness to accept the app (*app acceptance*); we additionally assessed willingness to use the app prior to its launch (*intentions to use*) or after its launch (*app use*; reflecting that people had already installed the app on their smartphones).

Taken together, we hypothesized that the more people identify (1) with other people in their social environment and/or (2) with members of the government, the greater their willingness to use the official contact tracing app should be. For the latter (H2), trust in the government was the assumed mediator. For the former (H1), we did not predict a mediating role of trust; notwithstanding, we explored whether trust would also serve as a potential mediator for identification with the social environment in predicting greater app acceptance. The general model is presented in Figure 1.

Figure 1. General model tested across the three surveys. The dashed line reflects an exploratory path and the solid lines reflect hypothesized paths. H: hypothesis.



As such, the broader aim of this study was to contribute to a better understanding of the preconditions of people's willingness to use this new technology from a motivational perspective. We sought to demonstrate, for the first time, the importance of (identification with) the social groups that people belong to as a driver toward greater app acceptance.

Methods

Procedure and Sample

At all three time points, participants were invited via a university email to complete a brief (7-10 minutes) survey on their perception of the current (COVID-19–related) situation. Participants received basic information about the respective study (eg, duration, compensation, purpose), provided informed consent, completed the main measures as indicated below, entered demographic information, were debriefed, confirmed their consent (or withdrew it) to use their survey data, and were finally given the chance to take part in a lottery of gift vouchers. Note that Surveys 2 and 3 included an experimental manipulation making identification with the government/environment salient (compared to control groups). This manipulation turned out to be ineffective; we thus analyzed and report results by means of correlations between measured constructs. More information on this aspect is certainly available from the authors. The local ethics committee provided ethical approval for all studies.

From May 25 to May 26, 2020, right before the official app was launched in Germany, 355 participants completed Survey 1 (268 women, 81 men, 6 diverse/nonspecified; mean age 23.53 years, SD 5.827, range 18-80 years); two additional participants of Survey 1 (none in Survey 2 and one in Survey 3) retracted their data after the debriefing, which were accordingly deleted prior to any analysis. Survey 2 (June 16, 2020, right after the app was launched) included 308 nonoverlapping participants (228 women, 74 men, 3 diverse/nonspecified, 3 missing; mean age 23.93 years, SD 5.79, range 18-73 years). Survey 3 (November 23-25, 2020, when additional app functions requiring more personal data were discussed) involved another separate sample of 381 participants (278 women, 100 men, 3 diverse/nonspecified; mean age 22.57 years, SD 4.83, range 18-69 years).

Measures

Source Data and Code

The exact order of measures in each study (including potential additional control measures) and the original materials for all surveys are reported in Multimedia Appendix 1 and Multimedia Appendix 2. Data [39] and code [40] are available at PsychArchives.

Identification

Identification with (1) people in their social environment and (2) members of the government, as our main predictors, were measured with six items each, adapted from McFarland et al [41] (eg, "How connected do you feel to the following groups?": social environment (1=*nothing* to 7=*very/very much*).

As a control variable, we assessed identification with two broader (in)groups that people may identify with to rule out that the relations we predict are driven by these aspects of identification. In Surveys 1 and 2, we assessed (with the same items) (3) identification with *people living in Germany* as a control variable, whereas in Survey 3, we assessed (4) identification with *people around the world* (humanity) as a control.

For exploratory purposes, we also assessed identification with *scientists in the health domain* as potential "users" of the app data; however, as this measure was strongly correlated with *identification with members of the government* (Survey 1, N=355: r=0.542, P<.001; Survey 2, N=308*:* r=0.473, P<.001), we refrained from further analyses with this measure.

Trust

Trust in the government as a mediator was measured with four items in all surveys to assess the trust dimensions as indicated by Mayer et al [42], adapted from Winter et al [43] (eg, "How trustworthy/honest/competent/credible do you perceive the government to be?": 1=*not at all* to 7=*very*).

App Acceptance

App acceptance as a first outcome was operationalized as low perceived privacy infringement. Participants in Survey 1 (prior to launch of the app) received a brief description that the government was currently *planning on launching* the app for voluntary use to limit the spread of the virus, potentially trace users' movements, and warn users in case of a high-risk contact; in Survey 2 (after the launch), this message stated that the government had just initiated the launch of the app (by June 16, with the same purpose as indicated for Survey 1). Participants in Surveys 1 and 2 indicated how they perceived the call to use such an app with six items, adapted from Alge et al [44] (eg, "I find it acceptable that such an app should be used": 1=*does not apply at all* to 7=*totally applies*).

In Survey 3, we assessed this outcome 5 months (November 2020) after the app had been launched and an extension by further functions was addressed to collect more extensive user data as a basis for new governmental measures. Accordingly, this outcome was operationalized in terms of people's acceptance of *more app functions* (meaning providing more personal data) with 8 items adapted from Surveys 1 and 2 (eg, "I would be willing to disclose more information from myself as a basis to decide on new measures": 1=*does not apply at all* to 7=*totally applies*).

Intention to Use the App and Use (Installation) of the App

The *intention* to use the app as a second outcome in Surveys 1 and 2 was assessed with one item: "To which extent would you be/are you willing to use this contact-tracing app?" (1=*not at all* to 7=*very much*). The *use* of the contact tracing app was assessed as an additional outcome in Survey 2; we operationalized use by asking participants whether they had already installed the official or another contact tracing app (0=no, 1=yes: 25.3% stating yes; mean 0.25, SD 0.44). To be better able to differentiate whether agreement in this case

referred to the official (ie, government-recommended) app, in Survey 3, we specifically assessed the use of this official app with the question: "Is the official contact tracing app installed on your smartphone?" (0=no, 1=yes, missing=unsure/different app: 57.7% stating yes, 41.2% no, and 1% unsure; mean 0.58, SD 0.49). As such, use was operationalized as *installation of the app* (ie, having installed the tracing app on their smartphones).

Data Analysis

We analyzed the data using structural equation modeling with Mplus 8.4 [45]. The tested model for each survey examined the relationship between the predictors (identification with the government, identification with the social environment), controlling for identification with people in Germany (Surveys 1-2) or for identification with humanity (Survey 3); the mediator trust; and the outcomes app acceptance (Surveys 1-3), intention to use the app (Surveys 1-2), and either the use of some contact tracing app (Survey 2) or the use of the official contact tracing app (Survey 3). Moreover, we tested for *indirect effects* of identification (with the government, with the social environment) in predicting more app acceptance, intention to use the app, and/or use of the app via greater trust. We hypothesized and tested an indirect effect of identification with the government via trust on the outcomes (H1); in addition, we explored if trust would also serve as a "linking mechanism" between identification with the social environment and the outcomes (ie, also for the prediction in H2).

Results

Descriptive Statistics

Descriptive statistics and Cronbach α values are shown in Table 1. The intercorrelations among measures for each study are reported in Tables 2-4. Items were originally in German and translated for this paper.

Table 1. Descriptive statistics and bivariate correlations for the three surveys.

Variable	Survey 1 (N=355)		Survey 2 (N=308)		Survey 3 (N=381)	
	Mean (SD)	Cronbach α	Mean (SD)	Cronbach α	Mean (SD)	Cronbach α
Identification with the social environment	6.29 (0.79)	.89	6.24 (0.83)	.91	6.38 (0.63)	.88
Identification with members of the government	3.11 (1.15)	.87	3.32 (1.19)	.91	3.01 (1.16)	.88
Identification with people in Germany[a] (control)	4.52 (1.03)	.87	4.57 (1.01)	.85	4.54 (1.11)	.90
Trust	4.35 (1.32)	.92	4.68 (1.28)	.93	4.73 (1.19)	.91
App acceptance[b]	4.33 (1.56)	.91	5.01 (1.41)	.89	4.44 (1.62)	.94
Intention to use the app (1 item)	4.50 (1.89)	N/A[c]	4.29 (2.14)	N/A	N/A	N/A
Installation of the app[d]	N/A	N/A	0.25 (0.44)	N/A	0.58 (0.49)	N/A

[a]For Survey 3, this control variable was identification with people around the world.

[b]Referred to more app functions for Survey 3.

[c]N/A: not applicable.

[d]0=no, 1=yes for Surveys 1-2; 0=no, 1=yes, unsure=missing for Survey 3.

Table 2. Correlation analysis (Pearson *r* and two-tailed *P* value) of study variables in Survey 1 (N=355).

Variable	Identification with social environment	Identification with government	Identification with people in Germany (control)	Trust	App acceptance	Intention to use app (1 item)
Identification with social environment						
r	1	0.21	0.35	0.26	0.08	0.11
P value	—[a]	<.001	<.001	<.001	.15	.045
Identification with government						
r	0.21	1	0.47	0.53	0.22	0.19
P value	<.001	—	<.001	<.001	<.001	<.001
Identification with people in Germany (control)						
r	0.35	0.47	1	0.28	0.08	0.08
P value	<.001	<.001	—	<.001	.15	.13
Trust						
r	0.26	0.53	0.28	1	0.50	0.49
P value	<.001	<.001	<.001	—	<.001	<.001
App acceptance						
r	0.08	0.22	0.08	0.49	1	0.81
P value	.15	<.001	.15	<.001	—	<.001
Intention to use app (1 item)						
r	0.11	0.19	0.08	0.49	0.81	1
P value	0.045	<.001	.13	<.001	<.001	—

[a]Not applicable.

Table 3. Correlation analysis (Pearson *r* and two-tailed *P* value) among variables in Survey 2 (N=308).

Variable	Identification with social environment	Identification with government	Identification with people in Germany (control)	Trust	App acceptance	Intention to use app (1 item)	Installation of app (0=no, 1=yes/no)
Identification with social environment							
r	1	0.26	0.46	0.26	0.07	0.11	–0.02
P value	—[a]	<.001	<.001	<.001	.22	.05	.76
Identification with government							
r	0.26	1	0.57	0.52	0.24	0.25	0.16
P value	<.001	—	<.001	<.001	<.001	<.001	.004
Identification with people in Germany (control)							
r	0.46	0.57	1	0.32	0.02	0.03	0.03
P value	<.001	<.001	—	<.001	.72	.62	.62
Trust							
r	0.26	0.52	0.32	1	0.47	0.46	0.23
P value	<.001	<.001	<.001	—	<.001	<.001	<.001
App acceptance							
r	0.07	0.24	0.02	0.47	1	0.78	0.51
P value	.22	<.001	.72	<.001	—	<.001	<.001
Intention to use app (1 item)							
r	0.11	0.25	0.03	0.46	0.78	1	0.54
P value	.05	<.001	.62	<.001	<.001	—	<.001
Installation of app (0=no, 1=yes/no)							
r	–0.02	0.16	0.04	0.23	0.51	0.54	1
P value	.76	.004	.45	<.001	<.001	<.001	—

[a]Not applicable.

Table 4. Correlation analysis (Pearson *r* and two-tailed *P* value) among variables in Survey 3 (N=381).

Variable	Identification with social environment	Identification with government	Identification with people around the world (control)	Trust	App acceptance (more app functions)	Installation of app (0=no, 1=yes/no, unsure=missing)
Identification with social environment						
r	1	0.08	0.25	0.18	0.09	0.08
P value	—[a]	.14	<.001	<.001	.07	.14
Identification with government						
r	0.08	1	0.36	0.57	0.26	0.14
P value	.14	—	<.001	<.001	<.001	.006
Identification with people around the world (control)						
r	0.25	0.36	1	0.22	0.05	–0.03
P value	<.001	<.001	—	<.001	.34	.54
Trust						
r	0.18	0.57	0.22	1	0.49	0.30
P value	<.001	<.001	<.001	—	<.001	<.001
App acceptance (more app functions)						
r	0.09	0.26	0.05	0.49	1	0.47
P value	.07	<.001	.34	<.001	—	<.001
Installation of app (0=no, 1=yes/no)						
r	0.08	0.14	–0.03	0.30	0.47	1
P value	.14	.006	.54	<.001	<.001	—

[a]Not applicable.

Testing Hypotheses

The results of the structural model for each survey are presented in Figures 2-4. For the interested reader, results from the measurement model (factor loadings for individual items) are presented in Table S1 in Multimedia Appendix 1. Overall, the results indicated a good model fit across studies (Table 5). Note that the models reported here tested the predictions as outlined above and were *not* optimized to improve model fit in any way (eg, based on modification indices); for the interested reader, an alternative model for each survey, which (1) excluded the respective control variable (identification with people in Germany/around the world) and (2) included correlations between specific error terms, did improve model fit across studies above 0.90 (see Table S3 in Multimedia Appendix 1).

Figure 2. Structural equation model tested in Survey 1 (prelaunch, May 2020; N=355). Indirect effects via trust are reported in Table 6. Coefficients are fully standardized (MPlus STDYX standardization). SE: Standard Error. **P*<.05; ***P*<.01; ****P*<.001.

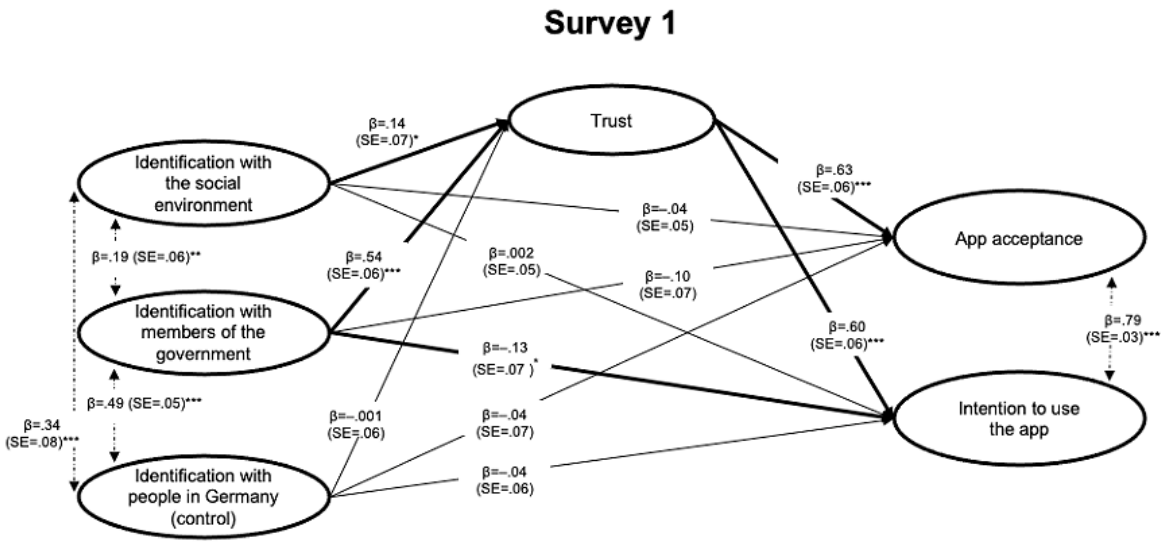


Figure 3. Structural equation model tested in Survey 2 (right after the app was launched, June 2020; N=308). Indirect effects via trust are reported in Table 6. Coefficients are fully standardized (MPlus STDYX standardization). SE: Standard Error. *P<.05; **P<.01; ***P<.001.

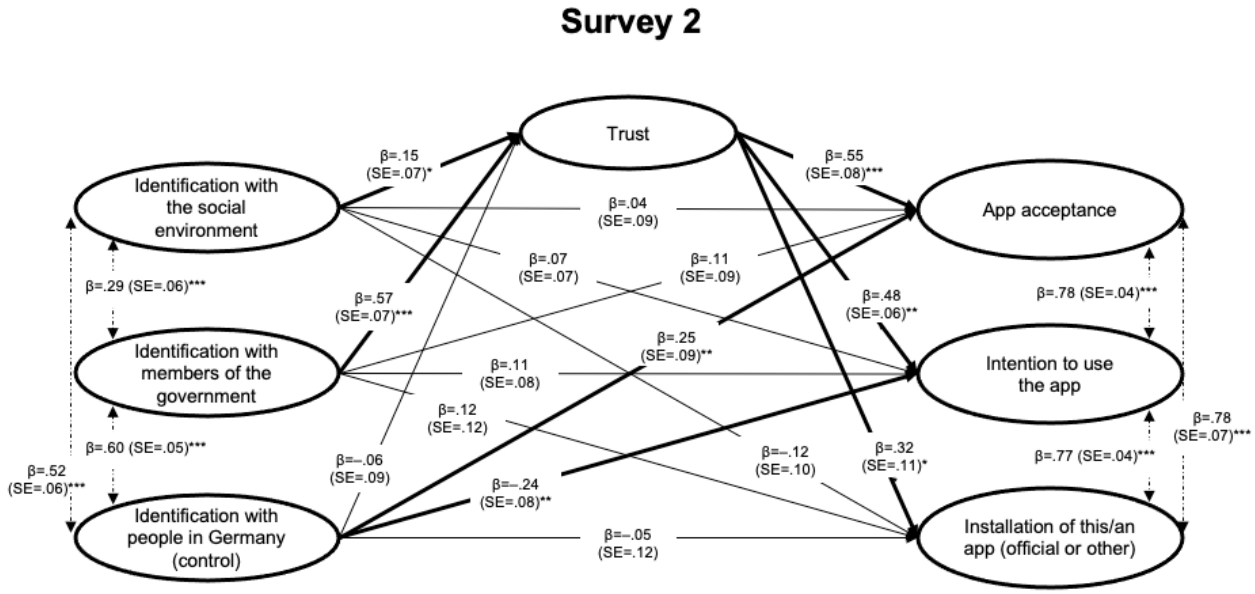


Figure 4. Structural equation model tested in Survey 3 (after launch of the app in December 2020, during discussion about adding more functions and collecting more personal data; N=381). Indirect effects via trust are reported in Table 6. Coefficients are fully standardized (MPlus STDYX standardization). SE: Standard Error. †P<.10; **P<.01; ***P<.001.

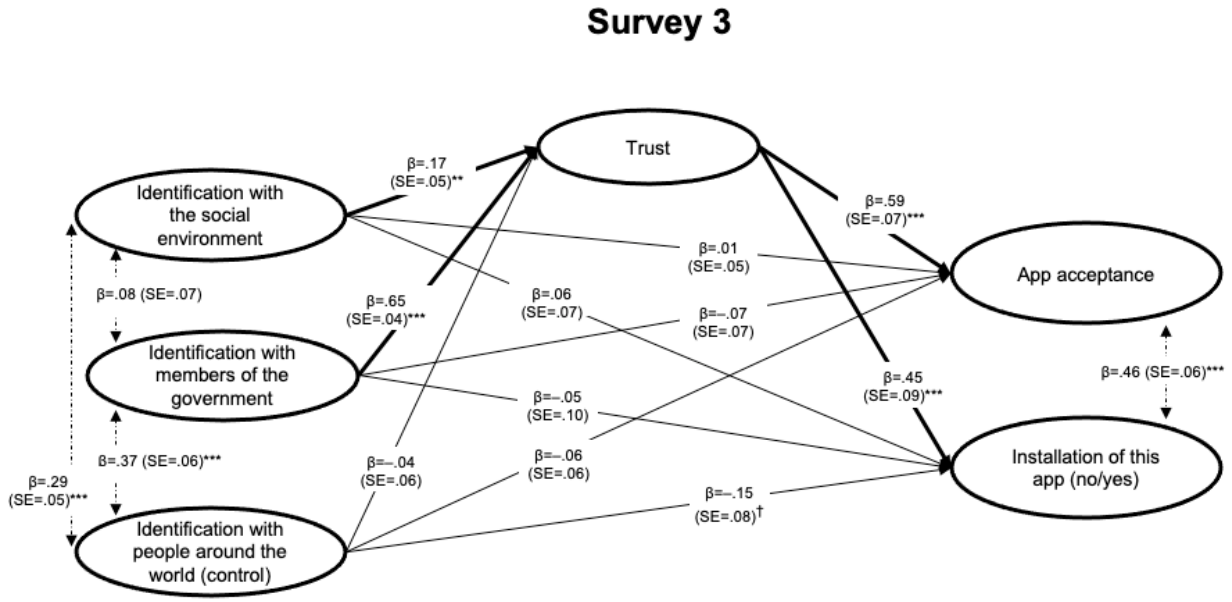


Table 5. Tests of model fit and fit indices for (nonoptimized) models tested across the three surveys.

Survey	$\chi 2$ (df)	P value	CFI[a]	TLI[b]	RMSEA[c]	SRMR[d]
1	1339.486 (363)	<.001	0.86	0.85	0.087	0.062
2	716.011 (386)	<.001	0.84	0.82	0.053	0.060
3	719.157 (420)	<.001	0.86	0.85	0.043	0.054

[a]CFI: comparative fit index.

[b]TLI: Tucker Lewis index.

[c]RMSEA: root mean square error of approximation.

[d]SRMR: standardized root mean square residual.

Table 6. Indirect effects of identification on app-related outcomes via trust across the three surveys.

Predictor	Outcome	Indirect effect via trust	
		b (SE)[a]	*P* value
Survey 1			
ID[b] with social environment	App acceptance	0.088 (0.044)	.046
ID with government	App acceptance	0.343 (0.051)	<.001
ID with social environment	Intention to use the app	0.083 (0.042)	.047
ID with government	Intention to use the app	0.325 (0.049)	<.001
Survey 2			
ID with social environment	App acceptance	0.081 (0.040)	.04
ID with government	App acceptance	0.314 (0.057)	<.001
ID with social environment	Intention to use the app	0.070 (0.035)	.04
ID with government	Intention to use the app	0.272 (0.046)	<.001
ID with social environment	Installation of this/an app	0.048 (0.029)	.10
ID with government	Installation of this/an app	0.185 (0.067)	.006
Survey 3			
ID with social environment	App acceptance	0.098 (0.032)	.002
ID with government	App acceptance	0.382 (0.049)	<.001
ID with social environment	Installation of this app	0.074 (0.025)	.004
ID with government	Installation of this app	0.288 (0.061)	<.001

[a]Coefficients are fully standardized (MPlus STDYX standardization).

[b]ID: identification.

Moreover, when testing for *indirect effects* of the identification measures (identification with the government and with the social environment) predicting outcomes regarding the app via greater trust, all indirect effects were supported (Table 6). This indicates that both types of identification (not only identification with members of the government but also identification with the social environment) predicted more trust and, accordingly, greater acceptance toward intentions to use or use (installation) of the app. In short, the results supported H1 and H2 across the three surveys and these central points in time, further demonstrating that the relation hypothesized in H2 was mediated via trust.

Discussion

Principal Results

In this study, we focused on the question: When are people motivated to use a contact tracing app? Results across three surveys at different points in time demonstrated the role of the social groups people belong to: the more people identified with their social environment (the beneficiaries) and the more they identified with members of the government (the source), the greater their app acceptance (ie, intentions and app installation). As predicted, identification with members of the government predicted greater app acceptance via more trust in the government; this outlines that trust in the source may be an important aspect that contributes to the acceptance of new technology, and that identification with the source may serve as a predictor of said trust. As such, the findings demonstrate the importance of these social groups beyond other target groups of identification (ie, people living in Germany or people around the world more generally).

Interestingly, the relation between identification with the social environment (ie, beneficiaries) and more app acceptance also seemed to be mediated via greater *trust*. To speculate about this exploratory finding, as highlighted in the Introduction, it is possible that identification with the social environment may not only make people more willing to trust those who explicitly belong to their ingroup (ie, social environment) but potentially even those who act as representatives for their larger ingroup (eg, society). However, this assumption awaits further confirmatory testing.

Limitations

Notably, our work is not without limitations. First, our study followed a cross-sectional design. Accordingly, the data presented here (including the mediation analyses) do not allow for conclusions about causality. An important step for future work is to go beyond this approach via collecting longitudinal data, which can not only help to investigate how the relations between these concepts unfold over time but can also enable examining how intentions and levels of social identification may change (eg, over the course of a pandemic or changes in the contact tracing app). Second, although we assessed data at different points in time and replicated the same model across three surveys, we did so only within a sample from one culture

(people living in Germany). Accordingly, replicating this work with experimental manipulations and in different societies would be desirable. Third, this work focused on intentions to use and use of the app, and the latter was operationalized via having installed the app (but not necessarily whether people constantly let it trace their movements); this served to be consistent across studies in our measures (as in Surveys 1 and 2, the required features of the upcoming app, such as keeping their Bluetooth on, were still unclear). Although these are important outcomes to study, it would be useful to build upon this work and extend it to, for instance, whether users do insert a (positive) test result.

Implications

Many new measures have been implemented since the outbreak of COVID-19 in 2020. The effectiveness of such measures greatly depends on people's willingness to adhere to them. Prior work suggests that one way to contribute to the acceptance of new measures is to appeal to *personal* benefits and keep the personal costs as low as possible. This study extends this prior work by adopting a focus on social relationships, namely, the extent to which people *identify with* (and trust in) the social groups that are relevant in this regard. The set of findings is relevant both in theoretical and practical terms.

From a theoretical point of view, these results add a crucial aspect to existing models on technology acceptance and health beliefs [15,46]. The results highlight that beyond known personal motives (eg, individual benefits and barriers), it is also important to take *social* aspects (eg, collective benefits or barriers) into account to understand when and why people will adopt a new technology. This seems especially relevant in interdependent contexts (eg, a pandemic) in which one person's health-related behavior (eg, social distancing, hand hygiene, or contact-tracing app use) more directly affects other people's situation. In line with prior work, we found that trust in the government constitutes a determinant of app acceptance [10,36] and, importantly, our findings show that identification predicts said trust.

This result is also important from a practical point of view. It suggests that to motivate people to adopt new technology such as the COVID-19 tracing app, one may not need to know all *personal* and facilitatory (which typically vary between people) barriers; rather, it seems important to foster their identification with social groups involved in the process. Indeed, promoting (the salience of) social identification with the source and/or beneficiaries could be achieved via several means. One example is to report more about "collective success" (eg, in fighting the pandemic; see related research on collective pride [47]). A second example is via governmental leaders engaging in identity leadership (creating a shared sense of "us" [48-50]) or politicians using consensual communication and/or "we"-referencing language (ie, referring more to "we," "us," and "ours," suggesting that they see themselves and act as "our leaders" [51,52]).

Conclusion

The results suggest that to motivate people to adhere to new measures in times of crisis (a global pandemic in this case), it is important to, in a responsible manner, take their relationships to social groups (ie, their *identification* with the source as well as the beneficiaries) into account (eg, potentially appealing to people's identification with these social groups). Doing so may not only contribute to a better understanding of what motivates people to accept new measures but also what contributes to their actual willingness to follow through with them.

Acknowledgments

We are very grateful to Julia Morad for her help in data collection for Survey 1.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Items, factor loadings, alternative models tested, and measures across all studies in their original order (Tables S1-S3).
[PDF File (Adobe PDF File), 492 KB - mhealth_v9i11e28146_app1.pdf]

Multimedia Appendix 2
Original materials (in German) for Surveys 1-3.
[PDF File (Adobe PDF File), 396 KB - mhealth_v9i11e28146_app2.pdf]

References

1. Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M. Features and functionalities of smartphone apps related to COVID-19: systematic search in app stores and content analysis. J Med Internet Res 2020 Aug 25;22(8):e20334 [FREE Full text] [doi: 10.2196/20334] [Medline: 32614777]
2. Corona Warn-App. Robert Koch Institut. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Archiv_Kennzahlen/Kennzahlen_05022021.pdf?__blob=publicationFile [accessed 2021-02-10]
3. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]

4. Kondylakis H, Katehakis DG, Kouroubali A, Logothetidis F, Triantafyllidis A, Kalamaras I, et al. COVID-19 mobile apps: a systematic review of the literature. J Med Internet Res 2020 Dec 09;22(12):e23170 [FREE Full text] [doi: 10.2196/23170] [Medline: 33197234]
5. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
6. Almalki M, Giannicchi A. Health apps for combating COVID-19: descriptive review and taxonomy. JMIR Mhealth Uhealth 2021 Mar 02;9(3):e24322 [FREE Full text] [doi: 10.2196/24322] [Medline: 33626017]
7. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
8. Montagni I, Roussel N, Thiébaut R, Tzourio C. Health care students' knowledge of and attitudes, beliefs, and practices toward the French COVID-19 app: cross-sectional questionnaire study. J Med Internet Res 2021 Mar 03;23(3):e26399 [FREE Full text] [doi: 10.2196/26399] [Medline: 33566793]
9. Blom AG, Wenz A, Cornesse C, Rettig T, Fikel M, Friedel S, et al. Barriers to the large-scale adoption of a COVID-19 contact tracing app in Germany: survey study. J Med Internet Res 2021 Mar 02;23(3):e23362 [FREE Full text] [doi: 10.2196/23362] [Medline: 33577466]
10. Zimmermann BM, Fiske A, Prainsack B, Hangel N, McLennan S, Buyx A. Early perceptions of COVID-19 contact tracing apps in German-speaking countries: comparative mixed methods study. J Med Internet Res 2021 Feb 08;23(2):e25525 [FREE Full text] [doi: 10.2196/25525] [Medline: 33503000]
11. Oldeweme A, Märtins J, Westmattelmann D, Schewe G. The role of transparency, trust, and social influence on uncertainty reduction in times of pandemics: empirical study on the adoption of COVID-19 tracing apps. J Med Internet Res 2021 Feb 08;23(2):e25893 [FREE Full text] [doi: 10.2196/25893] [Medline: 33465036]
12. Kaspar K. Motivations for social distancing and app use as complementary measures to combat the COVID-19 pandemic: quantitative survey study. J Med Internet Res 2020 Aug 27;22(8):e21613 [FREE Full text] [doi: 10.2196/21613] [Medline: 32759100]
13. Seto E, Challa P, Ware P. Adoption of COVID-19 contact tracing apps: a balance between privacy and effectiveness. J Med Internet Res 2021 Mar 04;23(3):e25726 [FREE Full text] [doi: 10.2196/25726] [Medline: 33617459]
14. Saw YE, Tan EY, Liu JS, Liu JC. Predicting public uptake of digital contact tracing during the COVID-19 pandemic: results from a nationwide survey in Singapore. J Med Internet Res 2021 Feb 03;23(2):e24730 [FREE Full text] [doi: 10.2196/24730] [Medline: 33465034]
15. Venkatesh V, Bala H. Technology Acceptance Model 3 and a research agenda on interventions. Decis Sci 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
16. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: a health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
17. Tajfel H, Turner JC. An integrative theory of intergroup conflict. In: Austin WG, Worchel S, editors. The social psychology of intergroup relations. Monterey, CA: Brooks Cole; 1979:33-37.
18. Tajfel H, Turner JC. The social identity theory of intergroup behaviour. In: Worchel S, Austin WG, editors. Psychology of intergroup relations. 2nd ed. Chicago: Nelson-Hall; 1986:7-24.
19. Behringer N, Sassenberg K, Scholl A. Knowledge contribution in organizations via social media: the interplay of identification and perceived usefulness. J Personnel Psychol 2017 Jan;16(1):12-24. [doi: 10.1027/1866-5888/a000169]
20. Brewer MB, Kramer RM. Choice behavior in social dilemmas: effects of social identity, group size, and decision framing. J Person Soc Psychol 1986 Mar;50(3):543-549 [FREE Full text] [doi: 10.1037/0022-3514.50.3.543]
21. Ouwerkerk JW, Ellemers N. The benefits of being disadvantaged: performance-related circumstances and consequences of intergroup comparisons. Eur J Soc Psychol 2002 Jan;32(1):73-91. [doi: 10.1002/ejsp.62]
22. O'Donnell AT, Jetten J, Ryan M. Who is watching over you? The role of shared identity in perceptions of surveillance. Eur J Soc Psychol 2009;40:135-147. [doi: 10.1002/ejsp.615]
23. Aron A, McLaughlin-Volpe T, Mashek D, Lewandowski G, Wright S, Aron E. Including others in the self. Eur Rev Soc Psychol 2004 Jan 04;15(1):101-132. [doi: 10.1080/10463280440000008]
24. Sassenberg K, Matschke C. The impact of exchange programs on the integration of the hostgroup into the self-concept. Eur J Soc Psychol 2009;40(1):148-159. [doi: 10.1002/ejsp.621]
25. Bass BM. Leadership and performance beyond expectations. New York: Free Press; 1985.
26. Conger JA, Kanugo RN. Charismatic leadership in organizations. Thousand Oaks, CA: Sage Publications; 1998.
27. House RJ. A 1976 theory of charismatic leadership. 1976 Presented at: Leadership: The Cutting Edge. Southern Ilinois University Fourth Biennial Leadership Symposium; October 26-28, 1976; Carbondale, IL p. 189-207. [doi: 10.2307/2523066]
28. Shamir B, Zakay E, Breinin E, Popper M. Correlates of charismatic leader behavior in military units: subordinates' attitudes, unit characteristics, and superiors' appraisals of leader performance. Acad Manag J 1998 Aug;41(4):387-409. [doi: 10.5465/257080]

29. Brewer MB. In-group bias in the minimal intergroup situation: a cognitive-motivational analysis. Psychol Bull 1979;86(2):307-324. [doi: 10.1037/0033-2909.86.2.307]
30. Brewer MB, Kramer RM. The psychology of intergroup attitudes and behavior. Annu Rev Psychol 1985 Jan;36(1):219-243. [doi: 10.1146/annurev.ps.36.020185.001251]
31. Abrams D, Hogg MA. Social identifications: A social psychology of intergroup relations and group processes. London: Routledge; Apr 26, 1990.
32. Messick DM, Mackie DM. Intergroup relations. Annu Rev Psychol 1989 Jan;40(1):45-81. [doi: 10.1146/annurev.ps.40.020189.000401] [Medline: 2648982]
33. Brewer MB. Ethnocentrismits role in interpersonal trust. In: Brewer MB, Collins BE, editors. Scientific inquiry and the social sciences. San Francisco, CA: Jossey-Bass; 1981:214-231.
34. Kramer RM, Brewer MB. Social group identitythe emergence of cooperation in resource conservation dilemmas. In: Wilke H, Messick D, Rutte C, editors. Experimental studies of social dilemmas. Frankfurt, Germany: Peter Lang; 1986:205-234.
35. Tyler TR, Degoey P. Collective restraint in social dilemmas: Procedural justice and social identification effects on support for authorities. J Person Soc Psychol 1995;69(3):482-497. [doi: 10.1037/0022-3514.69.3.482]
36. Blair RA, Morse BS, Tsai LL. Public health and public trust: Survey evidence from the Ebola Virus Disease epidemic in Liberia. Soc Sci Med 2017 Jan;172:89-97. [doi: 10.1016/j.socscimed.2016.11.016] [Medline: 27914936]
37. Morse B, Grépin KA, Blair RA, Tsai L. Patterns of demand for non-Ebola health services during and after the Ebola outbreak: panel survey evidence from Monrovia, Liberia. BMJ Glob Health 2016;1(1):e000007 [FREE Full text] [doi: 10.1136/bmjgh-2015-000007] [Medline: 28588907]
38. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
39. Dataset for Surveys 1-3. PsychArchives. 2021 Sep 21. URL: https://www.psycharchives.org/handle/20.500.12034/4543 [accessed 2021-11-12]
40. Code for Surveys 1-3. PsychArchives. 2021 Sep 21. URL: https://www.psycharchives.org/handle/20.500.12034/4542 [accessed 2021-11-12]
41. McFarland S, Brown D, Webb M. Identification with all humanity as a moral concept and psychological construct. Curr Dir Psychol Sci 2013 Jun 04;22(3):194-198. [doi: 10.1177/0963721412471346]
42. Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. Acad Manag Rev 1995 Jul;20(3):709-734. [doi: 10.2307/258792]
43. Winter K, Scholl A, Sassenberg K. A matter of flexibility: Changing outgroup attitudes through messages with negations. J Pers Soc Psychol 2021 Apr;120(4):956-976. [doi: 10.1037/pspi0000305] [Medline: 32614220]
44. Alge BJ, Ballinger GA, Tangirala S, Oakley JL. Information privacy in organizations: empowering creative and extrarole performance. J Appl Psychol 2006 Jan;91(1):221-232. [doi: 10.1037/0021-9010.91.1.221] [Medline: 16435952]
45. Mplus. URL: https://www.statmodel.com/ [accessed 2021-11-19]
46. Rosenstock IM. Historical origins of the Health Belief Model. Health Educ Monogr 1974 Dec 01;2(4):328-335. [doi: 10.1177/109019817400200403]
47. Leach CW, Ellemers N, Barreto M. Group virtue: the importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. J Pers Soc Psychol 2007 Aug;93(2):234-249. [doi: 10.1037/0022-3514.93.2.234] [Medline: 17645397]
48. Ellemers N, De Gilder D, Haslam SA. Motivating individuals and groups at work: a social identity perspective on leadership and group performance. Acad Manag Rev 2004 Jul;29(3):459-478. [doi: 10.5465/amr.2004.13670967]
49. Haslam SA, Reicher SD, Platow MJ. The new psychology of leadership: Identity, influence and power. Hove, UK: Psychology Press; 2011.
50. Hogg MA, van Knippenberg D, de Rast III DE. The social identity theory of leadership: Theoretical origins, research findings, and conceptual developments. Eur Rev Soc Psychol 2012 Mar;23(1):258-304. [doi: 10.1080/10463283.2012.741134]
51. Steffens NK, Haslam SA. Power through 'us': leaders' use of we-referencing language predicts election victory. PLoS One 2013;8(10):e77952 [FREE Full text] [doi: 10.1371/journal.pone.0077952] [Medline: 24194904]
52. Fladerer MP, Haslam SA, Steffens NK, Frey D. The value of speaking for “us”: the relationship between CEOs’ use of I- and We-referencing language and subsequent organizational performance. J Bus Psychol 2020 Feb 06;36(2):299-313. [doi: 10.1007/s10869-019-09677-0]

Edited by L Buis; submitted 23.02.21; peer-reviewed by K Liu, V Schury; comments to author 06.05.21; revised version received 17.05.21; accepted 17.09.21; published 24.11.21.

Please cite as:
Scholl A, Sassenberg K
How Identification With the Social Environment and With the Government Guide the Use of the Official COVID-19 Contact Tracing App: Three Quantitative Survey Studies
JMIR Mhealth Uhealth 2021;9(11):e28146
URL: https://mhealth.jmir.org/2021/11/e28146
doi:10.2196/28146
PMID:34662289

Original Paper

Detecting Tonic-Clonic Seizures in Multimodal Biosignal Data From Wearables: Methodology Design and Validation

Sebastian Böttcher[1,2], MSc; Elisa Bruno[3], PhD; Nikolay V Manyakov[4], PhD; Nino Epitashvili[1], MD; Kasper Claes[5], PhD; Martin Glasstetter[1], MSc; Sarah Thorpe[6], BSc Hons; Simon Lees[6]; Matthias Dümpelmann[1], PhD; Kristof Van Laerhoven[2], Prof Dr; Mark P Richardson[3,7], PhD; Andreas Schulze-Bonhage[1], Prof Dr; The RADAR-CNS Consortium[8]

[1]Epilepsy Center, Department of Neurosurgery, Medical Center - University of Freiburg, Freiburg im Breisgau, Germany

[2]Ubiquitous Computing, Department of Electrical Engineering and Computer Science, University of Siegen, Siegen, Germany

[3]Division of Neuroscience, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[4]Data Science Analytics & Insights, Janssen Research & Development, Beerse, Belgium

[5]UCB Pharma, Brussels, Belgium

[6]The RADAR-CNS Patient Advisory Board, King's College London, London, United Kingdom

[7]National Institute of Health Research Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, United Kingdom

[8]see Acknowledgements

Corresponding Author:
Sebastian Böttcher, MSc
Epilepsy Center
Department of Neurosurgery
Medical Center - University of Freiburg
Breisacher Str. 64
Freiburg im Breisgau, 79106
Germany
Phone: 49 76127052410
Email: sebastian.boettcher@uniklinik-freiburg.de

Abstract

Background: Video electroencephalography recordings, routinely used in epilepsy monitoring units, are the gold standard for monitoring epileptic seizures. However, monitoring is also needed in the day-to-day lives of people with epilepsy, where video electroencephalography is not feasible. Wearables could fill this gap by providing patients with an accurate log of their seizures.

Objective: Although there are already systems available that provide promising results for the detection of tonic-clonic seizures (TCSs), research in this area is often limited to detection from 1 biosignal modality or only during the night when the patient is in bed. The aim of this study is to provide evidence that supervised machine learning can detect TCSs from multimodal data in a new data set during daytime and nighttime.

Methods: An extensive data set of biosignals from a multimodal watch worn by people with epilepsy was recorded during their stay in the epilepsy monitoring unit at 2 European clinical sites. From a larger data set of 243 enrolled participants, those who had data recorded during TCSs were selected, amounting to 10 participants with 21 TCSs. Accelerometry and electrodermal activity recorded by the wearable device were used for analysis, and seizure manifestation was annotated in detail by clinical experts. Ten accelerometry and 3 electrodermal activity features were calculated for sliding windows of variable size across the data. A gradient tree boosting algorithm was used for seizure detection, and the optimal parameter combination was determined in a leave-one-participant-out cross-validation on a training set of 10 seizures from 8 participants. The model was then evaluated on an out-of-sample test set of 11 seizures from the remaining 2 participants. To assess specificity, we additionally analyzed data from up to 29 participants without TCSs during the model evaluation.

Results: In the leave-one-participant-out cross-validation, the model optimized for sensitivity could detect all 10 seizures with a false alarm rate of 0.46 per day in 17.3 days of data. In a test set of 11 out-of-sample TCSs, amounting to 8.3 days of data, the model could detect 10 seizures and produced no false positives. Increasing the test set to include data from 28 more participants without additional TCSs resulted in a false alarm rate of 0.19 per day in 78 days of wearable data.

Conclusions: We show that a gradient tree boosting machine can robustly detect TCSs from multimodal wearable data in an original data set and that even with very limited training data, supervised machine learning can achieve a high sensitivity and low false-positive rate. This methodology may offer a promising way to approach wearable-based nonconvulsive seizure detection.

(JMIR Mhealth Uhealth 2021;9(11):e27674) doi:10.2196/27674

KEYWORDS

wearables; epilepsy; seizure detection; multimodal data; mHealth; mobile health; digital health; eHealth

Introduction

Background

Epilepsy is one of the most common chronic neurological diseases, with a reported yearly worldwide incidence of more than 60 per 100,000 individuals [1]. Epilepsy also has a remarkably diverse set of indications, with several different types of symptoms and characteristic seizures of varying severity. Seizures are usually distinguished by their onset in the brain, focal or generalized. They can involve a variety of different combinations of symptoms, including impaired awareness or loss of consciousness; cognitive, emotional, or sensory abnormalities; sudden changes in the autonomic nervous system; or motor manifestations such as spasms, automatisms, or tonic and clonic movements of the limbs [2]. These convulsive seizures, particularly focal to bilateral or generalized tonic-clonic seizures (TCSs), are the most dangerous type of epileptic seizures. They imply loss of consciousness and loss of motor control with considerable risk for physical harm and can transition to life-threatening status epilepticus or sudden unexpected death in epilepsy [3]. For the diagnosis and treatment of epilepsy, clinicians rely on patient self-reporting and structured diaries, counting the number of seizures a patient had in a certain time frame. However, personal diaries filled out by the patients themselves have been proven to be very unreliable, with frequent undercounting because of a lack of awareness of seizures [4,5]. An objective seizure diary is therefore needed to obtain valid data on seizure occurrence, contributing to improved guidance for the treatment of people with epilepsy. Wearable nonelectroencephalography (non-EEG) devices (*wearables*) could provide data for such a diary. They are discreet and unobtrusive, contrary to many wearable EEG devices that are often cumbersome and stigmatizing [6], although some less obtrusive wearable EEG systems are in development [7,8]. Moreover, a robust detection of convulsive seizures with wearables, paired with identification of seizure-related risk factors [9], could be of great clinical importance and provide essential information for the identification of seizure-related sudden unexpected death in epilepsy risk factors.

Although seizure detection with non-EEG wearables is a relatively new field in epilepsy research, there have already been some studies that have demonstrated the viability of this kind of system. To date, most studies have concentrated on a single biosignal modality for training a seizure detection model, with a minority using a multimodal approach [10,11]. In essence, there are 4 main biosignal modalities that are recorded from non-EEG wearables used in epilepsy research: (1) *accelerometry* (ACC)—motion-based activity, (2) *electrodermal activity* (EDA)—changes in electrical properties of the skin, (3) *electrocardiography* (ECG) or *photoplethysmography* (PPG)—heart rate and heart rate variability estimation; and (4) *electromyography* electrical muscle activity. ACC is perhaps the most commonly used in related work because it is easy to integrate into wearable hardware and can provide relevant information, especially on movements during motor seizures. ACC signals have been used in both unimodal [12-14] and multimodal [15-17] seizure detection systems. EDA, also called galvanic skin response, has been used in some studies for seizure detection [16,18], as a large EDA change can occur especially in the postictal phase following TCSs [19]. Another modality that has been used is ECG, and its optical counterpart PPG, which uses light reflection to calculate the heart rate from blood volume changes in an unobtrusive manner. Although there have been some studies using ECG [20-23] or PPG [17,18,23,24] signals for epileptic seizure detection, the considerable movements during convulsive seizures frequently render this signal too noisy for accurate ictal heart rate determination. Finally, electromyography is a self-evident modality for detecting seizures with motor components, identifying ictal muscle contraction, and thus has been used for convulsive seizure detection as well [25-28].

Objective

In this study, we present an automatic seizure detection system for TCSs using supervised machine learning that is straightforward to implement and reproduce. We evaluated the detection model on a newly recorded data set from a multicenter clinical study with wearable non-EEG devices. Finally, we discuss the detection system, its performance, and its limitations and conclude with an outlook of possible further applications for this detection approach.

Methods

Data Set

During the course of the study, between July 2017 and February 2020, we collected wearable device data from 243 patients diagnosed with epilepsy: 70.7% (172/243) of patients were recruited at the epilepsy monitoring unit (EMU) in the Epilepsy Center, Medical Center, University of Freiburg, and 29.2% (71/243) of patients were recruited at the EMU in the neurophysiological department of King's College Hospital, London. Patients with a diagnosis of epilepsy in the age range of 7 to 80 years were recruited, unless they had vigorous involuntary nonepileptic movements. Consecutive patients were admitted to their respective EMU as part of their standard epilepsy clinical care, for differential diagnosis or for presurgical evaluation, and may have had their antiepileptic medication reduced during the recording. All patients were continuously

monitored via a video EEG system during their stay in the EMU. Clinical experts (EB and NE) manually reviewed the video and EEG data for all participants and labeled type, onset, and offset for all seizures. Specifically, they also labeled the onset and termination of every motor manifestation, including the tonic and clonic phases of each seizure. These labels were then used as the ground truth in the training and testing phases of the evaluation. Participants wore a variety of different wearable devices across the 2 sites; however, the only device worn by participants from both sites was a wrist-worn device (Empatica E4, Empatica Inc). The study and recording procedures were further described and discussed in the review by Bruno et al [29]. All recruited patients provided written informed consent, and the study procedures were approved by local ethics committees, the ethics committee at the University of Freiburg (538/16), and the London Fulham Research Ethics Committee (16/LO/2209; Integrated Research Application System project ID216316).

All data recorded at the 2 sites were live streamed from each wearable device to 1 base device per participant, running an Android operating system and a custom-developed app. The data were then transmitted from all base devices to a central server and stored for later analysis. The system was developed by the Remote Assessment of Disease and Relapse-*Central Nervous System* consortium and is available as an open-source project on GitHub [30].

Owing to battery limitations, each participant was assigned 2 devices, between which they changed twice daily to ensure continuous recordings. The wearable device recorded 3-axis ACC at a sample rate of 32 Hz, EDA at 4 Hz, and PPG at 64 Hz, which was processed on the device to a blood volume pulse signal. Participants generally wore the device on the arm that was most involved in motor semiology during seizures, that is, the arm that presented the most significant movements. In the set of 10 participants with TCSs included here, each wore the device on their nondominant hand, except for 2 participants who specified that they were ambidextrous.

Features

An extensive feature set was created from the ACC and EDA signals, encompassing 141 ACC and 10 EDA features, at sliding window sizes of 2, 10, and 20 seconds for the ACC features, and 5, 10, and 20 minutes for the EDA features. PPG signals were not analyzed because of major ictal movement artifacts. Although artifacts in PPG data can still convey information, in that the presence of noise itself can be information, we chose to omit it here in favor of focusing on the other 2 biosignals, because the information of PPG motion artifacts is naturally included in the ACC signal as well. The ACC features included a variety of different time and frequency domain features. The EDA features represented the skin conductance level (SCL), that is, tonic low-frequency EDA changes, and skin conductance response rate (SCRR), that is, phasic or higher-frequency EDA changes, calculated against a baseline.

As detection models usually perform most effectively with smaller feature sets, both in terms of computational cost and prediction performance [31], we aimed to reduce the number of used features significantly. For this feature selection, we first looked at related literature in the field of wearable seizure detection to narrow down window sizes that effectively capture relevant signal changes in time and identify feature types that were successfully used previously. Therefore, we selected a window of 10 seconds for the ACC features [13,14,16] and a longer window of 5 minutes for the EDA features to capture the tonic changes in the EDA signal that evolve over longer periods [19]. We then visualized the feature data in a period around the seizure, overlaid over each other, and for all features separately. In addition, we plotted the mean and SD for each data series. The data that were used for these graphs were taken only from the seizures of participants that were not included in the test set to be used in the out-of-sample performance evaluation (see *Results* section). Features showing recurrent typical ictal changes were then visually selected for further analysis (Figure 1). Variable seizure durations were handled by upsampling shorter seizures by linear interpolation to the length of the longest seizure among those plotted.

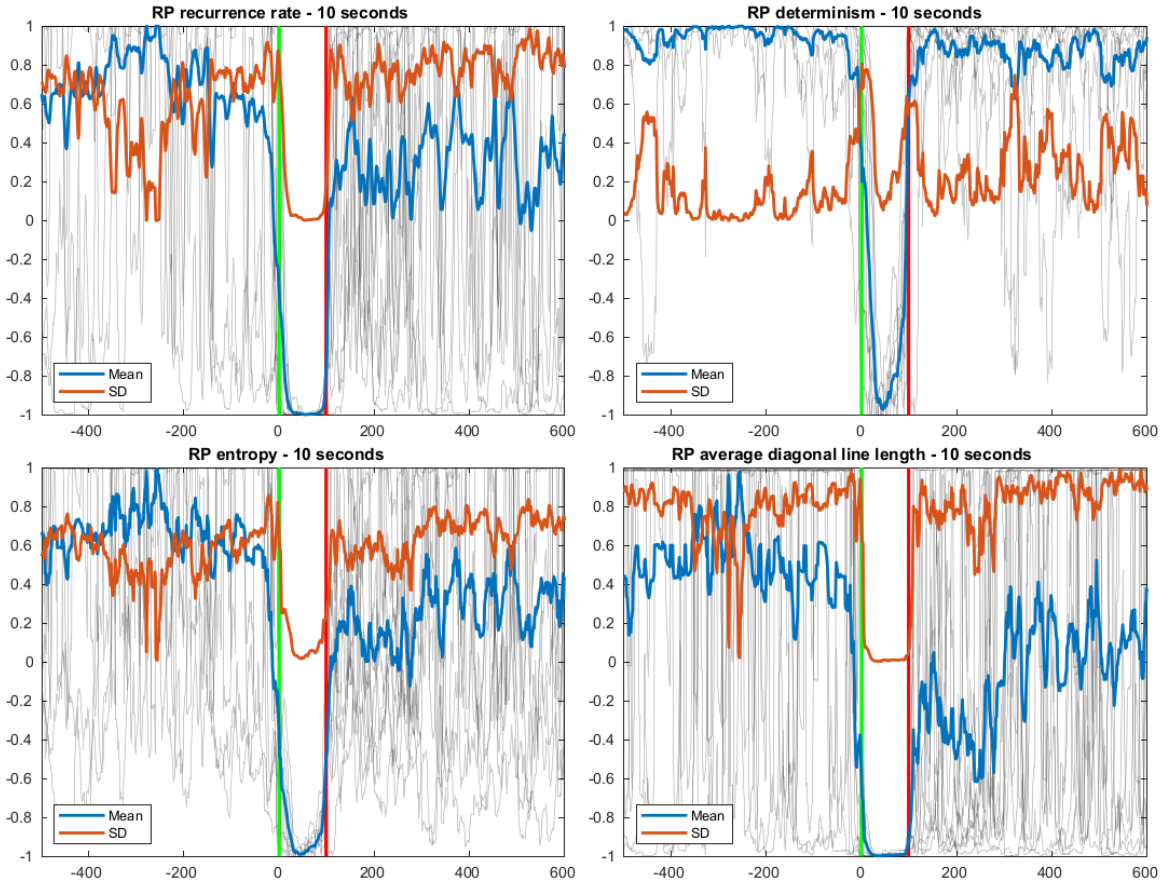
The resulting feature subset for the ACC modality consisted of the magnitude, zero crossing rate, and recurrence plot features (Figure 1) [32]. For the EDA features, the area under the curve and the maximum of the SCL within the window, and the SCRR were chosen, all corrected against a baseline, which is an interval of the same duration as the feature window, ending immediately before the beginning of the feature window. Thus, the resulting feature set can be divided into 4 main feature groups:

1. Magnitude of the ACC signal
   a. Raw ACC signal, over a 10-second window.
   b. Zero-phase band pass filtered ACC signal over a 10-second window. The band pass filter had a frequency band of 0.1 Hz to 10 Hz, representing the linear component of the ACC signal, and was applied before segmentation into windows.
   c. Zero-phase low-pass filtered ACC signal over a 10-second window. The low-pass filter had a cutoff frequency of 1 Hz, thus preserving only the gravitational component of the ACC signal, and was applied before segmentation into windows.
2. Zero crossing rate of the ACC signal over a 10-second window, for each of the 3 axes, respectively. The zero crossing rate is the number of times in a certain period the signal crosses the value 0 over the same period.
3. Four features calculated from the recurrence plot of the ACC signal:
   a. Determinism, that is, the percentage of points that form diagonal lines of a minimal length.
   b. The Shannon entropy of the probability that a line has a certain length.
   c. The average diagonal line length.
   d. Recurrence rate, that is, the density of recurrence points.
4. EDA-based features over a 5-minute window, minus the same value in the 5 minutes before the feature window
   a. The area under the curve of the SCL was calculated as the moving mean of the raw EDA signal over a 1-minute window.
   b. The maximum value of the SCL calculated as above.

c. The SCRR was calculated as the number of threshold crossings of the first derivative of the smoothed EDA signal within the window.

To accommodate the different window sizes over which the ACC and EDA features are calculated, a fixed interval between feature window applications was applied. This means that all features are calculated at fixed time points, with their respective windows centered on each consecutive point, creating the same number of feature vectors for both the ACC and EDA features over a segment of data. This enables the use of the complete, merged feature space as the single input into a detection model for training [11]. We chose this interval between the fixed time points for feature calculation as 2 seconds.

Figure 1. The overlaid feature value graphs for the recurrence plot features calculated from 10-second windows of the accelerometry data. Graphs representing feature values for each individual seizure (gray, background) are overlaid by the mean (blue) and SD (red). The green and red vertical bars represent the seizure onset and offset, respectively. The horizontal axis shows time in seconds related to seizure onset. All features are normalized between −1 and 1, independent from each other. RP: recurrence plot.



Seizure Detection

We used a gradient tree boosting machine (GTBM) [33] as the detection model for TCSs. Although similar to the well-known random forest (RF) method in being a set of trees that are grown with training data, a GTBM builds trees as weak learners in an additive manner. The model is improved with each new weak learner that is added to the ensemble, whereas the RF model trains all trees in parallel and independent of each other. Weak learners in this case are trees with a very low number of splits, down to decision stumps with just 1 split. This results in an overall lower bias and similar variance for GTBM models compared with RF models at the cost of higher parameter tuning effort. Therefore, gradient tree boosting models generally perform better than RF models if tuned sufficiently, and they have been successfully used in many machine learning problems [34]. To tackle this tuning effort, we performed hyperparameter optimization over several of the model parameters in a leave-one-participant-out (LOPO) manner. To this end, the data set was split into a training set and a test set. The training set consisted of the 10-minute peri-ictal data of 10 TCS from 8 patients with epilepsy recruited at the Freiburg site. The basic test set consisted of the complete data from 2 patients, 1 from the Freiburg site and 1 from the London site with 11 TCSs (see *Results* section). The hyperparameter optimization only used the training set to keep the test set unknown to the model before testing. All feature data were normalized between −1 and 1 before training and testing. For training, the combined feature input for the model, that is, the peri-ictal feature data of 10 TCS, were normalized, and for testing the complete feature data from the recordings for a participant were normalized independent from the feature data of the other participants in the test set.

The hyperparameter optimization was performed in a LOPO nested cross-validation manner on the training set. The data for 1 of the 8 participants in the training set were kept back as a validation set, and the model was trained on the seizures from the other 7 participants, using only 10-minute peri-ictal data for

each seizure. This reduction of the training data to only a small period around seizures helps with the large imbalance in the data set when comparing ictal and nonictal epochs. Once the model was trained, it was then tested on the complete data of the validation participant in the respective round, and the process was repeated 7 more times, cycling through the participants for validation. The mean score of the 8 validation runs was then saved as the performance of the current parameter combination, and the entire validation process was repeated for the next parameter combination. The parameters that were tuned in the optimization and their divisions are listed in Table 1, with the resulting optimal parameter combination highlighted. In total, 720 parameter combinations were evaluated in the hyperparameter optimization process.

Furthermore, the GTBM model also had some fixed parameters that were the same for all optimization runs. The boosting method used in the model was *adaptive boosting for binary classification* [35], and the misclassification cost for false negatives was always 1. The hyperparameter optimization resulted in an optimal set of parameters that were subsequently used in all the testing steps. The optimal parameter combination was chosen as the combination that achieved the highest sensitivity and lowest false alarm rate (FAR) during the LOPO validation run of the parameter combination, prioritizing sensitivity. Model parameters not specified here were left at their default values.

Table 1. Parameters optimized in the gradient tree boosting machine hyperparameter optimization and their optimization ranges.

Parameter	Value range	Description
Learning rate	1, *0.1*[a], 0.01, 0.001	The step size in the iterative learning process, also called shrinkage
Number of trees	25, 50, 100, *250*, 500, 750	The maximum number of trees to produce in the model
False positive cost	1, 10, 20, 30, 40, *50*	Specific misclassification cost for false positives when weighting during the learning process
Tree depth	*1*, 2, 4, 8, −1	The maximum number of splits in the decision tree, where −1 denotes one less than the number of samples in the training set, that is, the maximum possible value

[a]The chosen optimal parameter combination are italicized.

Evaluation

To process the model output and score its performance when compared with the ground truth, the same method was used both in the validation during hyperparameter optimization and later during the testing phase (see *Results* section). Owing to the method of feature extraction at fixed time intervals of 2 seconds described in the *Features* section, the output of the GTBM model is a prediction vector containing the predicted label every 2 seconds. The input labels, that is, the ground truth, and the predicted labels were binary, denoting the classification of each 2-second interval to either belong to a seizure or not. Comparing the ground truth and the prediction labels for evaluation can be done sample-wise by comparing each 2-second interval, or event-wise, by combining consecutive intervals of the positive class to distinct events. In our analysis, we chose the latter method, which requires postprocessing of the model output.

First, the prediction output of the model was smoothed with a hysteresis-like filter to avoid single-sample positives or gaps in consecutive positive predictions. To this end, all gaps between consecutive positive predictions smaller than 20 seconds in duration were filled out as positive, thus creating continuous, longer events from short neighboring positive predictions. Thereafter, all consecutive positive predictions of a certain length were discarded. We chose this value as 4 seconds, as it provides a good balance between discarding short, single-sample predictions and still keeping possible significant events. Thus, the prediction output of the model can be matched to the ground truth per participant by counting overlaps of predicted positive events with a positive ground truth event as true positives (TPs) and predicted positive events with no overlaps in the ground truth as false positives (FPs). The number of false negatives is then the difference between TPs and the number of seizures a participant recorded. The number of true negatives was not considered for this evaluation, as the sensitivity and are sufficient to evaluate a methodology for seizure detection. Unless otherwise stated, we report the sensitivity and FAR calculated across all relevant participants as a whole, not the mean over single participants.

All calculations for signal processing, feature extraction, and model development and evaluation were performed using *MATLAB 2020a* (MathWorks).

Results

Overview

For the study presented here, only study participants with focal to bilateral or generalized TCSs were included. This resulted in a data set of 21 TCSs from 10 participants, 9 from the Freiburg site with 19 seizures captured, and 1 from the London site with 2 seizures captured. The mean length of convulsive motor phenomena was 64 (SD 23) seconds. Table 2 lists the clinical and demographic information of the participants. They were 40% (4/10) female and on average 32.7 (SD 11.2) years old. The etiology of epilepsy for 2 participants was unknown at the time of recruitment. A total of 1 participant was diagnosed with generalized epilepsy, and the other 9 were diagnosed with focal epilepsy. For all captured seizures, wearable device data for at least 30 minutes before and after the ictal period were recorded in good quality; that is, the recorded data showed no major artifacts or intervals with constant 0 amplitude on visual inspection. A total of 612.6 hours of data were recorded for the included participants with seizures.

Table 2. Participants with recorded tonic-clonic seizure that were included in this study. Wearable data recorded from these participants were used in the evaluation of our seizure detection model. The recording duration is the duration that participants were wearing the device, without accounting for data loss.

Participant ID	Gender	Age (years)	Recording duration (days)	Epilepsy origin	Epilepsy type
FR1	Female	35	5	Unknown	Focal (TLE[a])
FR2	Female	26	6	Structural	Focal (TLE)
FR3	Male	22	4	Genetic	Generalized (IGE[b])
FR4	Female	34	4	Unknown	Focal (FLE[c])
FR5	Male	56	8	Structural	Focal (TLE)
FR6	Male	38	7	Structural	Focal (TLE)
FR7	Male	25	4	Structural	Focal (xTLE[d])
FR8	Male	16	7	Structural	Focal (FLE)
FR9	Male	37	12	Structural	Focal (xTLE)
LO1	Female	38	6	Structural	Focal (TLE)

[a]TLE: temporal lobe epilepsy.

[b]IGE: idiopathic generalized epilepsy.

[c]FLE: frontal lobe epilepsy.

[d]xTLE: extratemporal lobe epilepsy.

Cross-validation Training

The training set used for hyperparameter optimization included 10 seizures from 8 participants and covered 414.7 hours of wearable device data. With the best parameter combination, as described above, the LOPO cross-validation could detect all 10 seizures (sensitivity=100%) with a total of 8 FPs (FAR 0.46 per 24 hours). The FP rate was calculated as the ratio of total FPs across all participants to the number of hours of recordings multiplied by 24, and not the mean FAR across participants. In the training set LOPO cross-validation, 75% (6/8) of FPs were produced from the data of 1 participant and 2 by another. Thus, the other 6 participants were free of FPs. All 8 FPs detected by the model during the LOPO cross-validation occurred when the patient was off camera, for example, in the morning or evening when they were in the bathroom for their daily washing routine.

Out-of-Sample Testing

We also tested the model using a previously unseen test set from our overall data set. This test set included 11 seizures from 2 participants, 1 from the London site with 2 seizures recorded, and the other from the Freiburg site with 9 seizures recorded, for a total of 197.9 hours of test data. The choice of training and test set was deliberate: With the relatively low number of seizures and their distribution among participants in this data set, the goal was to train as many participants as possible but also having approximately the same number of seizures in the test set. This allocation ensures a model that is not patient specific while keeping the training and test sets balanced in terms of the number of seizures.

The GTBM model with the optimal parameters and trained with all 10 seizures from the training set could detect 10 of the 11 seizures in this test set (sensitivity=91%), without any FPs. However, this test set was rather limited as it was biased toward participants who had convulsive seizures; therefore, we expanded the test set to also include data from all 30 patients with epilepsy recruited at the London site that had data recorded with the wearable device. Although this does not add more seizures for the model to detect, it does add a considerable amount of data to assess the FP rate. The expanded test set thus encompasses 1935.9 hours of wearable device data from 31 participants, including the same 11 seizures as before. In this data set, the same model produces 30 FPs (0.37 per 24 hours). Further investigation of the FP distribution among the participants showed that 15 false detections resulted from a single participant who used a stepper during monitoring as physical activity to trigger her seizures. All FPs for that participant were related to this activity. Removing this participant performing unnatural repetitive movements from the expanded test set lowers the FP rate to 0.19 per 24 hours. Of the other participants in this expanded test set, the data of 2 participants produced 3 FPs, respectively, whereas 9 other participants each produced 1 FP, with the remaining 19 participants being free of FPs. Thus, the FAR, when calculated as the mean across all the included participants' individual FARs, was 0.45 (SD 1.1) per 24 hours, and 0.29 (SD 0.53) per 24 hours when excluding the participant with 15 FP. Table 3 provides a detailed overview of the results among the participants with recorded seizures.

Table 3. Per participant evaluation results, for participants with seizures recorded. The 3 totals given for the test set are (1) the total across the test set participants with seizures recorded (N=2), (2) the total when including all patients with epilepsy recruited at the London site with data recorded (not listed, N=31), and (3) the total when excluding 1 participant with an artificially disproportionate number of false positives (N=30).

Participant ID	Sensitivity, n (%)	FP[a], n	FAR[b] (per 24 hours)	PPV[c] (%)	Recording length (hours), n	Seizure type
Training set						
FR1	1 (100)	0	0	100	59.6	sGTCS[d]
FR2	1 (100)	6	1.56	14	92	sGTCS
FR3	2 (100)	0	0	100	35.5	GTCS[e]
FR4	1 (100)	2	1.34	33	35.8	sGTCS
FR5	1 (100)	0	0	100	36.3	sGTCS
FR6	1 (100)	0	0	100	88.5	sGTCS
FR7	1 (100)	0	0	100	40.7	sGTCS
FR8	2 (100)	0	0	100	26.2	sGTCS
Total	10 (100)	8	0.46	56	414.7	N/A[f]
Test set						
FR9	9 (100)	0	0	100	112.2	sGTCS
LO1	1 (50)	0	0	100	85.7	sGTCS
Total (1)	10 (91)	0	0	100	197.9	N/A
Total (2)	10 (91)	30	0.37	25	1935.9	N/A
Total (3)	10 (91)	15	0.19	40	1870.3	N/A

[a]FP: false positive.

[b]FAR: false alarm rate.

[c]PPV: positive predictive value.

[d]sGTCS: focal to bilateral tonic-clonic seizure.

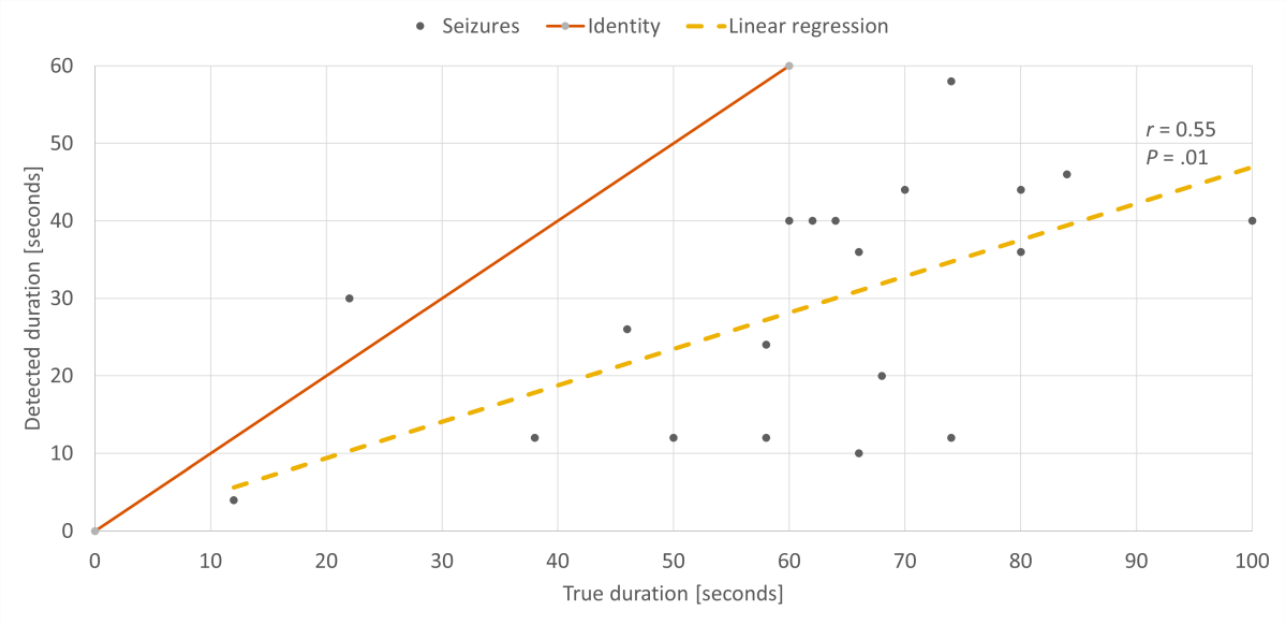
[e]GTCS: generalized tonic-clonic seizure.

[f]N/A: not applicable.

Seizure Duration

The duration of detected seizures was significantly correlated with the video EEG–based seizure duration, as labeled by clinical experts (Figure 2). The true seizure duration here is based on its clinical manifestation, that is, onset until offset of ictal motor phenomena related to TCS. In a Pearson correlation test, the correlation coefficient was r=.55, with P=.01. In general, the seizure duration was underestimated by the model as approximately half of the true duration, with a mean identified duration of 29 (SD 15) seconds versus the mean seizure duration of 64 (SD 23) seconds. This may reflect minor movement amplitudes during the tonic phase of the TCSs.

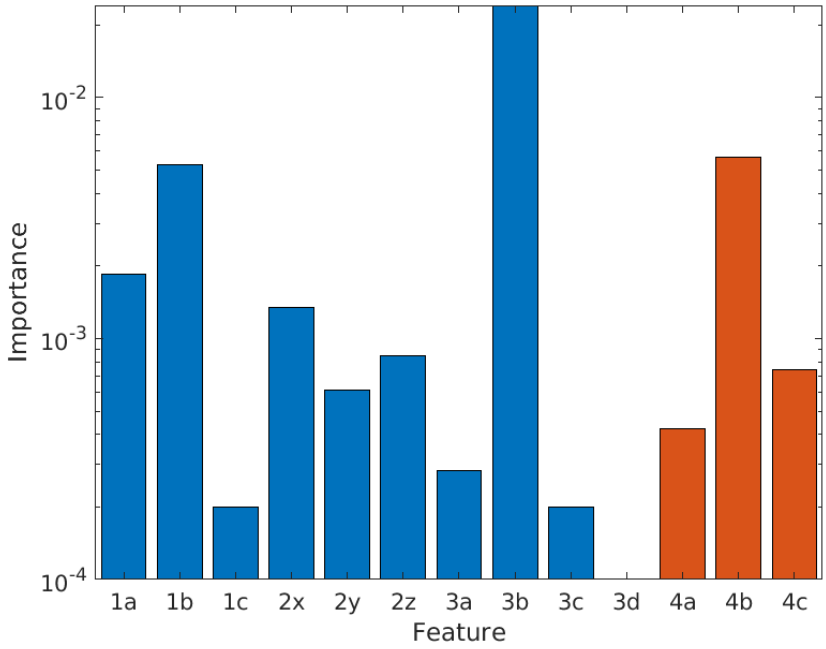
Figure 2. Correlation of the true seizure durations as labeled by clinical experts and the ictal durations detected by the gradient tree boosting machine model based on accelerometry and electrodermal activity. The dotted line shows the linear regression fit across the data points. The Pearson correlation coefficient was r=0.55, with *P*=.01. The identity line shows that the seizure duration is generally underestimated by the model.



Feature Importance

Furthermore, we analyzed the feature importance for our feature set, calculated as the mean feature importance over all trained GTBM models in the LOPO cross-validation (Figure 3), as a metric for the contribution of a specific feature to the performance of the model. The feature importance was based on the Gini impurity, calculated such that the smallest possible value was 0 [36]. Overall, all 4 feature groups, as described in the *Features* section, are represented in the resulting GTBM model to varying degrees of importance. The top 3 features among the feature set were that the Shannon entropy of the probability that a line in the recurrence plot had a certain length calculated over a 10-second window of the ACC signal, the magnitude of the band pass filtered ACC signal in a 10-second window, and the maximum of the SCL in a 5-minute window of the EDA signal, corrected for a baseline.

Figure 3. Feature importance, calculated as the mean feature importance of all models during a leave-one-participant-out cross-validation, with the optimal parameters of the gradient tree boosting machine as reported in the Seizure Detection section. All the features are shown as listed in the Features section (1: magnitude of accelerometry, 2: zero crossing rate of accelerometry, 3: recurrence plot features of accelerometry, and 4: electrodermal activity features). The feature importance is shown in logarithmic scale to better visualize smaller differences.



Discussion

Principal Findings

The results show that the GTBM model can robustly detect TCSs from non–EEG wearable device data. A sensitivity of 100% (10/10) on the training set during a LOPO cross-validation, a sensitivity of 91% (10/11) on the out-of-sample test set, and an FAR of less than 1 per 5 days in more than 1800 hours of data indicates a sufficient robustness of this methodology to consider it in designing an automated seizure diary. A large percentage of FPs occurred in a small percentage of participants, with most other participants showing between 0 and 0.5 FP per day. Furthermore, in participants who had TCS in our test set, no FPs were reported by the model. In addition, all true detections of our model occurred within the ictal period of the respective seizure, showing that the system has high accuracy. By evaluating a test set that includes data largely from 1 site (London), while the model was trained exclusively with data from the other site (Freiburg), we also showed the generalizability of our model.

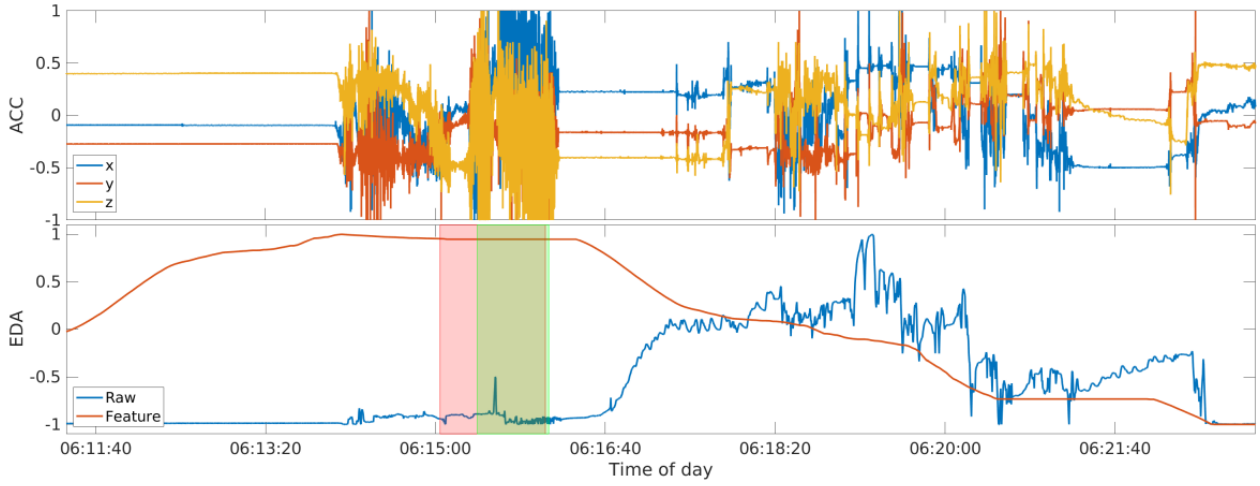
Although our data set contains continuous circadian data, most TCSs occurred during nighttime sleep. In the training set, 50% (5/10) of seizures occurred while the patient was awake, and in the test set, only 9% (1/11) occurred during wakefulness. Of these 6 awake seizures, 2 seizures occurred when the patient was outside the bed. All TP detections, both in the training set LOPO cross-validation and in the test set evaluation, occurred within the ictal phase of the respective seizure. Conversely, all FP detections occurred when the patient was awake and active, and most of them occurred during daytime. Patients were generally not confined to their beds but rather to their hospital rooms. They could freely perform a variety of activities of daily living, such as strolling across the room, going to the bathroom, brushing their teeth, eating and drinking, and washing themselves. Movement patterns during these activities, particularly if repetitive, could resemble those during convulsive seizures and may be a common source of FP detections. However, false alarms during these activities when the patient is awake could be ignored easily by way of patient validation and feedback to avoid inappropriate interventions.

Feature Importance

The distribution of feature contribution to the performance of the model shows that all selected features are used by the model to predict a seizure event, except for one, the recurrence rate in the recurrence plot of the ACC signal. The least amount of importance is assigned to the magnitude of the low-pass filtered ACC signal. This is an expected outcome, as this feature represents the gravitational component of the movement, which is minimal during convulsive seizures. During these seizures, almost all movements are part of the linear component, represented by the band pass filtered signal, which is also confirmed by this feature being one of the most important in the model.

Among the EDA-derived features, the highest importance was consistently assigned to the difference between the highest value in the feature and the baseline windows of the SCL. A typical EDA signal progression in the peri-ictal period is a steep increase from a low preictal baseline during the ictal phase, followed by a shallow decrease in the postictal phase, spanning multiple minutes. Thus, the feature based on the difference of the highest value between preictal, ictal, and postictal phases can sufficiently represent this trend, as evidenced by its high importance. Figure 4 shows the EDA signal progression and the respective maximum SCL feature during a seizure. The feature values are at their highest during the ictal phase, whereas the raw EDA signal shows the typical progression described above.

Figure 4. The seizure of participant LO1 that was detected by the model. The raw accelerometry signal is shown at the top, and the raw electrodermal activity signal as well as the best electrodermal activity feature (Section Features, Feature 4b) at the bottom; all are normalized between −1 and 1, independent from each other. The ictal tonic-clonic phase is overlaid in red, the true positive detection is overlaid in green. ACC: accelerometry; EDA: electrodermal activity.



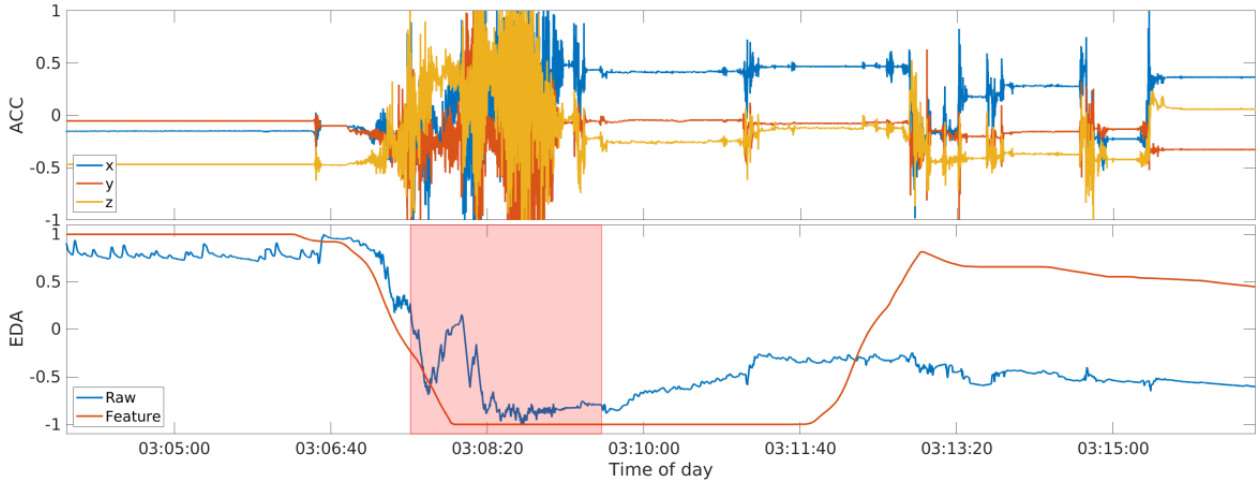
False Negatives

There was 1 seizure the model did not detect among the training and testing data sets (Figure 5). This false negative was produced by one of the participants recruited at the London site, and the seizure occurred during the night when the patient was asleep. The other seizure recorded for this participant was successfully detected by the model. To explain why the seizure was rejected

by the model, we examined the raw data before and after the seizure, specifically looking at the ACC response during the seizure, and the EDA trend going from the pre- to postictal phase. The motion response in the ictal phase of the rejected seizure was a typical progression from a short tonic phase at the beginning of the seizure to a longer, very pronounced, and violent clonic phase, stopping promptly with the seizure offset, followed by a short phase of postictal ACC silence. The raw EDA signal, however, follows a progression directly opposite to the signals from all other TCSs in the data set. The signal shows a steep decrease from a high baseline during the ictal phase and remains at a lower level in the postictal phase compared with the baseline in the preictal phase. Figures 4 and 5 show the comparison of data from the 2 recorded seizures from participant LO1, with the detected seizure being representative of all other TCS in the data set, especially those in the training set that created the model. Both seizures showed similar ACC data and a similar change in the ACC-based feature values. However, the EDA data and feature values were visibly opposite. This confirms that the model was trained properly on both the ACC and EDA features and that both modalities contributed to the model's classification of seizure occurrence. Thus, the misclassification of 1 event was due to atypical raw data and confirmed that the model included EDA features in its classification.

A possible explanation for the unusual EDA signal during this seizure could be that the EDA electrodes lost adequate contact with the skin, which was not fully re-established after the seizure. This could be caused by an improperly worn wearable device, or a loss of contact owing to the wearable device coming into contact with an external obstacle such as being pressed into the bed, slightly raising the EDA electrodes off the skin.

Figure 5. The seizure of participant LO1 that was not detected by the model and the single false negative that was produced during the evaluation. Note the differences in the electrodermal activity signal progression in comparison to Figure 4, which shows a typical response. The raw accelerometry signal is shown at the top, and the raw electrodermal activity signal and the best electrodermal activity feature (Section Features, Feature 4b) at the bottom; all are normalized between −1 and 1, independent from each other. The ictal tonic-clonic phase is overlaid in red. ACC: accelerometry; EDA: electrodermal activity.



Related Work

The research that is most closely related to our premise is certainly that of Onorati et al [16]. In their work, the Empatica research group developed a seizure detection model based on wearable data from the same device used in this study, *Empatica E4*. They used a support vector machine trained with 25 ACC as well as EDA features that were not further specified to detect convulsive seizures and achieve a very good performance, with their best classifier reaching a sensitivity of 94.5% and an FAR of 0.2 per day on 55 seizures from 22 patients. Our approach is on par with their results, and a contribution of the work presented here is to reinforce their findings. We show that the results of this quality can be achieved with a relatively basic methodology, and we describe this methodology in greater detail, making it fully accessible and reproducible. The methodology may even be transferrable to other diseases with convulsive attacks, such as paroxysmal dystonia or dissociative seizures. Thus, the study described here could be used as a stepping-stone for further work not only in epilepsy research but also in other medical fields.

In a further study, Kusmakar et al [13] used a monomodal support vector data description model on wearable ACC data to detect 21 generalized TCS from 12 patients, with a total recording length of 966 hours. The outlier classification model could achieve a sensitivity of 95% in a LOPO cross-validation, with a mean FAR of 0.72 per day. However, their model generated FP detections across almost all of the 12 included patients, showing a general trend toward FP detections independent of patient selection, whereas our model could achieve a generally lower FP rate on both the training and test sets, also revealing certain patients with a disproportionate FAR.

Arends et al [17] used the *LivAssured NightWatch* wearable device in a large ambulatory long-term monitoring study, collecting 908 convulsive seizures from 28 patients over more than 1800 nights. The device collects ACC and PPG signals from the patients' upper arm, specifically during the night. Their thresholding algorithm could detect 86% of the recorded seizures, with a positive predictive value of 49%, indicating that roughly half of all predictions were FPs. Although our methodology produces slightly worse results with respect to the overall FAR, studies differ in that the *NightWatch* study only

assessed nocturnal data with patients at rest, whereas our assessment, based on continuous data comprising wakefulness and sleep, showed the model's ability to correctly detect daytime seizures; notably, all our FPs were generated while the respective patient was awake and active.

In a more recent study, Johansson et al [14] used wrist-worn ACC sensors to detect 37 TCS from 11 patients with 666 hours of data. They evaluated 3 different types of models on a test set of 10 seizures and obtained the best result using an RF algorithm, detecting 9 of 10 seizures with an FAR of 0.24 per day. However, the evaluation of FPs is constrained in patients with TCS, introducing a certain bias in patient selection. In our evaluation, we added a control group of up to 29 participants without TCS recorded, with our model achieving a similar FAR, while also only producing FPs on these participants without seizures, whereas the participants with TCSs had no false alarms.

Limitations

The methodology for TCS detection described here also introduces some limitations, one of which is the long feature window used for the EDA feature computation. To include tonic changes in the EDA spanning over multiple minutes in the postictal phase, we used a 5-minute-long window, which automatically introduces an inherent detection delay, as a real-time system would need to first collect these data before being able to extract the EDA features and detect a potential seizure. Thus, this methodology would not be suited as a real-time warning system. Another limitation is the constraint of the model to detect only TCSs. As the model training process relies on data from the accelerometer sensor, nonmotor seizures cannot be detected with this set of modalities and features. Future work will be needed to assess the contribution of PPG and EDA sensors in detecting nonmotor seizures. Furthermore, the performance of the specific model trained here is likely not sufficient to be deployed directly as an automatic seizure diary, especially considering its constraint on TCS, which can be infrequent in everyday life. Additional work and more training data would be needed to create a system that is usable in clinical practice, possibly even shifting to a semipersonalized model that can be reinforced over time by patient feedback.

One of the most prevalent limitations in many studies in this field is the controlled in-hospital setting in which wearable device data are collected. Although patients in our study were able to perform some activities of daily living in and around their bed and were able to walk within their hospital room, the likelihood of FP generation can be assumed to be higher in an outpatient setting. False alarms during physical activity could be addressed by actively involving the patient through validation and feedback, for example, by giving them a chance to review seizure diary entries. Nevertheless, transferring this methodology to an ambulatory setting will require extensive modifications and reevaluation with data recorded in everyday living situations that include a gold standard for seizure labeling. In any case, a robust classifier that has a likelihood of working in the field must first be validated in an inpatient setting to progress to an ambulatory study, and the research presented here takes a clear step in that direction.

Acknowledgments

This project received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement no. 115902. This joint undertaking receives support from the European Union's Horizon 2020 Research and Innovation program and the European Federation of Pharmaceutical Industries and Associations. This communication reflects the views of the Remote Assessment of Disease and Relapse-*Central Nervous System* Consortium and neither Innovative Medicines Initiative nor the European Union and the European Federation of Pharmaceutical Industries and Associations are liable for any use that may be made of the information contained herein. The study processing fee was funded by the Baden-Wuerttemberg Ministry of Science, Research and Art, and the University of Freiburg in the funding program Open Access Publishing.

Conflicts of Interest

AS-B receives research funding from the German Ministry of Science, European Union, National Institute of Health, and from the companies BIAL, Precisis, and UNEEG; is an advisory board member of SEER Medical; and has received honoraria for lectures or advice from Arvelle, BIAL, EISAI, GW, Precisis, and UCB. MPR holds or coholds research funding from the UK Medical Research Council, UK National Institute for Health Research, Wellcome Trust, UK Engineering and Physical Sciences Research Council, Epilepsy Research UK, Epilepsy Foundation of America, European Commission, Canadian Institutes of Health Research, Xenon Pharma, and GW Pharma; has research collaborations with UNEEG Medical, SEER Medical, UCB Pharma, ANT Neuro, and IMEC; is a trustee of Epilepsy Research UK; and is an advisory board member of SUDEP Action. MPR does not receive personal remuneration from any of these sources. MPR holds a patent WO2013182848A1. NVM is an employee of Janssen Research & Development, Limited Liability Company, and holds company stocks or stock options. KC is an employee of UCB Pharma.

References

1. Beghi E. The epidemiology of epilepsy. Neuroepidemiology 2020 Dec 18;54(2):185-191 [FREE Full text] [doi: 10.1159/000503831] [Medline: 31852003]

2. Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE, et al. Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. Epilepsia 2017 Apr;58(4):522-530. [doi: 10.1111/epi.13670] [Medline: 28276060]
3. Ryvlin P, Nashef L, Lhatoo SD, Bateman LM, Bird J, Bleasel A, et al. Incidence and mechanisms of cardiorespiratory arrests in epilepsy monitoring units (MORTEMUS): a retrospective study. The Lancet Neurology 2013 Oct;12(10):966-977. [doi: 10.1016/S1474-4422(13)70214-X] [Medline: 24012372]
4. Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, et al. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. The Lancet Neurology 2013 Jun;12(6):563-571. [doi: 10.1016/S1474-4422(13)70075-9] [Medline: 23642342]
5. Hoppe C, Poepel A, Elger CE. Epilepsy: accuracy of patient seizure counts. Arch Neurol 2007 Nov;64(11):1595-1599. [doi: 10.1001/archneur.64.11.1595] [Medline: 17998441]
6. Schulze-Bonhage A, Sales F, Wagner K, Teotonio R, Carius A, Schelle A, et al. Views of patients with epilepsy on seizure prediction devices. Epilepsy Behav 2010 Aug;18(4):388-396. [doi: 10.1016/j.yebeh.2010.05.008] [Medline: 20624689]
7. Gu Y, Cleeren E, Dan J, Claes K, Van Paesschen W, Van Huffel S, et al. Comparison between Scalp EEG and Behind-the-Ear EEG for development of a wearable seizure detection system for patients with focal epilepsy. Sensors (Basel) 2017 Dec 23;18(1):29 [FREE Full text] [doi: 10.3390/s18010029] [Medline: 29295522]
8. Duun-Henriksen J, Baud M, Richardson MP, Cook M, Kouvas G, Heasman JM, et al. A new era in electroencephalographic monitoring? Subscalp devices for ultra-long-term recordings. Epilepsia 2020 Sep;61(9):1805-1817. [doi: 10.1111/epi.16630] [Medline: 32852091]
9. Bruno E, Böttcher S, Biondi A, Epitashvili N, Manyakov NV, Lees S, RADAR-CNS Consortium. Post-ictal accelerometer silence as a marker of post-ictal immobility. Epilepsia 2020 Jul;61(7):1397-1405. [doi: 10.1111/epi.16552] [Medline: 32459380]
10. Leijten FS, Dutch TeleEpilepsy Consortium. Multimodal seizure detection: a review. Epilepsia 2018 Jun;59 Suppl 1:42-47. [doi: 10.1111/epi.14047] [Medline: 29873832]
11. Böttcher S, Manyakov NV, Epitashvili N, Folarin A, Richardson MP, Dümpelmann M, et al. Using multimodal biosignal data from wearables to detect focal motor seizures in individual epilepsy patients. In: Proceedings of the 6th international Workshop on Sensor-based Activity Recognition and Interaction. 2019 Presented at: iWOAR '19: Proceedings of the 6th international Workshop on Sensor-based Activity Recognition and Interaction; Sep 16 - 17, 2019; Rostock Germany. [doi: 10.1145/3361684.3361687]
12. Nijsen T, Aarts R, Arends J, Cluitmans P. Automated detection of tonic seizures using 3-D accelerometry. In: Proceeding of the 4th European Conference of the International Federation for Medical and Biological Engineering. 2009 Presented at: 4th European Conference of the International Federation for Medical and Biological Engineering; Nov 23–27, 2008; Antwerp, Belgium. [doi: 10.1007/978-3-540-89208-3_47]
13. Kusmakar S, Karmakar C, Yan B, O'Brien T, Muthuganapathy R, Palaniswami M. Detection of generalized tonic-clonic seizures using short length accelerometry signal. In: Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2017 Presented at: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 11-15, 2017; Jeju, Korea (South). [doi: 10.1109/embc.2017.8037872]
14. Johansson D, Ohlsson F, Krýsl D, Rydenhag B, Czarnecki M, Gustafsson N, et al. Tonic-clonic seizure detection using accelerometry-based wearable sensors: a prospective, video-EEG controlled study. Seizure 2019 Feb;65:48-54 [FREE Full text] [doi: 10.1016/j.seizure.2018.12.024] [Medline: 30611010]
15. Poh M, Loddenkemper T, Reinsberger C, Swenson NC, Goyal S, Sabtala MC, et al. Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor. Epilepsia 2012 May;53(5):e93-e97 [FREE Full text] [doi: 10.1111/j.1528-1167.2012.03444.x] [Medline: 22432935]
16. Onorati F, Regalia G, Caborni C, Migliorini M, Bender D, Poh M, et al. Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors. Epilepsia 2017 Nov;58(11):1870-1879. [doi: 10.1111/epi.13899] [Medline: 28980315]
17. Arends J, Thijs RD, Gutter T, Ungureanu C, Cluitmans P, Van Dijk J, the Dutch Tele-Epilepsy Consortium. Multimodal nocturnal seizure detection in a residential care setting: a long-term prospective trial. Neurology 2018 Nov 20;91(21):e2010-e2019 [FREE Full text] [doi: 10.1212/WNL.0000000000006545] [Medline: 30355702]
18. Cogan D, Birjandtalab J, Nourani M, Harvey J, Nagaraddi V. Multi-biosignal analysis for epileptic seizure monitoring. Int J Neural Syst 2017 Feb;27(1):1650031. [doi: 10.1142/S0129065716500313] [Medline: 27389004]
19. Poh MZ, Loddenkemper T, Reinsberger C, Swenson NC, Goyal S, Madsen JR, et al. Autonomic changes with seizures correlate with postictal EEG suppression. Neurology 2012 Jun 05;78(23):1868-1876 [FREE Full text] [doi: 10.1212/WNL.0b013e318258f7f1] [Medline: 22539579]
20. De Cooman T, Varon C, Hunyadi B, Van Paesschen W, Lagae L, Van Huffel S. Online automated seizure detection in temporal lobe epilepsy patients using single-lead ECG. Int J Neural Syst 2017 Nov;27(7):1750022. [doi: 10.1142/S0129065717500228] [Medline: 28359222]

21. Jeppesen J, Fuglsang-Frederiksen A, Johansen P, Christensen J, Wüstenhagen S, Tankisi H, et al. Seizure detection using heart rate variability: a prospective validation study. Epilepsia 2020 Nov;61 Suppl 1:S41-S46. [doi: 10.1111/epi.16511] [Medline: 32378197]
22. De Cooman T, Vandecasteele K, Varon C, Hunyadi B, Cleeren E, Van Paesschen W, et al. Personalizing Heart Rate-Based Seizure Detection Using Supervised SVM Transfer Learning. Front Neurol 2020 Feb 26;11:145 [FREE Full text] [doi: 10.3389/fneur.2020.00145] [Medline: 32161573]
23. Vandecasteele K, De Cooman T, Gu Y, Cleeren E, Claes K, Paesschen W, et al. Automated epileptic seizure detection based on wearable ECG and PPG in a hospital environment. Sensors (Basel) 2017 Oct 13;17(10):2338 [FREE Full text] [doi: 10.3390/s17102338] [Medline: 29027928]
24. Mohammadpour Touserkani F, Tamilia E, Coughlin F, Hammond S, El Atrache R, Jackson M, et al. Photoplethysmographic evaluation of generalized tonic-clonic seizures. Epilepsia 2020 Aug;61(8):1606-1616. [doi: 10.1111/epi.16590] [Medline: 32652564]
25. Conradsen I, Beniczky S, Hoppe K, Wolf P, Sorensen HB. Automated algorithm for generalized tonic-clonic epileptic seizure onset detection based on sEMG zero-crossing rate. IEEE Trans Biomed Eng 2012 Feb;59(2):579-585. [doi: 10.1109/TBME.2011.2178094] [Medline: 22156944]
26. Halford JJ, Sperling MR, Nair DR, Dlugos DJ, Tatum WO, Harvey J, et al. Detection of generalized tonic-clonic seizures using surface electromyographic monitoring. Epilepsia 2017 Nov;58(11):1861-1869 [FREE Full text] [doi: 10.1111/epi.13897] [Medline: 28980702]
27. Fürbass F, Kampusch S, Kaniusas E, Koren J, Pirker S, Hopfengärtner R, et al. Automatic multimodal detection for long-term seizure documentation in epilepsy. Clin Neurophysiol 2017 Aug;128(8):1466-1472. [doi: 10.1016/j.clinph.2017.05.013] [Medline: 28622529]
28. Beniczky S, Conradsen I, Henning O, Fabricius M, Wolf P. Automated real-time detection of tonic-clonic seizures using a wearable EMG device. Neurology 2018 Jan 30;90(5):e428-e434 [FREE Full text] [doi: 10.1212/WNL.0000000000004893] [Medline: 29305441]
29. Bruno E, Böttcher S, Viana PF, Amengual-Gual M, Joseph B, Epitashvili N, et al. Wearable devices for seizure detection: practical experiences and recommendations from the Wearables for Epilepsy And Research (WEAR) International Study Group. Epilepsia 2021 Oct;62(10):2307-2321. [doi: 10.1111/epi.17044] [Medline: 34420211]
30. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, The Hyve, RADAR-CNS Consortium. RADAR-Base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR Mhealth Uhealth 2019 Aug 01;7(8):e11734 [FREE Full text] [doi: 10.2196/11734] [Medline: 31373275]
31. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003 Jan 3;3:1157-1182.
32. Großekathöfer U, Manyakov NV, Mihajlović V, Pandina G, Skalkin A, Ness S, et al. Automated detection of stereotypical motor movements in autism spectrum disorder using recurrence quantification analysis. Front Neuroinform 2017 Feb 16;11:9 [FREE Full text] [doi: 10.3389/fninf.2017.00009] [Medline: 28261082]
33. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001 Oct;29(5):1189-1232 [FREE Full text]
34. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013 Dec 4;7:21 [FREE Full text] [doi: 10.3389/fnbot.2013.00021] [Medline: 24409142]
35. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 1997 Aug;55(1):119-139. [doi: 10.1006/jcss.1997.1504]
36. Estimates of predictor importance for classification ensemble of decision trees. Mathworks. URL: https://mathworks.com/help/stats/compactclassificationensemble.predictorimportance.html [accessed 2021-09-29]

Abbreviations

ACC: accelerometry
ECG: electrocardiography
EDA: electrodermal activity
EEG: electroencephalography
EMU: epilepsy monitoring unit
FAR: false alarm rate
FP: false positive
GTBM: gradient tree boosting machine
LOPO: leave-one-participant-out
Non-EEG: nonelectroencephalography
PPG: photoplethysmography
RF: random forest
SCL: skin conductance level
SCRR: skin conductance response rate

TCS: tonic-clonic seizure
TP: true positive

Edited by L Buis; submitted 02.02.21; peer-reviewed by F Leijten, E Kuosmanen; comments to author 16.06.21; revised version received 23.07.21; accepted 03.09.21; published 19.11.21.

Please cite as:
Böttcher S, Bruno E, Manyakov NV, Epitashvili N, Claes K, Glasstetter M, Thorpe S, Lees S, Dümpelmann M, Van Laerhoven K, Richardson MP, Schulze-Bonhage A, The RADAR-CNS Consortium
Detecting Tonic-Clonic Seizures in Multimodal Biosignal Data From Wearables: Methodology Design and Validation
JMIR Mhealth Uhealth 2021;9(11):e27674
URL: https://mhealth.jmir.org/2021/11/e27674
doi:10.2196/27674
PMID:34806993

Original Paper

The Relationship Between Weight Loss Outcomes and Engagement in a Mobile Behavioral Change Intervention: Retrospective Analysis

Alissa Carey[1], MA; Qiuchen Yang[1], MS; Laura DeLuca[1,2], PhD; Tatiana Toro-Ramos[1,3], PhD; Youngin Kim[1,4], MD; Andreas Michaelides[1], PhD

[1]Academic Research, Noom Inc, New York, NY, United States

[2]Department of Clinical Psychology, Ferkauf Graduate School of Psychology, Yeshiva University, Bronx, NY, United States

[3]Amgen, Thousand Oaks, CA, United States

[4]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Andreas Michaelides, PhD
Academic Research
Noom Inc
Fl 9, 229 W 28th St
New York, NY, 10001
United States
Phone: 1 631 938 1248
Email: andreas@noom.com

Abstract

Background: There is large variance in weight loss outcomes of digital behavior change interventions (DBCIs). It has been suggested that different patterns of engagement in the program could be responsible for this variance in outcomes. Previous studies have found that the amount of engagement on DBCIs, such as the number of meals logged or articles read, is positively associated with weight loss.

Objective: This retrospective study extends previous research by observing how important weight loss outcomes (high weight loss: 10% or greater body weight loss; moderate weight loss: between 5% to 10%; stable weight: 0 plus or minus 1%) are associated with engagement on a publicly available mobile DBCI (Noom) from 9 to 52 weeks.

Methods: Engagement and weight data for eligible participants (N=11,252) were extracted from the Noom database. Engagement measures included the number of articles read, meals logged, steps recorded, messages to coach, exercise logged, weigh-ins, and days with 1 meal logged per week. Weight was self-reported on the program. Multiple linear regressions examined how weight loss outcome (moderate and high vs stable) was associated with each engagement measure across 3 study time periods: 9-16 weeks, 17-32 weeks, and 33-52 weeks.

Results: At 9-16 weeks, among the 11,252 participants, 2594 (23.05%) had stable weight, 6440 (57.23%) had moderate weight loss, and 2218 (19.71%) had high weight loss. By 33-52 weeks, 525 (18.21%) had stable weight, 1214 (42.11%) had moderate weight loss, and 1144 (39.68%) had high weight loss. Regression results showed that moderate weight loss and high weight loss outcomes were associated with all engagement measures to a significantly greater degree than was stable weight (all *P* values <.001). These differences held across all time periods with the exception of exercise for the moderate weight loss category at 1 time period of 33-52 weeks. Exercise logging increased from 9 to 52 weeks regardless of the weight loss group.

Conclusions: Our results suggest that these clinically important weight loss outcomes are related to the number of articles read, meals logged, steps recorded, messages to coach, exercise logged, weigh-ins, and days with 1 meal logged per week both in the short-term and long-term (ie, 1 year) on Noom. This provides valuable data on engagement patterns over time on a self-directed mobile DBCI, can help inform how interventions tailor recommendations for engagement depending on how much weight individuals have lost, and raises important questions for future research on engagement in DBCIs.

(JMIR Mhealth Uhealth 2021;9(11):e30622) doi:10.2196/30622

KEYWORDS

engagement; mHealth; obesity; weight management; Noom; application; app; behavioral change; digital behavior change interventions

Introduction

Engagement and Digital Behavior Change Interventions

Obesity and its potential comorbidities are a significant and increasing public health burden, with an estimated global cost of US $2 trillion per year due to economic loss of productivity and direct medical expenses stemming from weight-related issues [1]. Traditional dietary approaches to treat obese and overweight status have known shortcomings, calling for innovative solutions that involve behavioral management [2,3]. Digital behavior change interventions (DBCIs) such as mobile programs use technology to enhance availability and convenience compared to traditional in-person interventions, and these programs are growing in number [4-6]. These digital interventions are effective for weight loss and chronic disease prevention and management [7-10].

Body weight loss of 5%-10% is associated with improved risk of metabolic and cardiovascular conditions, and 10% or more loss is associated with even greater improvement [11]. Therefore, body weight loss of at least 5% is regarded as a clinically meaningful outcome [12]. However, there is wide variability in weight loss outcomes even when individuals use the same program [13,14]. It has been suggested that this variability could be due to differences in engagement with the program [15]. Engagement has been defined as “the extent (eg, amount, frequency, duration, depth) of usage” of the program [16]. Common measures of engagement include the amount of time spent on the platform, the number of times an individual has used a program feature such as weight or food logging, and the number of articles read [15].

Previous Work on Engagement

Previous work has found positive associations between engagement and weight loss outcomes [10,17-22]. In a digital commercial program, the number of weigh-ins per week, steps per day, active minutes per week, days logging meals per week, and the percentage of weeks with 5 or more meal logs were associated with weight loss at 6 months [19]. In the same study, weighing in at least 3 times a week, achieving 60 highly active minutes per week, and logging meals 3 times per week were associated with 5% or more body weight loss [19]. In our previous work on Noom, a commercial mobile DBCI, we found that the number of meal logs and group posts were associated with greater weight loss at 65 weeks, and the number of messages sent to the coach, exercises logged, and articles read were associated with weight gain [20]. We also previously found that the number of meal logs and weigh-ins were associated with weight loss at 6 months on Noom [10].

This Study

We extend this body of work in this retrospective study by examining how specific weight loss outcomes of interest are associated with engagement using a large sample and multiple time points. Multiple time points allow for the investigation of whether associations change over time, which is important because it is well established that engagement declines over time [23]. This study will allow for better understanding of how individuals who lost certain amounts of weight engaged in the program, which can inform future attempts to encourage engagement in specific and tailored ways based on current weight and goal weight loss. Specifically, we explored associations between weight loss outcomes of clinical importance (5%-10%, 10% and more, and stable weight) and various measures of engagement (the number of articles read, meals logged, steps recorded, messages to coach, exercise logged, weigh-ins, and days with 1 meal logged per week) from 9 to 52 weeks on Noom. Based on past work, it was hypothesized that associations between weight loss outcomes and the number of meal logs and weigh-ins would be stronger for moderate (5%-10%) or higher amounts of weight loss (10% or more) compared to stable weight loss outcomes [10,19,20], but it was unclear if that would be the case for all engagement measures because of mixed prior results [18]. We also hypothesized that the difference in engagement between these weight loss groups would hold over time [18].

Methods

Intervention

Noom is a behavior change and weight management mobile health intervention that provides users with self-monitoring features for food, exercise, and weight monitoring, as well as access to a virtual 1:1 behavior change coach, support group facilitated by a health coach, and a daily curriculum that includes diet, exercise, and psychoeducation. Noom’s theoretical foundation stems from cognitive behavioral therapy; third wave cognitive behavioral therapy, such as dialectical behavioral therapy and motivational interviewing techniques; and behavior change techniques, such as self-monitoring and social support [24-27].

Participants

Retrospective cohort data were extracted directly from Noom’s (Noom Inc) database in December 2019 and deidentified. Participants had all voluntarily signed up for the Noom Healthy Weight program online or through the app store (iTunes or Google Play). This study was approved by the Advarra Inc Institutional Review Board (Columbia, Maryland). As part of the approved protocol, at initial sign-up, all users were given an opportunity to consent to the use of all of their program data for research, and all users were given the opportunity to opt out and deny consent.

To be included in this study, individuals were required to be Noom users in the Healthy Weight program for up to 52 weeks beginning on December 1, 2018; had provided baseline weight, age, gender, and height; were 18 years or older; and fell into one of the 3 weight loss outcome categories used in the study.

A length of 52 weeks was chosen to be able to explore long-term weight loss and engagement [28]. Additionally, participants had to open the mobile health platform at least once after week 8 to be included in the study as a minimum threshold of activity. Week 8 was chosen because this would represent activity from week 9 onwards, which is when the first study time period began.

All participants were placed into 1 of 3 weight outcome categories based on their weight change from baseline: stable weight (0% plus or minus 1%), moderate weight loss (between 5% and 10%), and high weight loss (lost 10% or more body weight). These categories represent clinically meaningful weight loss outcomes, and labels were chosen following previous work [29,30].

The following time periods were chosen for analysis: 9-16 weeks, 17-32 weeks, and 33-52 weeks. The initial time period was chosen based on program length (16 weeks), and the starting point was set to halfway through the program (9 weeks) to prevent bias from early fluctuations in motivation or weight. The final time period (52 weeks) was designated based on previous work [31], and the middle time period (32 weeks) was chosen to represent an intermediate interval between the initial and final time periods. To be included at later time periods, participants had to fall into 1 weight change category and have opened the platform at least once during week 16 to be analyzed at weeks 17 to 32, and have opened the platform at least once in week 32 in order to be analyzed in weeks 33 to 52.

Measures

Weight, as well as baseline characteristics of gender, age, and height were self-reported by the users through the mobile interface.

The following engagement measures were used: number of days with at least 1 meal logged per week—a measure calculated based on participants' weekly self-reported food logs; number of articles read per week; number of meal logs per week—the number of meals participants logged in the platform per week; number of coach messages per week—the number of times participants messaged their coach per week; count of steps per week—the number of steps taken per week, either recorded by the participant's in-phone pedometer or supplemented by self-report in the platform; count of weigh-ins per week—the number of times participants self-reported their weight in the platform per week; and count of exercises per week—the number of times participants self-reported exercising in the platform per week. These engagement measures comprehensively included the possible ways users could actively participate on the platform.

No engagement measures were required as part of the intervention. The curriculum content (articles) functioned on a fixed schedule where participants were shown potential articles to read containing nutrition education, psychoeducation, and motivational information each day. They were encouraged at the beginning of the program to read these articles as part of a daily task list. Participants were also encouraged to perform weight logging at least once a week and to log all of their meals daily. Participants had the option of setting up push notifications to remind them to log their meals at certain times. Using this optional reminder system was not tracked as an engagement measure. Coaches were instructed to reply to user messages within 24 hours of receiving them, and, if the participant did not send the coach a message in 7 days, the coach would reach out with a weekly check-in to invite discussion over the participant's progress. Participants' engagement with coach messaging was calculated based on messages that they sent, not messages received.

Statistical Analysis

Descriptive statistics were calculated for users' baseline characteristics and are expressed in mean and SD for continuous variables and frequency and percentage for categorical variables (Table 1). These characteristics were self-reported, including users' weights, which were measured by the users with their own scales.

Table 1. Demographic characteristics for weight change groups across study periods.

Characteristic	Weight change groups			*P* value
	Stable (0 plus or minus1%)	Moderate loss (between 5%-10%)	High loss (10% or greater)	
Age (years), mean (SD)				
9-16 weeks	47.70 (12.27)	49.93 (12.59)	49.67 (12.4)	<.001
17-32 weeks	47.89 (12.16)	50.37 (12.51)	50.51 (12.33)	<.001
33-52 weeks	49.45 (11.71)	51.57 (12.32)	51.8 (12.18)	<.001
Gender, n/N (%)				
9-16 weeks	235/2594 (9.1)	738/6440 (11.5)	347/2218 (15.6)	<.001
17-32 weeks	166/1907 (8.7)	447/4369 (10.2)	356/2686 (13.2)	<.001
33-52 weeks	67/525 (12.8)	131/1214 (10.8)	126/1144 (11.0)	.47
Baseline BMI, mean (SD)				
9-16 weeks	26.80 (5.58)	26.57 (5.5)	27.19 (5.24)	<.001
17-32 weeks	26.97 (5.78)	26.73 (5.52)	27.43 (5.45)	<.001
33-52 weeks	27.10 (5.8)	27.38 (5.34)	28.11 (5.43)	.001

Overall Engagement by Weight Loss Group

Multivariate analysis of variance (MANOVA) was used to examine if overall engagement significantly differed across the 3 weight loss groups for each time period. Assumptions for MANOVA were checked and met.

Individual Engagement Measures by Weight Loss Group

Multiple linear regressions predicted the engagement associated with each weight outcome of interest. Regressions were conducted with overall mean weekly engagement (per time period) of each engagement measure as separate dependent variables in individual regression analyses and the weight loss category as the independent variable, with controlling for baseline characteristics of age, gender, and BMI. Self-reported engagement data were excluded from the overall mean calculation if missing data were found in any week during each time period, as it would unclear if the missing data indicated a lack of engagement or a lack of reporting. For automatically recorded measures (eg, steps), any missingness in data during a time point was kept in the total mean calculation, as this is an indicator of lack of engagement. Assumptions for linear regression were checked and met.

All statistical tests were 2-sided with significance set at a *P* value <.05 and were conducted through R version 3.6.0 (The R Foundation for Statistical Computing).

Results

Baseline Characteristics

Of the 11,252 participants observed at 9-16 weeks, 23.05% (n=2594) of participants were in the stable weight category, 57.23% (n=6440) were in the moderate weight loss category, and 19.71% (n=2218) were in the high weight loss category. Of the 8962 participants observed at weeks 17-32, 21.28% (n=1907) were in the stable weight category, with 48.75% (n=4369) in the moderate weight loss category and 29.97% (n=2686) in the high weight loss category. Finally, by 33-52 weeks, 18.21% (525/2883) were in the stable weight category, with 42.11% (1214/2883) in the moderate weight loss category and 39.68% (1144/2883) in the high weight loss category.

Of the participants observed at weeks 9-16, 88.45% (9952/11,252) were female, with a mean age of 49.43 (SD 12.54) and a mean BMI of 26.75 (SD 5.47). At weeks 17-32 and 33-52, the majority of users observed were female (17-32 weeks: 10036/11,252, 89.19%; 33-52 weeks: 9987/11,252, 88.76%), with a mean age of 49.93 (SD 12.46) during weeks 17-32 and a mean age of 51.24 (SD 12.20) during weeks 33-52. During weeks 9-16, the mean baseline BMI was 26.8 (SD 5.57) kg/m^2 for the stable group, 26.57 (SD 5.50) kg/m^2 for the moderate loss group, and 27.19 (SD 5.24) kg/m^2 for the high loss group. These baseline characteristics are included in Table 1.

Significant differences existed between the stable, moderate, and high weight loss groups during weeks 9-16 regarding age ($F_{2,11249}$=30.08; *P*<.001), gender (χ^2_2=50.78; *P*<.001), and BMI ($F_{2,11249}$=10.54; *P*<.001); and during weeks 17-32, age ($F_{2,8959}$ = 29.06; *P*<.001), gender (χ^2_2=30.69; *P*<.001), and BMI ($F_{2,8959}$ = 13.73; *P*<.001). These differences remained significant during weeks 32-52 for BMI ($F_{2,2880}$ = 8.38; *P*<.001) and age ($F_{2,2880}$ = 7.37; *P*<.001), but not for gender (χ^2_2 = 1.39; *P*=.50). As a result, we adjusted these demographic measures in the regression analysis. These overall means are described in Table 1.

Overall Engagement by Weight Loss Group

The multivariate analysis of variance test resulted in statistically significant differences in overall engagement among the 3 weight categories for weeks 9-16 ($F_{2,11249}$ = 197.43; *P*<.001), for weeks 17-32 ($F_{2,8959}$=153.50; *P*<.001), and for weeks 33-52 ($F_{2,2880}$ = 44.26; *P*<.001). Therefore, we concluded that for each study time period, engagement as a whole, consisting of mean days with at least 1 meal logged per week, mean articles read per week, mean meals logged per week, mean user messages, mean steps, mean weigh-ins per week, and mean exercises logged per week, significantly differed across the 3 weight categories. Engagement as a whole was highest for the high weight loss group, followed by the moderate weight loss group, with the lowest engagement in the stable group. The results of the MANOVA can be seen in Table S1 in Multimedia Appendix 1.

Individual Engagement Measures by Weight Loss Group

When examined individually, means of the following engagement measures decreased over time across all weight groups: days with at least 1 meal logged, articles read, number of meals logged, steps, coach messages, and weigh in variables. Logged exercise did not follow the same pattern, as total mean logged exercise increased through 52 weeks, regardless of the total amount of weight lost (see Table S2 in Multimedia Appendix 1).

The patterns of means suggest that the moderate and high weight loss groups had greater total engagement within each study time period across all engagement measures compared to the stable weight group. To confirm that differences between the weight loss groups and the stable group were significant and to examine the relationship between weight loss outcomes and each engagement measure, individual multiple regressions were conducted with the stable group as the reference group (see Table 2). Differences in engagement between the high weight loss group and moderate loss group were not examined due to a lack of a clinically meaningful difference between these 2 groups given that significant health improvements occur when weight loss exceeds 5% [12,29]. Therefore, we focused our statistical analysis on comparing the high loss group to the stable group and the moderate loss group to the stable group.

Table 2. Multiple regression results for weight change groups in each study period[a].

Engagement measures[b]	Time points								
	9-16 weeks			17-32 weeks			33-52 weeks		
	Estimate	SE	Adj[c] R^2	Estimate	SE	Adj R^2	Estimate	SE	Adj R^2
Days with at least 1 meal logged									
Moderate loss	0.20	0.01	N/A[d]	0.25	0.01	N/A	0.24	0.02	N/A
High loss	0.25	0.01	0.05	0.34	0.01	0.08	0.37	0.02	0.12
Articles read									
Moderate loss	0.52	0.03	N/A	0.89	0.05	N/A	0.63	0.07	N/A
High loss	0.68	0.04	0.05	1.20	0.05	0.06	1.01	0.07	0.05
Meals logged									
Moderate loss	5.83	0.15	N/A	5.25	0.19	N/A	4.58	0.37	N/A
High loss	8.06	0.19	0.16	8.27	0.21	0.20	8.22	0.37	0.17
Coach messages									
Moderate loss	0.06	0.01	N/A	0.07	0.01	N/A	0.08	0.01	N/A
High loss	0.09	0.01	0.01	0.10	0.01	0.01	0.12	0.01	0.03
Steps									
Moderate loss	4014.04	483.60	N/A	4452.65	579.76	N/A	4535.31	1165.38	N/A
High loss	8806.42	601.71	0.05	9821.59	633.01	0.06	8963.01	1176.49	0.05
Weigh ins									
Moderate loss	0.23	0.01	N/A	0.26	0.01	N/A	0.16	0.017	N/A
High loss	0.29	0.01	0.17	0.34	0.01	0.16	0.27	0.017	0.10
Exercises									
Moderate loss	0.25	0.07	N/A	0.28	0.08	N/A	0.17	0.16	N/A
High loss	0.51	0.09	0.01	0.77	0.09	0.01	0.87	0.16	0.02

[a]The stable group was used as the reference group.

[b]Results are the summary of 7 individual multiple linear regressions. Each engagement measure was a dependent variable of its own regression where weight change groups, gender, age and baseline BMI were independent variables.

[c]Adj: adjusted.

[d]N/A: not applicable.

When users' age, gender, and baseline BMI were controlled for, the moderate and high weight loss groups had significantly more days with at least 1 meal logged (*P* values <.001), articles read (*P* values <.001), meals logged (*P* values <.001), coach messages (*P* values <.001), steps (*P* values <.001), and weigh ins (*P* values <.001) compared to the stable group across all time points. The moderate and high weight loss groups had greater mean exercise per week compared to the stable weight group at 9-16 weeks and 17-32 weeks (*P* values <.001). For weeks 33-52, only the high loss group had significantly greater mean frequency of exercise compared to the stable group (β=.87; SE=0.16; *P*<.001).

Discussion

Principal Results

Using a large data set of more than 11,000 individuals, we sought to extend previous work by evaluating how specific weight loss outcomes (stable weight, 0% plus or minus 1%; moderate weight loss, between 5% to 10%; and high weight loss, 10% or greater) were associated with 7 different engagement measures across 3 time periods: 9-16 weeks, 17-32 weeks, and 33-52 weeks. Overall, our findings indicate significant differences in all 7 engagement measures among those with moderate and high weight loss compared to those with stable weight. These associations held over time, with the exception of exercise logging at 33-52 weeks for the moderate weight loss category.

Comparison With Prior Work

Our findings corroborate past work that found significant associations between the frequency of food logging and weigh-ins and weight loss outcomes [10,17-22]. Departing from some past studies, we found that weight loss outcomes were associated with all engagement measures over all time periods, with one exception. In contrast, some previous work has reported significant associations between engagement and weight loss

for some, but not all, engagement measures [18,21]. This discrepancy may be due to the fact that previous studies explored the amount of engagement necessary to achieve a certain amount of weight loss (ie, how engagement is associated with weight loss), whereas our study is concerned with understanding how achieving a certain level of weight loss is related to levels of engagement (ie, how weight loss is associated with engagement). Our findings raise the possibility that individuals who achieve successful weight outcomes tend to engage comprehensively in the program because they are generally more motivated, whereas only certain engagement measures are necessary to achieve greater weight loss. Future research should explore within-participant patterns of engagement across time as related to individual differences such as motivation or personality characteristics.

According to recent conceptual models of engagement, the content of the intervention (eg, availability of self-monitoring tools), contextual factors, and psychological characteristics like motivation and self-efficacy can influence engagement [16,32]. This study raises additional questions about weight loss outcomes, which could also influence engagement. It is possible that the factors driving individuals to successfully achieve certain levels of weight loss influence their engagement as well. We previously demonstrated that 5% or more weight loss on Noom was associated with psychosocial characteristics such as mental health quality of life and perceived work-life balance [33], and a systematic review reported that weight loss is associated with the expectations individuals have for their weight loss [14]. Future research should separate out the individual components involved in losing certain amounts of weight loss and investigate how each relates to engagement.

Through the large sample size and year-long time period, this study provides data of how engagement decreases over time. In general, the levels of engagement decreased over time for 6 of the 7 engagement measures (with the exception of exercise logging) within each group regardless of whether they achieved no loss, moderate loss, or high loss. This is consistent with past work showing declines in some engagement measures over time but not for physical activity logging [34]. Future research should investigate the possibility that sustained exercise habits are formed on this type of program. We also found that the differences between each weight loss category and stable weight were maintained even as engagement decreased over time. This aligns with a study showing that associations between engagement and weight change were consistent from 16 weeks through 52 weeks [18]. It is necessary for future research to test whether these results mean that long-term engagement accurately reveals true patterns of motivation and action, or whether late engagement is instead a marker of "user's behavior chang[ing] to an extent that digital engagement with the intervention is no longer needed" [35], or both.

Along these lines, we found that being in the moderate weight loss group at 33-52 weeks was not associated with the frequency of exercise logging. There are a few possible reasons for this. Perhaps as they were initially losing weight, individuals in the moderate group perceived that their modest weight loss was primarily due to dietary change rather than exercise, and then they were less likely to consistently log exercise. A previous survey study found that 71% of respondents assumed that exercise is an effective weight loss strategy, and this assumption was associated with feeling discouraged with exercise [36]. Alternatively, perhaps moderate weight loss is more associated with types of logging that require greater effort, such as meal logging and weight logging that users are encouraged to do daily, in contrast to exercise logging which only occurs after individuals have exercised. In our results, the weight loss category accounted for more of the variation in meal logging (adjusted R^2=0.16) and weight logging (adjusted R^2=0.17) than did other engagement measures (adjusted R^2=0.01-0.05). Similarly, in a previous study, meal logging and weight logging predicted changes in weight more than did exercise logging [37]. Future studies should investigate why and how meal and weight logging may differ from exercise logging.

Limitations

The study's strengths include exploring real-world engagement in a large sample on a publicly available mobile DBCI. Some limitations, however, should be noted. First, a convenience sample of individuals who had self-selected to sign up for the Noom Healthy Weight program was used. Thus, findings may not generalize to populations with less motivation to manage their weight. Given the retrospective design, causal relationships between participants' engagement measures and participants' weight loss outcomes cannot be determined. The correlational nature of analyses also prevents firm conclusions about the directionality of results. Finally, the users measured and reported their own weight in the platform, and their loss was calculated from these self-reported measures. Home scales may produce a considerable margin of error compared to ones used by health care professionals. Users were encouraged to use the same scale throughout the program so that their personal loss would be consistent with their individual scale. They were also encouraged to weigh in at the same time every day (ideally in the morning upon first waking up), but it was not possible to enforce these recommendations.

Weight loss was calculated based on self-reported weight measurements by participants, and baseline weight information was contextualized using BMI. Both measures are limited in a few ways. First, they do not adjust for the weight fluctuation that occurs during menstrual cycles or perimenopause and menopause. They also do not account for muscle mass or bone density. Individuals may have increased their muscle mass due to exercise, but this would not be adequately captured by these metrics. Finally, these measures could be subject to artificial inflation because of water retention due to excessive salt intake. Future studies should use a variety of self-reported and objective measurements to understand individuals' weight changes.

Conclusions

This retrospective study explored associations between important weight loss outcomes and engagement in a mobile DBCI over 1 year, which could help to inform tailoring interventions to encourage engagement based on achieved and goal weight loss outcomes or provide data that can be used to better understand variance in weight loss outcomes. This study also provides large-scale data on how individuals engage in a

self-directed mobile DBCI in the short- and long-term. We found that compared to stable weight, having achieved moderate weight loss or high weight loss was associated with higher engagement in the forms of the number of meals logged, articles read, steps logged, coach messages, weigh-ins, and days with at least 1 meal logged from 9 to 52 weeks. This raises the possibility that individuals who lose moderate or high amounts of weight actively engage in all possible aspects of the program, which future research should confirm. The one exception was that being in the moderate weight loss category at 33-52 weeks was more associated with exercise logging than was being in the stable weight category. The consistent associations over time suggests that these differences in engagement behavior are stable throughout both short-term and long-term weight loss. Future research can ascertain to what extent our results are generalizable to other intervention contexts.

Our results raise new questions for future studies which should seek to more fully understand the engagement of individuals who lose significant weight. In this study, participants who achieved moderate weight loss, on average at 17-32 weeks, logged 21 meals, read 3 articles, walked 31,600 steps, weighed in once, exercised 2.5 times within a week, and messaged their coach once every 2 weeks. Users who achieved high weight loss on average logged 24 meals, read 3.4 articles, walked 36,000 steps, weighed in once, exercised 3 times within a week, and messaged their coach once every 2 weeks. These overall means do not take into account variation within users, but future work should go further to, for instance, define profiles of engagement based on weight loss outcomes. This could provide insight into the large variance in weight loss outcomes observed in many interventions.

Future work can explore the following questions: what exactly is responsible for the association between certain weight loss outcomes and engagement in a weight loss program? Is it seeing results, personality factors, demographic factors, or past success or failure with weight loss, or some combination of these? Would telling someone that x level of engagement is related to x level of loss be enough to change their behavior to better engage the mobile health intervention? The next steps may involve exploring other individual or contextual factors in order to understand what guides the engagement seen on this DBCI.

Acknowledgments

TTR was affiliated with Noom while conducting this study, but her affiliation has since changed to Amgen. We would like to thank Heather Behr, Annabell S. Ho, and Ellen Siobhan Mitchell for their input with the final manuscript.

Conflicts of Interest

AC, QY, LD, TTR, YK, and AM are or were employees at Noom Inc and have received salary and stock options for their employment.

Multimedia Appendix 1
Supplementary Tables S1 and S2.
[DOCX File , 18 KB - mhealth_v9i11e30622_app1.docx]

References

1. Swinburn BA, Kraak VI, Allender S, Atkins VJ, Baker PI, Bogard JR, et al. The global syndemic of obesity, undernutrition, and climate change: the Lancet Commission Report. Lancet 2019 Feb 23;393(10173):791-846. [doi: 10.1016/S0140-6736(18)32822-8] [Medline: 30700377]
2. Dalcin AT, Jerome GJ, Fitzpatrick SL, Louis TA, Wang N, Bennett WL, et al. Perceived helpfulness of the individual components of a behavioural weight loss program: results from the Hopkins POWER Trial. Obes Sci Pract 2015 Oct;1(1):23-32 [FREE Full text] [doi: 10.1002/osp4.6] [Medline: 27668085]
3. Kopelman PG, Grace C. New thoughts on managing obesity. Gut 2004 Jul 01;53(7):1044-1053 [FREE Full text] [doi: 10.1136/gut.2003.021816] [Medline: 15194659]
4. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
5. Higgins JP. Smartphone applications for patients' health and fitness. Am J Med 2016 Jan;129(1):11-19. [doi: 10.1016/j.amjmed.2015.05.038] [Medline: 26091764]
6. Elbert NJ, van Os-Medendorp H, van Renselaar RW, Ekeland AG, Hakkaart-van RL, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
7. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]

8. Little P, Stuart B, Hobbs FR, Kelly J, Smith ER, Bradbury KJ, et al. An internet-based intervention with brief nurse support to manage obesity in primary care (POWeR+): a pragmatic, parallel-group, randomised controlled trial. Lancet Diabetes Endocrinol 2016 Oct;4(10):821-828 [FREE Full text] [doi: 10.1016/S2213-8587(16)30099-7] [Medline: 27474214]
9. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
10. Michaelides A, Raby C, Wood M, Farr K, Toro-Ramos T. Weight loss efficacy of a novel mobile Diabetes Prevention Program delivery platform with human coaching. BMJ Open Diabetes Res Care 2016;4(1):e000264 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000264] [Medline: 27651911]
11. Ryan DH, Yockey SR. Weight loss and improvement in comorbidity: differences at 5%, 10%, 15%, and over. Curr Obes Rep 2017 Apr 28;6(2):187-194. [doi: 10.1007/s13679-017-0262-y]
12. Williamson DA, Bray GA, Ryan DH. Is 5% weight loss a satisfactory criterion to define clinically significant weight loss? Obesity (Silver Spring) 2015 Dec;23(12):2319-2320 [FREE Full text] [doi: 10.1002/oby.21358] [Medline: 26523739]
13. Teixeira PJ, Going SB, Sardinha LB, Lohman TG. A review of psychosocial pre-treatment predictors of weight control. Obes Rev 2005 Feb;6(1):43-65. [doi: 10.1111/j.1467-789X.2005.00166.x] [Medline: 15655038]
14. Carraça EV, Santos I, Mata J, Teixeira PJ. Psychosocial pretreatment predictors of weight control: a systematic review update. Obes Facts 2018;11(1):67-82 [FREE Full text] [doi: 10.1159/000485838] [Medline: 29439252]
15. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
16. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
17. Smith N, Liu S. A systematic review of the dose-response relationship between usage and outcomes of online physical activity weight-loss interventions. Internet Interv 2020 Dec;22:100344 [FREE Full text] [doi: 10.1016/j.invent.2020.100344] [Medline: 32995302]
18. Sepah SC, Jiang L, Ellis RJ, McDermott K, Peters AL. Engagement and outcomes in a digital Diabetes Prevention Program: 3-year update. BMJ Open Diabetes Res Care 2017;5(1):e000422 [FREE Full text] [doi: 10.1136/bmjdrc-2017-000422] [Medline: 28948027]
19. Painter SL, Ahmed R, Hill JO, Kushner RF, Lindquist R, Brunning S, et al. What matters in weight loss? An in-depth analysis of self-monitoring. J Med Internet Res 2017 May 12;19(5):e160 [FREE Full text] [doi: 10.2196/jmir.7457] [Medline: 28500022]
20. Michaelides A, Major J, Pienkosz E, Wood M, Kim Y, Toro-Ramos T. Usefulness of a novel mobile diabetes prevention program delivery platform with human coaching: 65-week observational follow-up. JMIR Mhealth Uhealth 2018 May 03;6(5):e93 [FREE Full text] [doi: 10.2196/mhealth.9161] [Medline: 29724709]
21. Lin P, Grambow S, Intille S, Gallis JA, Lazenka T, Bosworth H, et al. The association between engagement and weight loss through personal coaching and cell phone interventions in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 18;6(10):e10471 [FREE Full text] [doi: 10.2196/10471] [Medline: 30341051]
22. Spaulding EM, Marvel FA, Piasecki RJ, Martin SS, Allen JK. User engagement with smartphone apps and cardiovascular disease risk factor outcomes: systematic review. JMIR Cardio 2021 Feb 03;5(1):e18834 [FREE Full text] [doi: 10.2196/18834] [Medline: 33533730]
23. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
24. Castelnuovo G, Pietrabissa G, Manzoni GM, Cattivelli R, Rossi A, Novelli M, et al. Cognitive behavioral therapy to aid weight loss in obese patients: current perspectives. PRBM 2017 Jun;Volume 10:165-173. [doi: 10.2147/prbm.s113278]
25. Lawlor E, Islam N, Bates S, Griffin SJ, Hill AJ, Hughes CA, et al. Third-wave cognitive behaviour therapies for weight management: A systematic review and network meta-analysis. Obes Rev 2020 Jul;21(7):e13013 [FREE Full text] [doi: 10.1111/obr.13013] [Medline: 32181957]
26. Armstrong MJ, Mottershead TA, Ronksley PE, Sigal RJ, Campbell TS, Hemmelgarn BR. Motivational interviewing to improve weight loss in overweight and/or obese patients: a systematic review and meta-analysis of randomized controlled trials. Obes Rev 2011 Sep;12(9):709-723. [doi: 10.1111/j.1467-789X.2011.00892.x] [Medline: 21692966]
27. 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]
28. Wing RR, Hill JO. Successful weight loss maintenance. Annu Rev Nutr 2001;21:323-341. [doi: 10.1146/annurev.nutr.21.1.323] [Medline: 11375440]
29. American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Expert Panel, 2013. Expert panel report: guidelines (2013) for the management of overweight and obesity in adults. Obesity (Silver Spring) 2014 Jul;22 Suppl 2:S41-410 [FREE Full text] [doi: 10.1002/oby.20660] [Medline: 24227637]

30. Postrach E, Aspalter R, Elbelt U, Koller M, Longin R, Schulzke J, et al. Determinants of successful weight loss after using a commercial web-based weight reduction program for six months: cohort study. J Med Internet Res 2013 Oct 14;15(10):e219 [FREE Full text] [doi: 10.2196/jmir.2648] [Medline: 24126250]
31. DeLuca L, Toro-Ramos T, Michaelides A, Seng E, Swencionis C. Relationship between age and weight loss in Noom: quasi-experimental study. JMIR Diabetes 2020 Jun 04;5(2):e18363 [FREE Full text] [doi: 10.2196/18363] [Medline: 32497017]
32. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
33. Mitchell ES, Yang Q, Behr H, Ho A, DeLuca L, May CN, et al. Psychosocial characteristics by weight loss and engagement in a digital intervention supporting self-management of weight. Int J Environ Res Public Health 2021 Feb 10;18(4):1733 [FREE Full text] [doi: 10.3390/ijerph18041712] [Medline: 33578975]
34. Butryn ML, Godfrey KM, Martinelli MK, Roberts SR, Forman EM, Zhang F. Digital self-monitoring: Does adherence or association with outcomes differ by self-monitoring target? Obes Sci Pract 2020 Apr;6(2):126-133 [FREE Full text] [doi: 10.1002/osp4.391] [Medline: 32313670]
35. Ainsworth B, Steele M, Stuart B, Joseph J, Miller S, Morrison L, et al. Using an analysis of behavior change to inform effective digital intervention design: how did the PRIMIT website change hand hygiene behavior across 8993 users? Ann Behav Med 2017 Jun;51(3):423-431 [FREE Full text] [doi: 10.1007/s12160-016-9866-9] [Medline: 27909944]
36. Thomas DM, Kyle TK, Stanford FC. The gap between expectations and reality of exercise-induced weight loss is associated with discouragement. Prev Med 2015 Dec;81:357-360 [FREE Full text] [doi: 10.1016/j.ypmed.2015.10.001] [Medline: 26500086]
37. Ross KM, Qiu P, You L, Wing RR. Week-to-week predictors of weight loss and regain. Health Psychol 2019 Dec;38(12):1150-1158 [FREE Full text] [doi: 10.1037/hea0000798] [Medline: 31566400]

Abbreviations

DBCI: digital behavior change intervention
MANOVA: multivariate analysis of variance

Edited by G Eysenbach; submitted 25.05.21; peer-reviewed by D Szinay; comments to author 18.06.21; revised version received 02.07.21; accepted 19.09.21; published 08.11.21.

Please cite as:
Carey A, Yang Q, DeLuca L, Toro-Ramos T, Kim Y, Michaelides A
The Relationship Between Weight Loss Outcomes and Engagement in a Mobile Behavioral Change Intervention: Retrospective Analysis
JMIR Mhealth Uhealth 2021;9(11):e30622
URL: https://mhealth.jmir.org/2021/11/e30622
doi:10.2196/30622
PMID:34747706

Original Paper

Improvements in Diet and Physical Activity–Related Psychosocial Factors Among African Americans Using a Mobile Health Lifestyle Intervention to Promote Cardiovascular Health: The FAITH! (Fostering African American Improvement in Total Health) App Pilot Study

Jissy Cyriac[1*], MD, MS; Sarah Jenkins[2*], MS; Christi A Patten[3*], PhD; Sharonne N Hayes[4*], MD; Clarence Jones[5*], MEd; Lisa A Cooper[6*], MD, MPH; LaPrincess C Brewer[4,7*], MD, MPH

[1]Department of Internal Medicine, Mayo Clinic School of Graduate Medical Education, Rochester, MN, United States

[2]Division of Clinical Trials and Biostatistics, Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, United States

[3]Department of Psychiatry and Psychology, Mayo Clinic College of Medicine, Rochester, MN, United States

[4]Department of Cardiovascular Medicine, Mayo Clinic College of Medicine, Rochester, MN, United States

[5]Hue-Man Partnership, Minneapolis, MN, United States

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

[7]Center for Health Equity and Community Engagement Research, Mayo Clinic, Rochester, MN, United States

[*]all authors contributed equally

Corresponding Author:
LaPrincess C Brewer, MD, MPH
Department of Cardiovascular Medicine
Mayo Clinic College of Medicine
200 First St. SW
Rochester, MN, 55905
United States
Phone: 1 507 266 1376
Email: brewer.laprincess@mayo.edu

Abstract

Background: African Americans continue to have suboptimal cardiovascular health (CVH) related to diet and physical activity (PA) behaviors compared with White people. Mobile health (mHealth) interventions are innovative platforms to improve diet and PA and have the potential to mitigate these disparities. However, these are understudied among African Americans.

Objective: This study aims to examine whether an mHealth lifestyle intervention is associated with improved diet and PA-related psychosocial factors in African Americans and whether these changes correlate with diet and PA behavioral change.

Methods: This study is a retrospective analysis evaluating changes in diet and PA-related self-regulation, social support, perceived barriers, and CVH behaviors (daily fruit and vegetable intake and moderate-intensity PA [MPA] per week) in 45 African American adults (mean age 48.7 years, SD 12.9 years; 33/45, 73% women) enrolled in the FAITH! (Fostering African American Improvement in Total Health) app pilot study. The intervention is a 10-week, behavioral theory–informed, community-based mHealth lifestyle intervention delivered through a mobile app platform. Participants engaged with 3 core FAITH! app features: multimedia education modules focused on CVH with self-assessments of CVH knowledge, self-monitoring of daily fruit and vegetable intake and PA, and a sharing board for social networking. Changes in self-reported diet and PA-related self-regulation, social support, perceived barriers, and CVH behaviors were assessed by electronic surveys collected at baseline and 28 weeks postintervention. Changes in diet and PA-related psychosocial factors from pre- to postintervention were assessed using paired 2-tailed *t* tests. The association of changes in diet and PA-related psychosocial variables with daily fruit and vegetable intake and MPA per week was assessed using Spearman correlation. Associations between baseline and 28-week postintervention changes in diet and PA-related psychosocial measures and CVH behaviors with covariates were assessed by multivariable linear regression.

Results: Participants reported improvements in 2 subscales of diet self-regulation (decrease fat and calorie intake, P=.01 and nutrition tracking, P<.001), one subscale of social support for healthy diet (friend discouragement, P=.001), perceived barriers

to healthy diet (P<.001), and daily fruit and vegetable intake (P<.001). Improvements in diet self-regulation (increase fruit, vegetable, and grain intake, and nutrition tracking) and social support for healthy diet (friend encouragement) had moderate positive correlations with daily fruit and vegetable intake (r=0.46, r=0.34, and r=0.43, respectively). A moderate negative correlation was observed between perceived barriers to healthy diet and daily fruit and vegetable intake (r=−0.25). Participants reported increases in PA self-regulation (P<.001). Increase in social support subscales for PA (family and friend participation) had a moderate positive correlation with MPA per week (r=0.51 and r=0.61, respectively).

Conclusions: Our findings highlight key diet and PA-related psychosocial factors to target in future mHealth lifestyle interventions aimed at promoting CVH in African Americans.

(JMIR Mhealth Uhealth 2021;9(11):e28024) doi:10.2196/28024

KEYWORDS

African Americans; cardiovascular health disparities; mHealth lifestyle intervention; diet; physical activity; mobile phone

Introduction

Background

African Americans have the highest rates of coronary heart disease and stroke-related deaths compared with White people and other racial and ethnic minority groups in the United States [1]. Cardiovascular risk factors such as suboptimal diet and physical activity (PA) have been identified as significant contributors to the disproportionate cardiovascular disease (CVD) burden among African Americans [1]. African Americans have an extremely low prevalence of individuals achieving ideal levels on 5 or more cardiovascular health (CVH) metrics, as outlined by the American Heart Association Life's Simple 7 (LS7) [1-3].

Suboptimal CVH in African Americans is rooted in structural racism, which has systematically limited their access to quality health care, employment opportunities, education, and safe neighborhood environments. The resulting scarcity of resources has led to elevated stress levels, food insecurity, and poor access to recreational spaces, all of which affect CVH in African Americans [4-7]. These structural inequities can further manifest as negative psychosocial factors that can influence health behaviors and CVH outcomes in African Americans [8-11]. In African Americans, increased self-reported stress and depressive symptoms have been associated with greater calorie consumption and lower levels of PA, respectively [10]. There is evidence to suggest that psychosocial factors affecting healthy diet and PA, such as cost and lack of time, are negatively associated with diet quality and regular PA, respectively [12-14]. Data on diet-related psychosocial factors indicate that self-regulatory behaviors, such as mindful food preparation, have been associated with healthier food acquisition and decreased purchase of preprepared foods among African American adults [15,16]. In addition, culturally tailored interventions integrating social support and addressing barriers have been identified as facilitators of PA in African American women [14,17]. African American women have also expressed a preference for interventions that promote self-regulatory behaviors (self-monitoring of PA) and incorporate strategies to overcome barriers to improve their PA patterns [17]. These findings provide compelling evidence that targeting diet and PA-related self-regulation, social support, and perceived barriers in mobile health (mHealth) lifestyle interventions for African Americans may facilitate improvements in diet and PA behaviors.

Mobile app–based health interventions have the potential to improve diet and PA patterns among African Americans. Research from the Pew Research Center has shown that African Americans are less likely than White people to have traditional broadband access at home but are equally likely to own cell phones and smartphones. They are also more likely to use smartphones to access the internet, web-based social networking sites, and health information [18-20]. Thus, the use of smartphones for health promotion has the potential to mitigate health disparities, and makes progress toward achieving health equity through technology-based interventions [21].

In addition, there is emerging evidence that interventions using mHealth apps may promote diet and PA behavior change [22-24] and may be effective interventional tools for underserved African American communities [21,25]. However, mHealth interventions have been significantly understudied in African American populations. James et al [26] found that <10% of identified mHealth studies within their systematic review included African American participants, and only 14% of those mHealth studies entirely comprised all African American participants. This is despite African Americans having high smartphone ownership, mobile technology use (including mobile apps), and eHealth literacy (EHL) [20,27-29]. EHL is defined as an individual's ability to search for and understand health information on the internet using computers and mobile devices [30]. Given high EHL and smartphone ownership in African Americans, mHealth lifestyle interventions may be particularly effective in promoting healthy diet and PA behaviors.

Current mHealth interventions to promote healthy diet and PA in African Americans suggest that they can be effective as stand-alone interventions [31-35] or adjuncts to in-person interventions [36-39]. Allicock et al [31] demonstrated that a stand-alone, mobile app–based intervention to encourage healthy diet and PA behaviors among African American survivors of breast cancer led to a significant reduction in sedentary time and fast food intake. Another culturally tailored, internet-based intervention for PA promotion as an adjunct to in-person PA sessions resulted in a significant reduction in sedentary behaviors among African American women [38]. Furthermore, mHealth interventions have demonstrated success in increasing diet and PA self-monitoring [39], promoting healthy diet and

regular PA [33,35] and facilitating weight loss in African Americans [32,34,36,37]. Although mHealth lifestyle interventions have demonstrated improvements in diet and PA outcomes, few have discussed the interplay between psychosocial factors such as underlying diet and PA-related self-regulation, social support, perceived barriers, and their associations with health behaviors [33,34,38,40].

Goal of This Study

We seek to provide insight into this current gap in the literature by exploring changes in diet and PA-related self-regulation, social support, perceived barriers, and their associations with diet and PA patterns among African Americans using our culturally tailored, community-based mHealth lifestyle intervention. The FAITH! (Fostering African American Improvement in Total Health) app pilot study tested a novel mHealth intervention aiming to improve CVH in African Americans through local African American church congregations [41]. Through the use of the FAITH! app, participants had statistically significant improvements in diet, PA, blood pressure, and overall composite LS7 score [35].

The primary aim of this study is to examine changes in diet and PA-related self-regulation, social support, and perceived barriers among participants in the FAITH! app pilot study. We also aim to assess whether any of the observed changes in these measures correlate with CVH behaviors (daily fruit and vegetable intake and moderate-intensity PA [MPA] per week). We hypothesize that participants would demonstrate improvements in diet and PA-related self-regulation, social support, perceived barriers, and CVH behaviors using the FAITH! app intervention.

Methods

Study Design

This study is a retrospective analysis examining changes in diet and PA-related psychosocial factors in the FAITH! app pilot study. Details of the intervention design, recruitment, implementation, and outcomes have been previously published [35,41,42]. Briefly, the pilot study was conducted using a single group, pretest-posttest intervention framework to assess CVH knowledge, behaviors, and biological factors among African Americans following the use of a mobile app–based intervention. The cohort was recruited from 5 African American churches within the Rochester and Minneapolis-St. Paul, Minnesota, metropolitan areas. The pilot study was registered with the Clinical Trials Registry (ClinicalTrials.gov NCT03084822) and approved by the Mayo Clinic institutional review board. Participants provided written informed consent before enrollment in the study.

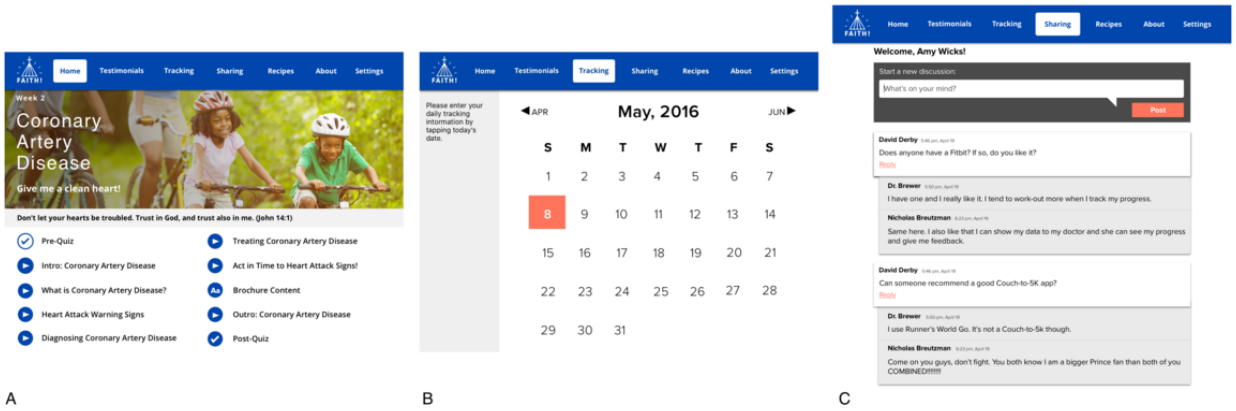
Data Collection

Baseline data were collected in July 2016, and follow-up data at 28 weeks were collected postintervention (April 2017). Electronic surveys were emailed directly to participants to assess their sociodemographic data, EHL, diet and PA-related self-regulation, social support, and perceived barriers, along with CVH behaviors (daily fruit and vegetable intake and MPA per week) at baseline and 28 weeks post intervention. Postintervention evaluation occurred at 28 weeks in an effort to align with the AHA LS7, the primary outcome of the parent study [35]. The LS7 comprises both biological and behavioral factors, and its evaluation was completed at the 28-week postintervention time point to assess for sustained changes in these variables. Thus, the present analysis examines changes from baseline to 28 weeks postintervention to concurrently assess for sustained changes in diet and PA-related self-regulation, social support, perceived barriers, and CVH behaviors. Outcome evaluation was designed based on psychosocial measures previously described in the literature, noted trends in a previous in-person iteration of the FAITH! program, and the theoretical foundation of the FAITH! app intervention [42-44]. Participants received up to US $150 in gift cards for their completion of survey assessments.

Theoretical Framework: Intervention and Psychosocial Measures

The 10-week intervention was delivered through a mobile app (FAITH! app; Figure 1) and consisted of 3 core features: (1) a multimedia education module series on CVH with pre- and postmodule self-assessments of CVH knowledge, (2) self-monitoring of daily fruit and vegetable intake and PA, and (3) a sharing board for networking with other participants. At study enrollment, participants were provided iPads with the FAITH! app installed and Fitbits (Charge 2, 2016 version, Fitbit Inc) for PA and step tracking. Within the FAITH! app, participants were expected to complete one educational module per week and complete the associated pre- and postmodule self-assessments, with completion of all educational modules by the end of the 10-week intervention. Participants entered daily entries for fruit and vegetable intake and PA during the course of 10 weeks. In addition, participants were encouraged to post on the sharing board for social networking with other study participants; no limits were imposed on the number of posts an individual participant could submit. Participants were expected to engage in all 3 core app features described above for the 10-week intervention delivery phase, and their engagement patterns with these features were closely monitored via Google Analytics [45]. Participants continued to have access to the app in the postintervention phase, but this was not monitored. Further details of the FAITH! app design and key features have been previously published [42].

Figure 1. FAITH! (Fostering African American Improvement in Total Health) app core features: (A) cardiovascular health education modules and self-assessments; (B) diet and physical activity self-monitoring; (C) sharing board.



The FAITH! app was developed using a community-based participatory approach grounded in behavioral theories (social cognitive theory, health belief model, and community mobilization model) [42]. Social cognitive theory posits that individual health behavior is dependent on their reciprocal interaction with observational learning and reinforcements. We applied this theory within the FAITH! app via the sharing board, where participants could see others from their own communities modeling positive health behaviors (eg, diet and PA). This in turn could motivate their adoption of these behaviors [42]. In addition, the sharing board allowed individuals to receive positive reinforcement from others in the study to maintain healthy diet and PA behaviors. Correspondingly, social support was assessed to gauge the degree of positive reinforcement and observational role modeling experienced by participants through the use of the intervention. Central to the health belief model is that interventions are more effective in changing health behaviors if they influence an individual's perception of susceptibility to illness or disease, severity of illness, potential positive benefits of healthy action, barriers to such action, and exposure to factors that prompt action (cues to action) [46]. Furthermore, a focus on perceived barriers and perceived benefits has been demonstrated to be the strongest predictor of behavior change [47]. On the basis of the principles of the health belief model, the CVH education modules were designed to convey the high CVD risk of African Americans and the severity of the consequences of CVD (heart attacks, heart failure, and strokes). In addition, the modules emphasized greater benefits than barriers to maintaining a healthy diet and regular PA and thus aimed to increase the likelihood that participants would take action to engage in these health behaviors. Participants' prompts to action and self-directed behavior were further reinforced by encouraged use of the diet and PA self-monitoring app features. Accordingly, changes in diet and PA self-regulation and perceived barriers to healthy diet and PA were assessed to determine how effectively the intervention addressed these components of the health belief model [42]. Community mobilization was leveraged with the involvement of community partners from participating African American churches in all stages of app design using a community-based participatory research approach to ensure that the app incorporated African American religious and spiritual beliefs, the social connectedness aspects of the African American church, and cultural traditions of the African American community [42,48]. African American churches have been the cornerstone of academic-community partnerships seeking to enact social justice through research to improve health and access to quality care among African Americans [49].

Measurement of Psychosocial Factors

Diet-Related Measures

Diet self-regulation was assessed using 14 items from the Health Beliefs Survey developed by Anderson et al [50], where each item was measured using a 5-point Likert scale (1=never to 5=always). Three subscales of diet self-regulation were assessed, including strategies to (1) increase fruit, vegetable, and grain intake (3 survey items; Cronbach α=.67), (2) decrease fat and calorie intake (6 survey items; Cronbach α=.78), and (3) plan and track nutrition (5 survey items; Cronbach α=.80). Responses for items in each category were averaged to calculate each score (range 1-5). Higher scores indicated greater dietary self-regulation. Social support for healthy diet was assessed using a survey instrument developed by Sallis et al [51], which consists of 20 items rated on a 5-point Likert scale (1=never to 5=very often) that assessed subscales of encouragement and discouragement from family and friends for healthy diet (4 subscales: family encouragement [5 items; Cronbach α=.81], friend encouragement [5 items; Cronbach α=.81], family discouragement [5 items; Cronbach α=.56], and friend discouragement [5 items; Cronbach α=.74]). Higher scores implied greater encouragement or discouragement, respectively. Perceived barriers to healthy diet were assessed using 15 items adapted from the Lose It Forever study questionnaire by Welsh et al [52]. The survey asked participants to evaluate the availability of healthy foods, the ability to control cravings for unhealthy foods, prepare healthy meals, and level of difficulty in the daily environment to choose healthy foods. Participants responded to items on a 5-point Likert scale (1=not at all true to me to 5=very true for me). Responses for each item were averaged to calculate a final score (range 1-5), with higher scores implying greater barriers (Cronbach α=.88).

PA-Related Measures

PA self-regulation was measured using the Self-Regulation Scale from the Health Beliefs Survey, which includes 10 items, using a 5-point Likert scale (1=never to 5=always) [33,50,53]. The scale assesses strategies to increase PA and the efforts used to track step counts over the past month. Scores across all items were averaged to calculate a final score (range 1-5), with higher scores indicating higher PA self-regulation (Cronbach α=.84). Social support for PA was measured using a survey instrument developed by Sallis et al [51] and comprised 23 items rated on a 5-point Likert scale (1=never to 5=very often) that assessed subscales of: family participation [10 items; Cronbach α=.92], family rewards or punishment [3 items; Cronbach α=.48], and friend participation [10 items; Cronbach α=.93]. Higher scores implied greater participation or reward or punishment, respectively. An adaptation of the Exercise Barriers Scale was used to assess barriers to exercise [54,55]. Participants rated a list of 20 items on a 4-point Likert scale (1=strongly agree to 4=strongly disagree) on how much they presented a barrier to exercise. Items included time for PA, convenience to complete PA, competing responsibilities, and concerns about hairstyle. The response to each item was reversed so that more agreement indicated more barriers. The items were averaged to calculate the final score (range 1-4), with higher scores implying greater barriers (Cronbach α=.91).

CVH Behaviors

Self-reported daily fruit and vegetable intake was assessed using 2 items adapted from previously developed instruments assessing fruit and vegetable intake [56-58]. Respondents reported daily servings of fruit and vegetables consumed. This instrument has been previously validated in a similar population of African Americans [59]. PA was assessed as self-reported total minutes of MPA per week using the short form of the International Physical Activity Questionnaire) survey. The International Physical Activity Questionnaire has also been validated among African Americans [60,61]. MPA per week was used as it is a part of the standardized LS7 components and associated metrics [2,35].

Covariates

Key covariates for assessment of associations with psychosocial factors as well as diet and PA behaviors included sociodemographic data (age, sex, income, education level, marital status, and employment status), EHL, and level of app engagement. Participant EHL was measured using the eHealth Literacy Scale, which consists of 8 items on a 5-point Likert scale (1=strongly disagree to 5=strongly agree; Cronbach α=.89) assessing self-reported skills in using eHealth information [30]. High versus low EHL was dichotomized at a score of 26, as previously used by Richtering et al [62]. The level of app engagement was categorized as high versus low. Participants with high app engagement met at least two of the following three criteria: (1) >70% completion of self-assessments within CVH education modules, (2) at least 7 entries into the diet and PA self-monitoring feature, and (3) at least one post on the sharing board. Participants who did not meet these criteria were categorized as having low app engagement. These parameters were determined by real-time monitoring of the participants' use patterns of key features within the FAITH! app throughout the intervention phase (via Google Analytics). As a pilot study of the newly created FAITH! app among a fairly understudied population, a priori engagement patterns were not available.

Statistical Analyses

Participant data were summarized with frequencies and percentages, means and SDs, or medians and IQRs, as appropriate. Changes in the psychosocial measures from baseline to 28 weeks postintervention (Δ) were assessed with paired 2-tailed *t* tests, with the exception of MPA per week, which was assessed with the signed-rank test. Effect sizes (Cohen *d*) were calculated as the average difference (28 weeks postintervention minus the baseline) divided by the SD of the difference, with the exception of MPA per week, for which the effect size was calculated as the difference in medians divided by the IQR. Effect sizes of <0.5 were categorized as *small*, 0.5 to <0.8 was categorized as *medium*, and ≥0.8 was categorized as *high*. Associations between baseline and 28-week postintervention changes in psychosocial measures and CVH behaviors with participant covariates (sex, employment status, EHL, and app engagement) were assessed using multivariable linear regression (or quantile regression at the median for MPA per week), including all covariates together. Given the small sample size, a focused set of covariates was included in the multivariable analysis. Associations between changes in diet and PA-related psychosocial measures with changes in daily fruit and vegetable intake and MPA per week were quantified using Spearman correlations (*r*). Correlations >0.2 were considered moderately associated, and correlations of >0.7 were considered highly associated. All analyses were performed using the SAS (version 9.4; SAS Institute, Inc). All statistical tests were 2-tailed, and statistical significance was defined as $P \leq .01$, and 99% CIs were reported along with the mean differences.

Results

Participant Demographics

The analytic sample included 45 African American participants (mean age 48.7 years, SD 12.9 years; 33/45, 73% women) who completed the 10-week mHealth lifestyle intervention and surveys at baseline and 28-week postintervention (Table 1).

Changes in Diet and PA-Related Psychosocial Measures and CVH Behaviors

Overview

Table 2 summarizes changes in the sample for all measured diet and PA-related psychosocial measures along with CVH behaviors from baseline to 28 weeks postintervention.

Table 1. Baseline participant characteristics (N=45).

Characteristics	Values
Age (years), mean (SD; range)	48.7 (12.9; 26.0-72.0)
Sex , n (%)	
Women	33 (73)
Men	12 (27)
Annual household income (n=40; US $) , n (%)	
<35,000	14 (35)
≥35,000	26 (65)
Employment status, n (%)	
Employed at least part time	34 (76)
Unemployed	11 (24)
Marital status, n (%)	
Unmarried	22 (49)
Married	23 (51)
Education level, n (%)	
No degree	15 (33)
Technical, associate's, college, or advanced degree	30 (67)
eHealth literacy score (n=40)	
Value, mean (SD; range)	30.5 (4.5; 21.0-40.0)
Low (<26), n (%)	6 (15)
High (≥26), n (%)	34 (85)
App engagement, n (%)	
Low	20 (44)
High	25 (56)

Table 2. Changes in diet and physical activity (PA)-related psychosocial measures and cardiovascular health behaviors, baseline to 28 weeks postintervention (N=45).

Characteristics	Baseline, mean (SD)[a]	28 weeks postintervention, mean (SD)[a]	Difference in score[a,b], mean (SD; 99% CI)	Effect size[c] (Cohen *d*)	*P* value
Diet-related psychosocial measures					
Diet self-regulation					
Increase fruit, vegetable, and grain intake	3.3 (0.7)	3.5 (0.7)	0.2 (0.8; −0.1 to 0.6)	0.32	.04
Decrease fat and calorie intake	3.0 (0.7)	3.3 (0.7)	0.3 (0.8; 0.0 to 0.6)	0.41	.01
Nutrition tracking	2.1 (0.8)	2.7 (0.8)	0.6 (0.9; 0.2 to 1.0)	0.61	<.001
Social support for healthy diet					
Family encouragement	13.4 (4.8)	12.9 (5.2)	−0.5 (3.6; −2.0 to 1.0)	−0.14	.36
Family discouragement	12.0 (3.8)	11.5 (4.7)	−0.5 (3.9; −2.1 to 1.1)	−0.14	.38
Friend encouragement	11.8 (4.6)	10.9 (4.7)	−1.0 (5.0; −3.2 to 1.3)	−0.19	.25
Friend discouragement	11.2 (4.6)	9.0 (3.2)	−2.1 (3.6; −3.7 to −0.5)	−0.60	.001
Perceived barriers to healthy diet	2.5 (0.7)	2.2 (0.6)	−0.4 (0.5; −0.6 to −0.2)	−0.76	<.001
Diet behavior					
Daily fruit and vegetable intake (servings per day)	3.4 (1.4)	4.5 (1.8)	1.2 (1.9; 0.4 to 1.9)	0.62	<.001
PA-related psychosocial measures					
PA self-regulation	2.3 (0.6)	2.7 (0.7)	0.4 (0.7; 0.2 to 0.7)	0.65	<.001
Social support for PA					
Family participation	19.7 (8.9)	20.1 (8.8)	0.3 (8.6; −3.2 to 3.8)	0.04	.80
Family rewards or punishment	3.7 (1.5)	4.0 (1.6)	0.4 (1.5; −0.2 to 1.0)	0.25	.10
Friend participation	18.5 (10.2)	21.9 (9.4)	3.4 (8.4; −0.8 to 7.6)	0.40	.03
Perceived barriers to PA	1.7 (0.4)	1.7 (0.5)	0.0 (0.5; −0.2 to 0.2)	−0.01	.93
PA behavior					
MPA[d] per week (minutes per week), median (IQR; 99% CI)[e]	35.0 (0.0 to 110.0)	75.0 (25.0 to 187.5)	30.0 (−12.5 to 122.5; −52.5 to 92.5)	0.22	.04

[a]Mean (SD) shown, unless otherwise specified.

[b]Difference in score calculated before rounding as change in score from baseline to postintervention.

[c]Effect size calculated before rounding as the mean difference divided by the SD of the difference, unless otherwise specified.

[d]MPA: moderate-intensity physical activity.

[e]99% CI for median difference estimated with quantile regression; effect size calculated as median difference divided by IQR; *P* value from signed-rank test.

Diet-Related Psychosocial Measures and Daily Fruit and Vegetable Intake

Participants reported statistically significant improvements in 2 subscales of diet self-regulation (decrease fat and calorie intake: Δ +0.3; *P*=.01; Cohen *d*=0.41 and nutrition tracking: Δ +0.6; *P*<.001; Cohen *d*=0.61), one subscale of social support (friend discouragement: Δ −2.1; *P*=.001; Cohen *d*=−0.60), and perceived barriers to healthy diet (Δ −0.4; *P*<.001; Cohen *d*=−0.76) from baseline to 28 weeks postintervention. The sample also showed statistically significant improvements in reported daily fruit and vegetable intake (Δ +1.2; *P*<.001; Cohen *d*=0.62).

PA-Related Psychosocial Measures and MPA Per Week

Participants reported statistically significant improvements in PA self-regulation (Δ +0.4; *P*<.001; Cohen *d*=0.65) from baseline to 28 weeks postintervention. Participants reported a slight improvement in MPA per week (Δ +30 minutes; *P*=.04; Cohen *d*=0.22); however, this did not reach statistical significance.

Correlation of Psychosocial Measures With CVH Behaviors

Tables 3 and 4 summarize the correlation of the measured changes in diet and PA-related self-regulation, social support, and perceived barriers to CVH behaviors. Improvements in subscales of diet self-regulation (increase fruit, vegetable, and

grain intake and nutrition tracking) had a moderate positive correlation (r=0.46 and r=0.34, respectively) with improvement in daily fruit and vegetable intake. Among social support for healthy diet subscales, increased friend encouragement for a healthy diet had a moderate positive correlation (r=0.43) with an increase in daily fruit and vegetable intake. A moderate negative correlation (r=−0.25) was seen between perceived barriers to healthy diet and daily fruit and vegetable intake (ie, greater barriers to healthy diet corresponded with less fruit and vegetable intake). For PA, 2 subscales of social support for PA (family and friend participation) had a moderate positive correlation (r=0.51 and r=0.61, respectively) with MPA per week.

Table 3. Correlations of changes in diet-related psychosocial measures to changes in cardiovascular health behaviors (N=45).

Diet-related psychosocial measures	Correlation to change in daily fruit and vegetable intake[a] (servings per day)
Diet self-regulation	
Increase fruit, vegetable, and grain intake	0.46
Decrease fat and calorie intake	0.03
Nutrition tracking	0.34
Social support for healthy diet	
Family encouragement	0.004
Family discouragement	0.19
Friend encouragement	0.43
Friend discouragement	0.05
Perceived barriers to healthy diet	−0.25

[a]Correlations >0.2 were considered moderately associated.

Table 4. Correlations of changes in physical activity (PA)-related psychosocial measures to changes in cardiovascular health behaviors (N=45).

PA-related psychosocial measures	Correlation to change in MPA[a] per week[b] (minutes per week)
PA self-regulation	0.17
Social support for PA	
Family rewards or punishment	0.09
Family participation	0.51
Friend participation	0.61
Perceived barriers to PA	−0.08

[a]MPA: moderate-intensity physical activity.

[b]Correlations >0.2 were considered moderately associated.

Comparisons With Covariates

In multivariable regression analyses, there were no statistically significant differences in pre- and postintervention score changes among any of the sociodemographics, EHL, or app engagement groups for diet and PA-related psychosocial measures and CVH behaviors.

Discussion

Principal Findings

Our findings demonstrate that a culturally tailored, community-based mHealth lifestyle intervention can improve key diet and PA-related psychosocial factors and CVH behaviors in African Americans. In addition, several of the improvements in these diet and PA-related psychosocial measures were associated with improvements in diet and PA, which to our knowledge has not been previously described in the setting of a mobile app–based lifestyle intervention. The changes noted in our study were of small to medium effect sizes. This is consistent with the findings of other studies with similar sample sizes evaluating changes in diet and PA-related psychosocial measures [40]. A longer intervention timeframe may be needed to see greater changes in the diet and PA-related psychosocial measures evaluated in this study.

Our findings illustrating the benefits of using an mHealth lifestyle intervention on diet and PA-related self-regulation in African Americans are consistent with those of other investigators. Participants in our study reported improvements in the subscales of diet self-regulation as well as PA self-regulation. Furthermore, improvements in diet self-regulation subscales were positively associated with increased daily fruit and vegetable intake. Ferrante et al [34] evaluated changes in diet and PA-related psychosocial variables and weight loss outcomes in a Fitbit plus mobile app lifestyle intervention versus a Fitbit only control group among African American survivors of breast cancer. Compared with the control

group, participants in the intervention arm demonstrated sustained improvement across a larger number of self-regulatory behaviors for healthy diet and PA. In the *Smart Walk* study, an mHealth app–based intervention aimed at increasing PA in African American women, participants demonstrated improvements in self-regulation for PA [40]. Similar to these cohorts, our analytic sample was predominantly composed of African American women. Qualitative data on the interplay between health and spirituality in African American women show that health self-management in African American women is deeply intertwined with their spiritual and faith connections [63]. Thus, it is possible that faith-based interventions, such as this study, may provide greater self-regulatory benefits for African American women.

Social support has been consistently identified as an important psychosocial variable for maintaining healthy diet [64] and PA, with family members and friends being identified as key sources of support [14,17]. Among the social support subscales for healthy diet, participants reported a decrease in friend discouragement in our study. In addition, improvements in friend encouragement were associated with increased daily fruit and vegetable intake. With respect to social support for PA, improving trends were noted for friend participation; however, this did not reach statistical significance. However, increases in family and friend participation were positively associated with increased MPA per week. Ferrante et al [34] reported improvements in social support for healthy diet but no significant changes in social support for PA, similar to our findings. In the *Smart Walk* study, there was no significant improvement in social support for PA; however, qualitative participant feedback suggested that increasing features for engagement on a sharing board with moderated discussion by the study team can facilitate participation and possibly increase social support [40]. In a recent qualitative analysis of the FAITH! app pilot study, participants demonstrated that they received encouragement and social support toward a healthy lifestyle from posts by other participants on the sharing board [48]. Thus, facilitating opportunities for discussion among study participants through a sharing board feature may foster meaningful engagement in mHealth interventions and increase social support in mHealth lifestyle interventions. Further investigation is necessary to better elucidate the most effective means of enhancing social support through greater opportunities for engagement within mHealth interventions.

Prior studies have pointed to multiple barriers such as fatigue, time, cost, and lack of social support leading to poor diet and PA among African Americans [13,14,16]. In addition, there are links between socioeconomic and educational inequities experienced by African Americans that influence perceived barriers to healthy diet and PA. Analysis of survey data from a predominantly African American population by Sharpe et al [65] found that food-secure households reported better diet-related psychosocial factors than food insecure households; however, both groups had largely similar dietary intake patterns. Another study by Wilcox et al [66] found that in a predominantly African American population, less than a high school education was associated with lower diet quality, whereas income and food security were positively associated with higher diet quality. Employed African Americans have been found to have a lower cooking frequency than unemployed African Americans, which indicates that time constraints may be a limiting factor in healthy diet among employed African Americans [16]. Our evaluation of participants' perceived barriers to healthy diet and PA encompassed similar factors as these studies but also included sociocultural considerations (hairstyle), ability to prepare healthy meals, environmental constraints (neighborhood), and support from family and friends. Overall, there was an improvement in perceived barriers to healthy diet, which further correlated with improvements in dietary intake. These findings suggest that our mHealth lifestyle intervention may offer support to African Americans in the navigation of perceived barriers stemming from longstanding structural inequities—or possibly, in spite of these existing inequities—by providing CVH education and highlighting practical strategies to incorporate healthy diet into daily life.

Strengths and Limitations

There are several strengths to this study. Our findings contribute further data on improvements in diet and PA-related psychosocial factors in African Americans participating in an mHealth lifestyle intervention. We further provide novel contributions by describing how changes in these underlying diet and PA-related psychosocial factors are associated with diet and PA behaviors. The intervention was co-designed with African American community members from participating churches who could give voice to the daily lived experiences of African Americans to ensure that the intervention emphasized African American faith, spirituality, and social connectedness. This study implemented an mHealth lifestyle intervention that was well-aligned with smartphone use patterns in African Americans and one that was well-received by participants [48].

Our study has several limitations. This pilot study included a small convenience sample of African Americans residing in Minnesota; thus, our study was limited by selection bias and is not representative of all African Americans. Furthermore, the small sample size limited our statistical power. As such, we were unable to run formal mediation analyses to probe causal relationships between diet and PA-related psychosocial factors and their respective behaviors. With respect to covariate comparisons, our sample was not adequately sized to detect meaningful patterns based on sociodemographics, EHL, or level of app engagement. The psychosocial factors and CVH behavior measures were self-reported by the participants, which could reflect social desirability bias. Our pretest-posttest, quasi-experimental design lacked a control group, and the study was of a relatively short duration. A longer intervention duration may demonstrate a greater effect on diet and PA-related psychosocial measures. Research is currently underway to further evaluate the preliminary findings presented in this pilot study with a larger, more representative sample of African Americans within a randomized controlled trial (NCT03777709).

Conclusions

Our preliminary findings indicate that the use of a culturally tailored mHealth lifestyle intervention can improve diet and PA behaviors as well as several underlying diet and PA-related psychosocial factors. Diet and PA-related self-regulation, social

support, and perceived barriers may be key psychosocial variables to target in future mHealth lifestyle interventions aiming to improve CVH behaviors among African Americans. In addition, we co-designed and implemented a mobile app–based intervention in partnership with an underserved African American community that was well-aligned with and complemented their mobile technology use patterns. It is important to understand the nuances of mobile technology use among African Americans compared with other populations in the United States and to design interventions that account for these differences to prevent the widening of the digital divide. Mobile app–based interventions may be powerful tools to address CVH disparities that disproportionately affect African Americans.

Acknowledgments

The authors thank all study participants and partnering church congregations for their unwavering support and commitment to the design and implementation of the intervention and their valuable contributions to this research.

LCB was supported by the American Heart Association–Amos Medical Faculty Development Program (grant 19AMFDP35040005) and the National Center for Advancing Translational Sciences (Clinical and Translational Science Awards grant KL2 TR002379), the National Institutes of Health and National Institute on Minority Health and Health Disparities (grant R21 MD013490-01), and the Centers for Disease Control and Prevention (grant CDC-DP18-1817) during the implementation and analysis of this study. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Center for Advancing Translational Science, National Institutes of Health, or Centers for Disease Control and Prevention. This publication was supported by Clinical and Translation Awards grant UL1 TR000135 from the National Center for Advancing Translational Science to the Mayo Clinic and the Mayo Clinic Center for Health Equity and Community Engagement in Research. The funding bodies had no role in the study design; collection, analysis, and interpretation of data; writing of the manuscript; and in the decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

References

1. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2021 update: a report from the American Heart Association. Circulation 2021 Feb 23;143(8):254-743. [doi: 10.1161/CIR.0000000000000950] [Medline: 33501848]
2. Djoussé L, Petrone AB, Blackshear C, Griswold M, Harman JL, Clark CR, et al. Prevalence and changes over time of ideal cardiovascular health metrics among African-Americans: The Jackson Heart Study. Prev Med 2015 May;74:111-116 [FREE Full text] [doi: 10.1016/j.ypmed.2015.02.006] [Medline: 25712326]
3. Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van HL, American Heart Association Strategic Planning Task Force and Statistics Committee. Defining and setting national goals for cardiovascular health promotion and disease reduction: The American Heart Association's strategic impact goal through 2020 and beyond. Circulation 2010 Feb 2;121(4):586-613 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.109.192703] [Medline: 20089546]
4. Cooper LA, Crews DC. COVID-19, racism, and the pursuit of health care and research worthy of trust. J Clin Invest 2020 Oct 01;130(10):5033-5035 [FREE Full text] [doi: 10.1172/JCI141562] [Medline: 32730230]
5. Williams DR, Lawrence JA, Davis BA, Vu C. Understanding how discrimination can affect health. Health Serv Res 2019 Dec;54 Suppl 2(Suppl 2):1374-1388 [FREE Full text] [doi: 10.1111/1475-6773.13222] [Medline: 31663121]
6. Carnethon MR, Pu J, Howard G, Albert MA, Anderson CA, Bertoni AG, American Heart Association Council on Epidemiology Prevention, Council on Cardiovascular Disease in the Young, Council on Cardiovascular and Stroke Nursing, Council on Clinical Cardiology, Council on Functional Genomics and Translational Biology, Stroke Council. Cardiovascular health in African Americans: a scientific statement from the american heart association. Circulation 2017 Nov 21;136(21):393-423. [doi: 10.1161/CIR.0000000000000534] [Medline: 29061565]
7. Churchwell K, Elkind MS, Benjamin RM, Carson AP, Chang EK, Lawrence W, American Heart Association. Call to action: structural racism as a fundamental driver of health disparities: a presidential advisory from the American Heart Association. Circulation 2020 Dec 15;142(24):454-468. [doi: 10.1161/CIR.0000000000000936] [Medline: 33170755]
8. Sims M, Glover LS, Gebreab SY, Spruill TM. Cumulative psychosocial factors are associated with cardiovascular disease risk factors and management among African Americans in the Jackson Heart Study. BMC Public Health 2020 Apr 28;20(1):566 [FREE Full text] [doi: 10.1186/s12889-020-08573-0] [Medline: 32345300]
9. Sims M, Diez-Roux AV, Gebreab SY, Brenner A, Dubbert P, Wyatt S, et al. Perceived discrimination is associated with health behaviours among African-Americans in the Jackson Heart Study. J Epidemiol Community Health 2016 Mar;70(2):187-194 [FREE Full text] [doi: 10.1136/jech-2015-206390] [Medline: 26417003]

10. Sims M, Lipford KJ, Patel N, Ford CD, Min Y, Wyatt SB. Psychosocial factors and behaviors in african americans: The Jackson Heart Study. Am J Prev Med 2017 Jan;52(1S1):48-55. [doi: 10.1016/j.amepre.2016.09.020] [Medline: 27989292]
11. Brewer LC, Redmond N, Slusser JP, Scott CG, Chamberlain AM, Djousse L, et al. Stress and achievement of cardiovascular health metrics: The American Heart Association life's simple 7 in blacks of the Jackson Heart Study. J Am Heart Assoc 2018 Jun 05;7(11):e008855 [FREE Full text] [doi: 10.1161/JAHA.118.008855] [Medline: 29871857]
12. Adams IK, Figueroa W, Hatsu I, Odei JB, Sotos-Prieto M, Leson S, et al. An examination of demographic and psychosocial factors, barriers to healthy eating, and diet quality among african american adults. Nutrients 2019 Feb 28;11(3):519 [FREE Full text] [doi: 10.3390/nu11030519] [Medline: 30823409]
13. Aycinena AC, Valdovinos C, Crew KD, Tsai WY, Mata JM, Sandoval R, et al. Barriers to recruitment and adherence in a randomized controlled diet and exercise weight loss intervention among minority breast cancer survivors. J Immigr Minor Health 2017 Feb;19(1):120-129. [doi: 10.1007/s10903-015-0310-1] [Medline: 26801931]
14. Burse NR, Bhuiyan N, Mama SK, Schmitz KH. Physical activity barriers and resources among black women with a history of breast and endometrial cancer: a systematic review. J Cancer Surviv 2020 Aug;14(4):556-577. [doi: 10.1007/s11764-020-00873-z] [Medline: 32240461]
15. Henry JL, Trude AC, Surkan PJ, Steeves EA, Hopkins LC, Gittelsohn J. Psychosocial determinants of food acquisition and preparation in low-income, urban african american households. Health Educ Behav 2018 Dec;45(6):898-907 [FREE Full text] [doi: 10.1177/1090198118760686] [Medline: 29589482]
16. Farmer N, Wallen GR, Yang L, Middleton KR, Kazmi N, Powell-Wiley TM. Household cooking frequency of dinner among non-hispanic black adults is associated with income and employment, perceived diet quality and varied objective diet quality, HEI (healthy eating index): NHANES analysis 2007-2010. Nutrients 2019 Sep 02;11(9):2057 [FREE Full text] [doi: 10.3390/nu11092057] [Medline: 31480746]
17. Joseph RP, Ainsworth BE, Mathis L, Hooker SP, Keller C. Utility of social cognitive theory in intervention design for promoting physical activity among African-American women: a qualitative study. Am J Health Behav 2017 Sep 01;41(5):518-533 [FREE Full text] [doi: 10.5993/AJHB.41.5.1] [Medline: 28760174]
18. Smith A. African Americans and technology use: a demographic portrait. Pew Research Center. 2014. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2014/01/African-Americans-and-Technology-Use.pdf [accessed 2021-09-13]
19. Anderson C, Cohn S. Mobile Technology and Home Broadband 2019. Pew Research Center. 2019 Jun 13. URL: https://www.pewresearch.org/internet/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2021-09-13]
20. James DC, Harville C. eHealth literacy, online help-seeking behavior, and willingness to participate in mhealth chronic disease research among African Americans, Florida, 2014-2015. Prev Chronic Dis 2016 Dec 17;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]
21. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020 Jan 14;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]
22. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May 15;8(3):449-462. [doi: 10.3945/an.116.014100] [Medline: 28507010]
23. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
24. 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]
25. Joseph RP, Royse KE, Benitez TJ. A systematic review of electronic and mobile health (e- and mhealth) physical activity interventions for african american and hispanic women. J Phys Act Health 2019 Mar 01;16(3):230-239. [doi: 10.1123/jpah.2018-0103] [Medline: 30782040]
26. James DC, Harville C, Sears C, Efunbumi O, Bondoc I. Participation of African Americans in e-health and m-health studies: a systematic review. Telemed J E Health 2017 May;23(5):351-364. [doi: 10.1089/tmj.2016.0067] [Medline: 27792475]
27. James DC, Harville C. Barriers and motivators to participating in mHealth research among African American men. Am J Mens Health 2017 Nov;11(6):1605-1613. [doi: 10.1177/1557988315620276] [Medline: 26634861]
28. James DC, Harville C, Whitehead N, Stellefson M, Dodani S, Sears C. Willingness of African American women to participate in e-health/m-health research. Telemed J E Health 2016 Mar;22(3):191-197 [FREE Full text] [doi: 10.1089/tmj.2015.0071] [Medline: 26313323]
29. James DC, Harville C. Smartphone usage, social media engagement, and willingness to participate in mHealth weight management research among African American women. Health Educ Behav 2018 Jun;45(3):315-322. [doi: 10.1177/1090198117714020] [Medline: 28606004]
30. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]

31. Allicock M, Kendzor D, Sedory A, Gabriel KP, Swartz MD, Thomas P, et al. A pilot and feasibility mobile health intervention to support healthy behaviors in African American breast cancer survivors. J Racial Ethn Health Disparities 2021 Feb;8(1):157-165. [doi: 10.1007/s40615-020-00767-x] [Medline: 32385847]
32. Cohen A, Perozich A, Rajan R, Persky S, Parisi J, Bowie J, et al. Framed, interactive theory-driven texting: effects of message framing on health behavior change for weight loss. Fam Community Health 2017;40(1):43-51 [FREE Full text] [doi: 10.1097/FCH.0000000000000128] [Medline: 27870754]
33. Joseph RP, Keller C, Adams MA, Ainsworth BE. Print versus a culturally-relevant Facebook and text message delivered intervention to promote physical activity in African American women: a randomized pilot trial. BMC Womens Health 2015 Mar 27;15:30 [FREE Full text] [doi: 10.1186/s12905-015-0186-1] [Medline: 25886945]
34. Ferrante JM, Devine KA, Bator A, Rodgers A, Ohman-Strickland PA, Bandera EV, et al. Feasibility and potential efficacy of commercial mHealth/eHealth tools for weight loss in African American breast cancer survivors: pilot randomized controlled trial. Transl Behav Med 2020 Oct 08;10(4):938-948. [doi: 10.1093/tbm/iby124] [Medline: 30535101]
35. Brewer LC, Hayes SN, Jenkins SM, Lackore KA, Breitkopf CR, Cooper LA, et al. Improving cardiovascular health among African-Americans through mobile health: The FAITH! app pilot study. J Gen Intern Med 2019 Aug;34(8):1376-1378 [FREE Full text] [doi: 10.1007/s11606-019-04936-5] [Medline: 30887434]
36. Allen JK, Stephens J, Dennison HC, Stewart KJ, Hauck S. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. J Obes 2013;2013:151597 [FREE Full text] [doi: 10.1155/2013/151597] [Medline: 24392223]
37. Lin M, Mahmooth Z, Dedhia N, Frutchey R, Mercado CE, Epstein DH, et al. Tailored, interactive text messages for enhancing weight loss among African American adults: The TRIMM randomized controlled trial. Am J Med 2015 Aug;128(8):896-904. [doi: 10.1016/j.amjmed.2015.03.013] [Medline: 25840035]
38. Joseph RP, Pekmezi D, Dutton GR, Cherrington AL, Kim Y, Allison JJ, et al. Results of a culturally adapted internet-enhanced physical activity pilot intervention for overweight and obese young adult African American women. J Transcult Nurs 2016 Mar;27(2):136-146 [FREE Full text] [doi: 10.1177/1043659614539176] [Medline: 24934566]
39. Wang J, Cai C, Padhye N, Orlander P, Zare M. A behavioral lifestyle intervention enhanced with multiple-behavior self-monitoring using mobile and connected tools for underserved individuals with type 2 diabetes and comorbid overweight or obesity: pilot comparative effectiveness trial. JMIR Mhealth Uhealth 2018 Apr 10;6(4):e92 [FREE Full text] [doi: 10.2196/mhealth.4478] [Medline: 29636320]
40. Joseph RP, Ainsworth BE, Hollingshead K, Todd M, Keller C. Results of a culturally tailored smartphone-delivered physical activity intervention among midlife African American women: feasibility trial. JMIR Mhealth Uhealth 2021 Apr 22;9(4):e27383 [FREE Full text] [doi: 10.2196/27383] [Medline: 33885368]
41. Brewer LC, Jenkins S, Lackore K, Johnson J, Jones C, Cooper LA, et al. mHealth intervention promoting cardiovascular health among African-Americans: recruitment and baseline characteristics of a pilot study. JMIR Res Protoc 2018 Jan 31;7(1):e31 [FREE Full text] [doi: 10.2196/resprot.8842] [Medline: 29386174]
42. Brewer LC, Hayes SN, Caron AR, Derby DA, Breutzman NS, Wicks A, et al. Promoting cardiovascular health and wellness among african-americans: community participatory approach to design an innovative mobile-health intervention. PLoS One 2019 Aug 20;14(8):e0218724 [FREE Full text] [doi: 10.1371/journal.pone.0218724] [Medline: 31430294]
43. Brewer LC, Balls-Berry JE, Dean P, Lackore K, Jenkins S, Hayes SN. Fostering African-American improvement in total health (FAITH!): an application of the american heart association's life's simple 7™ among midwestern African-Americans. J Racial Ethn Health Disparities 2017 Apr;4(2):269-281 [FREE Full text] [doi: 10.1007/s40615-016-0226-z] [Medline: 27059054]
44. Brewer LC, Morrison EJ, Balls-Berry JE, Dean P, Lackore K, Jenkins S, et al. Preventing cardiovascular disease: participant perspectives of the FAITH! program. J Health Psychol 2017 Oct;24(12):1710-1723. [doi: 10.1177/1359105317695878] [Medline: 28810418]
45. Google Analytics. Google. 2021. URL: https://analytics.google.com/analytics/web/provision/#/provision [accessed 2021-09-14]
46. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
47. Carpenter CJ. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. Health Commun 2010 Dec;25(8):661-669. [doi: 10.1080/10410236.2010.521906] [Medline: 21153982]
48. Brewer LC, Kumbamu A, Smith C, Jenkins S, Jones C, Hayes SN, et al. A cardiovascular health and wellness mobile health intervention among church-going african americans: formative evaluation of the FAITH! app. JMIR Form Res 2020 Nov 17;4(11):e21450 [FREE Full text] [doi: 10.2196/21450] [Medline: 33200999]
49. Brewer LC, Williams DR. We've come this far by faith: the role of the black church in public health. Am J Public Health 2019 Mar;109(3):385-386. [doi: 10.2105/AJPH.2018.304939] [Medline: 30726121]
50. Anderson ES, Winett RA, Wojcik JR, Williams DM. Social cognitive mediators of change in a group randomized nutrition and physical activity intervention: social support, self-efficacy, outcome expectations and self-regulation in the guide-to-health trial. J Health Psychol 2010 Jan;15(1):21-32. [doi: 10.1177/1359105309342297] [Medline: 20064881]

51. 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. [doi: 10.1016/0091-7435(87)90022-3] [Medline: 3432232]
52. Welsh EM, Jeffery RW, Levy RL, Langer SL, Flood AP, Jaeb MA, et al. Measuring perceived barriers to healthful eating in obese, treatment-seeking adults. J Nutr Educ Behav 2012;44(6):507-512 [FREE Full text] [doi: 10.1016/j.jneb.2010.06.005] [Medline: 21665549]
53. Anderson ES, Wojcik JR, Winett RA, Williams DM. Social-cognitive determinants of physical activity: the influence of social support, self-efficacy, outcome expectations, and self-regulation among participants in a church-based health promotion study. Health Psychol 2006 Jul;25(4):510-520. [doi: 10.1037/0278-6133.25.4.510] [Medline: 16846326]
54. Sechrist KR, Walker SN, Pender NJ. Development and psychometric evaluation of the exercise benefits/barriers scale. Res Nurs Health 1987 Dec;10(6):357-365. [doi: 10.1002/nur.4770100603] [Medline: 3423307]
55. Jackson H, Yates BC, Blanchard S, Zimmerman LM, Hudson D, Pozehl B. Behavior-specific influences for physical activity among African American women. West J Nurs Res 2016 Aug;38(8):992-1011. [doi: 10.1177/0193945916640724] [Medline: 27044446]
56. Luszczynska A, Tryburcy M, Schwarzer R. Improving fruit and vegetable consumption: a self-efficacy intervention compared with a combined self-efficacy and planning intervention. Health Educ Res 2007 Oct;22(5):630-638 [FREE Full text] [doi: 10.1093/her/cyl133] [Medline: 17060349]
57. Thompson FE, Subar AF, Smith AF, Midthune D, Radimer KL, Kahle LL, et al. Fruit and vegetable assessment. J Am Diet Assoc 2002 Dec;102(12):1764-1772. [doi: 10.1016/s0002-8223(02)90379-2] [Medline: 12487538]
58. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986 Sep;124(3):453-469. [doi: 10.1093/oxfordjournals.aje.a114416] [Medline: 3740045]
59. Kreuter MW, Sugg-Skinner C, Holt CL, Clark EM, Haire-Joshu D, Fu Q, et al. Cultural tailoring for mammography and fruit and vegetable intake among low-income African-American women in urban public health centers. Prev Med 2005 Jul;41(1):53-62. [doi: 10.1016/j.ypmed.2004.10.013] [Medline: 15916993]
60. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
61. Wolin KY, Heil DP, Askew S, Matthews CE, Bennett GG. Validation of the international physical activity questionnaire-short among blacks. J Phys Act Health 2008 Sep;5(5):746-760 [FREE Full text] [doi: 10.1123/jpah.5.5.746] [Medline: 18820348]
62. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
63. Harvey IS, Johnson L, Heath CD. Womanism, spirituality, and self-health management behaviors of African American older women. Women Gender Fam Color 2013 Jun;1(1):59. [doi: 10.5406/womgenfamcol.1.1.0059]
64. Watters JL, Satia JA, Galanko JA. Associations of psychosocial factors with fruit and vegetable intake among African-Americans. Public Health Nutr 2007 Jul;10(7):701-711. [doi: 10.1017/S1368980007662284] [Medline: 17381950]
65. Sharpe PA, Whitaker K, Alia KA, Wilcox S, Hutto B. Dietary intake, behaviors and psychosocial factors among women from food-secure and food-insecure households in the United States. Ethn Dis 2016 Apr 21;26(2):139-146 [FREE Full text] [doi: 10.18865/ed.26.2.139] [Medline: 27103763]
66. Wilcox S, Sharpe PA, Liese AD, Dunn CG, Hutto B. Socioeconomic factors associated with diet quality and meeting dietary guidelines in disadvantaged neighborhoods in the southeast United States. Ethn Health 2020 Nov;25(8):1115-1131. [doi: 10.1080/13557858.2018.1493434] [Medline: 29966432]

Abbreviations

CVD: cardiovascular disease
CVH: cardiovascular health
EHL: eHealth literacy
FAITH!: Fostering African American Improvement in Total Health
LS7: Life's Simple 7
mHealth: mobile health
MPA: moderate-intensity physical activity
PA: physical activity

Edited by L Buis; submitted 16.03.21; peer-reviewed by C Carrion, A Lucas, K Devine; comments to author 12.05.21; revised version received 27.07.21; accepted 01.09.21; published 12.11.21.

Please cite as:
Cyriac J, Jenkins S, Patten CA, Hayes SN, Jones C, Cooper LA, Brewer LC
Improvements in Diet and Physical Activity–Related Psychosocial Factors Among African Americans Using a Mobile Health Lifestyle Intervention to Promote Cardiovascular Health: The FAITH! (Fostering African American Improvement in Total Health) App Pilot Study
JMIR Mhealth Uhealth 2021;9(11):e28024
URL: https://mhealth.jmir.org/2021/11/e28024
doi:10.2196/28024
PMID:34766917

Original Paper

Identification of the Most Suitable App to Support the Self-Management of Hypertension: Systematic Selection Approach and Qualitative Study

Tourkiah Alessa[1,2], MSc; Mark Hawley[2], PhD; Luc de Witte[2], PhD

[1]Biomedical Technology Department, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia

[2]Centre for Assistive Technology and Connected Healthcare, School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom

Corresponding Author:
Tourkiah Alessa, MSc
Centre for Assistive Technology and Connected Healthcare
School of Health and Related Research
University of Sheffield
30 Regent St
Sheffield City Centre
Sheffield, S1 4DA
United Kingdom
Phone: 44 114 222 1726
Email: talessa@KSU.EDU.SA

Abstract

Background: Smartphone apps are increasingly being used to aid in hypertension self-management, and a large and ever-growing number of self-management apps have been commercially released. However, very few of these are potentially effective and secure, and researchers have yet to establish the suitability of specific hypertension apps to particular contexts.

Objective: The aim of this study is to identify the most suitable hypertension app in the context of Saudi Arabia and its health system.

Methods: This study used a 2-stage approach to selecting the most suitable app for hypertension self-management. First, a systematic selection approach was followed to identify a shortlist of the most suitable apps according to the criteria of potential effectiveness, theoretical underpinning, and privacy and security. Second, an exploratory qualitative study was conducted to select the most suitable from the shortlist: 12 doctors were interviewed, and 22 patients participated in 4 focus groups. These explored participants' attitudes towards self-management apps in general, and their views towards the apps identified via the systematic selection process. The qualitative data were analyzed using framework analysis.

Results: In the first stage, only 5 apps were found to be potentially effective while also having a theoretical underpinning and protecting users' data. In the second stage, both doctors and patients were generally interested in using hypertension apps, but most had no experience with these apps due to a lack of awareness of their availability and suitability. Patients and doctors liked apps that combine intuitive interfaces with a pleasant and clear visual design, in-depth features (eg, color-coded feedback accompanied with textual explanations), activity-specific reminders, and educational content regarding hypertension and potential complications. When the pros and cons of the 5 apps were discussed, 3 apps were identified as being more suitable, with Cora Health rated the highest by the participants.

Conclusions: Only 5 apps were deemed potentially effective and secure. Patients' and doctors' discussions of the pros and cons of these 5 apps revealed that 3 out of the 5 are clearly more suitable, with the Cora Health app being judged most suitable overall.

(JMIR Mhealth Uhealth 2021;9(11):e29207) doi:10.2196/29207

KEYWORDS

app; hypertension; self-management; mHealth; blood pressure; support; Saudi Arabia; cardiology; heart; effective; security

Introduction

Hypertension is one of the most common chronic diseases in adults, affecting 1 billion people worldwide and causing serious health complications, including stroke, heart disease, and renal failure [1-5]. Among Saudi adults over 30 years of age, 27.2% have been diagnosed with hypertension. Self-management can help control blood pressure (BP), mitigating complications arising from hypertension. However, patients commonly encounter substantial barriers to effectively self-managing their condition [6], and many fail to adequately self-manage their BP [2,4,7].

Smartphone use has expanded in recent years, including in Saudi Arabia, where there were 21.8 million smartphone users in 2018. This has resulted in increased access to health apps, which have the potential to assist patients' self-management, for example, by providing educational information and self-monitoring tools [8,9].

Alessa et al [10] have shown that smartphone apps with "comprehensive functionalities" are potentially effective. However, relatively few commercial apps meet these criteria [11], and most lack adequate security measures [11]. Inadequate privacy and security lead to potentially unacceptable risks to users' confidentiality. These authors also found that commercial apps generally lack a clear theoretical basis despite self-management aids having been shown to be more effective when they are theory based [12]. It is imperative that health care only implements interventions that are effective and safe. Privacy, security, and a sound theoretical underpinning should therefore be considered when selecting the most suitable self-management apps.

Although acceptance of an app positively influences its successful use in self-management [13,14], potential users were not consulted about their needs in the development of most of these apps [11,15]. In Saudi Arabia, most hypertension management takes place in hospitals and primary care centers, meaning doctors are the health care workers most actively involved in aiding patients' self-management [16]. However, very few studies have explored patients' or doctors' views toward these apps in general [17-19], and even fewer have examined the Saudi context or that of the other Gulf countries.

The aim of this study is thus to distinguish those hypertension self-management apps that are effective, secure, and underpinned by sound theory, and to identify the most suitable apps for the Saudi context by exploring their acceptance among Saudi doctors and patients. This study will offer a clear approach to selecting effective, secure, and acceptable apps among the many available on commercial app stores.

Methods

Study Design

This study consists of 2 stages. The first adopted a systematic approach using criteria of potential effectiveness, privacy and security, and theoretical underpinning to identify potentially effective and secure apps. The second stage consisted of a qualitative study assessing doctor and patient attitudes toward and acceptance of apps that meet these criteria.

Stage 1: Selection Process According to Existing Evidence

Alessa et al [11] found that 30 hypertension self-management apps out of 186 that were commercially available possess "comprehensive functionalities" and are therefore potentially effective. The present study assessed these 30 potentially effective apps against the criteria of privacy and security and theoretical underpinning.

Privacy and security were assessed based on the Online Trust Alliance [20] and the recommendations of the Information Commissioner's Office. Apps were assessed by 2 reviewers based on the availability of privacy policies, data collection and sharing practices, and data security [11]. Theoretical underpinning was assessed by the of coding each app according to the behavior change technique (BCT) taxonomy V1 through the identification of the number of BCTs present and their frequency. BCTs were then mapped to the mechanisms of action of the Theoretical Domains Framework [11].

Stage 2: Qualitative Study

An exploratory qualitative study was conducted to explore participants' experiences of self-management of hypertension, their attitudes toward self-management apps in general, and their views toward the apps identified via the selection process. This was done via patient focus groups and interviews with doctors. Participants were asked to watch videos providing standardized information about each app and were then asked for their opinions and to rate each app on a 1-to-5 scale (see Multimedia Appendix 1 and 2).

Participants

The qualitative study was conducted in Riyadh, Saudi Arabia. Convenience sampling was used to recruit doctors and patients [21] at 2 primary care centers and 2 hospitals via posters and flyers. Participants responded by email or phone and were sent an information sheet relevant to their involvement as either a doctor or patient. Suitable times were arranged for the focus groups and interviews. Before the commencement of each of these sessions, participants completed a consent form. The ethical approvals for this study were obtained from the ethical committee of the School of Health and Related Research at the University of Sheffield and the ethical committee of the Saudi Ministry of Health (reference #023341 and #18-56ZE, respectively).

To be eligible, focus group participants had to be 18 years or older, have hypertension as a primary disease for a minimum of 6 months, and be able to speak and give consent. Exclusion criteria were having a cognitive impairment or pregnancy. The eligibility criterion for doctors was having treated patients with hypertension for a minimum of 6 months. Interested participants were sent an information sheet and consent form. The interview and focus groups were conducted by the researcher (TA) in Arabic, which is the native language of participants and the researcher. The transcripts were translated into English by TA

and then back translated into Arabic by a professional translation service to ensure accuracy.

Data Analysis

Descriptive statistics were compiled from relevant quantitative data. All qualitative interviews were recorded, transcribed, and then checked for accuracy against the audio files before being translated. Framework analysis was used to analyze the transcripts using NVivo 12 software (QSR International). Framework analysis consists of 5 stages: (1) familiarization, (2) identifying a theoretical framework, (3) indexing, (4) charting, and (5) mapping [22,23]. Data familiarization was achieved by the researcher (TA) conducting interviews and focus groups, and transcribing and checking the transcriptions.

The analysis framework had 2 parts. The first part concerned participants' attitudes toward self-management apps in general. The second part of the framework examined participants' attitudes toward 5 specific apps. The a priori themes and subthemes were confirmed by discussion among the study researchers and summarized. Transcripts were indexed according to these themes and subthemes by TA. If emergent themes and subthemes were identified, TA would add them and recheck the other transcripts for this new theme. The final themes and subthemes were agreed upon through regular discussion between all of the study authors.

Results

Stage 1: Selection Process According to Existing Evidence

Table 1 shows the 30 apps previously identified as potentially effective [11]. All were found to have a theoretical underpinning. The BCTs in these apps linked to 10 out of 14 Theoretical Domains Framework mechanisms of action, with the number of mechanisms underlying each app ranging from 5 to 9.

Twenty-two apps were excluded because they did not have an available privacy policy (n=10) or because they insufficiently protected users' data (n=12).

Of the remaining 8 apps, 3 were duplicates, meaning they were identical versions of the app available for both Android and iPhone platforms. ESH Care (ESH) was also a duplicate, but the Android version had previously been excluded. Only one version of each of the apps was considered, leaving a total of 5 unique apps: Cora Health (Cora), ESH, LifeCourseHyTen (Hyten), Qardio, and Braun Healthy Heart (Braun).

Table 1. Privacy, security, and theoretical underpinning of the 30 potentially effective apps.

Number	App name	Version type	TDF[a] mechanisms of action, n	Privacy and security[b]
1	Blood pressure-Smart BP[c]	iPhone	7	No
2	Fast BP	iPhone	6	No
3	BP Wiz	iPhone	6	No
4	Blood pressure and plus diary	iPhone	7	No
5	BP Grapher simpler	iPhone	7	No
6	BP matters	iPhone	5	No
7	Braun Healthy Heart	iPhone	7	Yes
8	Braun Healthy Heart	Android	7	Yes
9	Qardio	iPhone	5	Yes
10	Qardio	Android	5	Yes
11	Blood Pressure (My Heart)	Android	7	No
12	Blood Pressure Diary	Android	5	No
13	Homedic	iPhone	7	No
14	Hemie	iPhone	4	No
15	LifeCourse HyTen	iPhone	5	Yes
16	LifeCourse HyTen	Android	5	Yes
17	Goal Achiever	Android	7	No
18	Cardio Journal – Blood Pressure diary	Android	6	No
19	Control tension	iPhone	6	No
20	Control tension	Android	6	No
21	ESH Care	iPhone	7	Yes
22	ESH Care	Android	7	No
23	Paracelsus (Pressure control)	Android	7	No
24	Blood Pressure Companion	iPhone	7	No
25	Cora Health	iPhone	9	Yes
26	HeartStar	iPhone	7	No
27	Kang BP	iPhone	6	No
28	BP Diary	Android	7	No
29	BP Diary	iPhone	7	No
30	Bprsseo pro	Android	7	No

[a]TDF: Theoretical Domains Framework.

[b]Apps that meet the criteria for data gathering, sharing, and security have “Yes” indicated, and those that do not have “No” indicated.

[c]BP: blood pressure.

Stage 2: Qualitative Study

Participant Characteristics

Twenty-two patients attended four focus groups, with five to six participants in each group. Twelve doctors were interviewed. The participant characteristics are displayed in Tables 2 and 3.

Table 2. Characteristics of the patient's sample (N=22).

Characteristic	Value
Age (years), mean (range)	50 (33-74)
18-30, n (%)	0 (0)
31-40, n (%)	4 (18)
41-50, n (%)	6 (28)
51-60, n (%)	8 (36)
>61, n (%)	4 (18)
Gender, n (%)	
Males	13 (59)
Females	9 (41)
Time since diagnosed with hypertension (years), n (%)	
<1	4 (18)
1-3	6 (27)
>3	12 (55)
Education level, n (%)	
Less than high school diploma, n (%)	3 (14)
High school diploma	5 (23)
Bachelor’s degree	8 (36)
Master’s degree	4 (18)
Doctorate	2 (9)
Smartphone users, n (%)	
Yes	20 (90)
No	2 (10)
Smartphone brand, n (%)	
iPhone	15 (75)
Android	5 (25)

Table 3. Characteristics of interviewed doctors (N=12).

Characteristics	Value
Age (years), mean (range)	40 (28-57)
Gender, n (%)	
Males	4 (33)
Females	8 (67)
Work experience with hypertension (years), n (%)	15.8 (4-39)
Profession, n (%)	
Resident doctor	2 (17)
Specialist doctor	6 (50)
Consultant doctor	4 (33)
Smartphone owner, n (%)	
Yes	12 (100)
No	0 (0)
Smartphone brand, n (%)	
iPhone	7 (58)
Android	5 (42)

General Views Toward and Experiences of Using Mobile Apps

Table 4 presents the themes and subthemes from the first part of the study framework. This is followed by a description of the results. Selected participants' quotations are provided in Multimedia Appendix 3.

Table 4. Identified themes and subthemes via framework analysis.

Theme and subthemes	Topics
Self-management experiences[a]	
Strategies used by patients and their compliance	Adherence to self-monitoring BP[b], taking required action, adherence to taking medication, adherence to lifestyle, and managing stress
Barriers and issues of using strategies for self-management	Lack of knowledge, busy life, lack of motivation, forgetting, acceptance of disease, asymptomatic patients affecting lack of patient initiative, beliefs about medication, and fear caused by high BP
Role of doctors	Education about and encouragement of self-management strategies
Patient knowledge and awareness about hypertension	Current patient knowledge and required information
Using health apps for self-management	
Doctors and patients experience in using health apps	Patients' experiences in using general apps and HTN[c] apps, and doctors' experiences in using health apps or recommending HTN apps
Expected useful features of smartphone apps	Self-monitoring and reminders, educational information, and feedback
Factors affecting uptake of the app	Demographic factors including age, education, and IT[d] literacy; app usability, app's language, and doctor support
Concerns about using health apps for self-management	Credibility and accuracy, company intentions, patient commitment in using the app, and app usability

[a]Italics indicate a priori themes.

[b]BP: blood pressure.

[c]HTN: hypertension.

[d]IT: internet technology.

Self-management Experience

The majority of doctors noted that most patients take their medication frequently, but some fail to monitor and record their BP. Most patients reported that they tried to monitor their BP and take medication regularly, and tried to stay healthy through diet, exercise, and managing stress. Patients and doctors acknowledged the role of doctors in encouraging patients to effectively self-manage their condition, for example, by setting strategies and goals together, and encouraging patients' adherence to these.

Several barriers to patients' involvement in self-management were mentioned. Doctors identified lack of patient initiative, acceptance of the disease, and inaccurate negative beliefs about medication as the most common barriers. However, patients reported barriers such as relying on impractical tools to record data, lack of knowledge relating to hypertension management, lack of motivation, forgetting, busy lifestyle, social pressures, and lack of exercise opportunities.

Doctors and patients believed that lack of patient knowledge negatively affected self-management. Doctors also expressed concern about patients accessing inappropriate or incorrect information. Doctors felt that younger patients and more educated patients tended to be better informed but would not necessarily take greater responsibility for their own health due to a lack of determination or concern.

Using Health Apps for Self-management

Most doctors reported having experience of using health apps themselves. Patients had experience of using apps for nonmedical purposes (eg, entertainment, socializing) but only 1 patient had ever used a hypertension self-management app before. The other patients were unaware of their availability or suitability. Doctors also had never recommended health apps to their patients. However, the data showed that participants were generally interested in using hypertension apps to support self-management and expected that these would have useful features, such as self-monitoring of BP.

Among doctors, users' ages and educational levels were considered the most influential factor affecting use of hypertension apps, whereas for patients, the most important factors were app language and usability.

Doctors expressed concerns about the credibility and accuracy of the apps, and doubt about their continued availability. They felt that they would be more willing to recommend apps that had been scientifically tested, were based on practice guidelines, or had been checked by doctors.

App Preference

Table 5 presents the themes and subthemes from the second part of the study framework. This is followed by a description of the results (a table showing the side-by-side data for each of the 5 apps is presented in Multimedia Appendix 4).

Table 5. A final framework developed to evaluate 5 apps after completing the analysis process.

Theme and subthemes	Topics
Adequacy of app content[a]	
User data collected	Accuracy and method of data inputting, and type of data collected
Feedback and tracking progress	Presentation of feedback and accuracy of feedback
Reminder	N/A[b]
Information provided	Level of details and type of information (information topics)
Social Support	Communication with others
Content credibility	Credibility
App usability	
How easy to use	App design, layout, and navigation
Training	Type and intensity of training required
Overall app assessment	
Factors affecting uptake and usage	Demographic factors including, age, education, and IT[c] literacy; app feature; language; price; privacy; and ads and promotion
Rating and recommendation	App rating, doctors' willingness to recommend apps, doctors' estimated uptake, patients' willingness to use and recommend apps, general recommendations
Potential benefits and drawbacks of app use	
Expected risks of inappropriate content	Difficulties, including stress, anxiety, and confusion; and decreased app use and poor self-management
Support patients' self-management	Controlled BP[d], empowered self-management, improved compliance and knowledge, and supportive doctors

[a]Italics indicate a priori themes.

[b]N/A: not applicable.

[c]IT: internet technology.

[d]BP: blood pressure.

Adequacy of App Content

Feedback and Tracking Progress

Most doctors and patients liked the 5 apps' method of presenting data in different formats, such as in graphs and tables. The data showed a preference for apps with high-quality graphs (Cora and ESH care), for feedback that used color coding and supplementary text (Cora), and for the automatic calculation of BMI (Qardio and ESH) or BP average (Qardio, ESH, and Cora). Doctors thought that the feedback of all of the apps could be improved if it offered the feature of setting goals (eg, for BP) that was tailored to patients' circumstances and demographic.

Reminders

Participants liked the reminder feature for self-management activities in all 5 apps. Both doctors and patients preferred apps, like Cora, that provide reminders for different tasks (eg, self-monitoring of BP) over apps that provide only a reminder for medication (ESH and Hyten) or a generic reminder for a nonspecified task (Braun and Qardio). A few doctors liked apps that allowed reminders for different medications and doses (ESH and Hyten).

Information Provided

Participants found Qardio's lack of educational information unhelpful. Opinions varied as to the usefulness of information offered by the other apps. Doctors generally criticized apps, like Braun and ESH, that lacked any information about medication and side effects, but also felt that detailed information about side effects of medication (Hyten) might be off-putting for patients. Participants thought that apps (eg, Cora) that have information about hypertension in general, as well as data on hypertension risks, BP readings, and how to measure BP, would benefit patients.

User Data Collected

Participants favored apps that collected detailed information that had easy and clear methods of data entry. They preferred apps that collect other data in addition to BP, such as exercise (Cora and Braun). They felt that some apps are not detailed enough to capture all relevant information (eg, entering the type of exercise) and found the way of entering data in some apps to be more difficult than that in others (Braun), not well organized (Hyten), or likely to lead to typo mistakes (ESH).

Social Support and Content Credibility

Patients had mixed opinions about the social support feature. Some found it useful while others found it unhelpful or unnecessary, given the increased access to social media

platforms. Doctors felt that the credibility of educational information should be ensured, either by assessing if the information was based on medical guidelines or by having apps reviewed by other doctors or medical companies. One doctor suggested that profit-motivated app development may not lead to the best quality information being included.

App Usability

Participants preferred interface designs with easy and clear layouts, where features of the app are easy to reach (eg, with app functions visible in the main menu like in Cora and ESH) rather than embedded in other functions (Qardio and Braun Health).

The muted color schemes of Hyten, Cora, and ESH were considered more user-friendly than were those with strong, bright colors (Braun).

Most doctors and patients thought that some level of training would be required for all 5 apps although they disagreed over the length and intensity that would be needed.

Potential Benefits and Drawbacks of App Use

Participants expressed several possible benefits of using these apps. They thought that reminders and monitoring would help to increase their engagement and that educational information could help to increase their awareness of their condition. However, some doctors were concerned that apps with too few functions (eg, Qardio and ESH) may lead to patients becoming bored, or, conversely, that too much detail (Hyten) or a poor layout (Hyten and Braun) would confuse patients.

Overall App Assessment

App Rating and Recommendation

The doctors' and patients' full rankings for all of the 5 apps, which was calculated by aggregating each group's 1-5 ratings. Cora was ranked highest by both doctors (total 51, mean 4.25) and patients (total 97.5, mean 4.4). Hyten was second among doctors (total 43, mean 3.5), while ESH was second among patients (total 85.5, mean 3.6). ESH was third among doctors (total 41.5, mean 3.4), while Hyten was third among patients (total 80, mean 3.8). Qardio and Braun were ranked lowest by patients (total 64, mean 2.9) and doctors (total 30, mean 2.5), respectively.

Doctors and patients made some recommendations for improvements of app features and content. Cora received the fewest suggestions. Some of the recommendations were common for all 5 apps, such as for the tracking of hospital appointments and other medical conditions. The suggestions are presented in full in Multimedia Appendix 4.

Factors Affecting Uptake and Usage

Doctors and patients identified different factors that may affect the use of the 5 apps. Age was a factor mentioned by several doctors who felt that 2 apps (Cora and Hyten) in particular may pose difficulties to older users. Some doctors stressed the importance of official endorsement by, for instance, the Ministry of Health, or public health campaigns to encourage patient uptake. Inexperience with smartphone technology was seen as another major potential barrier. Participants also mentioned the unavailability of apps in users' own language. Most patients did not express concern with privacy of the apps, but this was mentioned by doctors, particularly concerning high-profile individuals. Patients also expressed concern over app prices and the payment methods that might be required.

Discussion

Principal Findings

This study aimed to identify the most suitable hypertension app in the context of Saudi Arabia and its health system using a 2-stage approach: a systematic selection approach that assessed apps according to the criteria of potential effectiveness, theoretical underpinning, and privacy and security; and an exploratory qualitative study involving 12 doctors and 22 patients. The first stage found that only very few apps were deemed potentially effective and secure. The second stage showed that doctors and patients were generally interested in using hypertension apps. Their discussions of these 5 apps' pros and cons revealed that 3 out of the 5 are clearly more suitable, with Cora being judged the most suitable overall.

Comparison of the Study Findings With the Literature

The selection approach found that of the 30 apps previously identified as potentially effective [11], all 30 contained a theoretical underpinning but only 5 contained adequate privacy and security measures. This demonstrates the pitfalls of commercial app availability: most apps are unlikely to be effective and secure, leading to potentially serious effects on users' health and well-being. This suggests a lack of collaboration between researchers, experts, and developers, which would otherwise help in improving the potential effectiveness and quality of apps or provide clear evidence of effectiveness and safety [23-25].

The qualitative study found that both doctors and patients were interested in using hypertension apps but that most had never used these apps or been recommended them, due to a lack of awareness of their availability and suitability. This is in line with previous research, including that of Morrissey et al [18] who found that few hypertension apps were used by patients due to a lack of knowledge of these apps. This highlights the importance of identifying the most suitable apps and raising awareness of these among health care professionals and the public through official media and education channels [26].

Morrissey et al [18] and Vo et al [13] found that some patients expressed no interest in developing the digital competence required to use mobile health (mHealth) interventions. This contrasts with our study, which found that the majority of participants were keen to engage with self-management apps. One possible explanation for this is the relative age of the study populations. In Saudi Arabia, the average age of hypertension sufferers is lower than that in Europe, meaning the study population recruited for our study also had a younger average age and so was likely to have higher digital competence and greater willingness to engage with smartphone technology. Moreover, most participants in this paper had a higher education level and therefore were likely to have high digital competence. Bol et al [27] found that those with a higher level of education

were more likely to engage with mHealth interventions than were those with a lower education level. A number of patient participants for this present study had some preexisting medical knowledge, which may also partly explain the relatively high level of engagement.

When the pros and cons of the 5 apps were assessed, 3 apps were identified as being more suitable, with Cora rated the highest in participants' ratings. Patients and doctors liked these apps because they combine intuitive interfaces with pleasant and clear visual design, in-depth features (eg, color-coded feedback accompanied with textual explanations), activity-specific reminders, and educational content regarding hypertension and potential complications. Apps are more likely to be used and accepted if they include key components, such as pleasing visuals and the facility to personalize, and if they offer other broader functions, such as education [28]. Detailed features allow users to tailor the app to their circumstances and needs, and provide depth of information to support them [13]. Studies have found that apps that are designed to be easy to use lower the effort a user has to expend in using them [13,26], which could explain why users did not prefer the more complex apps. Our study's findings are also in line with those of Leong et al [29], who found that hypertension apps with an educational component scored higher on the study's quality checklist compared with those that did not.

Doctors and patients expressed somewhat different concerns in identifying the most suitable apps, with doctors generally being more concerned with medical accuracy and patients being generally more concerned with usability, interface, and visual design elements. This is similar to the findings of previous research revealing that doctors and patients often showed somewhat different priorities or preferences regarding mHealth apps even if they agreed to some extent on which is the best overall [30].

The self-management strategies identified in this present study were largely in line with those identified by Barlow et al [31]. Patients try to adopt a variety of self-management methods to stay healthy, such as self-monitoring BP and doing exercise. However, they face difficulties and barriers that affect or delay the adoption of these strategies. Lack of motivation, a busy life, lack of knowledge, and forgetting, were found to be the most common barriers to self-managing hypertension, which again is in line with other studies, including those specific to the Saudi context [6,30,32]. Although stress and anxiety have been identified as 2 of the most common barriers to effective self-management [6], these were not identified as significant barriers in this study. This may be because the main focus of approaches to self-management is on behavioral and medical management, with less focus placed on assisting patients in dealing with the emotional effects of chronic disease [33]. Participants may therefore not have been primed to discuss these topics. A meta-review found supporting self-management interventions with different components, including self-monitoring BP and provision of information, could be effective in controlling BP and improving adherence to adopted strategies [34]. Khatib et al [6] indicated that the barriers patients identified show that they have an interest in finding a solution to effectively self-manage their hypertension, and these authors call for a more targeted, multifaceted intervention to mitigate the identified barriers affecting self-management. Our study found that patients do indeed have an active interest in using mHealth interventions to support their self-management of hypertension, provided certain barriers can be overcome.

Previous research has shown that despite the many advantages of using apps in supporting self-management, certain concerns regarding their use persist, such as the accessibility and usability of the app and the effectiveness of these tools [13,18,19]. Our data are in line with these previous findings. Some participants felt that apps could be a helpful tool and felt motivated by functions that allowed them to track the entered data and their progress over a long period of time. Both patients and doctors raised concerns about the apps, including about the language, with patients also raising concerns about the apps' usability. App developers should consider the cultural preferences of target users (eg, language) and their technical preferences (eg, ease of use) to ensure the acceptance of and engagement with their apps in the future and to alleviate any hindrance affecting the use of health apps [13].

Previous research has found that doctors are in general less positive than are patients regarding the use of mHealth apps [35]. In this study, doctors were generally positive about the prospect of their use. However, they were generally more concerned than were patients about the credibility of the app and patients' ability to continue using it. They also questioned whether older users, who they felt are less competent users of the technology, can easily engage with these apps. Indeed, users' continued or ongoing use of apps and the credibility of health apps have become a major concern in recent years [13]. Vo et al [13] have suggested that app credibility could be increased if certain standards were developed to ensure that they only provide accurate and evidence-based information. Age and digital competence will become less of an issue as younger users, who have been immersed in smartphone culture, carry this competence with them into their old age. Meanwhile, the provision of training for new or older users could further mitigate these concerns [36,37].

Strengths and Limitations

The main strength of this paper lies in its development of a rigorous selection approach to identify the most suitable hypertension app(s), which has the potential to be transferred to apps targeting other conditions and in different contexts. There may be some limitations regarding the generalizability of these results. The study used a self-selecting sample of patients. Those who are more interested in and therefore probably more competent with smartphone technology might have been more likely to volunteer, and this might have impacted the results. A number of the patient participants had some preexisting medical knowledge, which may make the findings less generalizable. The number of older participants in the study sample was relatively low, which may further impact the generalizability, especially since the majority of those with hypertension are older people. The selection approach focused on privacy, security, and theoretical underpinning because these criteria were considered as the most important in implementing and using interventions in the health care field.

We did not consider other issues such as engagement due to the lack of available information about them. Finally, because none of the identified apps were available in Arabic, standardized video presentations were used to demonstrate how the apps worked, but this might have created a biased presentation of the apps' functionalities.

Conclusions

This study found that only 5 apps out of 30 could be deemed potentially effective and secure. It was also found that participants were favorable toward the idea of using health apps to aid in the self-management of hypertension. Through patients' and doctors' discussions of their pros and cons, 3 apps were identified as more suitable than the others, with the Cora Health app being the most suitable overall. In a next step, this app should be evaluated for its usability and effectiveness.

Acknowledgments

The authors thank all participants for their significant contributions. TA is an Assistant Professor at King Saud University, which provided funding for this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview topic guide.
[PDF File (Adobe PDF File), 116 KB - mhealth_v9i11e29207_app1.pdf]

Multimedia Appendix 2
Focus group topic guide.
[PDF File (Adobe PDF File), 163 KB - mhealth_v9i11e29207_app2.pdf]

Multimedia Appendix 3
Qualitative data.
[PDF File (Adobe PDF File), 246 KB - mhealth_v9i11e29207_app3.pdf]

Multimedia Appendix 4
App preference.
[DOCX File , 121 KB - mhealth_v9i11e29207_app4.docx]

References

1. Chow CK, Teo KK, Rangarajan S, Islam S, Gupta R, Avezum A, PURE (Prospective Urban Rural Epidemiology) Study investigators. Prevalence, awareness, treatment, and control of hypertension in rural and urban communities in high-, middle-, and low-income countries. JAMA 2013 Sep 04;310(9):959-968. [doi: 10.1001/jama.2013.184182] [Medline: 24002282]
2. El Bcheraoui C, Memish Z, Tuffaha M, Daoud F, Robinson M, Jaber S, et al. Hypertension and its associated risk factors in the kingdom of saudi arabia, 2013: a national survey. Int J Hypertens 2014;2014:564679 [FREE Full text] [doi: 10.1155/2014/564679] [Medline: 25170423]
3. Health Days 2017 - World Hypertension Day. Ministry of Health. 2017. URL: https://www.moh.gov.sa/en/HealthAwareness/healthDay/2017/Pages/HealthDay-2017-05-17.aspx [accessed 2021-04-06]
4. Al-Nozha MM, Al-Mazrou YY, Arafah MR, Al-Maatouq MA, Khalil MZ, Khan NB, et al. Smoking in Saudi Arabia and its relation to coronary artery disease. J Saudi Heart Assoc 2009 Jul;21(3):169-176 [FREE Full text] [doi: 10.1016/j.jsha.2009.06.007] [Medline: 23960568]
5. Aldiab A, Shubair MM, Al-Zahrani JM, Aldossari KK, Al-Ghamdi S, Househ M, et al. Prevalence of hypertension and prehypertension and its associated cardioembolic risk factors; a population based cross-sectional study in Alkharj, Saudi Arabia. BMC Public Health 2018 Nov 29;18(1):1327 [FREE Full text] [doi: 10.1186/s12889-018-6216-9] [Medline: 30497425]
6. Khatib R, Schwalm J, Yusuf S, Haynes RB, McKee M, Khan M, et al. Patient and healthcare provider barriers to hypertension awareness, treatment and follow up: a systematic review and meta-analysis of qualitative and quantitative studies. PLoS One 2014 Jan;9(1):e84238 [FREE Full text] [doi: 10.1371/journal.pone.0084238] [Medline: 24454721]

7. Saeed A, Al-Hamdan N, Bahnassy A, Abdalla A, Abbas M, Abuzaid L. Prevalence, awareness, treatment, and control of hypertension among Saudi adult population: a national survey. Int J Hypertens 2011;2011:174135 [FREE Full text] [doi: 10.4061/2011/174135] [Medline: 21912737]
8. Choi W, Wang S, Lee Y, Oh H, Zheng Z. A systematic review of mobile health technologies to support self-management of concurrent diabetes and hypertension. J Am Med Inform Assoc 2020 Jun 01;27(6):939-945 [FREE Full text] [doi: 10.1093/jamia/ocaa029] [Medline: 32357368]
9. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
10. Alessa T, Abdi S, Hawley MS, de Witte L. Mobile Apps to Support the Self-Management of Hypertension: Systematic Review of Effectiveness, Usability, and User Satisfaction. JMIR Mhealth Uhealth 2018 Jul 23;6(7):e10723 [FREE Full text] [doi: 10.2196/10723] [Medline: 30037787]
11. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone apps to support self-management of hypertension: review and content analysis. JMIR Mhealth Uhealth 2019 May 28;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]
12. Lepard MG, Joseph AL, Agne AA, Cherrington AL. Diabetes self-management interventions for adults with type 2 diabetes living in rural areas: a systematic literature review. Curr Diab Rep 2015 Jun;15(6):608 [FREE Full text] [doi: 10.1007/s11892-015-0608-3] [Medline: 25948497]
13. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
14. van der Weegen, Verwey R, Tange H, Spreeuwenberg M, de Witte L. Usability testing of a monitoring and feedback tool to stimulate physical activity. Patient Prefer Adherence 2014;8:311-322 [FREE Full text] [doi: 10.2147/PPA.S57961] [Medline: 24669188]
15. Jamaladin H, van de Belt TH, Luijpers LC, de Graaff FR, Bredie SJ, Roeleveld N, et al. Mobile apps for blood pressure monitoring: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e187 [FREE Full text] [doi: 10.2196/mhealth.9888] [Medline: 30429116]
16. Shnaimer J, Gosadi I. Primary health care physicians' knowledge and adherence regarding hypertension management guidelines in southwest of Saudi Arabia. Medicine (Baltimore) 2020 Apr;99(17):e19873 [FREE Full text] [doi: 10.1097/MD.0000000000019873] [Medline: 32332654]
17. Morrissey EC, Glynn LG, Casey M, Walsh JC, Molloy GJ. New self-management technologies for the treatment of hypertension: general practitioners' perspectives. Fam Pract 2018 May 23;35(3):318-322. [doi: 10.1093/fampra/cmx100] [Medline: 29088438]
18. Morrissey EC, Casey M, Glynn LG, Walsh JC, Molloy GJ. Smartphone apps for improving medication adherence in hypertension: patients' perspectives. Patient Prefer Adherence 2018;12:813-822 [FREE Full text] [doi: 10.2147/PPA.S145647] [Medline: 29785096]
19. Hallberg I, Ranerup A, Kjellgren K. Supporting the self-management of hypertension: Patients' experiences of using a mobile phone-based system. J Hum Hypertens 2016 Mar;30(2):141-146 [FREE Full text] [doi: 10.1038/jhh.2015.37] [Medline: 25903164]
20. Privacy in mobile apps - Guidance for app developers. Information Commissioner?s Office. 2013. URL: https://ico.org.uk/media/for-organisations/documents/1596/privacy-in-mobile-apps-dp-guidance.pdf [accessed 2021-04-04]
21. Bryman A. Socail Research Methods. Vol fourth. 4th ed. Oxford, UK: Oxford University Press; 2012:374-380.
22. Ritchie J, Lewis J, Lewis P, Nicholls C, Ormston R. Qualitative Research Practice: A Guide for Social Science Students and Researchers. Thousand Oaks, CA: SAGE Publications; 2013:297-298.
23. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for Evaluating the Content, Usability, and Efficacy of Commercial Mobile Health Apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
24. Scott IA, Scuffham P, Gupta D, Harch TM, Borchi J, Richards B. Going digital: a narrative overview of the effects, quality and utility of mobile apps in chronic disease self-management. Aust Health Rev 2020 Mar;44(1):62-82. [doi: 10.1071/AH18064] [Medline: 30419185]
25. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 01;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
26. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
27. Bol N, Helberger N, Weert J. Differences in mobile health app use: A source of new digital inequalities? The Information Society 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550] [Medline: 26281194]
28. Powell L, Parker J, Harpin V. ADHD: Is There an App for That? A Suitability Assessment of Apps for the Parents of Children and Young People With ADHD. JMIR Mhealth Uhealth 2017 Oct 13;5(10):e149 [FREE Full text] [doi: 10.2196/mhealth.7941] [Medline: 29030325]

29. Leong AY, Makowsky MJ. Quality of Blood Pressure Tracking Apps for the iPhone: Content Analysis and Evaluation of Adherence With Home Blood Pressure Measurement Best Practices. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e10809 [FREE Full text] [doi: 10.2196/10809] [Medline: 30977739]
30. Marshall IJ, Wolfe CDA, McKevitt C. Lay perspectives on hypertension and drug adherence: systematic review of qualitative research. BMJ 2012 Jul 09;345:e3953 [FREE Full text] [doi: 10.1136/bmj.e3953] [Medline: 22777025]
31. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
32. AlHadlaq R, Swarelzahab M, AlSaad S, AlHadlaq A, Almasari S, Alsuwayt S, et al. Factors affecting self-management of hypertensive patients attending family medicine clinics in Riyadh, Saudi Arabia. J Family Med Prim Care 2019 Dec;8(12):4003-4009 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_752_19] [Medline: 31879650]
33. Elissen A, Nolte E, Knai C, Brunn M, Chevreul K, Conklin A, et al. Is Europe putting theory into practice? A qualitative study of the level of self-management support in chronic care management approaches. BMC Health Serv Res 2013 Mar 26;13(1):117 [FREE Full text] [doi: 10.1186/1472-6963-13-117] [Medline: 23530744]
34. Shahaj O, Denneny D, Schwappach A, Pearce G, Epiphaniou E, Parke HL, et al. Supporting self-management for people with hypertension: a meta-review of quantitative and qualitative systematic reviews. J Hypertens 2019 Feb;37(2):264-279. [doi: 10.1097/HJH.0000000000001867] [Medline: 30020240]
35. Morton K, Dennison L, May C, Murray E, Little P, McManus RJ, et al. Using digital interventions for self-management of chronic physical health conditions: A meta-ethnography review of published studies. Patient Educ Couns 2017 Apr;100(4):616-635 [FREE Full text] [doi: 10.1016/j.pec.2016.10.019] [Medline: 28029572]
36. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. J Med Internet Res 2013 Nov 13;15(11):e235 [FREE Full text] [doi: 10.2196/jmir.2588] [Medline: 24225149]
37. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]

Abbreviations

BCT: behavior change technique
BP: blood pressure
Braun: Braun Healthy Heart
Cora: Cora Health
ESH: ESH Care
Hyten: LifeCourseHyTen
mHealth: mobile health

Edited by G Eysenbach; submitted 29.03.21; peer-reviewed by R Cruz Martínez; comments to author 20.04.21; revised version received 23.04.21; accepted 01.08.21; published 17.11.21.

Please cite as:
Alessa T, Hawley M, de Witte L
Identification of the Most Suitable App to Support the Self-Management of Hypertension: Systematic Selection Approach and Qualitative Study
JMIR Mhealth Uhealth 2021;9(11):e29207
URL: https://mhealth.jmir.org/2021/11/e29207
doi:10.2196/29207
PMID:34787586

Original Paper

Content, Behavior Change Techniques, and Quality of Pregnancy Apps in Spain: Systematic Search on App Stores

Aranzazu Muñoz-Mancisidor[1*], MSc; Ruben Martin-Payo[2,3*], PhD; Xana Gonzalez-Mendez[1,3*], MSc; María Del Mar Fernández-Álvarez[2,3], PhD

[1]Hospital Universitario San Agustin, Avilés, Spain

[2]Universidad de Oviedo, Oviedo, Spain

[3]Grupo de Investigación de Promoción de la Salud-Instituto de Investigación Sanitaria del Principado de Asturias, Oviedo, Spain

[*]these authors contributed equally

Corresponding Author:
Ruben Martin-Payo, PhD
Universidad de Oviedo
Calle Julián Clavería s/n, Campus del Cristo
Oviedo, 33006
Spain
Phone: 34 985 103 147
Email: martinruben@uniovi.es

Abstract

Background: Women consult information in mobile apps (apps) during pregnancy, and even obstetrics specialists highlight that pregnancy is the ideal moment for the use of apps as consultation sources. However, the high number of apps designed for pregnancy requires a careful assessment to determine their suitability before recommendation.

Objective: The aim of this study is to identify the apps available in Spanish that can be recommended based on their content, behavior change techniques (BCTs), and quality as a complementary tool during pregnancy.

Methods: A systematic search on app stores to identify apps was performed in the Apple App Store and Google Play with the subject term “pregnancy.” The apps meeting the following criteria were chosen: pregnancy-related content, free, and available in Spanish. An app was excluded if it was classified as a game or entertainment and thus lacking an educational or health aim and if it did not target the population under study. The selected apps were downloaded, and their quality was assessed using the Mobile Application Rating Scale (MARS), with the BCTs included evaluated using the BCT taxonomy version 1 and its content.

Results: A total of 457 apps were identified, 25 of which were downloaded for assessment (5.6%). The median for objective and subjective quality was 2.94 (IQR 2.71-3.46) and 1.75 (IQR 1.25-2.25), respectively. Regarding content, the median of topics included in the apps was 23 (IQR 16-23), with weight gain, nutrition, fetal development, and physical activity being the most common. The median number of BCTs was 12 (IQR 0.5-3.5). The most frequently identified BCTs in the apps were “Self-Monitoring of Outcomes,” followed by “Goal Behavior” and “Instructions.” Statistically significant correlations were observed between objective quality and content (ρ=0.624; *P*=.001), subjective quality and content (ρ=0.638; *P*=.001), objective quality and BCTs (ρ=0.672; *P*<.001), subjective quality and BCTs (ρ=0.623; *P*<.001), and BCTs and content (ρ=0.580; *P*=.002).

Conclusions: The results of this study suggest that only a small percentage of free pregnancy apps available in Spanish should be recommended. The apps with the best MARS scores were those that addressed a higher number of topics and included a higher number of BCTs. Those with the best content and quality, and a higher number of BCTs included could be recommended by health professionals.

(JMIR Mhealth Uhealth 2021;9(11):e27995) doi:10.2196/27995

KEYWORDS

pregnancy; mobile apps; behavior; technology assessment, biomedical; telemedicine

Introduction

It is estimated there are more than 360,000 pregnancies each year in Spain, with the median age of pregnant women being approximately 32 years [1], which implies that there is a generation used to digital technology [2]. This characteristic, as well as the need to obtain information, can influence the use of mobile apps as a source of information [3]. Recent studies suggest that apps are more frequently consulted during pregnancy to look up different types of information, such as behavior or body changes [4].

Off-site health care, mediated by digital technology, has emerged in the past years in obstetrics and gynecology [5]. According to Greiner [6], for the use of this technology to be successful, it must address the values and interests of all parties involved in said use. For example, and specifically related to apps, pregnant women value very highly the inclusion of evidence-based information, experts' opinions, and personalized tips in the app [4]. Therefore, the literature should highlight the importance of selecting appropriate behavioral change techniques (BCTs) in interventions with pregnant women, as not all interventions are equally effective [7]. There are several claims for the integration of apps as pregnancy-monitoring tools. For one, women usually respond quite positively to this integration, and they highlight the empowerment derived from these tools [3]. For another, some authors suggest that pregnancy is the best clinical time to use digital technology in terms of benefits [8]. Finally, there is an abundance of obstetrics and gynecology apps [9], and more specifically pregnancy apps [10].

This last aspect can also have a negative side that therefore must be considered. Carter et al's [11] review highlights the benefits of apps in the support of decision-making during pregnancy, but it also highlights a lack of rigorous evaluation reports about the use and content of the apps. According to this, inappropriate app usage could harm the woman or the fetus.

This calls for a review and assessment of apps before recommendation, and while there are no specific criteria to unequivocally identify which elements in the app to assess [12], previous reviews of apps suggest assessing at least content, quality, and BCTs. There are several studies that have assessed these parameters in pregnancy apps in other languages [13,14], but no other study presenting the results of an assessment of apps available in Spanish has been found. Therefore, the objective of this study is to identify which apps available in Spanish can be recommended based on their content, quality, and BCTs as a complementary tool during pregnancy.

Methods

Study Design

This review used a step-by-step systematic approach that included 2 steps: (1) identifying and selecting the apps with the function of "pregnancy monitoring" available in the Apple App Store and Google Play (Android) between November 2020 and December 2020; and (2) assessing their quality, content, and BCTs.

Step 1: Selection of Smartphone Apps

Our methods sought to replicate the way a patient might access a pregnancy app. Searches were performed in the Apple App Store and Google Play Stores using the word "pregnancy" in both stores. The searches were performed using an iPad Air (fourth generation; Apple Inc) and a Samsung Galaxy Tab A6 (Samsung Electronics Co, Ltd).

A first review of the apps based on their description in the digital stores was carried out. The apps meeting the following criteria were selected: content related to pregnancy, free, and available in Spanish. Apps were excluded if they were classified as a game or entertainment and therefore had no educational or health aim, or if they did not target the population under study (pregnant women).

The apps meeting the criteria specified above were downloaded, and a second review was carried out based on app usage. The same inclusion and exclusion criteria were used for the final selection, with malfunctioning or not working incorporated as exclusion criteria. The apps selected were labeled as recommended, and their quality, content, and BCTs were assessed.

Step 2: Assessment of Smartphone Apps (Quality, Content, and Techniques)

Quality Assessment

The objective and subjective quality of each app was assessed by consensus between 2 researchers (RMP and MFA) using the Spanish version of the Mobile Application Rating Scale (MARS) [15]. This tool was chosen because of the good metric qualities of both its original English version and its Spanish adaptation (internal consistency α>.77; temporal stability r>0.72; interrater reliability >0.76) [16]. MARS has been validated for its use in health apps, and it has been used in several studies related to our research focus, such as specific nutrition apps for pregnancy [13,14,17]. MARS includes 23 items distributed in 2 subscales, objective quality (19 items distributed in 4 dimensions: engagement, functionality, esthetics, and information quality) and subjective quality (4 items), and 6 specific and independent items for health apps (awareness, improvement of knowledge, improvement of behaviors, change intention, social support, and behavior change). All items are rated on a 5-point scale (1, inadequate; 2, poor; 3, acceptable; 4, good; 5, excellent) with possible total scores being 1 to 5 for objective, subjective, and specific items. Mean scores were calculated for each domain (engagement, functionality, esthetics, and information), and overall app quality was calculated by averaging the aggregated mean for all domains [15].

Content Assessment

A content analysis strategy was developed by a researcher (AMM) with the aim of reviewing which pregnancy topics each app could address, and a thematic content analysis was developed. Finally, the content identified related to the target of our study was classified into categories.

Assessment of BCTs

The BCTs included were evaluated independently by 2 researchers (RMP and MFA). No major differences were observed between both researchers, and the final BCTs included were accorded by consensus. A behavioral change technique was only coded when there was clear evidence of its inclusion in the app. The behavioral change techniques used in each app were assessed using the BCT taxonomy version 1 (BCTTV1), which was originally developed by Michie et al [18] and has been shown to be a comprehensive, valid, and reliable approach for assessing techniques for changing behavior in pregnancy apps [7]. Scheoppe et al [19] and Martín-Payo et al [20] have applied a dichotomous scoring system to BCTs to indicate the absence (absence=0) or presence (presence=1) of each technique, permitting a total BCT score per app (possible score 0-93) to be generated.

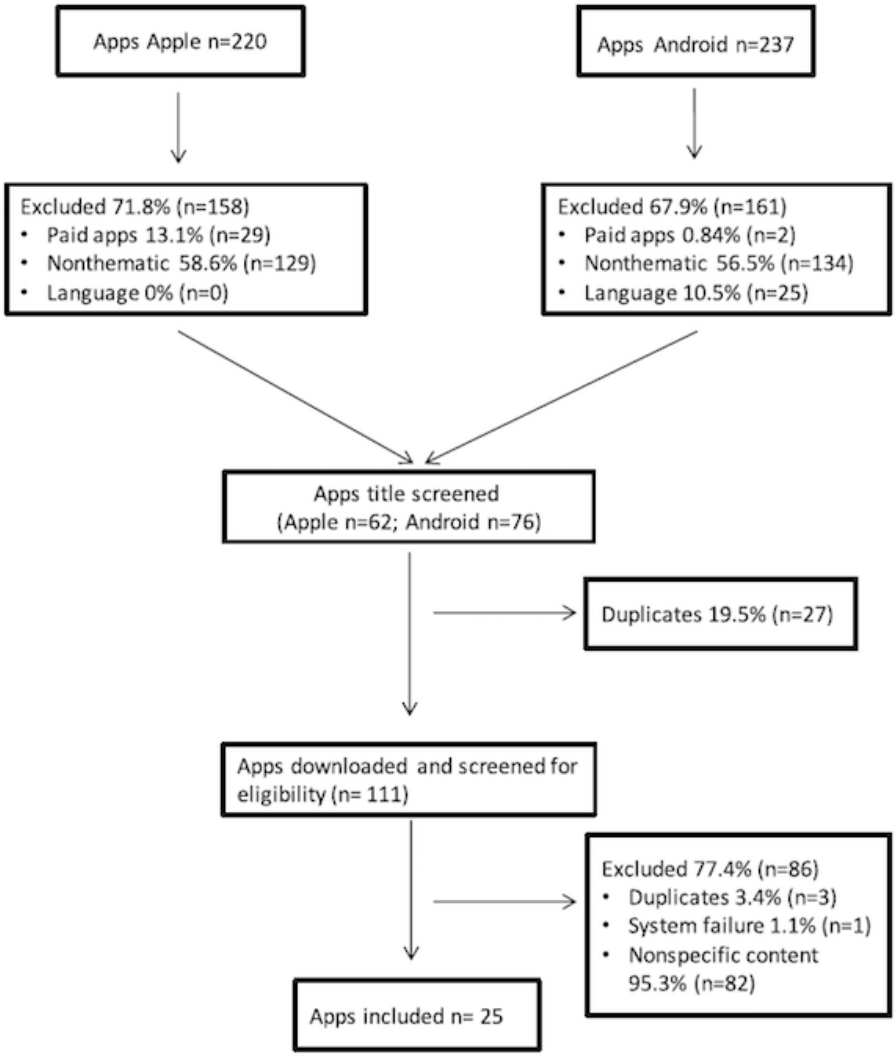
Data Analysis

Total scores, median, and IQR for each app on each domain of the MARS and the BCTTv1 were calculated. To determine if there was any relationship between app quality, BCTs, and content, Spearman's rank correlation was used to determine any associations between MARS total scores, the number of topics, and BCTs. All statistical analyses were conducted using SPSS version 24.0 (IBM Corp) with significance levels set at a *P* value <.05.

Results

A total of 220 apps were identified in the Apple store, and 237 were identified in the Google (Android) app store; of these apps, 71.8% (n=158) and 67.9% (n=161) were excluded, respectively, for not meeting inclusion criteria. More specifically, 27 were duplicates, and 111 were downloaded, of which 77.4% (n=87) were excluded, with 25 (5.6%) apps retrieved for quality, content, and BCT assessment (Figure 1).

Figure 1. Flowchart of the app search process.



Quality Assessment

The median in the different MARS dimensions was superior for objective quality than it was for subjective quality, with emphasis on functionality. The median for the specific part was lower than that for quality (Table 1).

Table 1. Mobile Application Rating Scale for app quality assessment (range 1 to 5).

Characteristic	Value, median (IQR)
Objective quality	2.94 (2.71-3.46)
Engagement	2.60 (2.20-3.60)
Functionality	4.00 (4.00-4.00)
Esthetics	3.00 (2.67-3.67)
Information	2.60 (2.00-3.00)
Subjective quality	1.75 (1.25-2.25)
Would recommend	2.00 (1.00-2.00)
Use after 12 months	2.00 (1.00-3.00)
Payment required	1.00 (1.00-1.00)
Rating	2.00 (1.00-3.00)
Awareness	2.00 (1.00-2.00)
Knowledge	2.00 (2.00-3.00)
Behavior	1.00 (1.00-3.00)
Change intention	2.00 (1.00-3.00)
Social support	2.00 (2.00-3.00)
Behavior change	1.00 (1.00-3.00)

Content Assessment

A total of 28 topics were identified (Multimedia Appendix 1) with the median being 23 (IQR 16-26). The more frequent topics included in the apps were “weight gain,” “balanced diet,” “fetal development,” “physical exercise,” and “changes during pregnancy.” The lower number of topics included in an app was 11 and the highest was 28 (Table 2).

Positive and significant correlations were observed between the MARS scores and the total number of topics included (Table 3).

Table 2. Number of topics included in each app.

App	Topics included in the app, n
Seguidor de mi embarazo: Preglife	24
Mi embarazo semana a semana en español	25
Mi Embarazo día a día	21
Mi embarazo día a día: Semanas de embarazo español	25
Embarazadas primerizas	15
Guía para Embarazadas Primerizas Gratis	21
Embarazo Mes a Mes	18
Tu Embarazo Semana a Semana	25
Mi EMBARAZO por SEMANAS Calendario Maternidad	21
Embarazo semana a semana español días y meses	11
Embarazo semana a semana español	17
Cuidados en el Embarazo	11
Embarazo saludable	14
Mi embarazo como prepararse día a día	11
Embarazo +	27
Babycenter	27
iNatal	28
Embarazo Semana a Semana app	26
Embarazo. Sprout	24
Tu Embarazo	20
Mi embarazo Doctissimo	26
Yo Embarazo Ribera Salud	13
Gestavida	27
Embarazo Óptimo	28
Mi embarazo al día	23

Table 3. Correlation between Mobile Application Rating Scale scores and total topics included in the apps.

Characteristic	Correlation of total of topics included in the app, ρ	*P* value
Objective quality	0.624	.001
Subjective quality	0.638	.001
Awareness	0.537	.006
Knowledge	0.727	<.001
Behavior	0.539	.005
Change intention	0.565	.003
Social support	0.684	<.001
Behavior change	0.734	<.001

BCT Assessment

A total of 12 different BCTs were identified, with a median of 2 (IQR 0.5-3.5).

The most frequently identified BCTs in the apps were "Self-Monitoring of Outcomes," followed by "Goal Behavior" and "Instructions" (Table 4).

The Spearman correlation analysis showed a significant and direct association between the number of BCTs included in the app, their quality, and the number of topics addressed (Table 5).

Table 4. Percentage of apps that included each BCT

BCTs[a]	Apps that included each BCT, n (%) (N=25)
Self-monitoring of outcomes of behavior	11 (44)
Goal-setting behavior	10 (40)
Prompts/cues	9 (36)
Instruction on how to perform a behavior	9 (36)
Action planning	6 (24)
Goal setting outcome	6 (24)
Social support unspecified	5 (20)
Demonstration of the behavior	4 (16)
Self-monitoring of behavior	2 (8)
Credible source	1 (4)
Graded tasks	1 (4)
Monitoring of emotional consequences	1 (4))

[a]BCT: behavior change technique.

Table 5. Correlation between Mobile Application Rating Scale scores and total number of topics included in the apps with BCTs.

Characteristic	Correlation of number of BCTs[a] included in the app, ρ	*P* value
Objective quality	0.672	<.001
Subjective quality	0.623	<.001
Awareness	0.510	.009
Knowledge	0.588	.002
Behavior	0.654	<.001
Change intention	0.572	.003
Social support	0.520	.008
Behavior change	0.668	<.001
Total number of topics included in the app	0.580	.002

[a]BCT: behavior change technique.

Discussion

Although many hundreds of pregnancy apps are commercially available, of those retrieved in this study, only 25 contained potentially suitable pregnancy-specific content to be recommended to pregnant women. This means that, according to the criteria used by the researchers based on quality and content, approximately 5.5% (25/457) of the apps could be recommended. Previous studies have drawn similar conclusions and highlight that not all obstetrics-gynecology commercialized apps can be recommended [21,22] as similar percentages have been observed in other studies despite different selection criteria being used [23].

Although popular app ratings in some digital shops can be useful on some occasions, they are not free from manipulation, and market research suggests that more than half of the reviews on iOS for apps are fake [23]. For this reason, and considering the results obtained, we consider that health apps must be more rigorously evaluated.

There is a previous study in which the MARS was used to evaluate apps specifically designed for pregnancy in Australia [14]. Although the scores in our study cannot be directly compared with the ones in the Australian study, it can be confirmed that the quality of the apps evaluated was similar to the quality observed in our study. The use of the MARS provides important information about the app's usability or user satisfaction, which is a side of apps usually ignored in health contexts but one that remains essential for the app's feasibility and the effectiveness of its use. Other studies have found the users' star rating to be an indicator of satisfaction, suggesting it as a predictor for app download and usage [24]. In general, the majority of the apps seem to work properly according to the objective quality scores (engagement, functionality, esthetics, and information) but not according to subjective quality scores. This may be related to the characteristics of the items that compose the app. For example, "using the app after 12 months" does not apply considering the planned used of the app, as pregnancy is shorter. On the other hand, the dislike for the "pay for the use of the app," which obtained the lowest score, can be attributed to the characteristics of the Spanish health system,

as a great part of the population perceives health care as free, and therefore, any payment related to health services is not popular.

The most common contents in the majority of apps were related to "weight gain," "nutrition," "fetal development," "physical activity," and "changes during pregnancy," and they tended to include self-monitoring and goal-setting behaviors. The inclusion of these topics in the apps can be potentially considered as very effective. Adoption of healthy behaviors during pregnancy can potentially improve maternal and child health. Adverse perinatal health outcomes are associated with maternal risk factors that may be modifiable through changes in maternal behavior [25,26]. Previous studies show that the use of apps has been effective in the improvement of women's knowledge or even in the promotion of healthy behaviors such as physical activity [27] or healthy eating [28]. Overdijkink et al [29] reached the same conclusion and highlighted the positive influence of using apps for gestational weight gain and increased vegetable and fruit intake, among others. Being able to recommend trustworthy apps as pregnancy tools may contribute to helping those population groups previously described by some authors as "at risk" due to their difficulties in adhering to behavioral recommendations [30].

This study discovered that generally a limited number of BCTs are used in apps specifically designed for pregnancy. The number of BCTs identified is consistent with the number identified by Brown et al [13], and it is superior to the number identified by Musgrave et al [14], at 11 and 5 BCTs, respectively. The consistency of results is not limited to the number of BCTs but also applies to the BCTs included despite the use of different taxonomies. As some authors suggest, the inclusion of BCTs contributes to improving the potential to promote behavior change [31]. This probably justifies the employment of different BCTs in digital behavior change interventions [32]. In this sense, Webb et al [33] conclude that the inclusion of BCTs is linked to the efficacy of interventions in which digital resources are used, a possible motivator for the potential incorporation of more specific BCTs by app developers. From a behavioral point of view, it makes sense to include more appropriate BCTs depending on the objectives [31]. However, it has been observed that a considerable number of apps do not include BCTs. In the literature consulted, these apps do not seem to be the more effective, and thus performing more analytical studies to prove this hypothesis appears warranted.

The results of this study are therefore extremely useful for clinical practice. As presented, apps specifically designed for pregnancy can be very positively evaluated by health professionals and more specifically, by midwives, who monitor low-risk pregnancies in the Spanish health system. Therefore, while health professionals are essential for appropriate pregnancy monitoring [34], apps can be used as complementary care.

Finally, although different criteria could be used to assess and evaluate the eligibility of the apps to be recommended [12], the choice of content selection, MARS scores, and BCTs seem to be adequate as a correlation has been observed among the 3 elements. This could suggest that a ranking could be established for app recommendation, with those apps with better quality, content, and BCTs scoring at the top.

Some limitations of this study should be noted. Our search was restricted to free apps. This was deliberate because we did not want to include those that would incur a cost to people. Another possible limitation is related to the search strategy used. The lack of standardized search terms may lead to the use of those apps considered more adequate by the researchers according to their own experience. Considering the dynamism of the app market, it is possible that future searches will identify different apps, and therefore some cannot be available for recommendation. Finally, no previous research assessing the effectiveness of apps could be found. Future research is needed for assessment because health professionals might not prescribe health apps due to distrust and a lack of knowledge about their efficacy [35].

The results of this study suggest that only a small percentage of free pregnancy apps available in Spanish should be recommended. The apps with the best MARS scores were those that addressed a higher number of topics and included a higher number of BCTs.

Those with best content, quality, and a higher number of BCTs included could be recommended by health professionals.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile Application Rating Scale and behavioral change techniques included in the apps.
[XLSX File (Microsoft Excel File), 15 KB - mhealth_v9i11e27995_app1.xlsx]

References

1. Indicadores demográficos. Instituto Nacional de Estadística. Madrid; 2019. URL: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177003&menu=ultiDatos&idp=1254735573002 [accessed 2021-01-23]
2. Wallwiener S, Müller M, Doster A, Laserer W, Reck C, Pauluschke-Fröhlich J, et al. Pregnancy eHealth and mHealth: user proportions and characteristics of pregnant women using Web-based information sources-a cross-sectional study. Arch Gynecol Obstet 2016 Apr 15;294(5):937-944. [doi: 10.1007/s00404-016-4093-y] [Medline: 27084763]

3. Goetz M, Müller M, Matthies LM, Hansen J, Doster A, Szabo A, et al. Perceptions of patient engagement applications during pregnancy: a qualitative assessment of the patient's perspective. JMIR Mhealth Uhealth 2017 May 26;5(5):e73 [FREE Full text] [doi: 10.2196/mhealth.7040] [Medline: 28550005]
4. Wang N, Deng Z, Wen LM, Ding Y, He G. Understanding the use of smartphone apps for health information among pregnant Chinese women: mixed methods study. JMIR Mhealth Uhealth 2019 Jun 18;7(6):e12631 [FREE Full text] [doi: 10.2196/12631] [Medline: 31215516]
5. Kern-Goldberger AR, Srinivas SK. Telemedicine in obstetrics. Clin Perinatol 2020 Dec;47(4):743-757. [doi: 10.1016/j.clp.2020.08.007] [Medline: 33153659]
6. Greiner A. Telemedicine applications in obstetrics and gynecology. Clin Obstet Gynecol 2017;60(4):853-866. [doi: 10.1097/grf.0000000000000328]
7. Flannery C, Fredrix M, Olander EK, McAuliffe FM, Byrne M, Kearney PM. Effectiveness of physical activity interventions for overweight and obesity during pregnancy: a systematic review of the content of behaviour change interventions. Int J Behav Nutr Phys Act 2019 Nov 01;16(1):97 [FREE Full text] [doi: 10.1186/s12966-019-0859-5] [Medline: 31675954]
8. DeNicola N, Marko K. Connected health and mobile apps in obstetrics and gynecology. Obstet Gynecol Clin North Am 2020 Jun;47(2):317-331. [doi: 10.1016/j.ogc.2020.02.008] [Medline: 32451020]
9. Xu W, Liu Y. mHealthApps: a repository and database of mobile health apps. JMIR Mhealth Uhealth 2015 Mar 18;3(1):e28 [FREE Full text] [doi: 10.2196/mhealth.4026] [Medline: 25786060]
10. Van Dijk MR, Huijgen NA, Willemsen SP, Laven JS, Steegers EA, Steegers-Theunissen RP. Impact of an mHealth platform for pregnancy on nutrition and lifestyle of the reproductive population: a survey. JMIR Mhealth Uhealth 2016 May 27;4(2):e53 [FREE Full text] [doi: 10.2196/mhealth.5197] [Medline: 27234926]
11. Carter J, Sandall J, Shennan AH, Tribe RM. Mobile phone apps for clinical decision support in pregnancy: a scoping review. BMC Med Inform Decis Mak 2019 Nov 12;19(1):219 [FREE Full text] [doi: 10.1186/s12911-019-0954-1] [Medline: 31718627]
12. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: A systematic review. J Telemed Telecare 2016 Oct 18;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
13. Brown HM, Bucher T, Collins CE, Rollo ME. A review of pregnancy iPhone apps assessing their quality, inclusion of behaviour change techniques, and nutrition information. Matern Child Nutr 2019 Jul;15(3):e12768 [FREE Full text] [doi: 10.1111/mcn.12768] [Medline: 30569549]
14. Musgrave LM, Kizirian NV, Homer CSE, Gordon A. Mobile phone apps in Australia for improving pregnancy outcomes: systematic search on app stores. JMIR Mhealth Uhealth 2020 Nov 16;8(11):e22340 [FREE Full text] [doi: 10.2196/22340] [Medline: 33196454]
15. Martin Payo R, Fernandez Álvarez MM, Blanco Díaz M, Cuesta Izquierdo M, Stoyanov S, Llaneza Suárez E. Spanish adaptation and validation of the Mobile Application Rating Scale questionnaire. Int J Med Inform 2019 Sep;129:95-99. [doi: 10.1016/j.ijmedinf.2019.06.005] [Medline: 31445295]
16. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
17. Brown HM, Bucher T, Collins CE, Rollo ME. A review of pregnancy apps freely available in the Google Play Store. Health Promot J Austr 2020 Sep;31(3):340-342. [doi: 10.1002/hpja.270] [Medline: 31225924]
18. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
19. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
20. Martín Payo R, Harris J, Armes J. Prescribing fitness apps for people with cancer: a preliminary assessment of content and quality of commercially available apps. J Cancer Surviv 2019 Jun;13(3):397-405. [doi: 10.1007/s11764-019-00760-2]
21. Farag S, Chyjek K, Chen K. Identification of iPhone and iPad applications for obstetrics and gynecology providers. Obstet Gynecol 2014;124(5):941-945. [doi: 10.1097/aog.0000000000000510]
22. Chyjek K, Farag S, Chen KT. Rating pregnancy wheel applications using the APPLICATIONS scoring system. Obstet Gynecol 2015 Jun;125(6):1478-1483. [doi: 10.1097/AOG.0000000000000842] [Medline: 26000520]
23. Siegel E. Fake reviews in Google Play and Apple App Store. Apptentive. 2014 May 27. URL: https://www.apptentive.com/blog/2014/05/27/fake-reviews-google-play-apple-app-store/ [accessed 2021-01-23]
24. Biviji R, Vest JR, Dixon BE, Cullen T, Harle CA. Factors related to user ratings and user downloads of mobile apps for maternal and infant health: cross-sectional study. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e15663 [FREE Full text] [doi: 10.2196/15663] [Medline: 32012107]

25. Aune D, Saugstad OD, Henriksen T, Tonstad S. Maternal body mass index and the risk of fetal death, stillbirth, and infant death: a systematic review and meta-analysis. JAMA 2014 Apr 16;311(15):1536-1546. [doi: 10.1001/jama.2014.2269] [Medline: 24737366]
26. Lawn JE, Blencowe H, Waiswa P, Amouzou A, Mathers C, Hogan D, et al. Stillbirths: rates, risk factors, and acceleration towards 2030. The Lancet 2016 Feb 6;387(10018):587-603. [doi: 10.1016/s0140-6736(15)00837-5]
27. Cawley C, Buckenmeyer H, Jellison T, Rinaldi JB, Vartanian KB. Effect of a health system-sponsored mobile app on perinatal health behaviors: retrospective cohort study. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e17183 [FREE Full text] [doi: 10.2196/17183] [Medline: 32628123]
28. van Dijk MR, Koster MPH, Oostingh EC, Willemsen SP, Steegers EAP, Steegers-Theunissen RPM. A mobile app lifestyle intervention to improve healthy nutrition in women before and during early pregnancy: single-center randomized controlled trial. J Med Internet Res 2020 May 15;22(5):e15773 [FREE Full text] [doi: 10.2196/15773] [Medline: 32412417]
29. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The usability and effectiveness of mobile health technology–based lifestyle and medical intervention apps supporting health care during pregnancy: systematic review. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e109 [FREE Full text] [doi: 10.2196/mhealth.8834] [Medline: 29691216]
30. Flannery C, Dahly D, Byrne M, Khashan A, McHugh S, Kenny LC, et al. Social, biological, behavioural and psychological factors related to physical activity during early pregnancy in the Screening for Pregnancy Endpoints (Cork, Ireland) cohort study. BMJ Open 2019 Jun 20;9(6):e025003 [FREE Full text] [doi: 10.1136/bmjopen-2018-025003] [Medline: 31227527]
31. McKay FH, Slykerman S, Dunn M. The app behavior change scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
32. Thomas Craig KJ, Morgan LC, Chen C, Michie S, Fusco N, Snowdon JL, et al. Systematic review of context-aware digital behavior change interventions to improve health. Transl Behav Med 2021 May 25;11(5):1037-1048. [doi: 10.1093/tbm/ibaa099] [Medline: 33085767]
33. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
34. Grimes HA, Forster DA, Newton MS. Sources of information used by women during pregnancy to meet their information needs. Midwifery 2014 Jan;30(1):e26-e33. [doi: 10.1016/j.midw.2013.10.007] [Medline: 24246969]
35. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among australian general practitioners: survey study. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: 10.2196/13199] [Medline: 31199343]

Abbreviations

BCT: behavior change technique
BCTTV1: behavior change technique taxonomy version 1
MARS: Mobile Application Rating Scale

Edited by L Buis, G Eysenbach; submitted 16.02.21; peer-reviewed by M Fu, C Cheng; comments to author 31.03.21; revised version received 09.04.21; accepted 03.10.21; published 17.11.21.

Please cite as:
Muñoz-Mancisidor A, Martin-Payo R, Gonzalez-Mendez X, Fernández-Álvarez MDM
Content, Behavior Change Techniques, and Quality of Pregnancy Apps in Spain: Systematic Search on App Stores
JMIR Mhealth Uhealth 2021;9(11):e27995
URL: https://mhealth.jmir.org/2021/11/e27995
doi:10.2196/27995
PMID:34787587

Original Paper

A Mobile Intervention for Self-Efficacious and Goal-Directed Smartphone Use in the General Population: Randomized Controlled Trial

Jan Keller[1], DPhil; Christina Roitzheim[2], MSc; Theda Radtke[3], Prof Dr; Konstantin Schenkel[4], MSc, DPhil; Ralf Schwarzer[1,5], Prof Dr

[1]Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

[2]Department of Psychology, Humboldt Universität zu Berlin, Berlin, Germany

[3]Institute of Psychology, University of Wuppertal, Wuppertal, Germany

[4]Department of Psychology, University of Zurich, Zurich, Switzerland

[5]SWPS University of Social Sciences and Humanities, Wroclaw, Poland

Corresponding Author:
Jan Keller, DPhil
Department of Education and Psychology
Freie Universität Berlin
Habelschwerdter Allee 45
Berlin, 14195
Germany
Phone: 49 30 8385 4906
Email: jan.keller@fu-berlin.de

Abstract

Background: People spend large parts of their everyday life using their smartphones. Despite various advantages of the smartphone for daily life, problematic forms of smartphone use exist that are related to negative psychological and physiological consequences. To reduce problematic smartphone use, existing interventions are oftentimes app-based and include components that help users to monitor and restrict their smartphone use by setting timers and blockers. These kinds of digital detox interventions, however, fail to exploit psychological resources, such as through promoting self-efficacious and goal-directed smartphone use.

Objective: The aim of this study is to evaluate the theory-based smartphone app "Not Less But Better" that was developed to make people aware of psychological processes while using the smartphone and to support them in using their smartphone in accordance with their goals and values.

Methods: In a randomized controlled trial, effects of a 20-day intervention app consisting of five 4-day training modules to foster a goal-directed smartphone use were evaluated. In the active control condition (treatment as usual), participants received a digital detox treatment and planned daily time-outs of at least 1 hour per day. Up to a 3-week follow-up, self-reported problematic smartphone use, objectively measured daily smartphone unlocks, time of smartphone use, self-efficacy, and planning towards goal-directed smartphone use were assessed repeatedly. Linear 2-level models tested intervention effects. Mediation models served to analyze self-efficacy and planning as potential mechanisms of the intervention.

Results: Out of 232 enrolled participants, 110 (47.4%; 55 participants in each condition) provided data at postintervention and 88 (37.9%; 44 participants in each condition) at 3-week follow-up. Both conditions manifested substantial reductions in problematic smartphone use and in the amount of time spent with the smartphone. The number of daily unlocks did not change over time. Further, modelling changes in self-efficacy as a mediator between the intervention and problematic smartphone use at follow-up fit well to the data and showed an indirect effect (b=–0.09; 95% bias-corrected bootstrap CI –0.26 to –0.01), indicating that self-efficacy was an important intervention mechanism. Another mediation model revealed an indirect effect from changes in planning via smartphone unlocks at postintervention on problematic smartphone use at follow-up (b=–0.029, 95% bias-corrected bootstrap CI –0.078 to –0.003).

Conclusions: An innovative, theory-based intervention app on goal-directed smartphone use has been found useful in lowering problematic smartphone use and time spent with the smartphone. However, observed reductions in both outcomes were not superior to the active control condition (ie, digital detox treatment). Nonetheless, the present findings highlight the importance

in promoting self-efficacy and planning goal-directed smartphone use to achieve improvements in problematic smartphone use. This scalable intervention app appears suitable for practical use and as an alternative to common digital detox apps. Future studies should address issues of high attrition by adding just-in-time procedures matched to smartphone users' needs.

Trial Registration: German Clinical Trials Register DRKS00017606; https://tinyurl.com/27c9kmwy

(JMIR Mhealth Uhealth 2021;9(11):e26397) doi:10.2196/26397

KEYWORDS

problematic smartphone use; smartphone unlocks; smartphone time; behavior change; self-efficacy; action planning; digital detox; time-out; randomized controlled trial

Introduction

After the first iPhone was released in 2007, smartphones have become an integral part of people's everyday life. Worldwide, in 2021, 3.8 billion persons are using a smartphone [1] and spend large parts of their leisure time with their smartphone (eg, reading news, social media, chatting with friends [2]). Smartphones are used for a variety of daily tasks, thereby simplifying life in many ways. However, there is an increasing scientific and public debate on problematic forms of smartphone use [3]. Empirical findings show links of problematic smartphone use with psychopathology, such as depression, anxiety, stress, and sleep disturbances, as well as negative physical consequences, such as forward neck posture and hand dysfunction [4-7]. Problematic smartphone use can be defined as the "inability to regulate one's use of the mobile phone, which eventually involves negative consequences in daily life" [8]. For instance, this "inability" or lack of control about one's smartphone use can manifest through habitual smartphone checking [4], which occurs on average 88 times per day [9]. Due to high and increasing prevalence rates, problematic smartphone use is considered to be an emerging public health problem [10].

To reduce problematic smartphone use, behavioral approaches focus on either complete abstention or moderating smartphone use by cutting it down—so called digital detox interventions [10]. Several technology-based solutions are available including smartphone apps which help users to monitor and restrict their use by setting timers and blockers. However, most apps lack a psychological underpinning and have not been evaluated by trial designs [10]. Another issue is that monitoring and restrictions alone might not be sufficient as indicated by several studies that examined digital detox interventions [11,12]. Empirical evidence shows that daily smartphone time-outs can indeed lead to decreases of smartphone use [13]; however, there are mixed findings regarding the effects on psychological outcomes [12]. Whereas some studies reveal that digital detox interventions are not related to psychological factors such as well-being or cognitive performance [12], some studies show even negative effects (eg, decreased life satisfaction, lowered affect, or an increase in loneliness) [14-16]. Digital detox interventions might not address useful psychological resources, such as those that can promote a self-efficacious and goal-directed smartphone use, which would be crucial to achieving sustainable behavioral changes [17].

Psychological resources were addressed in an existing group-based intervention app, which included self-monitoring, goal setting, social learning, and competition as active ingredients [18]. Findings from this intervention study showed that daily smartphone use in the intervention condition decreased from 234 to 177 minutes and smartphone-related self-efficacy beliefs were significantly promoted by the intervention [18]. Although these findings seem promising, more research is needed to investigate the mechanisms of these kinds of resource-oriented interventions. Given the previous literature, it remains unclear whether psychological resources (eg, self-efficacy) increased by such interventions would lead to improvements in target outcomes such as problematic smartphone use [4].

The intervention app "Not Less But Better" was developed which focuses on the promotion of psychological resources for goal-directed smartphone use within a 20-day program and is tailored to individuals' goals and values. The intervention offers techniques grounded in cognitive behavior therapy, acceptance and commitment therapy [19], and health behavior change theories like the health action process approach (HAPA) [20,21]. Acceptance and commitment therapy involves allowing unwanted thoughts, feelings, and urges to come and go without struggling with them, and setting value-based goals and achieving them. HAPA reflects a sequence of motivational and volitional constructs, in particular self-efficacy and planning, that are likely to support people in translating their behavioral goals into action. Based on these theoretical frameworks, several behavior change techniques (BCTs) [22]; that is, the smallest units of interventions that can induce behavior change, are applied by the intervention app. These BCTs include promoting self-efficacy beliefs to use the smartphone in accordance with personal goals (eg, focus on past success, BCT 15.3; or vicarious reinforcement, BCT 16.3) and planning when, where, and how to use the smartphone (ie, action planning, BCT 1.4) [22]. The active control condition comprises a 20-day digital detox intervention (ie, treatment as usual) with daily time-out restrictions of at least 1 hour per day (eg, not using the smartphone from 6 pm to 7 pm). This active control condition is in line with common procedures used in digital detox interventions [12,13].

The first aim of our study was to evaluate the effectiveness of the intervention condition in decreasing problematic smartphone use, daily smartphone unlocks (as an indicator for smartphone checking), and time of daily smartphone use. Extending previous studies on digital detox interventions [12], the second aim explored the psychological mechanisms of the intervention through comparison with the active control condition.

In relation to the primary outcome, we hypothesized that problematic smartphone use would show higher decreases in the intervention condition than in the active control condition.

In relation to other outcomes, changes over time of 2 behavioral indicators of goal-directed smartphone regulation were tested: (1) the frequency of daily smartphone unlocks and (2) the time of daily smartphone use. Persons in the intervention condition received psychological strategies to use their smartphone when in accordance with their goals, whereas persons in the active control condition were restricted to not use their smartphone within the self-set time-out interval. We hypothesized that daily smartphone unlocks and time of daily smartphone use would show reductions in both conditions and that no between-group differences would be present (equivalence hypothesis).

Regarding intervention mechanisms, possible pathways of how the intervention condition is related to reductions in problematic smartphone use via self-efficacy and planning of goal-directed smartphone use and reduced smartphone unlocks (as an indicator for smartphone checking behavior) were explored.

Methods

Study and Approval

This study reports primary findings from an app-based, 2-condition, randomized controlled trial (RCT) on healthy smartphone use among adults from the general population. The preregistration for the RCT can be accessed at the German Clinical Trials Register (DRKS00017606; date of registration: August 9, 2019; first participant enrolled: October 9, 2019; targeted sample size: 200). To provide a deeper focus on smartphone-related outcomes and intervention mechanisms, this paper reports findings on the primary outcome (problematic smartphone use), whereas findings on the secondary outcome of the RCT, psychological well-being, are not reported. The Ethics Committee of the Humboldt Universität zu Berlin granted ethics approval for this study (registration #2019-14R1).

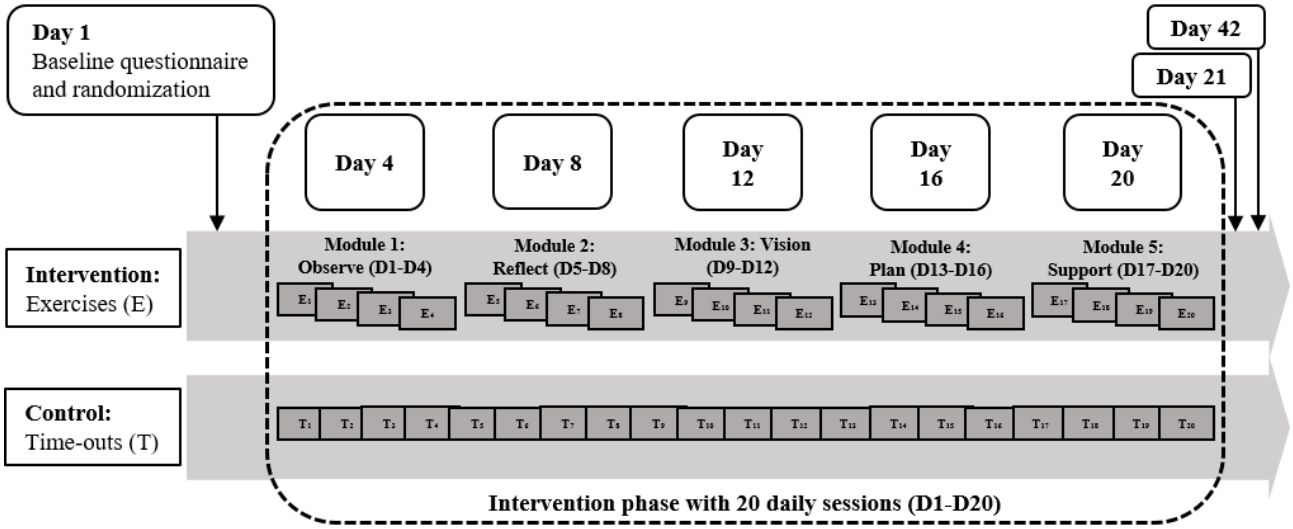
Recruitment and Design

Eligible participants were at least 18 years old, owned and used a smartphone with an iOS operating system (minimum Apple iPhone 5, iOS system 10+), and had sufficient visual ability and skills to understand and complete the English language study materials. Participants were recruited by using reactive strategies such as flyers, online postings, and email lists. As an incentive for study participation, participants took part in a lottery of 4 online shopping vouchers worth €25 (US $29) each and received course credits if needed. Data collection ranged between October 2019 and December 2019.

After downloading the study app and providing informed consent, participants responded to the baseline questionnaire. Subsequently, they were randomly assigned to either the intervention (intervention=1) or active control condition (control=0) using a simple (“flipping a coin”) randomization procedure via a web-based tool. No blinding procedures were used. Based on randomization and throughout the following 20 days (D; D1-D20), participants received daily app-based sessions on goal-directed smartphone use in 5 modules each spanning 4 days (intervention condition) or on defining daily time-outs (active control condition).

Throughout the 20-day intervention period, participants completed brief questionnaires on D4, D8, D12, D16, and D20, corresponding to the completion of the 4-day modules from the intervention condition. Moreover, participants responded to longer questionnaires at postintervention (D21) and at a 3-week follow-up (D42; Figure 1). Multimedia Appendix 1 (Figure S1) provides a Consolidated Standards of Reporting Trials (CONSORT) flow diagram.

Figure 1. Study design with measurement points and 20 daily sessions of the intervention and active control conditions.



Intervention

To foster usability and acceptability, the development of the content in the intervention condition followed a user-centered design and person-based approach [23,24]. Initially, a large pool of brief daily exercises was developed based on elements used in cognitive behavior therapy, acceptance and commitment therapy[16], and behavior change interventions [22]. The pool of exercises was tested with a total of 44 volunteering smartphone users. Moreover, guidance from 7 experts was received, which

resulted in a final set of 20 exercises. The material of 5 out of 20 daily exercises (duration: 2 to 10 minutes per day) from the intervention app can be found in Multimedia Appendix 2 and 3.

The five 4-day modules were the following: Observe (D1-D4), Reflect (D5-D8), Vision (D9-D12), Plan (D13-D16), and Support (D17-D20). In the Observe module, participants observed their physical reactions (eg, posture, impulses of checking behavior) to the smartphone in exercises on impulse control and mindfulness. The Reflect module included educational elements regarding a better understanding of habitual and problematic forms of participants' smartphone use [10]. In the Vision module, smartphone-related exercises on mindfulness, value-related committed action [19], and goal-setting were conducted [20]. The Plan module focused on strategies toward goal-directed smartphone use, such as developing if-then plans on when, where, and how to use the smartphone (action planning) or executing alternative behavioral responses in critical situations (coping planning) [25]. In the Support module, practical tips to support sustainable behavior change (eg, redesign of the home screen) were provided.

Similar to earlier digital detox intervention studies [12], participants in the active control condition received a daily time-out treatment. Across the intervention phase (D1-D20), they were asked to plan a smartphone time-out of at least 1 hour within the next 24 hours. Participants could freely choose whether and how long they executed their planned smartphone time-out; that is, nonaccess to the smartphone was not technically enforced.

Measures

Problematic Smartphone Use

As the primary outcome of the RCT, self-reported problematic smartphone use was measured with 8 items from the Mobile Phone Problem Use Scale [26]) at baseline, postintervention (D21), and follow-up (D42). Items such as "In the past 7 days, I felt anxious if I have not checked for messages or switched on my smartphone for some time" or "In the past 7 days, I have been told that I spend too much time on my smartphone" were answered on a 6-point scale (1="not at all true" to 6="exactly true"). Internal consistency across measurement points and conditions ranged between Cronbach's α=.68 and α=.88.

Daily Smartphone Unlocks and Daily Minutes of Smartphone Use

The frequencies of daily smartphone unlocks and daily minutes of smartphone use from the previous 7 days were assessed by asking participants to transfer objectively measured values from the iOS app "Screen Time" (on iOS phones by default) into the study app. At baseline, postintervention, and follow-up, participants responded to the items "What is your average daily number of unlocks of the last 7 days?" (daily smartphone unlocks) and "What is your average screen time per day of the last 7 days?" (daily minutes of smartphone use). Univariate outliers (z>3.29) of smartphone unlocks and minutes per day were winsorized to 1 unit higher than the next highest value in the distribution [27].

Planning and Self-Efficacy Toward Goal-Directed Smartphone Use

Participants responded to items on planning and self-efficacy toward goal-directed smartphone use on a 6-point scale (1="not at all true" to 6="exactly true") at baseline and throughout the intervention period on D4, D8, D12, D16, and D20. With the instruction "Please refer to today and the past 3 days," responses referred to days when respective modules were conducted in the intervention condition. Planning and self-efficacy items were adapted from scales that were previously validated in various health behavior settings (eg, dietary behavior, physical activity) [28].

Self-reported planning of goal-directed smartphone use was measured with 5 items using the stem "I have made a detailed plan regarding…" followed by statements such as "when to use my smartphone consciously (eg, "on the way to work)".

Self-reported self-efficacy toward goal-directed smartphone use was assessed using 3 items with the stem "I am confident that I can…" followed by statements such as "I use my smartphone consciously even if I first have to find a way to integrate this into my daily routine." Across measurement points and conditions, the internal consistency of the planning scales ranged between Cronbach's α=.87 and α=.97, whereas self-efficacy scales showed a range of Cronbach's alpha between α=.84 and α=.95.

Perceived Impact of the 20-Day Program and Covariates

To evaluate intervention fidelity, perceived impact of the 20-day program was measured at postintervention using an adapted version (eg, "The app increased my intentions/motivation to address my smartphone use") of a validated scale developed to assess the quality of mobile health apps [29]. Internal consistency of the 6-item perceived impact scale was α=.87 in the intervention and α=.91 in the active control condition.

The list of covariates comprised participants' sex, baseline age, smartphone-related action control (scale adapted; eg, "I have tried hard to use my smartphone consciously" [30]), and problematic smartphone use. As a result of attrition analyses (see the Results section), the latter 3 measures were added to the list of covariates to control for selective attrition [31].

Statistical Analysis

Power analysis with G*Power version 3.1 revealed that 34 participants per group would be needed to detect a significant within-between interaction (f=0.25) in problematic smartphone use (α=.05, power=0.80, and r=0.40 among repeated measures [26]) across 3 assessments.

Data were analyzed based on the intention-to-treat approach. For applied analyses, Mplus 8 and its full information maximum likelihood procedure were used to account for missing data [32].

Linear 2-level models with 3 time points (D1, D21, and D42; within level) nested in participants (between level) were computed (for a conceptual model see Figure S2, Multimedia Appendix 1). For problematic smartphone use as the outcome (model A), time (linear day trend, centred at 0) x experimental condition (0=active control condition; 1=intervention condition)

interactions were estimated. For daily smartphone unlocks and daily time of smartphone use as outcomes, the equivalence hypothesis was tested by comparing a null model (ie, with the day trend as predictor; model B and model C) with a nested alternative model (ie, with addition of the condition and the linear day x condition interaction as predictors). Using log-likelihood parameters of both models, a chi-square difference test was run [33]. A nonsignificant chi-square value would indicate that the null model was better fit to the data, confirming the equivalence hypothesis. Moreover, grand-mean centered covariates were added as between-level predictors. Unless models did not converge, the linear day trend and the linear day trend x experimental condition interaction were modeled as random effects predictors [34].

Regarding mediation models, a simple mediation model (model A) was specified, in which self-efficacy toward goal-directed smartphone use was a putative postintervention (D20) mediator between experimental conditions and follow-up problematic smartphone use (D42). By using the measure of self-efficacy on D20, the effects of all sessions of the active control and intervention conditions on changes in self-efficacy were tested. In a second mediation model (model B), planning toward goal-directed smartphone use and the frequency of smartphone unlocks were tested as putative sequential mediators between experimental conditions and follow-up problematic smartphone use (D42). To assure temporal order of the sequential mediators planning and unlocks (at postintervention; D21), planning reports on D16 and D20 were used to compute a mean score, reflecting planning levels that referred to the days of the fourth and fifth module. In each mediation model, we controlled for the set of covariates and for baseline levels of the mediators and outcome. Model fit was evaluated using the χ^2 test statistic, the comparative fit index (CFI), root mean square error of approximation (RMSEA), and the root mean square residual (SRMR), with nonsignificant *P* values of the χ^2 test, CFI levels >0.95, and RMSEA and SRMR levels <0.05 indicating good fit [35]. The 95% bias-corrected bootstrap CIs (CI_{bc}) of direct and indirect effects were generated by bootstrapping with 5000 resamples.

Results

Sample Characteristics, Randomization, and Attrition Check

Randomization to the 2 experimental arms was based on 232 enrolled individuals (205 women, 23 men, 1 diverse, 3 missing values) with a mean age of 29.62 years (SD 8.09, range 18-60 years). Further baseline sample characteristics are displayed in Multimedia Appendix 1 (Table S1).

After providing informed consent and responding to the baseline questionnaire, 114 participants were assigned to receive the intervention on goal-directed smartphone use, and 118 participants were assigned to receive the time-out treatment in the active control condition on D1. A randomization check ($\chi 2$ and *t* tests, followed by logistic regressions) using the experimental condition variable as the outcome revealed no unique between-condition differences in baseline variables, pointing to a successful randomization.

A subsample of 110 (47% out of 232; n=55 in each condition) participants provided data at postintervention, and 88 (38% out of 232; n=44 in each condition) participants did so at the follow-up. Attrition rates within the range of this study are normal for online interventions because researchers do not have much control over the attrition of anonymous participants [36]. Participants from the longitudinal sample (n=88) showed a high response rate to questionnaires between D1 and D42 with a mean response rate of 93% (7.47 out of 8 assessments; SD 0.96, range 4-8).

To examine attrition bias, $\chi 2$ tests, *t* tests, and logistic regressions were performed across baseline variables as well as baseline variable x experimental condition interactions, with a dummy-coded attrition variable (0=dropped out; 1=remained in the study) as the outcome. A significant, unique difference emerged for age (dropped out: mean 31.11 years, SD 8; remaining in the study: mean 27.28 years, SD 7.7), baseline problematic smartphone use (dropped out: mean 3.74, SD 0.77; remaining in the study: mean 3.44, SD 0.75), and baseline action control (dropped out: mean 2.21, SD 1.01; remaining in the study: mean 2.55, SD 1.08). This indicates that participants in the longitudinal sample were younger and demonstrated lower problematic smartphone use and higher action control at baseline when compared to those who dropped out. Subsequent analyses therefore controlled for attrition variables of age, problematic smartphone use, and action control [31].

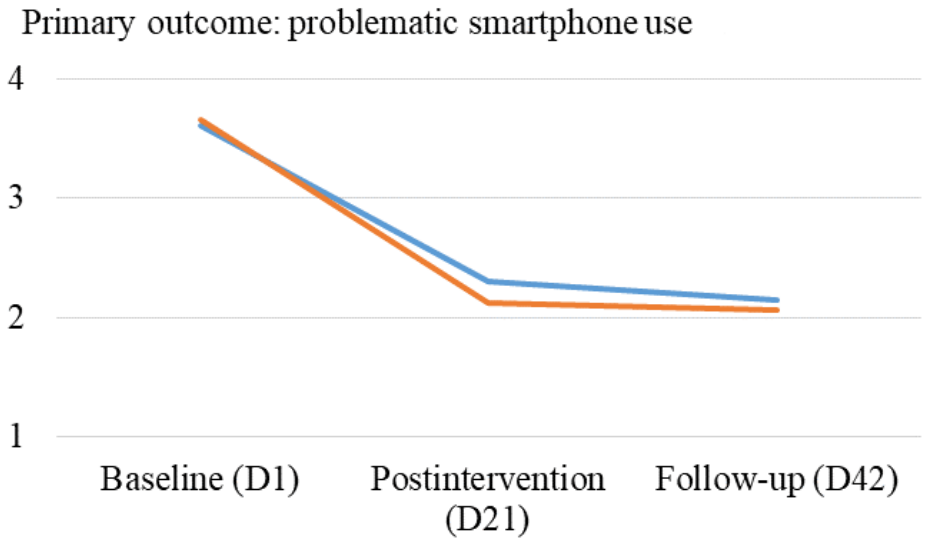
User Engagement and Perceived Impact

Regarding user engagement of the total sample (N=232), participants executed on average 11.34 (SD 7.87) daily time-out sessions in the active control condition compared to 12.84 (SD 7.41) daily exercise sessions in the intervention condition (Table S2, Multimedia Appendix 1). No between-conditions differences were found ($F_{1,230}$=2.24; *P*=.14; η^2=0.01). A high user engagement across those who were retained for analyses on D21 (n=110) was found, with an average completion rate of 93% of the sessions (18.64 out of 20; SD 3.42) in the active control condition and 97% of the exercises (19.40 out of 20; SD 1.55) in the intervention condition. Between-condition differences in postintervention levels of perceived impact revealed that participants reported a higher impact of exercises in the intervention condition (mean 4.88, SD 0.80) as opposed to the active control condition (mean 3.76, SD 1.21; $F_{1,101}$=30.53; *P*<.001; η^2=0.23).

Changes in Study Outcomes Over Time

In a first step, mean levels of problematic smartphone use (Figure 2) and additional study variables (Figure S3, Multimedia Appendix 1) over time and across both conditions were visualized.

Figure 2. Mean levels of problematic smartphone use in both experimental conditions up to the follow-up. D: study day.



In the 2-level model with problematic smartphone use as the outcome, a negative linear day prediction was found, indicating that problematic smartphone use decreased over time in the active control condition (Table 1; model A: *b*=–0.04; 95% CI –0.04 to –0.03; intraclass correlation [ICC]=0.21). Not in line with our hypothesis on the primary outcome, the nonsignificant linear day x condition prediction indicated that problematic smartphone use showed a similar decrease over time in the intervention (vs control) condition. Descriptive analyses (Table S2; Multimedia Appendix 1) showed that problematic smartphone use changed from mean 3.60 to mean 2.30 in the active control condition (ie, a decrease of 36%) and from mean 3.65 to mean 2.12 in the intervention condition (ie, a decrease of 42%) throughout the intervention period.

Moreover, daily smartphone unlocks did not change over time (model B: *b*=–0.13; 95% CI –0.29 to 0.03; ICC=0.65), whereas daily time of smartphone use decreased over time (model C: *b*=–0.77; 95% CI –1.12 to –0.43; ICC=0.67). In the testing of the equivalence hypothesis, alternative models with the intervention condition variable as an additional moderator did not yield a better fit to the data when compared to null models (model B: $\Delta\chi 2_{3.15}$, Δdf=4, *P*=.53; model C: $\Delta\chi 2_{2.14}$, Δdf=4, *P*=.71) [33]. In the alternative models, linear day trend x condition interactions were nonsignificant predictors of daily smartphone unlocks (*b*=–0.15; 95% CI –0.49 to 0.20) and daily time of smartphone use (*b*=0.36; 95% CI –0.32 to 1.05). This indicates that the intervention condition showed similar patterns of change over time in daily smartphone unlocks and daily time of smartphone use when compared to the active control condition.

Regarding significant predictions of covariates, daily smartphone unlocks were more likely when participants reported higher levels of problematic smartphone use at baseline or when they were younger or male; note that only a small group of men (n=23) participated in the RCT. Moreover, a longer duration of smartphone use was more likely when participants reported higher levels of action control and problematic smartphone use at baseline.

Table 1. Estimates for 2-level models predicting changes in study outcomes up to a 3-week follow-up (N=228).

Predictors	Model A[a]: problematic smartphone use		Model B[b,c]: smartphone unlocks per day		Model C[a,c]: smartphone use (minutes per day)	
	b	95% CI_{BC}[d]	*b*	95% CI_{BC}	*b*	95% CI_{BC}
Fixed effects						
Intercept at baseline	*3.54* [e]	3.39 to 3.70	*77.94*	72.70 to 83.19	*219.94*	208.67 to 231.22
Intervention (vs active control)	0.01	–0.19 to 0.21	N/A[f]	N/A	N/A	N/A
Linear day trend	*–0.04*	–0.04 to –0.03	–0.13	–0.29 to 0.03	*–0.77*	–1.12 to –0.43
Linear day trend x intervention	–0.01	–0.01 to 0.01	N/A	N/A	N/A	N/A
Age	–0.01	–0.01 to 0.01	*–1.04*	–1.86 to –0.23	–1.16	–2.51 to 0.19
Sex (0=female; 1=male)	–0.09	–0.36 to 0.18	*27.66*	*6.43 to 48.89*	7.81	–23.15 to 38.77
Acton control at baseline	–0.05	–0.15 to 0.04	–1.01	–5.18 to 3.16	*11.31*	2.17 to 20.46
PSU[g] at baseline	N/A	N/A	*10.99*	4.98 to 17.00	*42.60*	30.76 to 54.45
Random effect variances						
Level 2 (between person)						
Intercept	0.25	0.10 to 0.39	1014.34	518.36 to 1510.32	3938.15	2432.40 to 5443.91
Linear day trend	0.01	–0.01 to 0.01	0.06	–0.46 to 0.58	0.15	–1.26 to 1.56
Level 1 (within person)						
Residual variance	0.41	0.33 to 0.48	564.95	331.07 to 798.83	2205.84	1372.03 to 3039.64

[a]Based on 684 observations.

[b]Based on 683 observations.

[c]Based on the equivalence hypothesis, this model was estimated without a linear day x intervention moderation.

[d]CI_{BC}: bias-corrected bootstrap CI.

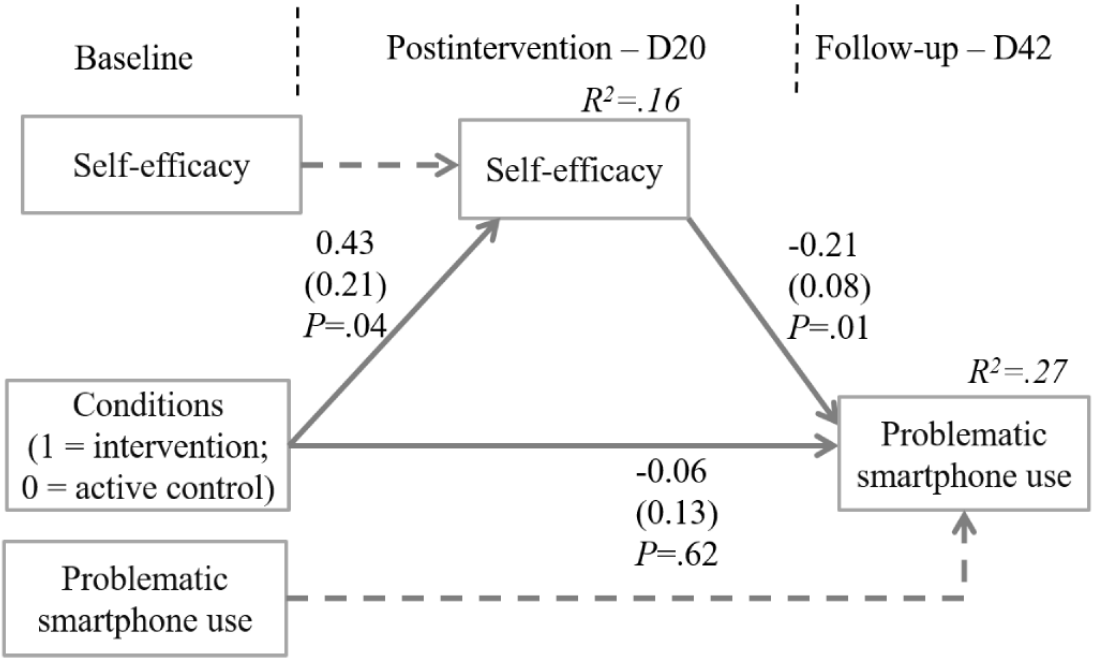
[e]Italics indicate significant fixed effects predictions.

[f]N/A: not applicable.

[g]PSU: problematic smartphone use.

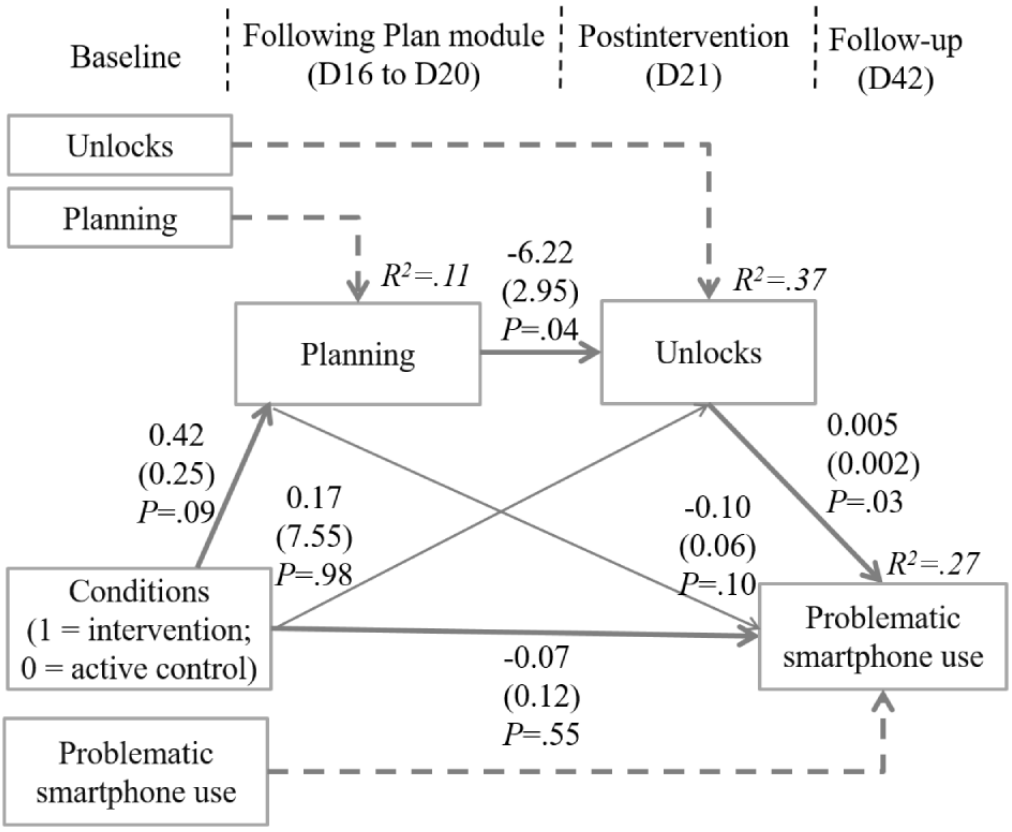
A simple mediation analysis involving experimental conditions, self-efficacy as the mediator at postintervention (D20), and problematic smartphone use as the outcome at follow-up (D42) was run (Figure 3). This mediation model including data from 231 participants fit well with the data (χ^2_8=7.06, *P*=.53; CFI=1.00; RMSEA <0.01; SRMR=0.03). The intervention (vs active control) condition was positively related to changes in self-efficacy at postintervention (*b*=0.43; SE=0.21; *P*=.04; 95% CI_{bc} 0.01-0.85) which, in turn, were negatively linked to changes in follow-up problematic smartphone use (*b*=–0.21; SE=0.08; *P*=.01; 95% CI_{bc} –0.36 to –0.05). The mediation yielded a significant indirect effect of *b*=–0.09; 95% CI_{bc} –0.26 to –0.01). Thus, self-efficacy changes at the end of the intervention translated into substantial reductions in problematic smartphone use at 3 weeks following the intervention. Of the variance of self-efficacy and problematic smartphone use, 16% and 27% were accounted for by the joint set of predictors, respectively.

Figure 3. Self-efficacy as a mediator between the intervention condition and problematic smartphone use (unstandardized coefficients).



A second mediation model was specified with planning after the Plan module (D16 to D20) and smartphone unlocks at postintervention (D21) as sequential mediators (Figure 4) that fit well with the data (χ^2_{15}=19.00; *P*=.21; CFI=0.98; RMSEA=0.03; SRMR=0.05). No between-condition differences for changes in planning were found (*b*=0.42; SE=0.25; *P*=.09). Changes in planning across both conditions were related to unlocks at postintervention (*b*=–6.22; SE=2.95; *P*=.04); that is, with each unit of higher planning toward goal-directed smartphone use, the frequency of smartphone unlocks decreased by approximately 6 units per day. Moreover, changes in daily smartphone unlocks across both conditions were significantly related to problematic smartphone use at follow-up (*b*=0.005; SE=0.002; *P*=.03). For the sequential mediation between intervention condition and problematic smartphone use via planning and daily smartphone unlocks, the 95% CI included 0.000 (95% CI_{bc} –0.050 to 0.000), whereas the 90% CI did not include 0 (*b*=–0.012; 90% CI_{bc} –0.043 to –0.002). When testing the simple mediation between planning and problematic smartphone use via daily smartphone unlocks, we found a significant indirect effect (*b*=–0.029; 95%CI_{bc} –0.078 to –0.003). This indicates that higher levels of planning in both conditions at the end of the intervention phase were associated with lower daily smartphone unlocks which, in turn, were connected with problematic smartphone use. The joint set of predictors explained 11%, 37%, and 27% of the variance in planning, daily smartphone unlocks, and problematic smartphone use, respectively.

Figure 4. Planning and smartphone unlocks as sequential mediators between the intervention condition and problematic smartphone use (unstandardized coefficients).



Discussion

Principal Results

An innovative, theory-guided intervention app was field-tested in an RCT by examining a sample of 232 participants up to a 3-week follow-up. The purpose of the app was to make people aware of their (problematic) smartphone use and to support them in establishing self-efficacious and goal-directed smartphone use in their daily life.

The results indicated that the intervention app was useful in lowering problematic smartphone use (primary outcome) as well as time spent with the smartphone. However, observed reductions in both outcomes were not superior to the active control condition (ie, digital detox treatment).

Findings on reductions of problematic smartphone use in both conditions are in line with previous evidence that interventions on psychological resources [18], along with digital detox interventions [37], can be beneficial for health-related outcomes. The findings revealed selective attrition related to higher levels of problematic smartphone use, and thus future interventions targeting problematic forms of smartphone use should add elements so that persons with severe smartphone-related issues receive support matched to their needs. Other strategies to prevent selective attrition in such online-based study designs include scheduling reminders and following up participants with lower study engagement using just-in-time messages or phone calls [38]. Regarding the time of smartphone use, our findings are consistent with evidence from interventions on goal-directed smartphone use [18] and smartphone time-outs [13], thereby testing both forms of interventions concurrently. Although planning a time-out of at least 1 hour, participants in the active control condition reduced their smartphone time by only 43 minutes per day (from 218 to 175 minutes per day; Table S2, Multimedia Appendix 1). Possibly, participants did not fully adhere to their time-out interval or compensated the abstinence from the smartphone by, for instance, catching up with new messages after their time-out [39]. Participants in the intervention condition, who learned and exercised on goal-directed smartphone use, reduced their smartphone time by 32 minutes per day (from 228 to 196 minutes per day; Table S2, Multimedia Appendix 1), an interesting finding given that less smartphone use was not primarily focused on by the intervention app.

To explain the observed changes in problematic smartphone use, intervention mechanisms were systematically examined. According to theories such as the HAPA model [21], self-efficacy toward goal-directed smartphone use should play a role in the process of behavior change. Similar to previous evidence on resource-oriented interventions [18], the present intervention was successfully fostering self-efficacy beliefs which, in turn, are a relevant resource for reductions in problematic smartphone use. The finding that fostering self-efficacious smartphone use is important to improving problematic smartphone use highlights how resource-oriented interventions can even outperform common digital detox interventions. Future interventions targeting problematic forms of smartphone use should follow up on present evidence by enabling persons to control their smartphone use by themselves.

Moreover, precise planning on when, where, and how to use one's smartphone should also make a difference in behavior change because habitual checking behavior might be reduced [4]. Participants made plans on their smartphone use during the fourth training module (after D13), which was related to postintervention smartphone unlocks; that is, 1 unit of higher planning was linked to a lower daily unlock frequency of 6 units. Planning of smartphone use might result in the planning of smartphone sessions, in which persons take their time to use their phone for current smartphone-related tasks or leisure time activities. This, in turn, might reduce urges towards smartphone-checking behavior and thereby reduce the amount of unlocking of one's phone. Moreover, the results indicate that daily unlocks at postintervention were linked with problematic smartphone use at follow-up. Next to addressing the time of smartphone use, future research should additionally focus on daily smartphone unlocks—as an indicator of checking behavior—and link smartphone unlocks to clinical-, health-, and work-related outcomes [4,40].

Overall, in terms of practical implications, the findings suggest following up with the currently existing version of the app as a means to change problematic smartphone use by scaling the app.

Strengths, Limitations, and Future Research

This study has several strengths. A comprehensive 20-day intervention app was evaluated by contrasting it with a treatment-as-usual control condition with various smartphone-related outcomes and testing intervention mechanisms being examined. However, some limitations need to be considered. First, the substantial reductions in problematic smartphone use were confirmed at a 3-week follow-up, but one cannot be sure whether there would be long-term maintenance of the improved behavior. Second, the primary outcome was a self-report assessment because objective data could not be obtained for this criterion. Third, further smartphone-related outcomes and mechanisms should be examined, such as smartphone-related impulsivity or habitual smartphone use [4]. Fourth, although the intervention condition included comprehensive theory-based content, this condition was not found to be superior in the final evaluation compared to the active control condition. Active control conditions benefit from the attention they received from the researchers or the software, and one can assume that volunteering participants come along with a high level of curiosity and motivation to succeed. However, a passive control condition was intentionally missing in this research design. It should be noted that the effects of a digital detox intervention, as opposed to a passive control condition, were examined by prior research [13]. Fifth, regarding sample characteristics, reactive recruitment procedures (eg, online postings and email lists) resulted in high participation rates of women and younger persons. Thus, associations of study variables with gender and age should be interpreted with caution. Moreover, the distribution of gender and age does not allow for inferences regarding the general population to be made. Further studies need to find representative samples for defined populations, for instance, by using proactive recruitment strategies. Finally, high attrition rates were observed, which is a general issue in online-based research [36]. Future studies could add just-in-time procedures matched to smartphone users' needs to maintain user engagement [38].

Conclusions

An innovative, theory-based intervention app was successfully evaluated as being capable of changing problematic smartphone use. The app was found to be useful for lowering problematic smartphone use and daily time spent with the device. However, observed reductions in both outcomes were not superior to those in the active control condition (ie, digital detox treatment). As an intervention mechanism, the intervention condition developed increased self-efficacy toward goal-directed smartphone use, which was linked to a reduction in problematic smartphone use. Further, planning of smartphone use at the end of the intervention phase was connected with a lower frequency of daily smartphone unlocks, which, in turn, was related to less problematic smartphone use. Further research could build on theories of behavior change and identify more psychological intervention content to improve these types of intervention apps. The app in its current form appears suitable for practical use as an alternative to common digital detox apps.

Acknowledgments

The authors are grateful for the contributions of Selcuk Aciner (study administration), Eléonore Filhol (development of the study app), and Andrea Bedoya (design of the study app).

CR was funded by the Berlin Start-Up Scholarship of the Land of Berlin for the European Social Fund (project # 2018010008).

KS was funded by the Swiss National Science Foundation (grant #100019-169781/1).

Authors' Contributions

CR and JK developed the study concept, study design, and performed data collection, which was supervised by RS, TR, and KS. The data analysis and interpretation were performed by JK under the supervision of RS. Subsequently, CR, JK, and RS drafted the first version of the manuscript, which was followed by TR and KS revising the manuscript and preparing its second version. In the final step, all authors iteratively provided critical revisions. All authors approved the final version of the manuscript for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supporting information on the study design and results.
[DOCX File , 179 KB - mhealth_v9i11e26397_app1.docx]

Multimedia Appendix 2
Exemplary study material from the intervention app.
[PDF File (Adobe PDF File), 4884 KB - mhealth_v9i11e26397_app2.pdf]

Multimedia Appendix 3
Audio-based exercise of Day 2.
[AVI File , 2271 KB - mhealth_v9i11e26397_app3.avi]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1234 KB - mhealth_v9i11e26397_app4.pdf]

References

1. Newzoo. Number of smartphone users from 2016 to 2021 (in billions). statista. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2021-03-15]
2. Chan M. Mobile-mediated multimodal communications, relationship quality and subjective well-being: An analysis of smartphone use from a life course perspective. Computers in Human Behavior 2018 Oct;87:254-262. [doi: 10.1016/j.chb.2018.05.027]
3. Panova T, Carbonell X. Is smartphone addiction really an addiction? J Behav Addict 2018 Jun 01;7(2):252-259 [FREE Full text] [doi: 10.1556/2006.7.2018.49] [Medline: 29895183]
4. Elhai JD, Dvorak RD, Levine JC, Hall BJ. Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. J Affect Disord 2017 Jan 01;207:251-259. [doi: 10.1016/j.jad.2016.08.030] [Medline: 27736736]
5. Thomée S, Härenstam A, Hagberg M. Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults--a prospective cohort study. BMC Public Health 2011;11:66 [FREE Full text] [doi: 10.1186/1471-2458-11-66] [Medline: 21281471]
6. İnal EE, Demİrcİ K, Çetİntürk A, Akgönül M, Savaş S. Effects of smartphone overuse on hand function, pinch strength, and the median nerve. Muscle Nerve 2015 Aug;52(2):183-188. [doi: 10.1002/mus.24695] [Medline: 25914119]
7. Jung SI, Lee NK, Kang KW, Kim K, Lee DY. The effect of smartphone usage time on posture and respiratory function. J Phys Ther Sci 2016 Jan;28(1):186-189 [FREE Full text] [doi: 10.1589/jpts.28.186] [Medline: 26957754]
8. Billieux J. Problematic use of the mobile phone: A literature review and a pathways model. CPSR 2012 Oct 01;8(4):299-307. [doi: 10.2174/157340012803520522]
9. Markowetz A, Błaszkiewicz K, Montag C, Switala C, Schlaepfer TE. Psycho-informatics: Big Data shaping modern psychometrics. Med Hypotheses 2014 Apr;82(4):405-411. [doi: 10.1016/j.mehy.2013.11.030] [Medline: 24529915]
10. van Velthoven MH, Powell J, Powell G. Problematic smartphone use: Digital approaches to an emerging public health problem. Digit Health 2018;4:2055207618759167 [FREE Full text] [doi: 10.1177/2055207618759167] [Medline: 31463071]
11. Mac Cárthaigh S. The effectiveness of interventions to reduce excessive mobile device usage among adolescents: A systematic review. Neurology, Psychiatry and Brain Research 2020 Feb;35:29-37. [doi: 10.1016/j.npbr.2019.11.002]
12. Radtke T, Apel T, Schenkel K, Keller J, von Lindern E. Digital detox: An effective solution in the smartphone era? A systematic literature review. Mobile Media & Communication 2021 Jul 15:205015792110286. [doi: 10.1177/20501579211028647]
13. Hanley SM, Watt SE, Coventry W. Taking a break: The effect of taking a vacation from Facebook and Instagram on subjective well-being. PLoS One 2019;14(6):e0217743 [FREE Full text] [doi: 10.1371/journal.pone.0217743] [Medline: 31170206]
14. Vanman EJ, Baker R, Tobin SJ. The burden of online friends: the effects of giving up Facebook on stress and well-being. J Soc Psychol 2018;158(4):496-507. [doi: 10.1080/00224545.2018.1453467] [Medline: 29558267]
15. Vally Z, D'Souza CG. Abstinence from social media use, subjective well-being, stress, and loneliness. Perspect Psychiatr Care 2019 Oct;55(4):752-759. [doi: 10.1111/ppc.12431] [Medline: 31402459]
16. Radtke, Schmocker D, Hoppe A, Berli C. Digital detoxing? Results of a randomized control trial on the effectiveness of a timeout from smartphone use in employees. Submitted for publication .

17. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016 Sep;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]
18. Ko M, Yang S, Lee J. A group-based intervention app for improving self-regulation of limiting smartphone use. 2015 Presented at: 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; 14-18 March, 2015; Vancouver, Canada p. 1235-1245. [doi: 10.1145/2675133.2675244]
19. Gloster AT, Walder N, Levin ME, Twohig MP, Karekla M. The empirical status of acceptance and commitment therapy: A review of meta-analyses. Journal of Contextual Behavioral Science 2020 Oct;18:181-192. [doi: 10.1016/j.jcbs.2020.09.009]
20. Zhang C, Zhang R, Schwarzer R, Hagger MS. A meta-analysis of the health action process approach. Health Psychol 2019 Jul;38(7):623-637. [doi: 10.1037/hea0000728] [Medline: 30973747]
21. Schwarzer R. Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. Applied Psychology 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
22. 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]
23. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
24. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
25. Hagger MS, Luszczynska A. Implementation intention and action planning interventions in health contexts: state of the research and proposals for the way forward. Appl Psychol Health Well Being 2014 Mar;6(1):1-47. [doi: 10.1111/aphw.12017] [Medline: 24591064]
26. Foerster M, Roser K, Schoeni A, Röösli M. Problematic mobile phone use in adolescents: derivation of a short scale MPPUS-10. Int J Public Health 2015 Feb;60(2):277-286. [doi: 10.1007/s00038-015-0660-4] [Medline: 25645102]
27. Tabachnick B, Fidell L. Using Multivariate Statistics. Fifth Edition. Boston, MA: Allyn & Bacon/Pearson Education; 2007.
28. Schwarzer R, Schuz B, Ziegelmann JP, Lippke S, Luszczynska A, Scholz U. Adoption and maintenance of four health behaviors: theory-guided longitudinal studies on dental flossing, seat belt use, dietary behavior, and physical activity. Ann Behav Med 2007 Apr;33(2):156-166. [doi: 10.1007/bf02879897] [Medline: 17447868]
29. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
30. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: Planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychology & Health 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]
31. Graham JW. Missing data analysis: making it work in the real world. Annu. Rev. Psychol 2009 Jan 01;60(1):549-576. [doi: 10.1146/annurev.psych.58.110405.085530]
32. Muthén B, Muthén L. Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén; 2017.
33. Satorra A, Bentler PM. Ensuring positiveness of the scaled difference chi-square test statistic. Psychometrika 2010 Jun;75(2):243-248 [FREE Full text] [doi: 10.1007/s11336-009-9135-y] [Medline: 20640194]
34. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. J Mem Lang 2013 Apr;68(3):A [FREE Full text] [doi: 10.1016/j.jml.2012.11.001] [Medline: 24403724]
35. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. Methods of Psychological Research 2003;8(2):23-74. [doi: 10.4135/9781526421036946001]
36. 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]
37. Hunt MG, Marx R, Lipson C, Young J. No more FOMO: Limiting social media decreases loneliness and depression. Journal of Social and Clinical Psychology 2018 Dec;37(10):751-768. [doi: 10.1521/jscp.2018.37.10.751]
38. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
39. Curtaz K, Hoppe A, Nachtwei J. Bewusste Auszeiten vom Smartphone tun gut! Eine Interventionsstudie zeigt die Wirksamkeit der (OFFTIME) - App in Hinblick auf Erholung und Arbeitsengagement. HR Performance 2015;1:112-114.
40. Duke, Montag C. Smartphone addiction, daily interruptions and self-reported productivity. Addict Behav Rep 2017 Dec;6:90-95 [FREE Full text] [doi: 10.1016/j.abrep.2017.07.002] [Medline: 29450241]

Abbreviations

BCT: behavior change technique
CFI: comparative fit index
CIbc: bias-corrected bootstrap CI
CONSORT: Consolidated Standards of Reporting Trials
D: Day
HAPA: health action process approach
ICC: intraclass correlation
RCT: randomized controlled trial
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual

Edited by L Buis; submitted 10.12.20; peer-reviewed by SL Tan, S Haug, R Paz Castro; comments to author 06.01.21; revised version received 15.03.21; accepted 10.09.21; published 23.11.21.

Please cite as:
Keller J, Roitzheim C, Radtke T, Schenkel K, Schwarzer R
A Mobile Intervention for Self-Efficacious and Goal-Directed Smartphone Use in the General Population: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(11):e26397
URL: https://mhealth.jmir.org/2021/11/e26397
doi:10.2196/26397
PMID:34817388

Original Paper

Text Message Analysis Using Machine Learning to Assess Predictors of Engagement With Mobile Health Chronic Disease Prevention Programs: Content Analysis

Harry Klimis[1,2], MBBS, BMedSci, GradCertClinEpi, PhD; Joel Nothman[3], BSc, BA, PhD; Di Lu[3], BEng; Chao Sun[3], BEng, BSc, MEng, PhD; N Wah Cheung[1,4,5], MBBS, PhD; Julie Redfern[1], BAppSc, PhD; Aravinda Thiagalingam[1,2], MB ChB, PhD, FRACP, FCSANZ; Clara K Chow[1,2], MBBS, PhD

[1]Faculty of Medicine and Health, Westmead Applied Research Centre, The University of Sydney, Westmead, Australia

[2]Department of Cardiology, Westmead Hospital, Westmead, Australia

[3]Sydney Informatics Hub, The University of Sydney, Camperdown, Australia

[4]Department of Endocrinology, Westmead Hospital, Westmead, Australia

[5]Western Sydney Integrated Care Program, Sydney, Australia

Corresponding Author:
Harry Klimis, MBBS, BMedSci, GradCertClinEpi, PhD
Faculty of Medicine and Health, Westmead Applied Research Centre
The University of Sydney
Westmead
Australia
Phone: 61 88903125
Email: harry.klimis@sydney.edu.au

Abstract

Background: SMS text messages as a form of mobile health are increasingly being used to support individuals with chronic diseases in novel ways that leverage the mobility and capabilities of mobile phones. However, there are knowledge gaps in mobile health, including how to maximize engagement.

Objective: This study aims to categorize program SMS text messages and participant replies using machine learning (ML) and to examine whether message characteristics are associated with premature program stopping and engagement.

Methods: We assessed communication logs from SMS text message–based chronic disease prevention studies that encouraged 1-way (SupportMe/ITM) and 2-way (TEXTMEDS [Text Messages to Improve Medication Adherence and Secondary Prevention]) communication. Outgoing messages were manually categorized into 5 message intents (informative, instructional, motivational, supportive, and notification) and replies into 7 groups (stop, thanks, questions, reporting healthy, reporting struggle, general comment, and other). Grid search with 10-fold cross-validation was implemented to identify the best-performing ML models and evaluated using nested cross-validation. Regression models with interaction terms were used to compare the association of message intent with premature program stopping and engagement (replied at least 3 times and did not prematurely stop) in SupportMe/ITM and TEXTMEDS.

Results: We analyzed 1550 messages and 4071 participant replies. Approximately 5.49% (145/2642) of participants responded with *stop*, and 11.7% (309/2642) of participants were engaged. Our optimal ML model correctly classified program message intent with 76.6% (95% CI 63.5%-89.8%) and replies with 77.8% (95% CI 74.1%-81.4%) balanced accuracy (average area under the curve was 0.95 and 0.96, respectively). Overall, *supportive* (odds ratio [OR] 0.53, 95% CI 0.35-0.81) messages were associated with reduced chance of stopping, as were *informative* messages in SupportMe/ITM (OR 0.35, 95% CI 0.20-0.60) but not in TEXTMEDS (for interaction, *P*<.001). *Notification* messages were associated with a higher chance of stopping in SupportMe/ITM (OR 5.76, 95% CI 3.66-9.06) but not TEXTMEDS (for interaction, *P*=.01). Overall, *informative* (OR 1.76, 95% CI 1.46-2.12) and *instructional* (OR 1.47, 95% CI 1.21-1.80) messages were associated with higher engagement but not *motivational* messages (OR 1.18, 95% CI 0.82-1.70; *P*=.37). For *supportive* messages, the association with engagement was opposite with SupportMe/ITM (OR 1.77, 95% CI 1.21-2.58) compared with TEXTMEDS (OR 0.77, 95% CI 0.60-0.98; for interaction, *P*<.001). *Notification* messages were associated with reduced engagement in SupportMe/ITM (OR 0.07, 95% CI 0.05-0.10) and TEXTMEDS (OR 0.28, 95% CI 0.20-0.39); however, the strength of the association was greater in SupportMe/ITM (for interaction *P*<.001).

Conclusions: ML models enable monitoring and detailed characterization of program messages and participant replies. Outgoing message intent may influence premature program stopping and engagement, although the strength and direction of association appear to vary by program type. Future studies will need to examine whether modifying message characteristics can optimize engagement and whether this leads to behavior change.

(JMIR Mhealth Uhealth 2021;9(11):e27779) doi:10.2196/27779

KEYWORDS

mHealth; machine learning; chronic disease; cardiovascular; text messaging; SMS; digital health; mobile phone; engagement; prevention

Introduction

Mobile health (mHealth) is increasingly being used to support individuals with chronic diseases in novel ways that leverage the mobility and capabilities of mobile phones [1]. Owing to the global ubiquity of mobile phones, the predominant perceived benefit of mHealth is the potential to rapidly scale and bridge geographical, financial, and cultural access barriers to health care and, thus, provide population health benefits. SMS text message–based interventions, in particular, may have greater potential for reaching lower-income groups and those with poor health compared with smartphone app–based interventions [2]. There is evidence suggesting that SMS text message–based interventions can result in improvement in multiple behavior-related risk factors, including smoking, physical activity, blood pressure, weight, and diabetes mellitus [3-7]. However, intervention effect sizes are modest, and the duration of the effect is uncertain [8].

Understanding how SMS text message–based program content affects participant replies may aid in optimizing future SMS text message–based programs. For example, SMS text message content that have been associated with premature program withdrawal can be avoided, and content that participants engage most with can be used more frequently. Engagement with mHealth programs has been considered an important factor in their effectiveness [9] and has been most commonly defined by frequency (ie, how often contact is made) and dropouts [10,11]. However, there are knowledge gaps, including how to maximize engagement with mHealth programs and whether mHealth content affects engagement. Incorporating instructional behavior change techniques within the mHealth program content, such as action plans [12] and setting goals [13], may help promote and sustain healthy lifestyle behaviors in patients with chronic diseases, including cardiovascular disease (CVD). The degree of interactivity (ie, 1-way vs 2-way flow of information) may also affect engagement [14], although, to date, it is poorly understood.

Our team has previously developed and supported patients with SMS text message–based mHealth programs who have chronic diseases, including CVD [15-18], diabetes [17], renal disease [19], and chronic obstructive pulmonary disease [18]. Across our programs, we developed a database of different SMS text messages and participant replies. The first step in understanding the complex interaction between mHealth content and participant replies is to categorize message content into *themes* (ie, groups of messages that share common features, such as reporting struggle, or common goals, such as motivation). To do this manually is time consuming, not practicable, and would limit scalability. As a solution, machine learning (ML) models can be trained to automatically categorize text [20], potentially saving time and resources, and is reproducible, improving the scalability of the program. In addition, this would allow contextualized assessment of participant replies (ie, assessment of replies in context with the program message participants are replying to) and engagement. Therefore, the aims of this research are (1) to develop ML models to categorize program SMS text messages and participant replies and (2) to examine whether message characteristics were associated with premature program stopping and engagement.

Methods

Study Population

We analyzed our combined communication logs from 3 Australian SMS text message–based digital health programs (SupportMe, ACTRN12616001689460 [17]; TEXTMEDS (Text Messages to Improve Medication Adherence and Secondary Prevention), ACTRN12613000793718 [16]; and ITM ACTRN12616001167459 [18]). In brief, with respect to the original studies (Table 1), SupportMe was a 6-month single-blinded, multicenter randomized controlled trial (RCT) for participants from community and hospital settings with a history of CVD or type 2 diabetes mellitus, and the primary outcome was systolic blood pressure at 6 months [17]. The TEXTMEDS study was a 12-month single-blinded, multicenter RCT delivered to patients following an acute coronary syndrome, with the primary end point being the percentage of patients who are adherent to cardioprotective medications [16]. ITM was a 6-month multicenter, single-blinded RCT targeting patients attending cardiac or pulmonary rehabilitation, with the primary outcome being exercise capacity, as measured by the 6-minute walk test [18]. For both SupportMe [17] and ITM [18], participants were not encouraged to reply to messages they received (ie, 1-way communication), although they were still able to reply, and replies were monitored. TEXTMEDS [16] encouraged replies from participants (ie, 2-way communication) and included the opportunity to liaise with a health counselor. Analysis of SupportMe and ITM were grouped together and referred throughout the manuscript as SupportMe/ITM, as both programs were similar (1-way communication not involving a health counselor) and recruited community and hospital participants with stable chronic conditions compared with TEXTMEDS (2-way with a health counselor and following a hospital admission for an acute cardiovascular event). Despite differences in the setting that participants were recruited from

(ie, community vs hospital), the programs were delivered during the stable chronic phase of the illness (ie, as an outpatient) to deliver appropriate secondary prevention recommendations. The TEXTMEDS and SupportMe studies had primary ethics approval from the Western Sydney Local Health District Human Research Ethics Committee (TEXTMEDS: HREC2012/12/4.1 (3648) AU RED HREC/13/WMEAD/15; SupportMe: AU RED HREC/16/ WMEAD/331). The ITM study received primary ethics approval from the Sydney Local Health District Hospital Human Research Ethics Committee and associated governance committees at the sites.

Table 1. SMS text message–based prevention programs for metabolic disease.

Project	Duration	2-way communication encouraged[a]	Population	Recruitment number				Number of replies
				Total	Lost to follow-up	Withdrawn consent	Deaths	
TEXTMEDS[b] [16]	12 months	Yes	CVD[c] (recruited from hospital post-ACS[d])	1424 (716 in intervention arm; 1:1 allocation)	39 (9 in intervention)	6 (1 in intervention)	15 (10 in intervention)	2356
ITM (support for patients with respiratory disease and CVD via integrated SMS text messaging) [18]	6 months	No	CVD and respiratory disease (recruited from community with one or more chronic conditions)	316 (236 in intervention arm, 80 in control arm; 3:1 allocation)	26 (22 in intervention)	19 (12 in intervention)	4 (3 in intervention)	417
SupportMe (SMS text messaging support for patients with chronic disease) [17]	6 months[e]	No	CVD and diabetes (recruited from community and hospital with one or more chronic conditions)	902 (454 in intervention arm; 1:1 allocation)	15 (9 in intervention)	7 (4 in intervention)	9 (5 in intervention)	1298

[a]Two-way communication was possible with all the included SMS text message–based programs but only encouraged for TEXTMEDS (Text Messages to Improve Medication Adherence and Secondary Prevention).

[b]TEXTMEDS: Text Messages to Improve Medication Adherence and Secondary Prevention.

[c]CVD: cardiovascular disease.

[d]ACS: acute coronary syndrome.

[e]A total of 7 patients in SupportMe at the conclusion of the 6-month intervention continued into a 6-month maintenance phase, which consisted of receiving texts at half the original frequency.

Developing ML Models to Characterize Text Messages

A total of 2 health professionals (HK and Anu Indrawansa, Westmead Hospital, Sydney, New South Wales, Australia) manually categorized all outgoing program SMS text messages in our SMS text message bank, all replies from SupportMe/ITM, and 829 TEXTMEDS replies.

Outgoing SMS text message intent (Textbox 1) was categorized as *informative* (provides health facts or education), *instructional* (provides tips or recommendations), *motivational* (provides feedback to encourage healthy behavior), *supportive* (provides contact details or referral to support groups or websites), and *notification* (notifies the patient about matters that are not educational, such as the welcome and exit messages). These message intent categories were chosen as they align with the dominant behavioral techniques used to develop the SMS text message bank, that is, provision of information about behavior health link and consequences (information-motivation-behavioral skills model), provision of instructions (social cognitive theory), provision of general encouragement (social cognitive theory), and relapse prevention (support to manage potential failure) [21].

Textbox 1. Example SMS text messages from each message intent category.

Informative

- "[person_name] by switching from full fat to low fat milk in tea & coffee you could remove 1 kg of saturated fat from your diet a year!" [TEXTMEDS; Text Messages to Improve Medication Adherence and Secondary Prevention]
- "There are many ways to increase your activity levels [person_name]. Try Tai Chi, pilates, gardening, yoga or dancing" [ITM]
- "Did you know a blood test called HbA1c measures your average blood sugar over the last 3 months? Ask your doctor for a check every 3-6 months" [SupportMe]

Instructional

- "Cardiac drugs are safe but if you have any side effects discuss with your Dr - there are many medication options [person_name]." [TEXTMEDS]
- "If you are feeling more breathless than usual, try to relax, rest and practice your breathing techniques" [ITM]
- "[person_name], use up vegies by mixing them with herbs, spices & water to cook up a hearty soup" [SupportMe]

Supportive

- "Are you having a good week [person_name]? Just reminding you that you can text us if we can be of help" [TEXTMEDS]
- "Staying calm when you are breathless really helps. Is there someone in your household who can help you stay calm when you feel uptight?" [ITM]
- "Hi [person_name], you may need extra carbohydrates before, during, or after exercise to prevent low blood sugars - discuss with your healthcare team" [SupportMe]

Motivational

- "Dont worry [person_name] if you have a bad day. Remember that there is another chance tomorrow to choose the healthy option." [TEXTMEDS]
- "Hi [person_name], when you are quitting smoking - if you have a bad day, don't worry & keep trying" [ITM]
- "Hi [person_name], did you exercise today?" [SupportMe]

Notification

- "Hi [person_name], you are now halfway through TEXTMEDS. Soon we will ring to check how you are, but don't tell us you have been receiving messages" [TEXTMEDS]
- "Hi [person_name], welcome to the ITM study. We hope you enjoy the messages. Respond STOP to opt out" [ITM]
- "Hi [person_name], welcome to the SupportMe study. You are in the group that will not receive regular messages. We will contact you at 6 months" [SupportMe]

Participant replies were categorized as follows: *stop* (replies indicating participants wish to stop receiving further messages), *thanks* (replies showing appreciation), *question* (replies prompting a response from the health counselor or research team), *reporting healthy* (replies indicating participants are complying with health recommendations), *reporting struggle* (replies indicating difficulties with complying with health recommendations), *general comment* (general replies regarding the program or health), and *other* (replies not fitting the above categories or not study related, for example, blank messages, inadvertent replies, and invalid numbers).

If an SMS text message could belong to 2 different categories or *intents* (eg, *informative* and *instructional*), the majority category or *intent* was favored.

Before developing the ML models, the outgoing messages were grouped into *clusters* of duplicate and highly similar messages in an attempt to support the automatic classification of related and novel future messages (Lancaster Stemmer method [22]). The similarity between 2 messages was measured by calculating the proportion of tokens (eg, words or punctuations) that were common between them (Jaccard similarity) [23]. We used 0.5 as the similarity threshold (ie, any 2 texts with a Jaccard similarity score >0.5 were put into the same cluster). Each ML model was tasked with using only the given *input features* to classify each item into a single best *output category.* For the program message (ie, outgoing messages) predictive model, the input features were the SMS text message, and the output category was the message intent. For the participant replies, the input features were the reply message (DistilBERT embeddings as detailed below) and the message intent of the outgoing message (1-hot encoded), and the output categories were the participant reply categories listed above.

Each ML model was created using a DistilBERT model pretrained on the Toronto Book Corpus and full English Wikipedia to encode word meaning and sentence structure (*distilbert-base-uncased*) [24]. We then applied an L2-regularized logistic regression model to weigh the 768 real-valued features produced by DistilBERT, generating a classification. We used grid search with 10-fold cross-validation to optimize the model hyperparameters (*inner cross-validation*). The selected hyperparameters were those that maximized balanced accuracy, which is defined as the mean sensitivity across all classes. For evaluation, we performed this procedure

within a 10-fold cross-validation (*outer cross-validation*) using a technique known as nested cross-validation. Specifically, for each fold, 10% of all data were held out for testing, and within each of these, a 10-fold cross-validation was applied again with a similar 90% training to 10% validation split to select the best hyperparameters. To avoid biases in evaluation, we constrained the sampling of held out data sets such that near-identical outgoing messages (ie, those in the same Jaccard similarity cluster) could not appear in both training and test data set intent classifications. Similarly, for reply classification, we controlled for idiosyncratic language in the cross-validation procedure by ensuring that no participant replies appeared in both the training set and its corresponding test set.

Associations of Message Characteristics to Premature Program Stopping and Engagement

We assessed associations using univariate logistic regression between outgoing message characteristics (outgoing message intents: informative, instructional, motivational, supportive, and notification) to the outcomes (1) reply type *stop* and (2) engagement. Engagement is difficult to define in the setting of SMS text message–based prevention programs; however, the frequency of participant replies does give a quantifiable measure of engagement as it reflects that the participant is repeatedly interacting with the program. Thus, engagement in this study was defined as a patient who replied at least 3 times and did not prematurely withdraw from the study or stop the intervention. An interaction analysis was performed to assess whether the associations were affected by program type (SupportMe/ITM vs TEXTMEDS) as the programs encouraged different levels of interaction (ie, 1-way vs 2-way communication; Table 1). To do this, univariate logistic regression models (between outgoing message characteristics to the outcomes as described above) were used, incorporating program type (SupportMe/ITM and TEXTMEDS) as the interaction term.

Statistical Analysis

ML model accuracy was assessed using balanced accuracy, the area under the receiver operating characteristic (ROC) curve, and multiclass classification evaluators. This was done using the Scientific Python stack (Scikit-learn 0.22, Pandas 1.1, and Matplotlib 3.3) on Python 3.7 (Python Software Foundation). Associations of message characteristics to program termination and engagement were examined using SPSS Statistics (version 26.0; SPSS Inc). Chi-square tests of independence were performed to determine associations between the ML-derived program message intent and the outcomes (1) reply type *stop* or (2) engagement. Two-sided $P<.05$ were considered statistically significant unless otherwise stated.

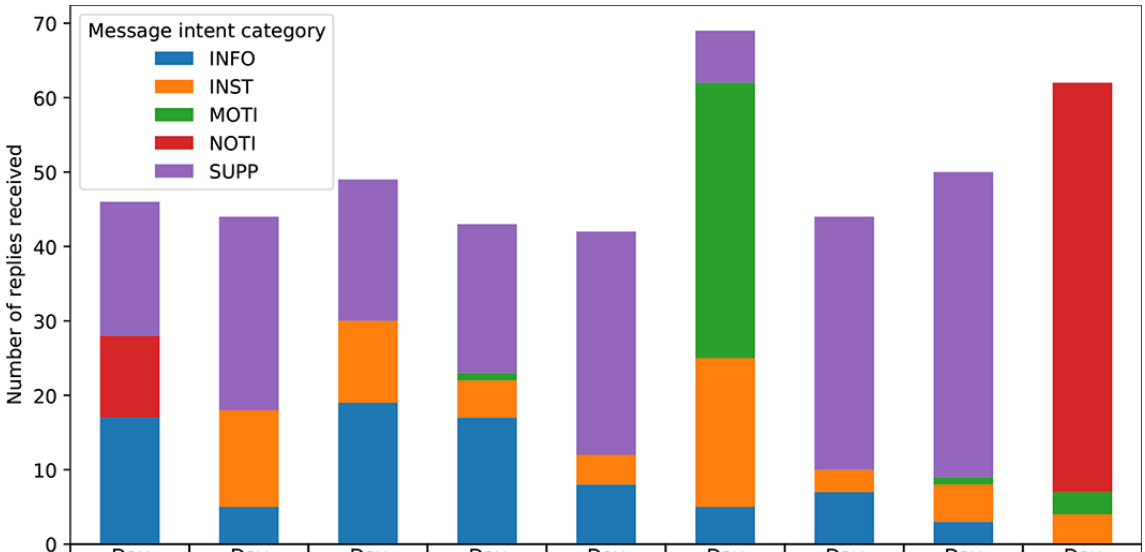
Results

Descriptive Analyses of Program Messages and Replies

We analyzed a total of 1550 program messages and 4071 participant replies. The total number of patients in each group for each study and the received responses are shown in Table 1. Approximately, 30.01% (793/2642) of participants replied at least once, and 5.83% (154/2642) sent a *stop* reply type to opt out of future messages. For SupportMe/ITM, 32.35% (394/1218) participants replied to at least one message (384/394, 97.5% were from the intervention arm and 10/394, 2.5% were from the control arm). Of these, 20.8% (82/394) sent a *stop* reply type, of which 7 were from the control group. For TEXTMEDS, 28.02% (399/1424) of participants replied to at least one message (all were from the intervention arm), of which 18% (72/399) sent a *stop* reply type. For TEXTMEDS, most replies were in response to *supportive* messages (794/2356, 33.7% of all replies), and for SupportMe/ITM, it was in response to *informative* messages (661/1715, 38.54% of all replies; Multimedia Appendix 1). The majority reply type was *general comment* for both TEXTMEDS (635/2356, 26.95%) and SupportMe/ITM (574/1715, 33.47%; Multimedia Appendix 2). The types of responses elicited by each message intent for SupportMe/ITM and TEXTMEDS are detailed in the Multimedia Appendices 3 and 4.

Approximately 11.43% (302/2642) of participants met our definition of being engaged. For SupportMe/ITM, 8.05% (98/1218) of participants were engaged and contributed to 78.6% (1348/1715) of the total replies in SupportMe/ITM. For TEXTMEDS, 14.33% (204/1424) of participants engaged with the program and contributed to 89.13% (2100/2356) of the total replies in TEXTMEDS. Most replies during TEXTMEDS were received during the middle of the program, with the least number shouldering this period in response to motivational and instructional message intents (Figure 1). In contrast, most replies during SupportMe/ITM were received at the beginning of the program (first 50 days) and in response to instructional and informative message intents (data not shown).

Figure 1. Distribution of participant reply categories by program message intent during the 12-month TEXTMEDS (Text Messages to Improve Medication Adherence and Secondary Prevention) program. The 9 peak periods (4-day duration each) were defined as those which received >40 replies within each peak period. INFO: Informative; INST: Instructional; MOTI: Motivational; NOTI: Notification; TEXTMEDS: Text Messages to Improve Medication Adherence and Secondary Prevention; SUPP: Supportive.



	Day 0-3	Day 28-31	Day 56-59	Day 112-115	Day 116-119	Day 196-199	Day 260-263	Day 344-347	Day 372-375
General comment	10	10	16	9	7	37	10	16	4
Question	4	10	3	2	2	3	3	7	2
Reporting healthy	6	2	7	17	20	17	15	17	2
Reporting struggle	4	7	7	3	8	2	4	4	2
Stop	3	1	2	1	0	0	1	0	1
Thanks	15	12	10	7	5	7	10	2	50
Other	4	2	4	4	1	3	1	4	1

ML Models to Classify Outgoing Message Intent and Reply Type

Table 2 shows the classification report for the program messages according to message intent. Altogether, the ML model correctly classified the intent of program messages as 76.6% (95% CI 63.5%-89.8%; balanced accuracy) of the time (Table 2). Average specificity was 93.2%, positive predictive value 76.3%, and negative predictive value 93.4%. The average area under the curve from the ROC curves was 0.95 (Figure 2). Sensitivity was lowest for the *supportive* message intent (69.7%), and specificity was lowest for the *informative* message intent (86.8%).

Table 2. Machine learning performance for program message intent.

Message intent	Sensitivity (SD)	Specificity (SD)	PPV[a] (SD)	NPV[b] (SD)	FPR[c] (SD)	FNR[d] (SD)	F1-score (SD)
INFO[e]	0.797 (0.144)	0.868 (0.072)	0.850 (0.070)	0.840 (0.099)	0.132 (0.072)	0.203 (0.144)	0.815 (0.089)
INST[f]	0.761 (0.169)	0.885 (0.093)	0.795 (0.124)	0.887 (0.064)	0.115 (0.093)	0.239 (0.169)	0.759 (0.118)
MOTI[g]	0.778 (0.242)	0.968 (0.033)	0.671 (0.248)	0.986 (0.016)	0.032 (0.033)	0.222 (0.242)	0.702 (0.221)
NOTI[h]	0.800 (0.400)	0.999 (0.002)	0.900 (0.300)	0.994 (0.015)	0.001 (0.002)	0.200 (0.400)	1.000 (0.000)
SUPP[i]	0.697 (0.296)	0.940 (0.046)	0.635 (0.251)	0.962 (0.036)	0.060 (0.046)	0.303 (0.296)	0.741 (0.138)
Average[j]	0.766 (0.175)	0.932 (0.027)	0.763 (0.148)	0.934 (0.027)	0.068 (0.027)	0.234 (0.175)	0.782 (0.100)

[a]PPV: positive predictive value.
[b]NPV: negative predictive value.
[c]FPR: false positive rate.
[d]FNR: false negative rate.
[e]INFO: Informative.
[f]INST: Instructional.
[g]MOTI: Motivational.
[h]NOTI: Notification.
[i]SUPP: Supportive.
[j]Macroaveraged.

Figure 2. Receiver operating characteristic curves for predicting program message intent. Generated under one-vs-rest assumption (ie, each curve is generated assuming a binary scenario with the selected class against all other classes). AUC: area under the curve; INFO: Informative; INST: Instructional; MOTI: Motivational; NOTI: Notification; SUPP: Supportive.

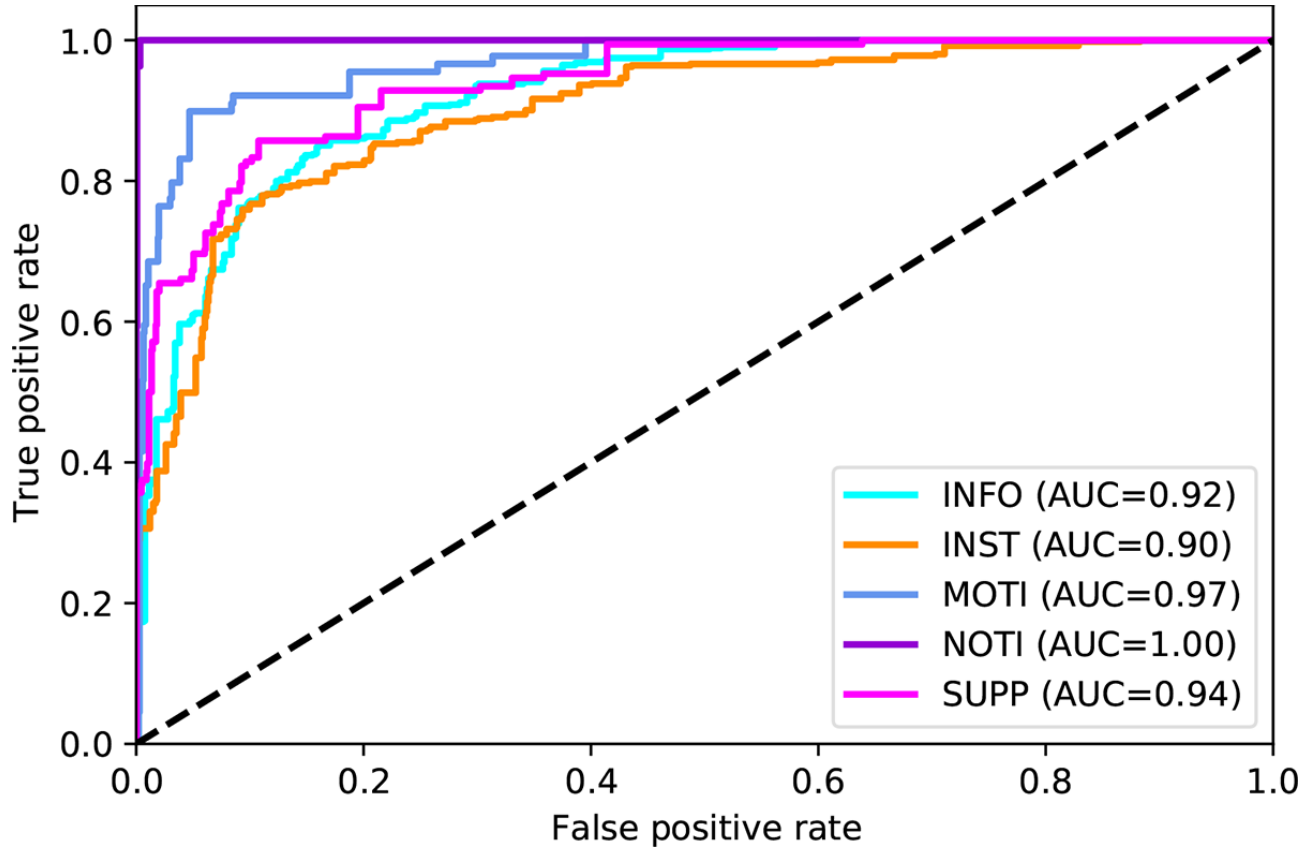


Table 3 shows the classification report for the participant replies. Altogether, the ML model correctly categorized the replies with 77.8% (95% CI 74.1%-81.4%) balanced accuracy (Table 3). Average specificity was 95.7%, positive predictive value 72.6%, and negative predictive value 95.5%. Sensitivity was lowest with *reporting struggle* category (64.9%), and specificity was lowest with the *general comment* category (89.3%). The average area under the curve from the ROC curves was 0.96 (Figure 3).

Table 3. Machine learning performance for participant reply categories.

Participant replies	Sensitivity (SD)	Specificity (SD)	PPV[a] (SD)	NPV[b] (SD)	FPR[c] (SD)	FNR[d] (SD)	F1-score (SD)
General comment	0.684 (0.121)	0.893 (0.055)	0.817 (0.074)	0.815 (0.073)	0.107 (0.055)	0.316 (0.121)	0.737 (0.079)
Thanks	0.911 (0.050)	0.959 (0.026)	0.863 (0.090)	0.972 (0.027)	0.041 (0.026)	0.089 (0.050)	0.771 (0.099)
Question	0.815 (0.213)	0.976 (0.014)	0.474 (0.157)	0.995 (0.007)	0.024 (0.014)	0.185 (0.213)	0.592 (0.174)
Reporting healthy	0.707 (0.097)	0.940 (0.037)	0.601 (0.198)	0.960 (0.040)	0.060 (0.037)	0.293 (0.097)	0.623 (0.111)
Reporting struggle	0.649 (0.167)	0.979 (0.012)	0.696 (0.136)	0.976 (0.013)	0.021 (0.012)	0.351 (0.167)	0.658 (0.106)
Stop	0.860 (0.147)	0.993 (0.008)	0.888 (0.131)	0.992 (0.009)	0.007 (0.008)	0.140 (0.147)	0.866 (0.116)
Other	0.818 (0.082)	0.956 (0.029)	0.740 (0.132)	0.972 (0.016)	0.044 (0.029)	0.182 (0.082)	0.885 (0.065)
Average[e]	0.778 (0.048)	0.957 (0.012)	0.726 (0.071)	0.955 (0.013)	0.043 (0.012)	0.222 (0.048)	0.733 (0.054)

[a]PPV: positive predictive value.

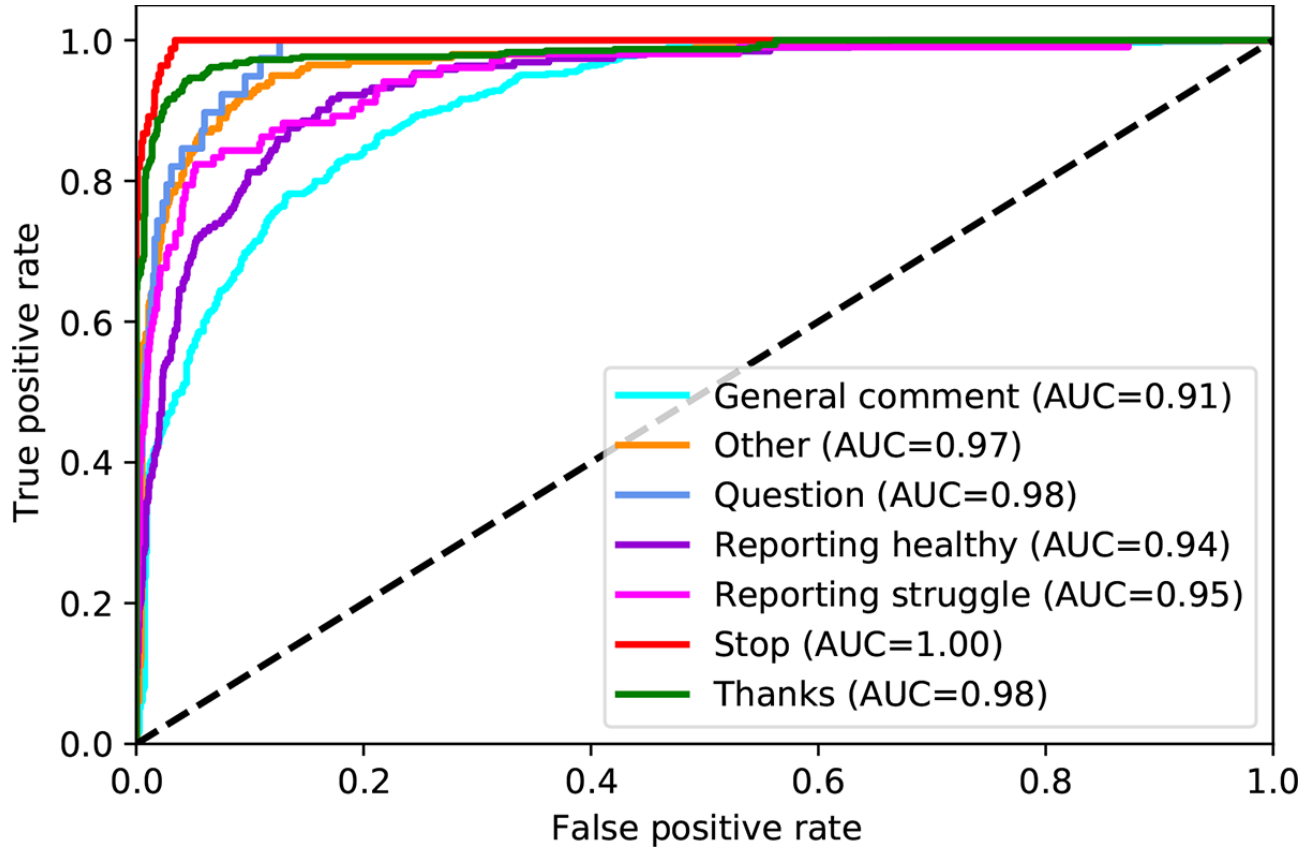
[b]NPV: negative predictive value.

[c]FPR: false positive rate.

[d]FNR: false negative rate.

[e]Macroaveraged.

Figure 3. Receiver operating characteristic curves for predicting participant reply type. Generated under one-vs-rest assumption (ie, each curve is generated assuming a binary scenario with the selected category against all other categories). AUC: area under the curve.



Associations of Message Characteristics to Premature Program Stopping and Engagement

Overall, *supportive* message intent was associated with a reduced chance of premature program stopping (odds ratio [OR] 0.53, 95% CI 0.35-0.81; *P*=.003). For SupportMe/ITM, participants were less likely to reply *stop* following *informative* messages (OR 0.35, 95% CI 0.20-0.60; *P*<.001) and more likely to reply *stop* following *notification* messages (OR 5.76, 95% CI 3.66-9.06; *P*<.001). This differed from TEXTMEDS, where there was no significant association with *informative* (OR 1.25, 95% CI 0.78-2.02; *P*=.35) and *notification* messages (OR 1.89, 95% CI 0.95-3.74; *P*=.07; Table 4).

Table 4. Univariate logistic regression with program message intent and outcome variables (premature program stopping and engagement).

Message intent, outcome variables	Total			SupportMe/ITM			TEXTMEDS[a]			*P* value for interaction[b]
	OR[c] (95% CI)	β coefficient	*P* value	OR (95% CI)	β coefficient	*P* value	OR (96% CI)	β coefficient	*P* value	
Premature program stopping (ie, reply type "stop")										
INFO[d]	0.69 (0.49-0.98)	–0.37	.04	0.35 (0.20-0.60)	–1.05	<.001	1.25 (0.78-2.02)	0.23	.35	<.001
INST[e]	0.98 (0.69-1.40)	–0.02	.93	0.86 (0.54-1.38)	–0.15	.54	1.03 (0.60-1.76)	0.03	.92	.63
SUPP[f]	0.53 (0.35-0.81)	–0.64	.003	0.60 (0.29-1.26)	–0.51	.18	0.58 (0.34-0.99)	–0.54	.05	.94
MOTI[g]	1.00 (0.52-1.91)	–0.00	.99	1.08 (0.43-2.73)	0.08	.87	0.98 (0.39-2.47)	–0.02	.97	.89
NOTI[h]	4.01 (2.80-5.75)	1.39	<.001	5.76 (3.66-9.06)	1.75	<.001	1.89 (0.95-3.74)	0.64	.07	.01
Engagement (ie, 3 replies and did not prematurely stop the program)										
INFO	1.76 (1.46-2.12)	0.56	<.001	2.16 (1.67-2.78)	0.77	<.001	1.62 (1.21-2.16)	0.48	<.001	.14
INST	1.47 (1.21-1.80)	0.39	<.001	1.68 (1.29-2.18)	0.52	<.001	1.51 (1.10-2.08)	0.42	.01	.63
SUPP	1.22 (1.01-1.49)	0.20	.04	1.77 (1.21-2.58)	0.57	.003	0.77 (0.60-0.98)	–0.27	.04	<.001
MOTI	1.18 (0.82-1.70)	0.17	.37	0.82 (0.50-1.34)	–0.20	.43	1.64 (0.92-2.93)	0.50	.10	.07
NOTI	0.14 (0.11-0.17)	–1.99	<.001	0.07 (0.05-0.10)	–2.61	<.001	0.28 (0.20-0.39)	–1.28	<.001	<.001

[a]TEXTMEDS: Text Messages to Improve Medication Adherence and Secondary Prevention.

[b]*P* interaction refers to the comparison of the associations between each message intent with program type (SupportMe/ITM vs TEXTMEDS) and adjusted for message intent.

[c]OR: odds ratio.

[d]INFO: Informative.

[e]INST: Instructional.

[f]SUPP: Supportive.

[g]MOTI: Motivational.

[h]NOTI: Notification.

Overall, *informative* (OR 1.76, 95% CI 1.46-2.12; *P*<.001) and *instructional* (OR 1.47, 95% CI 1.21-1.80; *P*<.001) message intents were associated with increased engagement but not *motivational* (OR 1.18, 95% CI 0.82-1.70; *P*=.37) message intent (Table 4). *Supportive* message intent was associated with increased engagement in SupportMe/ITM (OR 1.77, 95% CI 1.21-2.58; *P*=.003) but reduced engagement in TEXTMEDS (OR 0.77, 95% CI 0.60-0.98; *P*=.04; for interaction, *P*<.001). *Notification* messages were associated with reduced engagement in both SupportMe/ITM (OR 0.07, 95% CI 0.05-0.10; *P*<.001) and TEXTMEDS (OR 0.28, 95% CI 0.20-0.39; *P*<.001); however, the strength of the association was greater for SupportMe/ITM (for interaction, *P*<.001).

Discussion

Principal Findings

In this study, ML models were created to categorize program message intent and participant replies from SMS text message–based programs with good accuracy. Thus, they can enable the monitoring and detailed characterization of program messages and participant replies, which can be used to further customize SMS text message–based programs. Furthermore, this study found that program message type can influence premature program discontinuation and encourage participant engagement. However, some of these associations varied or were attenuated by program type. This suggests that participant engagement may be maximized by adjusting program message characteristics, that program type and patient population type are important to consider, and that larger studies to examine these interactions will enable further program refinement.

We have previously assessed the accuracy of ML to *triage* participant text replies in an attempt to identify those requiring health professional review with only 1.4% false negatives and categorized replies according to *themes* with modest accuracy (0.723 weighted average accuracy) [20]. However, unlike this study, this was done without knowledge of the message sent that the participant is replying to or its intent. This limits the accuracy and contextualization of the reply modeling, and thus no conclusion regarding the characteristics of texts that elicit replies can be drawn. In this study, by contextualizing the participant replies to the program messages, we allowed a detailed assessment of which program message characteristics are likely to elicit different types of replies. Furthermore, the use of a *reporting struggle* category in this study, in addition to a *reporting healthy* category, which has been contextualized with the program message intent, allows the automated identification of participants who may need more or less tailored support, respectively.

A recent review of systematic reviews identified 3 reviews and 10 studies (clinical trials and feasibility studies) that measured engagement [25]. Of the included studies, only 1 quantified engagement with an SMS text message–based intervention, and the measure of engagement was the frequency of replies to reminders for blood glucose measurement [25]. There are no SMS text message–based studies reporting engagement in a population with CVD. Smartphone- and internet-based programs were the most common mHealth types in the studies assessing engagement, and although there was variability in the measure of engagement used, most measures focused on the way users interact with the program (eg, number of visits to a web-based program or frequency of responses). In this study, participants were more likely to engage with *informative* and *instructional* program message intents. These message intents use the behavior change techniques of goal setting and self-monitoring, both of which have been associated with higher engagement with digital health interventions (DHIs) [12,26]. Importantly, the pattern of participant interaction throughout the course of a DHI is dynamic and involves different levels of use over time [27]. Highlighting this in our study, the type of replies and message intent prompting replies changed throughout the duration of the TEXTMEDS program (Figure 1). In contrast to our study, a mixed methods study assessed the utility of either email or text prompts in encouraging engagement (as measured by the number of log-ins) in a web-based intervention for diabetes and found that email prompts increased engagement but not text prompts, and this was affected by email content [28]. However, over a 15-month period, only 7 text prompts were sent compared with 49 email prompts, which may explain the lack of significance.

Features enabling remote contact with a health care professional can positively influence engagement with DHIs [12,29,30], as can interactivity [31]. In this study, the associations between some message intents and premature program stopping or engagement differed by program type. For instance, with TEXTMEDS, *notification* messages were not associated with premature program termination, unlike SupportMe/ITM. In addition, the odds of *notification* messages decreasing engagement, although being elevated in both programs, were significantly higher in SupportMe/ITM than in TEXTMEDS (Table 4). These differences may be related to the encouraged flow of information—2-way communication was encouraged with TEXTMEDS (and with health counselor support) and not encouraged with SupportMe/ITM. Supporting this, Redfern et al [32] demonstrated, using qualitative methods, including reviewing user surveys and focus groups, that support felt from program participation and from health staff are important factors that influence engagement. Differences in patient populations may also explain the interaction effects with program type. Participants in SupportMe/ITM had stable chronic diseases compared with TEXTMEDS, where participants were recruited following an acute coronary syndrome, and it is reasonable to suspect that participants who were recruited in the context of symptoms may engage differently compared with participants without symptoms. This is consistent with Redfern et al [32], who also determined that initiating the mHealth program close to the time of a cardiovascular event was associated with increased engagement.

Unexpectedly, *supportive* messages were associated with decreased engagement in TEXTMEDS but increased engagement in SupportMe/ITM. However, the highest proportion of responses overall were elicited following *supportive* messages in TEXTMEDS and *informative* messages in SupportMe/ITM (Multimedia Appendix 1). Furthermore, TEXTMEDS was twice as long as SupportMe/ITM, and most of the responses received were in the first half of the program (Figure 1). Thus, it is possible that engagement decreased in the second half of the program and differentially affected supportive message intent compared with the others, as a larger number of similar messages may have been sent to participants and considered repetitive over the longer program duration.

Clinical Implications

The results of this study contribute to the field of SMS text message–based interventions by (1) demonstrating that using ML can automatically and accurately categorize SMS text messages sent to and from participants in an SMS text message–based program to support their health, and (2) providing new knowledge on how participants engage with

SMS text messages and factors associated with engagement and premature program termination. Overall, our ML models for characterizing program message intent and user replies enable the ability to monitor and describe the way participants interact with different SMS text message–based prevention programs. This has implications for the optimized development of future SMS text message–based programs, as the results suggest that participant engagement may be maximized (and premature program termination avoided) by adjusting message characteristics, that is, the clinical implications of message content affecting participant withdrawal and engagement are the potential of using this knowledge to alter future messages automatically in real time to sustain engagement throughout the intervention duration. This could minimize participant withdrawal and maximize the likelihood of behavior change. This has not been assessed in previous studies, and a lack of knowledge of participant-program interactions has limited the utility of existing SMS text message–based programs.

As there may be differences in the degree of interactivity encouraged (ie, 1-way vs 2-way communication), when assessing engagement and premature program stopping, we performed an interaction analysis (in addition to analyzing the programs separately and combined) to assess if the associations were affected by program type (Table 4). From these results, it is clear that program type can affect engagement. This is an important finding as there is large heterogeneity in existing SMS text message–based programs with respect to the degree of interactivity allowed, and recognizing that participants may interact differently with 1-way versus 2-way programs also potentially informs different approaches to optimizing engagement, depending on the program type.

Validation of these models with different SMS text message–based programs delivered to different population groups (different clinical and geographical settings across high-, middle-, and low-income countries) would assist in increasing generalizability and utility. Future research should explore the association between program engagement and intervention success or behavioral change. In addition, there is a need to determine whether modifying message characteristics can maximize participant engagement and whether this can lead to sustained behavior change.

Limitations

Although manual categorization was done by health professionals, it is entirely possible that manual categorization may have differed if performed by a different group and, thus, affected the final ML models. In addition, some of the SMS text messages could be categorized into more than one category, which can also affect the final model; however, to minimize bias, we selected the majority category within each group of similar messages. Although there were differences in the populations between studies (Table 1), message content was similar between programs, allowing the clustering of messages into similar categories as described in the methods. Furthermore, although there was good overall accuracy with the ML models, the number of program messages and participant replies was relatively small for ML modeling, and the accuracy could be further improved by using a larger sample. This emphasizes the importance of validating these models against larger data sets.

As discussed in the *Methods* section, it was necessary to combine all studies to develop ML models to increase the accuracy of the ML models. Although there were differences in the clinical setting that the patients across the studies were recruited in (TEXTMEDS from hospital, ITM from community, and SupportMe from community and hospital), all patients had chronic diseases and were managed in the chronic phase of their illness (if recruited from the hospital, the messages were not initiated until the patient was discharged into the community). Thus, we do not believe that this would be a significant confounding variable. However, as discussed above, patients in TEXTMEDS were recruited following a symptomatic episode, which can affect engagement [32]. In addition, although all studies recruited patients with CVD, 2 of them (SupportMe and ITM) recruited patients with diabetes and chronic respiratory disease, and it is possible that the type of chronic disease influences engagement. Although all patients were recruited in the chronic phase of their illness, illness severity was not assessed. Although not previously assessed in the context of mHealth, there is survey evidence that severity of chronic diseases can affect perceived health attitudes [33]; thus, if applied to mHealth, it is possible that illness severity can affect engagement. Thus, the validation of these models against different types of chronic illnesses in future research is prudent.

This study compared a measure of engagement between 2 SMS text message–based programs that differ in the encouraged level of interaction (ie, 2-way vs 1-way communication). Thus, it is possible that participants replied depending on whether they were encouraged or not, and it is also possible that participants who did not engage with messages using our definition engaged with behavior change. However, almost one-third of the participants in SupportMe/ITM replied compared with one-fifth in TEXTMEDS, and, thus, using the frequency of replies (and excluding withdrawals) as a surrogate marker of engagement is a reasonable method of quantifying engagement, which is consistent with previous studies [25]. It is interesting that the proportion of participants who replied in SupportMe/ITM (1-way) is comparable with TEXTMEDS (2-way). In both types of programs, all messages were monitored by a health professional, and the patients were informed of this. Thus, although patients were not encouraged to reply by recruitment staff, we believe they did as they felt their replies would still be seen (even if they were not answered). Many of the texts (as seen in Textbox 1) would ask questions regardless of whether the program was 1-way or 2-way, and although these were meant to be rhetorical, participants replied. The major benefit of our definition of engagement is that we provide a simple quantifiable marker of engagement that can potentially be used across different SMS text message–based programs, allowing comparisons between studies. As TEXTMEDS was twice as long as SupportMe/ITM, participants had more opportunities to become engaged according to our definition, and this may explain the numerically higher proportion of participants who were engaged in TEXTMEDS compared with SupportMe/ITM (204/1424, 14.33% vs 98/1218, 8.05%).

The association between participant sociodemographics and engagement with DHIs and behavior change was outside the scope of this study and should be explored in future research. The timing of messages sent and message length may affect engagement but was outside the scope of this study and will be assessed in future research.

Conclusions

In our study, using ML, we categorized outgoing and incoming messages from different SMS text message–based programs to support people with chronic diseases with good accuracy, enabling monitoring and detailed characterization of program messages and participant replies. Message intent can influence adherence (ie, not stopping the intervention) and participant engagement, although we suspect this association is affected by the type of interaction encouraged (ie, 1-way vs 2-way communication) and possibly the setting in which participants are recruited. The clinical implications include optimization of future SMS text message–based programs by using program message characteristics that maximize participant engagement and potentially behavior change.

Acknowledgments

The authors wish to acknowledge the Sydney Informatics Hub at the University of Sydney for running the machine learning models, Tara Czinner for assisting with developing the project scope with the Sydney Informatics Hub, and Anu Indrawansa for assisting with managing the SMS text message bank database. HK is supported by the Royal Australian College of Physicians Fellows Research Entry Scholarship. JR is funded by the National Health and Medical Research Council (NHMRC) Career Development Fellowship (APP1143538). CKC is supported by the National Health and Medical Research Council Career Development Award (APP1105447), cofunded by a Future Leader Fellowship from the National Heart Foundation. The TEXTMEDS study was funded by an NHMRC project grant (ID APP1042290). The SupportMe study was funded by the Translational Research Grants Scheme of NSW Health. The ITM study was funded by the National Heart Foundation, NSW Cardiovascular Network Research Development project grant (101134).

Authors' Contributions

HK and CKC contributed to the conception and study design. HK, CKC, AT, JR, and NWC contributed to the acquisition of raw data. HK, JN, DL, CS, and CKC contributed to the analysis and interpretation of data. HK and CKC drafted the manuscript. All authors critically reviewed the manuscript and gave final approval.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution of participant replies for each program message intent for TEXTMEDS (Text Messages to Improve Medication Adherence and Secondary Prevention) and SupportMe/ITM.
[PNG File , 28 KB - mhealth_v9i11e27779_app1.png]

Multimedia Appendix 2
Distribution of response categories for TEXTMEDS (Text Messages to Improve Medication Adherence and Secondary Prevention) and SupportMe/ITM.
[PNG File , 36 KB - mhealth_v9i11e27779_app2.png]

Multimedia Appendix 3
Participant reply category by program message intent for SupportMe/ITM.
[DOCX File , 58 KB - mhealth_v9i11e27779_app3.docx]

Multimedia Appendix 4
Participant reply category by program message intent for TEXTMEDS (Text Messages to Improve Medication Adherence and Secondary Prevention).
[DOCX File , 58 KB - mhealth_v9i11e27779_app4.docx]

References

1. Klimis H, Thakkar J, Chow CK. Breaking barriers: mobile health interventions for cardiovascular disease. Can J Cardiol 2018 Jul;34(7):905-913. [doi: 10.1016/j.cjca.2018.02.012] [Medline: 29887218]
2. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]

3. Klimis H, Khan ME, Kok C, Chow CK. The role of text messaging in cardiovascular risk factor optimization. Curr Cardiol Rep 2017 Jan;19(1):4. [doi: 10.1007/s11886-017-0811-8] [Medline: 28102482]
4. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: a meta-analysis. J Subst Abuse Treat 2015 Sep;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]
5. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. J Am Med Assoc 2015 Sep;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
6. Bobrow K, Farmer AJ, Springer D, Shanyinde M, Yu L, Brennan T, et al. Mobile phone text messages to support treatment adherence in adults with high blood pressure (SMS-text adherence support [StAR]): a single-blind, randomized trial. Circulation 2016 Feb 09;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]
7. Ricci-Cabello I, Bobrow K, Islam SM, Chow CK, Maddison R, Whittaker R, et al. Examining development processes for text messaging interventions to prevent cardiovascular disease: systematic literature review. JMIR Mhealth Uhealth 2019 Mar 29;7(3):e12191 [FREE Full text] [doi: 10.2196/12191] [Medline: 30924790]
8. Islam SM, Farmer AJ, Bobrow K, Maddison R, Whittaker R, Pfaeffli Dale LA, et al. Mobile phone text-messaging interventions aimed to prevent cardiovascular diseases (Text2PreventCVD): systematic review and individual patient data meta-analysis. Open Heart 2019;6(2):e001017 [FREE Full text] [doi: 10.1136/openhrt-2019-001017] [Medline: 31673381]
9. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
10. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
11. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
12. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011 Jan 25;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]
13. Kristjansdottir OB, Børøsund E, Westeng M, Ruland C, Stenberg U, Zangi HA, et al. Mobile app to help people with chronic illness reflect on their strengths: formative evaluation and usability testing. JMIR Form Res 2020 Mar 04;4(3):e16831 [FREE Full text] [doi: 10.2196/16831] [Medline: 32130126]
14. Park LG, Beatty A, Stafford Z, Whooley MA. Mobile phone interventions for the secondary prevention of cardiovascular disease. Prog Cardiovasc Dis 2016;58(6):639-650 [FREE Full text] [doi: 10.1016/j.pcad.2016.03.002] [Medline: 27001245]
15. Chow CK, Redfern J, Thiagalingam A, Jan S, Whittaker R, Hackett M, et al. Design and rationale of the tobacco, exercise and diet messages (TEXT ME) trial of a text message-based intervention for ongoing prevention of cardiovascular disease in people with coronary disease: a randomised controlled trial protocol. BMJ Open 2012 Jan 19;2(1):e000606 [FREE Full text] [doi: 10.1136/bmjopen-2011-000606] [Medline: 22267690]
16. Chow CK, Thiagalingam A, Santo K, Kok C, Thakkar J, Stepien S, et al. TEXT messages to improve MEDication adherence and Secondary prevention (TEXTMEDS) after acute coronary syndrome: a randomised clinical trial protocol. BMJ Open 2018 Jan 27;8(1):e019463 [FREE Full text] [doi: 10.1136/bmjopen-2017-019463] [Medline: 29374674]
17. Cheung NW, Redfern J, Thiagalingam A, Hng T, Islam SM, Haider R, SupportMe Investigators. Text messaging support for patients with diabetes or coronary artery disease (SupportMe): protocol for a pragmatic randomised controlled trial. BMJ Open 2019 Jun 19;9(6):e025923 [FREE Full text] [doi: 10.1136/bmjopen-2018-025923] [Medline: 31221870]
18. Redfern J, Hyun K, Singleton A, Hafiz N, Raeside R, Spencer L, et al. ITM support for patients with chronic respiratory and cardiovascular diseases: a protocol for a randomised controlled trial. BMJ Open 2019 Mar 01;9(3):e023863 [FREE Full text] [doi: 10.1136/bmjopen-2018-023863] [Medline: 30826759]
19. Stevenson J, Campbell KL, Brown M, Craig J, Howard K, Howell M, et al. Targeted, structured text messaging to improve dietary and lifestyle behaviours for people on maintenance haemodialysis (KIDNEYTEXT): study protocol for a randomised controlled trial. BMJ Open 2019 May 05;9(5):e023545 [FREE Full text] [doi: 10.1136/bmjopen-2018-023545] [Medline: 31061013]
20. Lowres N, Duckworth A, Redfern J, Thiagalingam A, Chow CK. Use of a machine learning program to correctly triage incoming text messaging replies from a cardiovascular text-based secondary prevention program: feasibility study. JMIR Mhealth Uhealth 2020 Jun 16;8(6):e19200 [FREE Full text] [doi: 10.2196/19200] [Medline: 32543439]
21. Redfern J, Thiagalingam A, Jan S, Whittaker R, Hackett ML, Mooney J, et al. Development of a set of mobile phone text messages designed for prevention of recurrent cardiovascular events. Eur J Prev Cardiol 2014 Apr;21(4):492-499. [doi: 10.1177/2047487312449416] [Medline: 22605787]
22. Paice CD. Another stemmer. ACM SIGIR Forum 1990 Nov;24(3):56-61. [doi: 10.1145/101306.101310]

23. The Jaccard Similarity algorithm. Neo4j Labs. 2019. URL: https://neo4j.com/docs/graph-algorithms/current/labs-algorithms/jaccard/#labs-algorithms-similarity-jaccard-context [accessed 2021-09-23]
24. Wolf TD, Sanh V, Chaumond J, Delangue C, Moi A, Pierric C, et al. Transformers: state-of-the-art natural language processing. In: Proceedings of the Conference on Empirical Methods in Natural Language Processingystem Demonstrations.: Association for Computational Linguistics; 2020 Presented at: Conference on Empirical Methods in Natural Language Processingystem Demonstrations; October 2020; Online p. 38-45. [doi: 10.18653/v1/2020.emnlp-demos.6]
25. Cheikh-Moussa K, Mira JJ, Orozco-Beltran D. Improving engagement among patients with chronic cardiometabolic conditions using mHealth: critical review of reviews. JMIR Mhealth Uhealth 2020 Apr 08;8(4):e15446 [FREE Full text] [doi: 10.2196/15446] [Medline: 32267239]
26. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020 Mar 24;6:2055207620914427 [FREE Full text] [doi: 10.1177/2055207620914427] [Medline: 32269830]
27. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
28. Alkhaldi G, Modrow K, Hamilton F, Pal K, Ross J, Murray E. Promoting engagement with a digital health intervention (HeLP-Diabetes) using email and text message prompts: mixed-methods study. Interact J Med Res 2017 Aug 22;6(2):e14 [FREE Full text] [doi: 10.2196/ijmr.6952] [Medline: 28829328]
29. Bennett GG, Steinberg D, Askew S, Levine E, Foley P, Batch BC, et al. Effectiveness of an app and provider counseling for obesity treatment in primary care. Am J Prev Med 2018 Dec;55(6):777-786 [FREE Full text] [doi: 10.1016/j.amepre.2018.07.005] [Medline: 30361140]
30. Patel ML, Wakayama LN, Bass MB, Breland JY. Motivational interviewing in eHealth and telehealth interventions for weight loss: a systematic review. Prev Med 2019 Sep;126:105738. [doi: 10.1016/j.ypmed.2019.05.026] [Medline: 31153917]
31. Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in web-based health behavior change programs. J Med Internet Res 2006 Aug 30;8(3):e15 [FREE Full text] [doi: 10.2196/jmir.8.3.e15] [Medline: 16954125]
32. Redfern J, Santo K, Coorey G, Thakkar J, Hackett M, Thiagalingam A, et al. Factors influencing engagement, perceived usefulness and behavioral mechanisms associated with a text message support program. PLoS One 2016 Oct 14;11(10):e0163929 [FREE Full text] [doi: 10.1371/journal.pone.0163929] [Medline: 27741244]
33. Sassenberg K, Greving H. Internet searching about disease elicits a positive perception of own health when severity of illness is high: a longitudinal questionnaire study. J Med Internet Res 2016 Mar 04;18(3):e56 [FREE Full text] [doi: 10.2196/jmir.5140] [Medline: 26944335]

Abbreviations

CVD: cardiovascular disease
DHI: digital health intervention
mHealth: mobile health
ML: machine learning
NHMRC: National Health and Medical Research Council
OR: odds ratio
RCT: randomized controlled trial
ROC: receiver operating characteristic
TEXTMEDS: Text Messages to Improve Medication Adherence in Secondary Prevention

Edited by L Buis; submitted 06.02.21; peer-reviewed by M Bientzle, A Bermon, S Baee, R Mpofu; comments to author 05.04.21; revised version received 07.05.21; accepted 03.09.21; published 10.11.21.

Please cite as:
Klimis H, Nothman J, Lu D, Sun C, Cheung NW, Redfern J, Thiagalingam A, Chow CK
Text Message Analysis Using Machine Learning to Assess Predictors of Engagement With Mobile Health Chronic Disease Prevention Programs: Content Analysis
JMIR Mhealth Uhealth 2021;9(11):e27779
URL: https://mhealth.jmir.org/2021/11/e27779
doi:10.2196/27779
PMID:34757324

Original Paper

Implementing Automated Text Messaging for Patient Self-management in the Veterans Health Administration: Qualitative Study Applying the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability Framework

Vera Yakovchenko[1], MPH, MS; D Keith McInnes[1,2], ScD, MSc; Beth Ann Petrakis[1], MPA; Chris Gillespie[1], PhD; Jessica M Lipschitz[3,4], PhD; Megan B McCullough[1,5], PhD; Lorilei Richardson[1], PhD, MS; Brian Vetter[6], RD; Timothy P Hogan[1,7], PhD

[1]Center for Healthcare Organization and Implementation Research, VA Bedford Healthcare System, Bedford, MA, United States
[2]Department of Health Law Policy and Management, Boston University School of Public Health, Boston, MA, United States
[3]Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, United States
[4]Harvard Medical School, Boston, MA, United States
[5]Department of Public Health, University of Massachusetts, Lowell, MA, United States
[6]Office of Connected Care, Veterans Health Administration, Washington, DC, United States
[7]Department of Population and Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX, United States

Corresponding Author:
Vera Yakovchenko, MPH, MS
Center for Healthcare Organization and Implementation Research
VA Bedford Healthcare System
200 Springs Road
Bedford, MA, 01730
United States
Phone: 1 781 687 3374
Email: vera.yakovchenko@va.gov

Abstract

Background: The Veterans Health Administration (VHA) is deploying an automated texting system (aTS) to support patient self-management.

Objective: We conducted a qualitative evaluation to examine factors influencing national rollout of the aTS, guided by the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework, which is intended to support the evaluation of novel technologies.

Methods: Semistructured interviews were conducted with 33 staff and 38 patients who were early adopters of the aTS. Data were analyzed following deductive and inductive approaches using a priori codes and emergent coding based on the NASSS.

Results: We identified themes across NASSS domains: (1) Condition: The aTS was considered relevant for a range of patient needs; however, perceptions of patient suitability were guided by texting experience and clinical complexity rather than potential benefits. (2) Technology: Onboarding of the aTS presented difficulty and the staff had different opinions on incorporating patient-generated data into care planning. (3) Value: Supply-side value relied on the flexibility of the aTS and its impact on staff workload whereas demand-side value was driven by patient perceptions of the psychological and behavioral impacts of the aTS. (4) Adopters: Limited clarity on staff roles and responsibilities presented challenges in incorporating the aTS into clinical processes. (5) Organization: Staff were willing to try the aTS; however, perceptions of leadership support and clinic readiness hindered usage. (6) Wider system: Staff focused on enhancing aTS interoperability with the electronic medical record. (7) Embedding and adaptation over time: The interplay of aTS versatility, patient and staff demands, and broader societal changes in preferences for communicating health information facilitated aTS implementation.

Conclusions: VHA's new aTS has the potential to further engage patients and expand the reach of VHA care; however, patients and staff require additional support to adopt, implement, and sustain the aTS. The NASSS highlighted how the aTS can be better

embedded into current practices, which patients might benefit most from its functionality, and which aspects of aTS messages are most relevant to self-management.

Trial Registration: ClinicalTrials.gov NCT03898349; https://clinicaltrials.gov/ct2/show/NCT03898349

(JMIR Mhealth Uhealth 2021;9(11):e31037) doi:10.2196/31037

KEYWORDS

implementation science; implementation facilitation; texting; veterans; eHealth; self-management; digital health; digital medicine

Introduction

Although use of virtual care technologies is expanding, automated text-messaging systems supporting patient self-management remain underused [1]. With the ubiquity of mobile phones, automated texting systems have the potential for wide reach and sustained use, representing a promising alternative to telephone calls, traditional mail, and emails [2-5]. Because of its simplicity, flexibility, and low cost, texting is increasingly recognized as a tool to reach those who may be less engaged in health care services and may thus narrow health disparities [6]. Studies show positive results for texting interventions, with demonstrated effects on chronic illness self-management, medication adherence, missed appointments, and behavior change, including weight loss and smoking cessation [2,7-10]. Despite promising evidence, the determinants of adoption, implementation, and sustainment of texting interventions are not well understood [11-15].

In 2016, the Veterans Health Administration (VHA) launched an automated texting system (aTS) called "Annie," modeled after the Florence aTS developed by the United Kingdom's National Health Service [16,17]. The aTS texting protocols are intended to promote, motivate, and enhance self-management by helping patients understand, track, and monitor their own health through 1- and 2-way messages. The aTS is condition-agnostic and able to support preprogrammed texting protocols for a variety of health conditions and behaviors. At present, there are over 100 aTS texting protocols available for use with VHA patients. These protocols fall into three categories: (1) nonpatient-specific protocols pertaining to conditions that do not require clinical diagnosis and treatment and those that patients registered with the aTS can self-subscribe (eg, tobacco cessation, coronavirus precautions); (2) nontreatment protocols intended to deliver educational and motivational messages, and generic reminders; and (3) treatment protocols that pertain to conditions requiring diagnosis and treatment by a clinician, and are intended to assist patients with self-management. Any VHA staff member regardless of licensure can register and confirm patient participation in the aTS as well as consent a patient to participate in a nontreatment aTS protocol; however, only licensed VHA clinicians can consent a patient for aTS treatment protocols. When initiating use, patients are asked to acknowledge that they understand the privacy implications of texting, which is not a secure or encrypted form of communication; they must understand that are texting with a computer system, and VHA staff may not regularly read or review the messages, making the system inappropriate for urgent issues or emergencies. Although VHA's aTS was designed as a patient-facing self-management tool, clinical team members can choose to view patient message exchanges and track responses over time through a staff-facing aTS portal.

Some of the earliest aTS protocols addressed issues ranging from physical activity engagement, medication adherence, and blood glucose monitoring to colonoscopy preparation, diabetes management and foot care, hypertension management, smoking cessation, and weight management. More recent protocols have focused on COVID-19 precautions and vaccination support, HIV treatment, chronic pain, insomnia, and maternity care. Staff can tailor protocols (eg, content, periodicity, timing) based on patient needs and preferences. A multidisciplinary group of VHA content and technical subject matter experts maintains the aTS protocol library and guides staff from across VHA facilities on how to adapt existing protocols, and to create, test, and implement new protocols. In an implementation-effectiveness cluster randomized trial of the aTS, Yakovchenko et al found that test sites with enhanced aTS implementation support not only had more patients using the aTS, but these patients also reported better adherence to treatment and lower distress about failing treatment compared to usual aTS implementation sites [1].

This qualitative evaluation examines the implementation experiences of early aTS adopters in VHA to inform national rollout of the system and improve its design and functionality. Findings from this evaluation can provide other health care systems with an understanding of the implementation challenges they might face when introducing and expanding their own texting systems.

We used the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework to organize and synthesize our findings [18,19]. The NASSS framework identifies 7 domains influencing the uptake and use of patient-facing technologies: condition, technology, value proposition, adopter system, organization, broader surrounding context, and interaction among these domains. An advantage of a complexity-informed framework like NASSS is its acknowledgment that complex adaptive systems develop and behave in unpredictable, dynamic, and nonlinear fashions.

Methods

Design

We used a qualitative interpretive phenomenological approach to explore staff and patient experiences with VHA's aTS. Semistructured interviews were conducted between June 2016 and February 2018 as part of a larger evaluation of the implementation and effectiveness of the aTS. Verbal consent

was obtained prior to interviews. The VHA Bedford Healthcare System Institutional Review Board reviewed the evaluation and determined it to be a program evaluation for quality improvement purposes, thereby exempting it from further Institutional Review Board oversight (VHA Program Guide 1200.21). The study was registered at ClinicalTrials.gov (NCT03898349).

Setting and Participants

We conducted semistructured interviews with purposively sampled participants from 14 VHA medical centers. The sites represented diversity in geography, rurality and urbanicity, patient volume, sample size, and complexity. Staff participants were VHA clinical team members (hereafter referred to as staff), including physicians, nurses, dieticians, pharmacists, and social workers. Patient participants were individuals who received health care services at these sites.

Data Collection

Staff interview guides focused on the staff perceptions regarding the aTS, implementation barriers and facilitators, setup and enrollment procedures, implications for care delivery, clinical workload, patient-provider relationship, and experiences using the system. Patient interview guides explored the patient perceptions of the aTS, including its influence on relationships with staff and engagement in care, usefulness of the aTS for supporting self-management, and factors that might influence use and perceived usefulness. Interview guides were reviewed by VHA's aTS designers for clarity, pilot-tested among the evaluation team members, and iteratively revised.

Interviews were conducted in person at VHA medical centers and over telephone, typically lasting up to 60 minutes. We gathered participant demographic data immediately prior to the start of the interviews using a brief questionnaire. Most interviews were one-on-one. In some instances where 2 staff members were interviewed simultaneously to accommodate availability, a primary interviewer led the discussion and a secondary interviewer assisted and took field notes. There were 3 male and 3 female interviewers (DKM, BAP, CG, JML, MBM, and TPH) who were masters- and doctoral-level public health, public administration, psychology, and anthropology professionals. Reflexivity was considered during the interviews and throughout the evaluation. All interviews were audio-recorded and transcribed verbatim. Patients received a gift card of a local store as compensation for their time. Staff were not eligible for compensation due to VHA regulations.

Analysis

Descriptive statistics were used to characterize participant demographics. Interview data were analyzed using deductive and inductive approaches with the NVivo 12 Pro software (QSR International). Coding and analyses were performed by 6 trained masters- and doctoral-level researchers with extensive qualitative research experience (VY, DKM, BAP, CG, LR, TPH). We drew upon the NASSS domains to create a preliminary codebook, and additional codes were inductively added if they were not otherwise reflected in the framework. A subset of interviews was examined by all evaluation team members to formulate coding rules through a process of critical review and consensus building. Recurring meetings were held to compare coder interpretations and discuss coding discrepancies. As the final step, we used the NASSS complexity assessment tool (CAT) to inform categorization of each domain as simple (straightforward, predictable, few components), complicated (remains predictable but with multiple interacting components or issues), or complex (dynamic, unpredictable, not easily disaggregated into constituent components) [20].

Results

Participant Characteristics

We conducted a total of 71 interviews with 38 VHA patients and 33 VHA staff representing a range of experiences with the aTS, as shown in Table 1. Most patients were male (n=30, 80%) and White non-Hispanic (n=29, 76%), with a median age of 56 years. Nearly all had a smartphone (n=37, 97%) and texted daily (n=30, 79%). Most also used computers (n=33, 86%) and the internet (n=34, 90%) daily. However, patients had varying experiences with VHA's aTS, with some having only registered with the system, others having received 1-way messages, and others having sent and received daily messages over multiple months. Among staff, 63% (n=21) were clinical (MD, NP, RN), 20% (n=7) were clinical pharmacists, and 22% (n=7) were other types of staff (eg, dietician, social worker). Most were women (n=25, 76%), with a median age of 36 years and 8 years of VHA work experience. Most staff (n=27, 81%) used mobile phones several times a day to send or receive messages. At the time of the interviews, staff reported varying experiences with the aTS, including some who had used the aTS from 1 day to 6 months and had enrolled 0 to over 20 patients.

We present our findings organized based on the NASSS framework domains, as displayed in Figure 1 and Table 2. Quotes are attributed to patients (pt) or staff (s).

Table 1. Participant characteristics.

Participants and characteristics	Value, n (%)
Patients (N=38)	
Median age, years	59
Male	32 (85)
Race/ethnicity: White/non-Hispanic	26 (68)
Daily texting	30 (79)
Staff (N=33)	
Staff type	
MD[a], NP[b], RN[c]	21 (64)
PharmD[d]	5 (15)
Other (dietician, social worker)	7 (21)
Median age, years	40
Median VHA[e] tenure, years	8
Male	8 (24)
Daily texting	28 (84)

[a]MD: Doctor of Medicine.

[b]NP: nurse practitioner.

[c]RN: registered nurse.

[d]PharmD: Doctor of Pharmacy.

[e]VHA: Veterans Health Administration.

Figure 1. Nonadoption, Abandonment, Scale-up, Spread, and Sustainability framework in the Veterans Health Administration (adapted from Greenhalgh [19]). VHA: Veterans Health Administration.

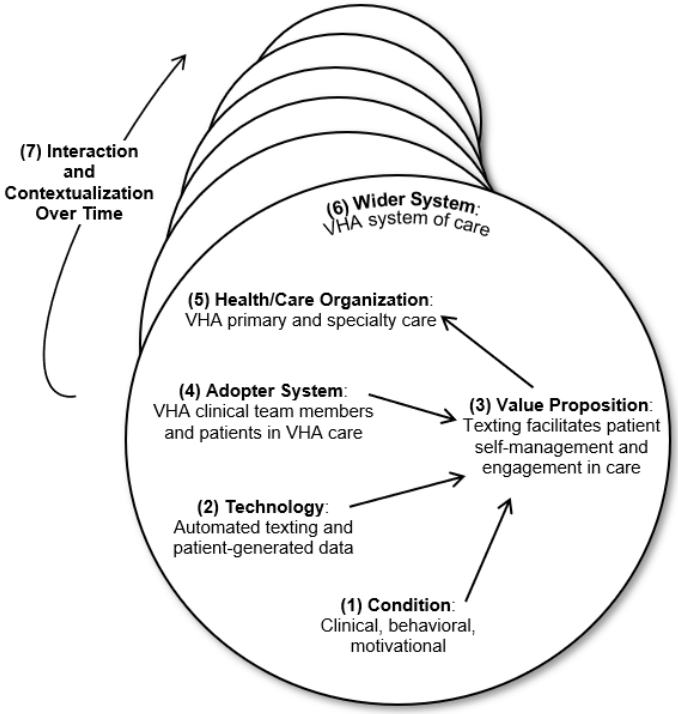


Table 2. Themes and quotes based on Nonadoption, Abandonment, Scale-up, Spread, and Sustainability domains.

NASSS[a] domain	Theme	Provider quote	Patient quote
Domain 1: Condition or illness	Patient suitability	"The patient population here tends to really need a lot of hand holding and a lot of TLC[b]."	"…made the treatment a heck of a lot easier, because I remembered to take the pills on time."
Domain 2: Technology	Knowledge, support features, and functionality; patient-generated data	"It took me a little while to familiarize myself because the training versus actually doing it yourself, you know, there's a learning curve…"	"It meets my needs and probably in the future, as my conditions change or my needs increase, it will be there."
Domain 3: Value proposition	Supply-side and demand-side values	"We're actually really excited to use it with our patients. This is something that we had talked about doing or developing something like this…"	"…would tend to get a little overwhelmed at times, but Annie helped alleviate that."
Domain 4: Adopter system	Staffing, roles, and skills; complex decisions	"I thought a lot of people are going to be able to participate but I guess when we started offering…some patients don't, I guess they're not used to it, most the patients that I offer decline to participate."	"I don't even know who would be doing it. Would it be the doctor? Or the nurses, or? Whose responsibility would it be to actually implement that? So, I don't know how practical that would actually be, but I think it would be better if they did gear it more toward that individual patient and whatever problems they're having."
Domain 5: Organization	Leadership and readiness to innovate; workflows and routines	"Whenever you're using new technology and new approaches with the technology component it's just good to have somebody that you can, who's very responsive and can find out the answer for you in a timely fashion…"	"And my understanding was this was to alleviate a lot of paperwork. They just send me a text message and that way I can write it in my date book or keep it in my phone till the time for the appointment."
Domains 6 and 7: Wider context and embedding; adaptation over time	Fiscal interoperability, digitalization, and marketing	"But in order for this to work this has to be easy for the provider and easy for the Veteran. Otherwise, it's not gonna work, it's not gonna help, we're not going to be effective."	"I use Annie for other medications as well…I just group them all in together."

[a]NASSS: Nonadoption, Abandonment, Scale-up, Spread, and Sustainability.

[b]TLC: tender loving care.

NASSS Domain 1: Condition

Staff described the aTS as a tool to support patients in activating and maintaining health-related behaviors across various conditions and care contexts. The staff and patients described the relevance of the aTS to a range of needs, including multidose vaccinations, HIV prevention and care, birth control, breast cancer screening, chemotherapy, dialysis, postoperative support, mindfulness, yoga, and anxiety management. For some, the options at times could feel overwhelming. The aTS protocols for disease processes that are "pretty well mapped out" were especially appealing to staff, as were the protocols for medication management, appointment reminders, and laboratory reminders before appointments. Patients believed the aTS could "start the habit" (pt140) to activate health behavior change and influence patterns of self-care.

Patient Suitability

Although staff were using the aTS under different conditions, they expressed similar perspectives on patient suitability to use the system. Although they generally understood that the aTS was "open to pretty much anybody," considerations about who might be appropriate for the aTS converged around several preconceived, largely nonclinical criteria that informed their offering of the aTS to patients. These criteria included patients being younger in age, perceived to have greater texting savviness and technological literacy, and a higher baseline motivation to change. Staff preferred to engage patients whom they predicted would agree to use the aTS and to avoid having conversations about the aTS with others: "…there's no sense in even bringing (the aTS) up because it just wouldn't be for them" (s224).

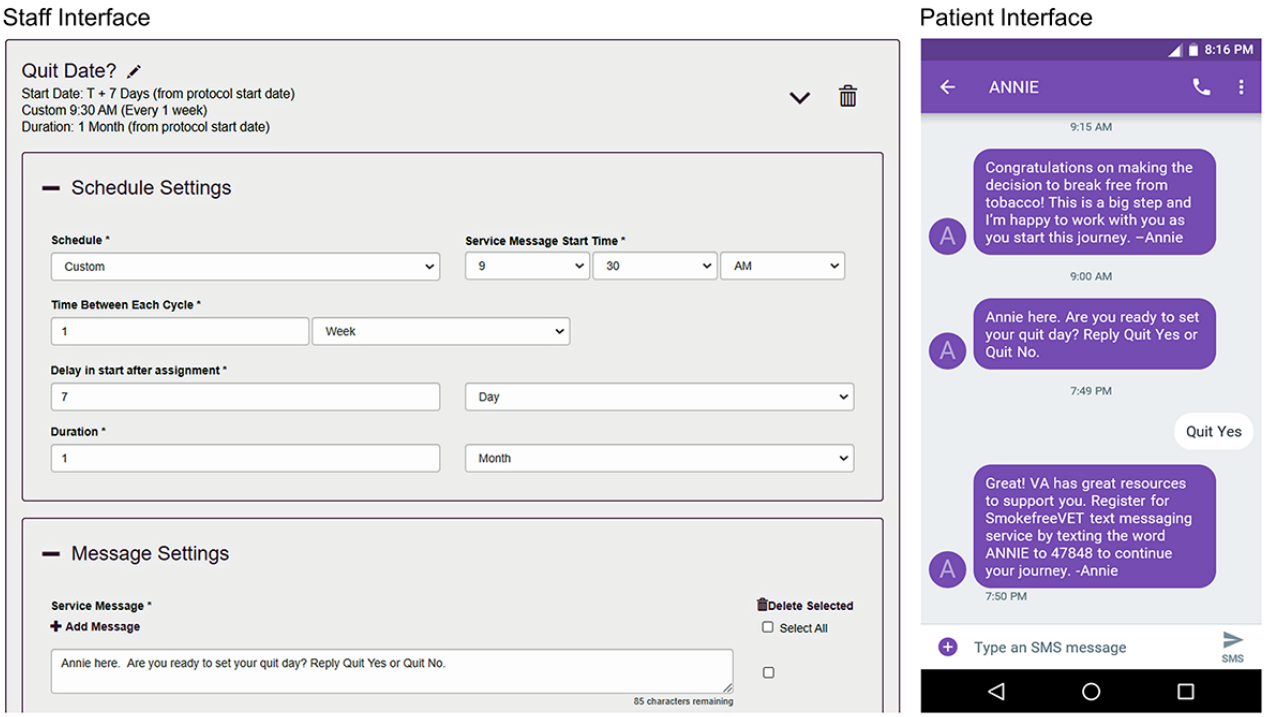
Staff further emphasized that they thought the aTS could complement traditional care for patients who might be prone to neglecting their own health or are facing various health and social challenges like memory issues and other cognitive impairments, limited social support, homelessness, and low health and technological literacy, with whom they must "work creatively" to engage in care. For more complex patients, an option was to consider caregivers as potential recipients of aTS messages. Patients and staff emphasized the importance of patients being motivated and having baseline self-efficacy to benefit from the aTS: "If you want to do it, be serious about it," (pt120, male 50 years old). Staff shared their observations that once patients started using the aTS, they tended to continue using it, and attrition was low. Patients expressed varying levels of interest in the system, from those who were actively interested ("I want it to be that thing overlooking my shoulder that gives me a little extra discipline," [pt124]) to those who were simply not interested at all ("I have no desire to play with the

telephone," [pt119]). Several staff shared their concern that some patients would agree to use the aTS to appease them during an in-person appointment and then not follow through once the protocol started. Staff mentioned needing support in "finding the correct patients" (s224). One staff member suggested a targeted recruitment approach to address issues of patients who were struggling. They suggested, "look for outliers with high blood pressure, outliers with diabetes, so they could easily focus on people and they can sell it" (s205). However, such a clinically focused patient outreach for the aTS was less commonly described than the practice of targeting patients based on the nonclinical characteristics noted above.

NASSS Domain 2: Technology

The aTS has patient- and staff-facing components that differ in terms of their interface, functionality, and complexity. The left panel in Figure 2 displays the aTS interface as seen by staff when assigning protocols, whereas the right panel displays a text exchange as seen by patients.

Figure 2. Screenshots of the staff and patient interfaces of the automated text messaging system for tobacco cessation protocol.



Knowledge and Support

Because staff invited patients with previous texting experience to use the aTS, there was no pressing need for staff-to-patient training. Conversely, staff acknowledged a steep learning curve that influenced their willingness to use and continue using the aTS, even with training, access to written materials, and live support from the aTS designers. Many commented about the time lag between training and the first enrollment of patients, which necessitated refresher training and technical support. However, once comfortable with the aTS, many deemed it "fairly straightforward and seamless" (s303), although some still found it "clunky" and "a little bit slow."

Perceptions on the time needed to introduce, consent, and enroll patients in the aTS were strong deterrents of use for busy staff. The process of signing into the system and assigning a protocol to patients was considered complicated and a point at which patients may lose interest or get frustrated; however, once this front-end "logjam" of tasks was completed, demands on staff time were manageable. One nurse practitioner commented that "my efficiency of enrolling (patients) and following through makes a big difference in time…(once) you're comfortable with (the aTS) you are more apt to continue to do it because you know it's not going to take you a lot of time" (s312). Several staff suggested that the aTS designers must consider opportunities to automate enrollment, perhaps through other existing technologies like waiting room kiosks, to bypass the extensive front-end tasks.

Features and Functionality

Regarding functionality, many patients reported difficulties with the proper syntax required to send responses to the aTS (eg, incorrectly responding "Yes" instead of "Med Yes"). Although some learned to use the correct syntax as indicated in the aTS messages, a few abandoned the system out of frustration. However, most patients commented on the simplicity of the aTS: "Easy to use, it was no problem…quick and to the point" (pt318). Some patients struggled to understand that the aTS was not a direct line of communication with their clinical team (ie, unlike secure messaging through VHA's online patient portal). Several patients indicated that they were interested in the ability to text staff directly, but the staff overwhelmingly did not want such a functionality.

The ability to tailor aspects of the aTS interaction was considered a positive system feature among patients and staff, although there were some who were unaware this was possible. The potential to create texting protocols and tailor messages for individual patients enhanced the staff's sense of ownership and

perceived value of the aTS. Some staff saw the process of tailoring the aTS as shared decision-making with their patients, and others appreciated the system's support of patient autonomy, allowing patients to use the aTS when they want to engage. As one physician noted, patients might experience texting fatigue. "The patients can pause it if they need a break. And it might be that they pause and 2 weeks down the road they can reactivate it" (s214).

Patient-Generated Data

There were mixed opinions among staff on how best to use patients' aTS responses. The patients and staff were interested in visualizing patients' numerical responses to the aTS messages over time, such as blood pressure and weight values. Staff appreciated the ability to longitudinally record such data in between visits but lamented that real-time data were not recorded directly in the electronic medical record (EMR). Some staff were less concerned about transferring this patient-generated data to the EMR and embraced reviewing the aTS portal's message dashboard, occasionally together with patients. Reviewing a patient's message logs prior to an appointment was considered helpful to guide conversation, titrate medication, and revise health plans. However, others were opposed to reviewing patient messages, yet another source of information to manage, as they believed the aTS fell entirely in the realm of self-management and was not intended for staff monitoring. As for those with more favorable opinions of the patient-generated data, some staff were unclear how, if at all, to use it: "What am I going to do with that information? Am I going to schedule an appointment with the patient? Am I going to call the patient, which is now extra work?" (s224).

NASSS Domain 3: Value Proposition

Supply-Side Value

Staff perceived the aTS as versatile, amenable to supporting simple (eg, 1-way education and motivation), complicated (eg, 2-way appointment and lab reminders), and complex (eg, procedure preparation) messaging. The aTS was viewed as a welcome complement to education provided during in-person visits and offered patients flexibility to interact with, reflect on, and apply new learning about their health at their preferred pace. Staff felt the system was aligned with VHA's larger goals of empowering patients, promoting self-management, and staying connected with patients between visits. Nevertheless, for some, perceptions of the workload associated with the aTS impeded their interest in using it.

Staff indicated several benefits of using the aTS. First, some stated the aTS was suitable for younger patients, who often prefer technology-mediated communication. Second, they recognized that investing in the aTS might have administrative benefits such as reducing appointment no-shows and increasing patient preparation for appointments (eg, necessary lab tests not completed). Third, there may be workload improvements from reducing repetitive clinical tasks. Such changes could enable other clinical team members to work at the top of their licenses. As one physiologist noted, the aTS "would open up time for the nurses to really focus their energy on patients who need a lot of tender loving care instead of just doing reminder phone calls on a weekly basis to track down patients…" (s206).

Demand-Side Value

Patients endorsed value in several areas. First, patients felt the aTS supported new learning about their health and saw the aTS as a means of promoting closer connections to their clinical team. As one patient commented, "It's kind of like the little guy whispering in your ear that this is important. Your doctor is concerned about what you're doing" (pt144). To this end, patients reported aTS messages were more credible compared to those that might come from other non-VHA technologies. Second, patients reported that texting had a psychological and behavioral impact, making them feel accountable, safer, and more comfortable. One patient who was participating in a VHA weight management program commented, "It's eye opening to me and forces me to be honest with myself about the things I've done and the things I've eaten, so that's a huge benefit because without having to do that every day it's very easy to just mindlessly go on from day to day without thinking about it specifically" (pt122). Third, patients felt texting was desirable because it offers more control when one wants to absorb information. As one patient described, "You can accept a text message when you're ready on your terms. That kind of stuff is not quite as invasive as a phone call…I think it's a great way to present nonthreatening education, reminders, guidance" (s214). Finally, patients reported leveraging the aTS messages to support their self-management efforts in unintended ways. For example, reminders from the aTS to take one's morning medications for a specific condition were used in practice by some patients as a reminder to take all their morning medications.

NASSS Domain 4: Adopters

The characteristics and experiences of the intended aTS adopters, the staff and patients, constituted the most complex domain in our analysis, largely owing to ill-defined staff roles and responsibilities during system implementation and persistent concerns about patients' understanding of self-management.

Staffing, Roles, and Skills

Despite having learned about the aTS functionality, many staff reported feeling unprepared to implement it after 1 training session, citing insufficient guidance on how to practically incorporate the aTS in their clinic and in-patient care planning. There was consensus that staff roles and interdependencies would need to change to accommodate aTS uptake and use, and many noted that without dedicated personnel, the aTS might have limited success. Staff recommended that one or several individuals, rather than an entire team, assume aTS responsibilities to ensure greater uptake and minimize disruption. Others proposed distributing enrollment steps across staff to promote teamwork.

Readiness and willingness to implement the aTS differed across staff. Pharmacists and nurses conveyed a higher readiness than physicians, many of whom were resistant. Some staff cautioned that having physicians play a central role would be a "rate-limiting step" in the system's spread. As one physician commented, "If providers are expected to have a lot of

involvement with this, it's gonna flop" (s211). Specifically, physicians sought a degree of separation from aTS education and enrollment processes, suggesting other staff as more appropriate for these tasks. However, many nurses saw the aTS as an extension of their work, including one who commented that the aTS "harmonizes perfectly" with their clinical role (s215). Staff described how nonlicensed medical support assistants could contribute to aTS enrollment processes after verbal consent was obtained from a licensed clinician. Finally, staff suggested that champions, such as existing telehealth coordinators, could help "unburden the staff from having to be the tech expert" (s217).

Complex Decisions

Staff often expressed hesitation about using 2-way texting because of concerns about interpreting and acting on the content of the patients' replies (ie, reporting high blood pressure). Although patients were made aware during consent that messages were unmonitored, staff remained concerned about liability.

Patients voiced a parallel concern about not knowing who was overseeing the messages they exchanged through the aTS and the extent to which those messages reflected an understanding of their unique health situation. As one patient remarked, "Don't get me wrong, but I don't want a clerk sending these kinds of messages where a health provider should at least be seeing it" (pt111).

NASSS Domain 5: Organization

According to the participants, the organization domain included various factors impacting aTS uptake, including leadership, readiness to innovate, and logistics of workflows and routines.

Leadership and Readiness to Innovate

At the time of executing this project, publicity for the system was highly localized, with no coordinated plan to raise awareness across leadership levels, stakeholders, or other staff. Most staff who tried the aTS were keen to learn about it and did so without a wider team or organizational involvement, citing a "willing (ness) to try new things" and openness to "new technology to help us with the Veterans" (s302). Staff cautioned that the leadership was "very contemplative," "skeptical," and required strong evidence to buy into new technology like the aTS. Although leadership buy-in was initially described as a "stumbling block," most staff felt that with sufficient evidence, the aTS would with time become "an easy sell" (s216). Notably, clinic willingness to use the aTS varied within facilities, and staff at smaller facilities tended to view the aTS as a potential time-saving tool given their limited workforce. Interfacility variation stemmed from leadership and readiness, whereas intrafacility (between clinics within a facility) variation was more adopter-oriented.

Workflows and Routines

Staff uniformly recommended that the most important consideration should be "how are we going to work this into our flow" (s217). Staff were conflicted, noting that although the aTS is "a good thing for the patient…I don't see where in the workday we can be checking this" (s203). Others expected that the aTS could offset some more administrative activities: "It would open up time for the nurses to focus their energy on patients who need a lot of tender loving care…to help streamline their work, so they could use their clinical skills more effectively" (s206). Nevertheless, at the time of our evaluation, facility-wide shared visions of how best to use the aTS were lacking, as were ideas regarding how best to coordinate aTS use to support its broader spread.

NASSS Domain 6: Wider System, Interaction Between Domains, and Adoption Over Time

Wider sociocultural forces including financial, political, legal, and regulatory factors posed hurdles to aTS uptake and sustained use. The system's long-term adoption and use was related to its design and flexibility, as well as the ability of staff, clinics, and facilities to monitor and respond to the system.

Fiscal

Staff wanted to be recognized for the time they were devoting to the aTS as well as be held accountable for and have dedicated time to use it with patients. However, staff were neither able to track workload credit to account for time spent on the aTS nor externally incentivized to use it.

Interoperability

Staff perceptions on the scale-up, spread, and sustainability of the aTS were largely based on two key system characteristics: (1) interoperability with the EMR and (2) potential to be used in conjunction with other existing technologies. Because of limited interoperability with the EMR, if staff wanted to include information in an aTS protocol specific to a patient (eg, appointment or medication refill dates), they had to manually enter the information into the protocol. This was not only laborious, but it also raised concerns about accuracy problems, particularly in the case of longer and more complex aTS protocols. For these reasons, some staff viewed the aTS as unsustainable. Some described the aTS as a "step down" from more intensive, clinician-directed initiatives such as VHA's MOVE! weight management program, and its home telehealth program with remote monitoring and case management services for chronic health conditions. One dietician commented that the aTS could help patients "to still be aware of their health goals but not so dependent on us as clinicians to really be involved in that care" (s210).

Digitalization and Marketing

Patients and staff alike viewed the aTS as part of a larger nationwide digitalization initiative. Staff suggested national-level marketing to improve the visibility of the aTS and reduce the educational burden at the clinic level. Staff likened the aTS to the rollout of VHA's online patient portal nearly 20 years earlier, suggesting that there may be transferable lessons. There was recognition that younger patients may be more amenable to adopting the aTS than the patients most staff were currently seeing for care, and although many younger patients may not yet be facing health problems, a focus on health promotion and disease prevention could help spread aTS use.

Discussion

Principal Findings

Guided by the NASSS, we evaluated the implementation of an aTS in the nation's largest integrated health care system. The perceived value of the aTS derives from its versatility, patient and staff demand, and growing societal comfort and familiarity with technology that aids health-related communication. The aTS has considerable potential to complement traditional, in-person care as well as usage of other patient-facing technologies. Nevertheless, implementation posed challenges related to the system's limited functionality, mixed user experiences, inadequately defined workflows, and limited interoperability with other systems like the EMR.

Although staff became competent in using the aTS, many were unable to integrate the system into their workflow given the limited duration of patient visits, technical challenges, and distribution of other clinical tasks. Staff appeared to assume that once a critical number of patients started using the aTS and the burden of educating, consenting, and enrolling large numbers of patients had passed, they would thereafter have the relatively easy task of assigning protocols. Despite being a plausible scenario, this can place a heavy burden on early staff adopters who may be expected to perform extra short-term work in the belief they will receive long-term benefits.

VHA and other health care systems are investing in technologies for remote delivery of health care services, and texting systems are arguably one of the most efficient forms of communication. However, we found that staff assumptions about many of their patients precluded universal offering of the aTS. Moving forward, safeguards must be established to protect against such bias. More targeted patient outreach was recommended by some participants and is supported by literature [21]. Ways to increase aTS use despite staff concerns could include streamlining the aTS enrollment process by offering the system to all new patients at the time of VHA health care enrollment and providing specially trained staff for this, thereby removing clinical staff from time-consuming front-end processes.

We found that some patients were skeptical about texting, whereas others felt little need to improve their self-management and therefore declined the aTS. Our data highlight how the process of tailoring aTS message content and timing not only encouraged patient use but also enhanced patient autonomy, which aligns with patient-centered care principles [22,23]. Although many staff reported that the responses patients sent to the system would rarely change the care they provided, most of the staff wanted access to the patient messages within the aTS and related system reports [24]. As the potential value of patient-generated data grows, future efforts may involve using the aTS to gather patient-reported outcomes, including satisfaction, comprehension of instructions, and postdischarge follow-up assessments [25].

In cases of modest patient interest in the aTS system as well as patient memory and cognitive issues, patients and staff suggested that informal caregivers might be greater beneficiaries of the aTS than patients. Indeed, Wagner et al found text messaging caregivers directly was significantly associated with changes in diabetes outcomes for patients [26]. Extending the reach of the aTS to caregivers requires further study as an implementation strategy and an approach to improving patient outcomes. Moreover, given the heterogeneity of past texting studies, more research is needed to determine associations between texting intervention characteristics (eg, frequency, timing, duration, interactivity) and outcomes [2].

The NASSS is quickly becoming a prominent meta-framework for identifying complexities and their interactions in studies of technology implementation [27,28]. The framework posits 3 levels of complexity—simple, complicated, and complex—which may predict technology adoption and nonadoption. New tools, including the NASSS-CAT, are particularly useful in explaining our data and could help guide technology implementation initiatives [20]. We used the NASSS-CAT retrospectively; future implementation work may consider using such tools prospectively over the duration of a project.

From our application of the NASSS framework, we determined that overall, the aTS was perceived as easy to use by patients (simple in NASSS terminology) and difficult by staff (complicated in NASSS terminology), a situation that could threaten scalability. In our analysis, the aspect most likely to hamper the uptake and spread of aTS was related to adopters (characteristics and experiences of intended users), principally because there was insufficient staff training, and staff roles and responsibilities for implementation tasks were poorly defined. We determined it as the most unpredictable and abstruse domain (complex in NASSS terminology). The only other complex domain was value (users' perceived benefit of the technology), rated as such because staff could detect the potential of the aTS but could not always detect a relative advantage given workload demands. Such staff views likely fueled their doubts that aTS would be sustainable. Several NASSS domains presented moderate difficulty to staff and patients (ie, they were rated as complicated). These were condition (health issue being addressed), technology (innovation characteristics such as system usability and data generated), organization (characteristics of the health care system), and wider system (societal elements such as political and regulatory concerns). A frequently cited challenge was the lack of the interoperability of the aTS with the EMR and existing patient-facing technologies. Interoperability helps drive spread and sustainability by facilitating documentation of aTS enrollment, easily transferring patient responses to clinical notes, and autopopulating text messages unique to an individual patient's situation. Integrating the aTS into the EMR has implications for workflows, potentially reducing staff burden. There may also be opportunities for integration of the aTS into population health management tools that are commonly used in VHA [29-32]. It is not surprising that no domains were rated as simple given the aTS is a dynamic platform that can address the needs of a range of health conditions, behaviors, and patients.

The challenges noted above need not hamper the uptake and spread of the aTS, and since the completion of this evaluation, VHA has made iterative refinements in the aTS, some of which address these challenges. A variety of practices can help with

implementation of new technologies in large complex health care systems: (1) At the facility level, local champions can be identified or assigned, who can assist with multiple aspects of text-messaging promotion, such as marketing and education, technical assistance, and work groups and communities of practice [33]. (2) At central levels, an expert panel can be created to oversee text-messaging protocol governance, and, depending on its expertise and capacity, such a panel could also work with individual facilities to develop and test new protocols. (3) Marketing and educational efforts can be effective in producing materials, supporting websites and helplines, creating training platforms, and convening a community of practice. (4) Finally, dashboards can be created to track new enrollment in text-messaging systems in real time at the national, regional, and facility levels.

Strengths and Limitations

One strength of this evaluation is the identification of diverse perspectives of VHA patients and staff from a range of disciplines and VHA settings. VHA is unlike many other health care systems, and veterans differ from nonveteran patients; therefore, our findings may not be entirely generalizable to other health care systems and patient populations. It is important to note that at the time of our work, the aTS was in a beta testing stage, and the system was actively being updated. In addition to impacting the user experience, staff who engaged with the system at this time could generally be considered early adopters and may differ from users who encounter the aTS at later stages of its implementation. Data were collected in 2 waves over a period of 1.5 years. We used the NASSS-CAT retrospectively, thus likely not taking full advantage of its potential. Future implementation work may do well to consider using such tools over the duration of a project prospectively and retrospectively.

Conclusions

This is the first paper to report qualitative findings from the perspectives of patients and staff on factors affecting the adoption, implementation, and spread of VHA's new aTS. NASSS, a meta-framework for identifying complex elements and their interactions, was an important tool to help classify these factors and recognize their interplay [27,28]. As health care systems implement new technologies to deliver high-quality, effective, patient-centered care, the multilevel complexities of adoption (or nonadoption), implementation, and sustainment must be studied. Insights gained from such evaluations continue to inform improvements in VHA's aTS system and its national rollout and use, and they can aid in the scaling of texting interventions in other health care systems.

Acknowledgments

The views expressed in this article are those of the authors and do not necessarily reflect the position and policy of the Department of Veterans Affairs or the United States Government. This work was supported by the US Department of Veterans Affairs, Office of Connected Care, and Office of Research and Development, Health Services Research and Development Service, Quality Enhancement Research Initiative Program (PEC 15-470; PI: Hogan).

Authors' Contributions

VY, DKM, and TPH conceived and designed the study together. DKM, BAP, CG, JML, MBM, LR, and TPH collected the data, and VY, DKM, BAP, CG, LR, and TPH conducted the analysis. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Yakovchenko V, Hogan TP, Houston TK, Richardson L, Lipschitz J, Petrakis BA, et al. Automated text messaging with patients in Department of Veterans Affairs specialty clinics: cluster randomized trial. J Med Internet Res 2019 Aug;21(8):e14750 [FREE Full text] [doi: 10.2196/14750] [Medline: 31444872]
2. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
3. Vodopivec-Jamsek V, de Jongh T, Gurol-Urganci I, Atun R, Car J. Mobile phone messaging for preventive health care. Cochrane Database Syst Rev 2012 Dec 12;12(1):CD007457-CD007443 [FREE Full text] [doi: 10.1002/14651858.CD007457.pub2] [Medline: 23235643]
4. Schickedanz A, Huang D, Lopez A, Cheung E, Lyles CR, Bodenheimer T, et al. Access, interest, and attitudes toward electronic communication for health care among patients in the medical safety net. J Gen Intern Med 2013 Jul;28(7):914-920 [FREE Full text] [doi: 10.1007/s11606-012-2329-5] [Medline: 23423453]
5. Pew Research Center. Smartphone ownership is growing rapidly around the world, but not always equally. URL: https://www.pewresearch.org/global/wp-content/uploads/sites/2/2019/02/Pew-Research-Center_Global-Technology-Use-2018_2019-02-05.pdf [accessed 2021-06-06]
6. Hirshberg A, Sammel MD, Srinivas SK. Text message remote monitoring reduced racial disparities in postpartum blood pressure ascertainment. Am J Obstet Gynecol 2019 Sep;221(3):283-285. [doi: 10.1016/j.ajog.2019.05.011] [Medline: 31121137]

7. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec;12:CD007459 [FREE Full text] [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
8. Fischer HH, Moore SL, Ginosar D, Davidson AJ, Rice-Peterson CM, Durfee MJ, et al. Care by cell phone: text messaging for chronic disease management. Am J Manag Care 2012 Feb;18(2):e42-e47 [FREE Full text] [Medline: 22435883]
9. Junod Perron N, Dao MD, Righini NC, Humair J, Broers B, Narring F, et al. Text-messaging versus telephone reminders to reduce missed appointments in an academic primary care clinic: a randomized controlled trial. BMC Health Serv Res 2013 Apr;13:125 [FREE Full text] [doi: 10.1186/1472-6963-13-125] [Medline: 23557331]
10. Foreman KF, Stockl KM, Le LB, Fisk E, Shah SM, Lew HC, et al. Impact of a text messaging pilot program on patient medication adherence. Clin Ther 2012 May;34(5):1084-1091. [doi: 10.1016/j.clinthera.2012.04.007] [Medline: 22554973]
11. Househ M. The role of short messaging service in supporting the delivery of healthcare: an umbrella systematic review. Health Informatics J 2016 Jun;22(2):140-150 [FREE Full text] [doi: 10.1177/1460458214540908] [Medline: 25038203]
12. Schwebel FJ, Larimer ME. Using text message reminders in health care services: a narrative literature review. Internet Interv 2018 Sep;13:82-104 [FREE Full text] [doi: 10.1016/j.invent.2018.06.002] [Medline: 30206523]
13. Houston TK, Richardson LM, Cotten SR. Patient-directed digital health technologies: Is implementation outpacing evidence? Med Care 2019 Feb;57(2):95-97. [doi: 10.1097/MLR.0000000000001068] [Medline: 30601341]
14. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014 Oct;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442] [Medline: 25326646]
15. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
16. McInnes DK, Houston TK, Woods SS, Frisbee KL, Evans NC. Disseminating information systems across the Atlantic: collaboration between U.K. National Health Service and U.S. Department of Veterans Affairs. Implementation Sci 2015 Aug;10(S1):1-1. [doi: 10.1186/1748-5908-10-S1-A62]
17. Cottrell E, Chambers R, O'Connell P. Using simple telehealth in primary care to reduce blood pressure: a service evaluation. BMJ Open 2012 Oct;2(6):1-10 [FREE Full text] [doi: 10.1136/bmjopen-2012-001391] [Medline: 23117563]
18. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the NASSS framework. BMC Med 2018 May;16(1):66 [FREE Full text] [doi: 10.1186/s12916-018-1050-6] [Medline: 29754584]
19. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
20. Greenhalgh T, Maylor H, Shaw S, Wherton J, Papoutsi C, Betton V, et al. The NASSS-CAT tools for understanding, guiding, monitoring, and researching technology implementation projects in health and social care: protocol for an evaluation study in real-world settings. JMIR Res Protoc 2020 May;9(5):e16861 [FREE Full text] [doi: 10.2196/16861] [Medline: 32401224]
21. Head KJ, Noar SM, Iannarino NT, Grant HN. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
22. Snyder CF, Wu AW, Miller RS, Jensen RE, Bantug ET, Wolff AC. The role of informatics in promoting patient-centered care. Cancer J 2011 Jul;17(4):211-218 [FREE Full text] [doi: 10.1097/PPO.0b013e318225ff89] [Medline: 21799327]
23. Steiner JF, Zeng C, Comer AC, Barrow JC, Langer JN, Steffen DA, et al. Factors associated with opting out of automated text and telephone messages among adult members of an integrated health care system. JAMA Netw Open 2021 Mar;4(3):e213479 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.3479] [Medline: 33769509]
24. Nundy S, Lu CE, Hogan P, Mishra A, Peek ME. Using patient-generated health data from mobile technologies for diabetes self-management support: provider perspectives from an academic medical center. J Diabetes Sci Technol 2014 Jan;8(1):74-82. [doi: 10.1177/1932296813511727] [Medline: 24876541]
25. Woods SS, Evans NC, Frisbee KL. Integrating patient voices into health information for self-care and patient-clinician partnerships: Veterans Affairs design recommendations for patient-generated data applications. J Am Med Inform Assoc 2016 May;23(3):491-495. [doi: 10.1093/jamia/ocv199] [Medline: 26911810]
26. Wagner DV, Barry SA, Stoeckel M, Teplitsky L, Harris MA. NICH at its best for diabetes at its worst: texting teens and their caregivers for better outcomes. J Diabetes Sci Technol 2017 May;11(3):468-475 [FREE Full text] [doi: 10.1177/1932296817695337] [Medline: 28745094]
27. James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, scale-up, and sustainability of video consulting in health care: systematic review and synthesis guided by the NASSS framework. J Med Internet Res 2021 Jan;23(1):e23775 [FREE Full text] [doi: 10.2196/23775] [Medline: 33434141]
28. Kadesjö Banck J, Bernhardsson S. Experiences from implementation of internet-delivered cognitive behaviour therapy for insomnia in psychiatric health care: a qualitative study applying the NASSS framework. BMC Health Serv Res 2020 Aug;20(1):729 [FREE Full text] [doi: 10.1186/s12913-020-05596-6] [Medline: 32771024]

29. Carmichael JM, Meier J, Robinson A, Taylor J, Higgins DT, Patel S. Leveraging electronic medical record data for population health management in the Veterans Health Administration: successes and lessons learned. Am J Health Syst Pharm 2017 Sep;74(18):1447-1459. [doi: 10.2146/ajhp161048] [Medline: 28887346]
30. Fischer MJ, Kourany WM, Sovern K, Forrester K, Griffin C, Lightner N, et al. Development, implementation and user experience of the Veterans Health Administration (VHA) dialysis dashboard. BMC Nephrol 2020 Apr;21(1):136 [FREE Full text] [doi: 10.1186/s12882-020-01798-6] [Medline: 32299383]
31. Lau MK, Bounthavong M, Kay CL, Harvey MA, Christopher MLD. Clinical dashboard development and use for academic detailing in the U.S. Department of Veterans Affairs. J Am Pharm Assoc 2019 Apr;59(2S):S96-S103.e3. [doi: 10.1016/j.japh.2018.12.006] [Medline: 30713078]
32. Barnes GD, Sippola E, Dorsch M, Errickson J, Lanham M, Allen A, et al. Applying population health approaches to improve safe anticoagulant use in the outpatient setting: the DOAC Dashboard multi-cohort implementation evaluation study protocol. Implement Sci 2020 Sep;15(1):83 [FREE Full text] [doi: 10.1186/s13012-020-01044-5] [Medline: 32958020]
33. Nazi KM, Hogan TP, Wagner TH, McInnes DK, Smith BM, Haggstrom D, et al. Embracing a health services research perspective on personal health records: lessons learned from the VA My HealtheVet system. J Gen Intern Med 2010 Jan;25 Suppl 1:62-67 [FREE Full text] [doi: 10.1007/s11606-009-1114-6] [Medline: 20077154]

Abbreviations

aTS: automated texting system
CAT: complexity assessment tool
EMR: electronic medical record
NASSS: Nonadoption, Abandonment, Scale-up, Spread, and Sustainability Framework
VHA: Veterans Health Administration

Edited by G Eysenbach; submitted 07.06.21; peer-reviewed by J Sussman, Y Chu; comments to author 01.07.21; revised version received 23.08.21; accepted 13.09.21; published 15.11.21.

Please cite as:
Yakovchenko V, McInnes DK, Petrakis BA, Gillespie C, Lipschitz JM, McCullough MB, Richardson L, Vetter B, Hogan TP
Implementing Automated Text Messaging for Patient Self-management in the Veterans Health Administration: Qualitative Study Applying the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability Framework
JMIR Mhealth Uhealth 2021;9(11):e31037
URL: https://mhealth.jmir.org/2021/11/e31037
doi:10.2196/31037
PMID:34779779

Original Paper

Mobile Text Messaging for Tobacco Risk Communication Among Young Adult Community College Students: Randomized Trial of Project Debunk

Alexander V Prokhorov[1*], MD, PhD; Karen Sue Calabro[1], DrPH; Ashish Arya[1], MD; Sophia Russell[2], MPH; Katarzyna W Czerniak[1], MLA; Gabrielle C Botello[3], BA; Minxing Chen[4], MS; Ying Yuan[4], PhD; Adriana Perez[5], MS, PhD; Damon J Vidrine[6], MS, DrPH; Cheryl L Perry[7], PhD; Georges Elias Khalil[8*], MPH, PhD

[1]Department of Behavioral Science, MD Anderson Cancer Center, The University of Texas, Houston, TX, United States

[2]Department of Health Disparities, MD Anderson Cancer Center, The University of Texas, Houston, TX, United States

[3]Department of Health Services Research, MD Anderson Cancer Center, The University of Texas, Houston, TX, United States

[4]Department of Biostatistics, MD Anderson Cancer Center, The University of Texas, Houston, TX, United States

[5]Department of Biostatistics and Data Science, School of Public Health, University of Texas Health Science Center, Austin, TX, United States

[6]Department of Health Outcomes and Behavior, Moffitt Cancer Center, Tampa, FL, United States

[7]School of Public Health, Health Science Center, University of Texas, Austin, TX, United States

[8]Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, United States

[*]these authors contributed equally

Corresponding Author:
Georges Elias Khalil, MPH, PhD
Department of Health Outcomes and Biomedical Informatics
College of Medicine
University of Florida
Clinical and Translational Science Building
2004 Mowry Road Office 2252
Gainesville, FL, 32610
United States
Phone: 1 3522948415
Email: gkhalil@ufl.edu

Abstract

Background: The use of new and emerging tobacco products (NETPs) and conventional tobacco products (CTPs) has been linked to several alarming medical conditions among young adults (YAs). Considering that 96% of YAs own mobile phones, SMS text messaging may be an effective strategy for tobacco risk communication.

Objective: Project Debunk is a community-based randomized trial aiming to identify specific types of messages that effectively improve perceived NETP and CTP risk among YAs in community colleges.

Methods: With YAs recruited offline from 3 campuses at the Houston Community College (September 2016 to July 2017), we conducted a 6-month randomized trial with 8 arms based on the combination of 3 message categories: framing (gain-framed vs loss-framed), depth (simple vs complex), and appeal (emotional vs rational). Participants received fully automated web-based SMS text messages in two 30-day campaigns (2 messages per day). We conducted repeated-measures mixed-effect models stratified by message type received, predicting perceived CTP and NETP risks. Owing to multiple testing with 7 models, an association was deemed significant for $P<.007$ (.05 divided by 7).

Results: A total of 636 participants completed the baseline survey, were randomized to 1 of 8 conditions (between 73 and 86 participants per condition), and received messages from both campaigns. By the 2-month post campaign 2 assessment point, 70.1% (446/636) completed all outcome measures. By the end of both campaigns, participants had a significant increase in perceived NETP risk over time ($P<.001$); however, participants had a marginal increase in perceived CTP risk ($P=.008$). Separately for each group, there was a significant increase in perceived NETP risk among participants who received rational messages ($P=.005$), those who received emotional messages ($P=.006$), those who received simple messages ($P=.003$), and those who received gain-framed messages ($P=.003$).

Conclusions: In this trial, YAs had an increase in perceived NETP risk. However, with stratification, we observed a significant increase in perceived NETP risk upon exposure to rational, emotional, simple, and gain-framed messages. In addition, YAs generally had an increase in perceived CTP risk and presented nonsignificant but observable improvement upon exposure to emotional, complex, and loss-framed messages. With the results of this study, researchers and practitioners implementing mobile health programs may take advantage of our tailored messages through larger technology-based programs such as smartphone apps and social media campaigns.

Trial Registration: ClinicalTrials.gov NCT03457480; https://clinicaltrials.gov/ct2/show/NCT03457480

International Registered Report Identifier (IRRID): RR2-10.2196/10977

(JMIR Mhealth Uhealth 2021;9(11):e25618) doi:10.2196/25618

KEYWORDS

tobacco use; risk communication; text messaging; message framing; regulatory science; young adults; vaping; mobile phone

Introduction

Background

Tobacco use in many forms, such as combustible, vaporized, or smokeless, has been linked to several alarming medical conditions among young adults (YAs; aged 18-25 years). These include nicotine dependence [1], psychiatric disorders [2], and developing pulmonary [3,4] and cardiovascular diseases [5]. Of more recent concern are the findings of deleterious health effects associated with inhalation of nicotine-containing aerosol (vaping) [6,7] and longitudinal associations between vaping and future use of conventional tobacco products (CTPs; including combustible cigarettes, cigars, pipes, and chewing tobacco, dip, or snuff) [8-11]. However, approximately 15% of American YAs are current cigarette smokers, and 36% of American YAs are current users of vaping products, one of many new and emerging tobacco products (NETPs), including vapes, hookahs, and snus in the United States [12].

YA tobacco use can be partly attributed to the relatively low perceived risk of products when compared with other adult age groups [13-15]. This is particularly the case for NETPs such as vaping products and hookahs, which are believed to be safer than CTPs [13,16-18]. According to national reports, YAs tend to have a lower perception of the harm from vaping products [19] and hookah [20] compared with combustible cigarettes. YAs tend to demonstrate a lack of knowledge about the ingredients in vaping products [14,21,22]. Ultimately, lower risk perception, among other factors, has led to health-compromising behaviors among YAs [23], including experimentation with various nicotine and tobacco products as well as other substances [24-26].

Responding to Tobacco Marketing

Despite known health consequences, aggressive tobacco marketing to young people has been found to reduce risk perception and promote continued tobacco use [27,28]. YAs form a highly vulnerable population that continues to be targeted as a potentially profitable market segment for the tobacco industry [27-30]. The tobacco industry broadly disseminates modern advertising on the radio, television, the internet, in print, through direct mail, in nightclubs and pubs, and at the point of sale [31-36].

Tobacco advertising to YAs has become particularly successful through mobile media channels [37,38]. Today, tobacco companies depend on mobile strategies for marketing, considering that 96% of YAs own smartphone devices [39]. Mobile marketing forums constitute the next generation of marketing strategies, notably with demonstrations and invitations through social media websites [40-43] and a variety of protobacco smartphone apps advertised under *kids* and *games* categories [44].

Health promotion experts and activists ought to respond to tobacco marketing by communicating tobacco risk to YAs as delineated by the educational mission and research priorities of the US Food and Drug Administration [45,46]. Considering that nearly all YAs (96%) own mobile phones, mobile phone SMS text messaging is likely to be an effective strategy for tobacco risk communication [39,47]. Although YAs tend to use a variety of mobile phone apps for communication (eg, WhatsApp), SMS text messaging remains a universal and practical method of risk communication. SMS text messaging programs have been successfully implemented for preventive behavioral treatment, including smoking cessation [48]. However, there have been no published accounts for its application in communicating tobacco risk to YAs [49-53].

Project Debunk: A Text Messaging Program

The goal of our project (Project Debunk) was to develop a library of risk communication messages. Our message design was based on a combination of 3 main message categories, each with 2 message types: (1) framing (gain-framed or loss-framed messages), (2) depth (ie, simple or complex messages), and (3) appeal (ie, emotional or rational messages). Framing and appeal were supported by previous research [54-57], whereas depth was an original category proposed by our investigative team. Messages from each category were also developed to communicate the harm of CTPs and NETPs (Multimedia Appendix 1). Messages describing CTPs included information regarding combustible cigarettes, cigars, and pipes. Messages describing NETPs included information regarding vaping devices, snus, little cigars, cigarillos, and hookah (products that were becoming increasingly prevalent at the time of this study [12]). The resulting 976 SMS text messages were designed through focus group discussions with YAs and feedback from experts in public health, health communication, and behavioral science [58]. The results indicate that YAs find the messages

interesting and appropriate. They described the messages as informative, interesting, easy to understand, straight to the point, and at an appropriate character limit [58]. As a result, these messages are of interest to YAs who would like to be informed regarding tobacco. These SMS text messages have been validated using linguistic inquiry and word counting, indicating an adequate design based on framing, depth, and appeal [59]. In addition, early analysis from a Project Debunk randomized trial further validated the messages [60]. Loss-framed messages were more likely to be perceived as presenting a loss than gain-framed messages, complex messages were reported to be more complex than simple messages, and emotional messages were perceived to be more emotionally engaging than rational messages [60]. In addition, there were no differences among the message types with respect to reported message credibility, message enjoyment, perceived message relevance, or message readability level [60]. A detailed description of the trial and baseline characteristics has been published under the study protocol [60].

Theoretical Framework

Available theoretical frameworks have described that the success of message characteristics depends on individual differences in the way they process information. First, according to the elaboration likelihood model (ELM) [61,62], the effectiveness of message characteristics depends on one's cognitive effort used to engage with the message content. When centrally processing information, individuals put more effort into paying attention to message content (eg, complex and rational messages [63]). On the other hand, when peripherally processing information, individuals put less cognitive effort by paying attention to more peripheral cues, such as emotional features of the message [64]. Second, according to the prospect theory of message framing, gain and loss framing can have an effect on health behavior depending on whether the individual is risk-aversive or risk-taking [65]. Regardless, different message characteristics can be effective for different audience members. Results from previous meta-analyses of relevant research have not favored one message type over another when improving health outcomes [66-68]. As a result, it is essential to explore the effects of different message characteristics on perceived risk.

Study Objective

The objective of this paper on Project Debunk is to present the results of a community-based randomized trial. The trial aims to identify specific types of messages that are effective in increasing the perceived risk of NETP use and CTP use among YAs in community colleges. Considering the limited research, we cannot predict or anticipate differences among message types in improving perceived tobacco risk [60]. For this reason, we will test the success of improving perceived risk over time among participants exposed to each message type alone. Our central hypothesis is that controlling for all other message types, campaign participants receiving each message type will have an increase in perceived NETP risk and perceived CTP risk over time. A message type is deemed impactful if it improves YAs' perceived risk over time. Once impactful SMS text messages have been identified, they can subsequently be introduced into an advanced digital intervention.

Methods

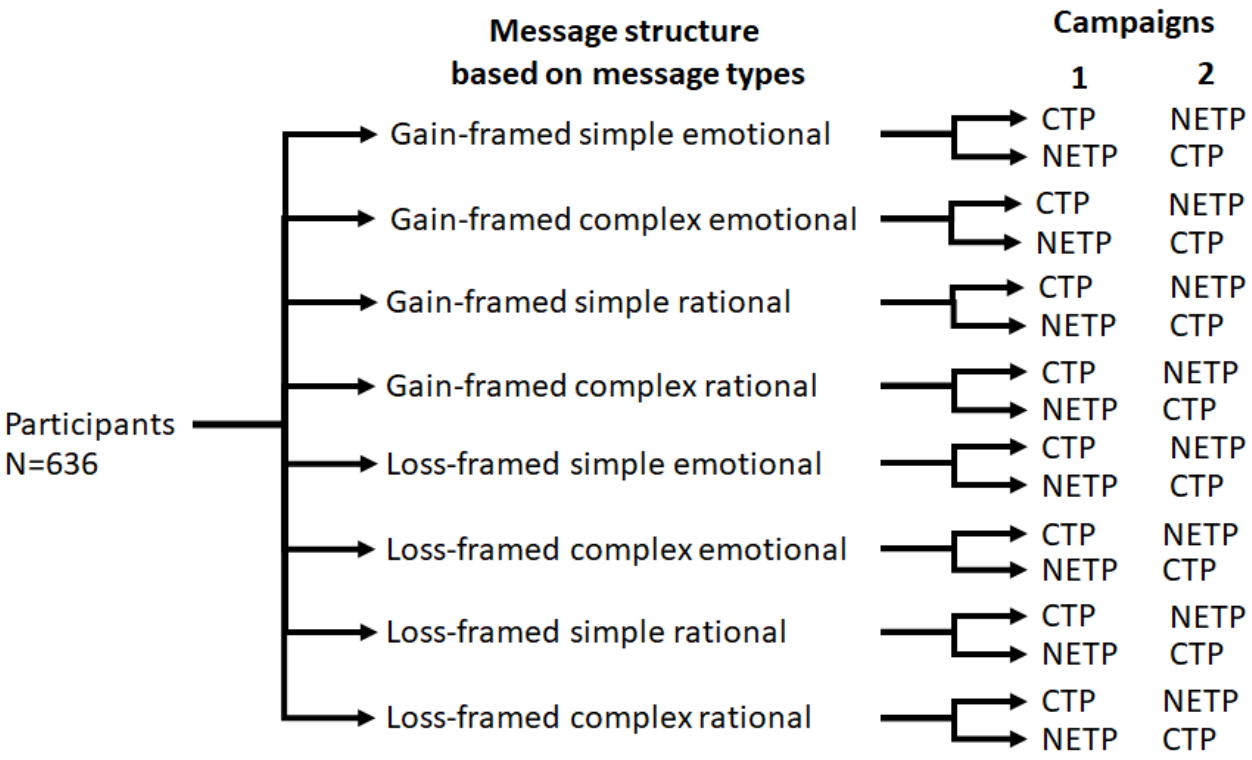
Study Design

A detailed description of the study design has been presented elsewhere [60]. Briefly, we conducted a 6-month randomized trial (September 2016 to July 2017) with 8 arms based on a combination of 3 message categories: framing, depth, and appeal (NCT03457480). Each category included 2 message types, leading to a 2 (framing: gain vs loss) ×2 (depth: simple vs complex) ×2 (appeal: rational vs emotional) factorial design. Randomization in this design allowed us to control for receiving different message types, as we tested changes over time for each message type. Our objective was to examine changes in perceived risk over time. As a result, comparison with a control group was not conducted.

Participants received SMS text messages on their mobile phones for free in 2 waves or campaigns. Each campaign comprised 2 SMS text messages per day for 30 days (ie, 60 text messages). This resulted in a total of 120 messages considering both campaigns. Participants were randomized into 8 arms, with 8 permutations based on message types. As a result, there was a total of 960 messages disseminated for this study. The 2 campaigns were performed 1 week apart. The development process and content of messages were *frozen* during the trial.

For ethical reasons, all participants received both NETP and CTP messages. To control for the order in which SMS text messages were received, the study included a crossover design. Participants within each arm were randomly divided into 2 groups: the first group received messages on CTPs during the first campaign and then NETPs during the second campaign, whereas the second group received messages on NETPs during the first campaign and then CTPs during the second campaign (Figure 1). The study adhered to the CONSORT (Consolidated Standards of Reporting Trials) and CONSORT-EHEALTH (CONSORT–Electronic and Mobile Health Applications and Online TeleHealth) guidelines (Multimedia Appendix 2) [69,70].

Figure 1. Study design and randomization to 8 conditions followed by 2 crossover conditions (total of 16 conditions). CTP: conventional tobacco product; NETP: new and emerging tobacco product; G: gain-framed messages, L: loss-framed messages, C: complex messages, S: simple messages, R: rational messages, and E: emotional messages.



Population

YAs were recruited from 3 campuses at the Houston Community College. The campuses were selected based on their ethnically diverse populations [71]. YAs were eligible for the study if they were aged between 18 and 25 years, enrolled in a community college, possessed a mobile phone, regularly used SMS text messaging, were willing to provide their phone number, were capable of receiving SMS text messages from our messaging system, and were able to read and speak English. The study protocol was approved by the institutional review boards of the University of Texas MD Anderson Cancer Center and Houston Community College.

Recruitment and Enrollment

At each of the 3 campuses, recruitment took place face to face at highly visible recruitment stations or booths, and printed materials announcing the study (posters and fliers) were displayed in high traffic areas. Research staff screened interested students for eligibility, and eligible students provided informed consent to participate in the study. During the face-to-face consent process, participants received information regarding the objective of the study, study procedures, potential risks, potential benefits, compensation information, and contact information. Following consent, participants completed a 20-minute web-based baseline survey on their phones. Participants received a US $25 gift card for completing the baseline survey and each of the 2 postcampaign surveys.

After 3 days of receiving the baseline survey, YAs began to receive SMS text messages through the MD Anderson Cancer Center resource called assessment, intervention, and measurement. A password-protected allocation sequence was generated by the assessment, intervention, and measurement resource, automatically sending SMS text messages on the basis of allocation and keeping the research team blind to participant allocation. Participants were blinded to the type of message they received. Research assistants were available over the phone in case of usability issues.

Measures

Through web-based skip-pattern surveys, we assessed a series of previously validated and pretested measures [72]. The measures have been validated for web-based use. We listed all measures within our research protocol for the current trial [60] and adhered to the Checklist for Reporting Results of Internet E-Surveys (previously published as a supplementary material [60]). At baseline, participants provided information regarding their age, gender, race, ethnic group, basic expenses, education attainment, numeracy level [73], and current use of CTP and NETP (ie, past 30 days) [74,75]. Immediately after each campaign, we assessed the self-reported attention level to the messages (2 items, such as “When I was reading the text messages, I paid attention to the messages more than to what was happening around me.”). Participants were asked if this was true for none of the messages (1), 1-2 messages (2), some of the messages (3), a lot of the messages (4), or all the messages (5). At baseline, 2 months post campaign 1, and 2 months post campaign 2, we assessed perceived CTP risk and perceived NETP risk [76]. Using a validated scale [76], the perceived risk of using cigarettes, cigars, little cigars, cigarillos, pipes, chewing tobacco, and dip or snuff, hookah, vaping products, and snus was measured separately for each product. This allowed

respondents to distinguish between the risks of different tobacco products. The measure included 5 items; with the first item on a 4-point Likert scale from no risk to great risk, we asked respondents how much they think people risk harming themselves if they use each of the tobacco products. On a 4-point Likert scale from strongly disagree to strongly agree, the 4 remaining items presented statements such as, "The following products increase the risk for medical problems such as reproductive problems, respiratory problems, or heart disease" [76]. Depending on product type, Cronbach α scores ranged between .72 and .87. Perceived NETP risk was measured as the average score for vaping products, hookah, little cigars or cigarillos, and snus. Perceived CTP risk was measured as the average score for cigarettes, cigars, pipes, and chewing tobacco, dip, or snuff. The use of hookah, little cigars, or cigarillos, snus, and electronic cigarettes became increasingly prevalent at the time of this study, making such products new and emerging [12]. This was particularly the case in Texas [77]. For this reason, we chose to treat them as NETPs.

Statistical Analysis

Sample size determination has been previously described [60]. First, chi-square tests and 1-way analyses of variance were conducted to check for differences among groups with respect to the digital divide (ie, gaps in access to mobile phones), confounders, and message reception. Then, we examined 2 study outcomes (perceived CTP risk and perceived NETP risk) and their change over time (from baseline to 2 months post campaign 2). To test the overall campaign success, we conducted 2 repeated-measures mixed-effect models for all participants, examining changes over time in perceived CTP risk and perceived NETP risk. This pair of models included the time effect and main effects of message types: assignment to gain-framed messages, assignment to emotional messages, and assignment to simple messages.

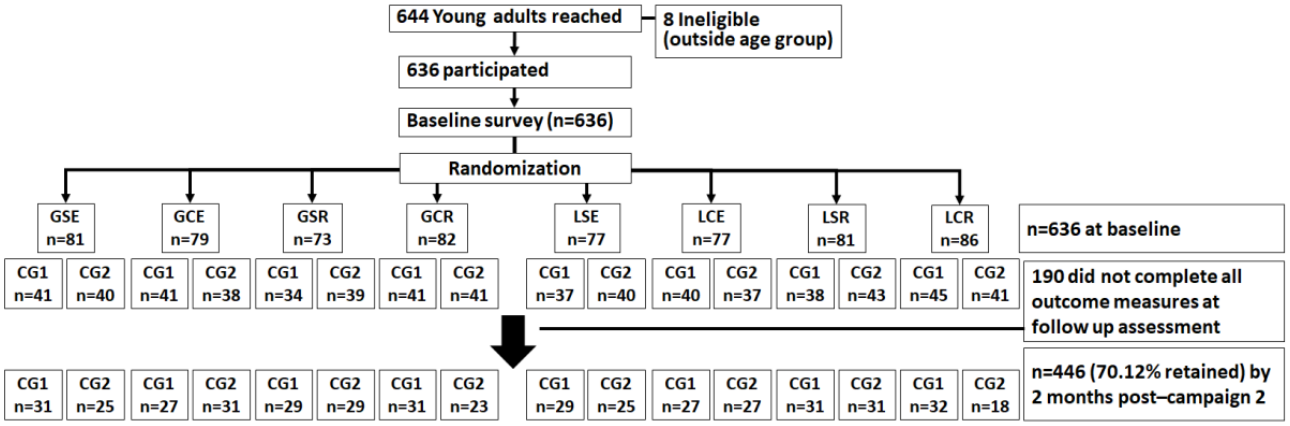
Our central hypothesis is that, controlling for all other message types, campaign participants receiving each message type will have an increase in perceived NETP risk and perceived CTP risk over time. To test the success of improving perceived risk over time among participants exposed to specific message types, 6 models were used for participants receiving (1) rational messages, (2) emotional messages, (3) complex messages, (4) simple messages, (5) loss-framed messages, and (6) gain-framed messages. These 6 models were conducted to predict perceived CTP risk and then perceived NETP risk. All 7 models that predicted each main outcome controlled for crossover group assignment and the differential effect of crossover group assignment on time. In addition, after examining potential covariates through a series of regression analyses, all models controlled for age, gender, having a child, basic expenses, education plan, numeracy level, and past 30-day tobacco use at baseline (Multimedia Appendix 3). All models were fitted with restricted or residual maximum likelihood estimation. Considering 7 models for each main outcome, a P value <.007 (.05 divided by 7) was considered significant, and a P value <.008 was considered marginal [78-81]. We also report the direction of relationships for predictions with P<.05 when they are concerned with our study aim. We used STATA (version 14; StataCorp LLC) for our analyses.

Results

Attrition

Figure 2 presents the study flow diagram. Of the 644 YAs who agreed to participate, we excluded 8 (1.2%) YAs who did not meet the age criterion (aged > 25 years). All 636 participants completed the baseline survey, were randomized to 1 of the 16 conditions, and received the SMS text messages of campaign 1 as prescribed. All participants continued until 2 months post campaign 2; however, 29.9% (190/636) did not complete all outcome measures at the 2-month post campaign 2 assessment (70.1% completion rate).

Figure 2. Study flow diagram. CG: crossover group; CG1 receive messages about new and emerging tobacco products (NETP) during campaign 1 and then messages about conventional tobacco product (CTP) during campaign 2; CG2 receive messages about CTP during campaign 1 and then messages about NETP during campaign 2; GCE: gain-framed, complex, emotional; GCR: gain-framed, complex, rational; GSE: gain-framed, simple, emotional; GSR: gain-framed, simple, rational; LCE: loss-framed, complex, emotional; LSE: loss-framed, simple, emotional; LCR: loss-framed, complex, rational; LSR: loss-framed, simple, rational.



Demographic Characteristics

Demographic characteristics for the entire sample and stratified by group at baseline were described under the trial protocol [60]. In summary, the average age of our study sample was 20.78 (SD 2.18) years, and approximately two-thirds of participants (430/636, 67.6%) were male. Participants exhibited a numeracy level of 4.75 (SD 1.87) out of 8, and 29.1% (185/636) aimed to continue their education to receive a doctorate degree. In addition, 35.1% (223/636) reported that they *just meet* basic expenses. The perceived CTP risk for the sample at baseline was 2.56 (SD 0.69), and it became 2.64 (SD 0.71) out of 3 at follow-up. The perceived NETP risk for the sample at baseline was 2.16 (SD 0.77), and it became 2.41 (SD 0.76) out of 3 at follow-up. Multimedia Appendix 4 presents the demographic characteristics of the sample by group.

G stands for gain-framed, L stands for loss-framed, C stands for complex, S stands for simple message, R stands for rational, and E stands for emotional. CG1 stands for crossover group 1 (receiving messages about NETPs during campaign 1 and then messages about CTPs during campaign 2). CG2 stands for crossover group 2 (receiving messages about CTPs during campaign 1 and then messages about NETPs during campaign 2). Retention is based on completing all survey questions pertaining to perceived NETP and CTP risk.

Checking for the Digital Divide

The results indicate that the groups did not differ with respect to their preferred method of communication (χ^2_7=22.1; *P*=.07), willingness to receive health SMS text messages (χ^2_7=5.9; *P*=.56), or frequency of carrying a phone (χ^2_7=3.4; *P*=.84).

Message Reception

A series of 1-way analyses of variance indicated no significant difference in attention scores between emotional and rational messages ($F_{1,421}$<0.001; *P*=.99), simple and complex messages ($F_{1,421}$=1.09; *P*=.29), and gain-framed and loss-framed messages ($F_{1,421}$=0.86; *P*=.35). Only approximately 2.8% (12/423) of participants indicated not paying attention to any of the messages. The average score for attention to the messages (mean 3.16, SD 0.93) was significantly higher than 3 out of 5, indicating paying attention to more than *some of the messages* (*P*<.001).

Overall Campaign Outcomes

Table 1 presents the overall changes in the main outcomes over time. By the end of both campaigns, as indicated by the coefficient for time, participants had a significant increase in perceived NETP risk (*P*<.001; Table 1). In this model, although not significant, NETP users generally exhibited lower scores in perceived NETP risk compared with nonusers (*P*=.009). By the end of both campaigns, participants had a marginal increase in perceived CTP risk (*P*=.008; Table 1).

Table 1. Change over time in perceived risk of using NETP[a] and CTP[b] for the sample (N=636).[c]

Characteristics	Perceived NETP risk		Perceived CTP risk	
	B (SE)	*P* value	B (SE)	*P* value
Time[d]	0.23 (0.06)	<.001	0.13 (0.05)	.008
Crossover group	−0.04 (0.06)	.49	−0.04 (0.05)	.45
Crossover group by time	0.02 (0.08)	.79	0.05 (0.07)	.50
Age	0.02 (0.01)	.08	−0.001 (0.01)	.93
Being female	−0.01 (0.06)	.78	0.01 (0.05)	.77
Having a child	−0.05 (0.09)	.55	0.09 (0.08)	.23
Basic expenses				
Just meet	0.08 (0.10)	.44	0.08 (0.09)	.37
Meet adequately	0.14 (0.11)	.19	0.17 (0.09)	.06
Meet comfortably	0.10 (0.11)	.38	0.02 (0.09)	.84
Cannot meet (reference)	__[e]	—	—	—
Education plan				
Associate degree	−0.09 (0.14)	.52	0.12 (0.12)	.33
Bachelor's degree	−0.14 (0.13)	.28	0.003 (0.10)	.19
Master's degree	−0.11 (0.13)	.38	0.06 (0.10)	.08
Doctorate degree	−0.09 (0.13)	.49	0.16 (0.10)	.13
Certificate (reference)	—	—	—	—
Numeracy level	0.01 (0.02)	.48	0.01 (0.12)	.28
Baseline use of NETP[f]	−0.16 (0.06)	.009	—	—
Baseline use of CTP[f]	—	—	−0.10 (0.06)	.10
Receive gain-framed messages[g]	0.10 (0.05)	.05	0.05 (0.04)	.21
Receive emotional messages[g]	0.01 (0.05)	.84	0.02 (0.04)	.57
Receive simple messages[g]	0.13 (0.05)	.01	0.07 (0.04)	.11

[a]NETP: new and emerging tobacco product.

[b]CTP: conventional tobacco product.

[c]Two models are presented in this table. Unstandardized coefficients are presented, and the significance level is examined at .007. Multimedia Appendix 5 presents 95% CIs for each coefficient.

[d]The unadjusted time effect predicting perceived NETP risk was B (SE)=0.24 (0.04), *P*<.001, and the unadjusted time effect predicting perceived CTP risk was B (SE)=0.16 (0.03), *P*<.001.

[e]For reference factors (eg, cannot meet), this indicates that data is not applicable. For actual variables (eg, baseline NETP use), this indicates that the variable was not included in the model.

[f]Baseline use of NETP or CTP indicates past 30-day use of NETPs and CTPs at baseline.

[g]These variables compare receiving 1 message type with its counterpart (gain-framed vs loss-framed, emotional vs rational, and simple vs complex).

Checking for Confounders

To check for potential demographic confounders of perceived CTP risk, we determined whether intervention effects varied by demographic characteristics, particularly those identified as covariates. Overall, the results failed to identify effects as a moderating function of age (*P*=.31), sex (*P*=.35), having a child (*P*=.90), basic expenses (*P*=.26), education level (*P*=.06), numeracy level (*P*=.06), or current tobacco use (*P*=.41). A similar analysis for perceived NETP risk indicated no effects as a function of age (*P*=.51), sex (*P*=.64), having a child (*P*=.75), basic expenses (*P*=.14), education level (*P*=.87), numeracy level (*P*=.05), or current tobacco use (*P*=.26).

Change in Perceived NETP Risk by Type of Message Received

As presented in Table 2, there was a significant increase in perceived NETP risk among participants receiving emotional messages regarding NETPs (*P*=.006). We also observed a nonsignificant increase in perceived NETP risk among participants who received complex messages (*P*=.01) and those who received loss-framed messages (*P*=.01).

Table 3 shows a significant increase in perceived NETP risk among participants who received rational messages (*P*=.005), simple messages (*P*=.003), and gain-framed messages regarding NETPs (*P*=.003).

Table 2. Change in perceived risk of using NETP[a] among participants receiving emotional messages, those receiving complex messages, and those receiving loss-framed messages.[b]

Characteristics	Emotional (n=314)		Complex (n=324)		Loss-framed (n=321)	
	B (SE)	*P* value	B (SE)	*P* value	B (SE)	*P* value
Time[c]	0.24 (0.09)	.006	0.21 (0.08)	.01	0.20 (0.08)	.01
Crossover group	0.03 (0.09)	.72	−0.09 (0.09)	.32	−0.04 (0.09)	.63
Crossover group by time	−0.11 (0.13)	.41	0.11 (0.12)	.34	0.19 (0.12)	.11
Age	0.02 (0.02)	.23	0.02 (0.02)	.15	0.02 (0.02)	.18
Being female	−0.12 (0.08)	.14	0.04 (0.08)	.61	0.08 (0.08)	.29
Having a child	0.04 (0.13)	.76	−0.22 (0.13)	.09	0.16 (0.13)	.23
Basic expenses						
Just meet	0.06 (0.15)	.66	0.06 (0.17)	.71	0.09 (0.14)	.54
Meet adequately	0.08 (0.15)	.59	0.11 (0.17)	.51	0.04 (0.14)	.78
Meet comfortably	0.09 (0.16)	.55	0.04 (0.17)	.84	0.05 (0.15)	.72
Cannot meet (reference)	__[d]	—	—	—	—	—
Education plan						
Associate degree	−0.25 (0.21)	.23	−0.04 (0.20)	.84	−0.29 (0.21)	.17
Bachelor's degree	−0.26 (0.19)	.17	−0.22 (0.18)	.22	−0.22 (0.19)	.24
Master's degree	−0.23 (0.19)	.20	−0.22 (0.18)	.22	−0.26 (0.19)	.15
Doctorate degree	−0.23 (0.19)	.22	−0.19 (0.18)	.28	−0.25 (0.19)	.18
Certificate (reference)	—	—	—	—	—	—
Numeracy level	0.03 (0.02)	.15	−0.02 (0.02)	.36	−0.02 (0.02)	.43
Baseline NETP use[e]	−0.22 (0.09)	.009	−0.16 (0.08)	.06	−0.02 (0.08)	.01
Receive simple messages[f]	0.06 (0.07)	.41	—	—	0.15 (0.07)	.04
Receive gain-framed messages[f]	0.16 (0.07)	.04	0.13 (0.07)	.08	—	—
Receive emotional messages[f]	—	—	0.09 (0.07)	.20	−0.02 (0.07)	.78

[a]NETP: new and emerging tobacco product.

[b]Three models are presented in this table. Unstandardized coefficients are presented with the significance level at .007. Multimedia Appendix 5 presents 95% CIs for each coefficient.

[c]The unadjusted time effects for participants receiving emotional, complex, and loss-framed messages were B(SE)=0.18 (0.06), *P*=.004, B(SE)=0.26 (0.06), *P*<.001, and B(SE)=0.30 (0.06), *P*<.001, respectively.

[d]For reference factors (eg, cannot meet), this indicates that data is not applicable. For actual variables (eg, receive simple messages), this indicates that the variable was not included in the model.

[e]Baseline NETP use indicates past 30-day use of NETPs at baseline.

[f]These variables compare receiving 1 message type with its counterpart (gain-framed vs loss-framed, emotional vs rational, and simple vs complex).

Table 3. Change in perceived risk of using NETP[a] among participants receiving rational messages, those receiving simple messages, and those receiving gain-framed messages.[b]

Characteristics	Rational (n=322)		Simple (n=312)		Gain-framed (n=315)	
	B (SE)	*P* value	B (SE)	*P* value	B (SE)	*P* value
Time[c]	0.22 (0.08)	.005	0.25 (0.08)	.003	0.25 (0.08)	.002
Crossover group	−0.12 (0.09)	.16	−0.02 (0.09)	.85	−0.04 (0.09)	.65
Crossover group by time	0.14 (0.11)	.21	−0.08 (0.12)	.50	−0.14 (0.12)	.24
Age	0.02 (0.02)	.16	0.02 (0.02)	.27	0.02 (0.02)	.17
Being female	0.07 (0.08)	.36	−0.07 (0.08)	.39	−0.10 (0.08)	.21
Having a child	−0.16 (0.13)	.22	0.12 (0.13)	.33	−0.22 (0.13)	.10
Basic expenses						
Just meet	0.01 (0.15)	.52	0.10 (0.14)	.45	0.04 (0.16)	.80
Meet adequately	0.16 (0.15)	.28	0.15 (0.14)	.28	0.20 (0.16)	.22
Meet comfortably	0.09 (0.16)	.58	0.16 (0.14)	.27	0.13 (0.16)	.43
Cannot meet (reference)	—[d]	—	—	—	—	—
Education plan						
Associate degree	0.04 (0.20)	.82	−0.17 (0.20)	.39	0.07 (0.19)	.71
Bachelor's degree	−0.02 (0.18)	.88	−0.04 (0.18)	.83	−0.09 (0.17)	.60
Master's degree	−0.001 (0.18)	.99	0.03 (0.18)	.87	−0.01 (0.17)	.96
Doctorate degree	0.04 (0.18)	.80	0.03 (0.17)	.87	0.04 (0.17)	.82
Certificate (reference)	—	—	—	—	—	—
Numeracy level	−0.003 (0.02)	.86	0.04 (0.02)	.04	0.03 (0.02)	.13
Baseline NETP use[e]	−0.09 (0.09)	.31	−0.16 (0.09)	.06	−0.10 (0.09)	.26
Receive simple messages[f]	0.21 (0.07)	.003	—	—	0.07 (0.07)	.30
Receive gain-framed messages[f]	0.07 (0.07)	.32	0.09 (0.07)	.04	—	—
Receive emotional messages[f]	—	—	−0.07 (0.07)	.34	0.03 (0.07)	.67

[a]NETP: new and emerging tobacco product.

[b]Three models are presented in this table. Unstandardized coefficients are presented, and the significance level is examined at .007. Multimedia Appendix 5 presents 95% CIs for each coefficient.

[c]The unadjusted time effects among participants receiving rational, simple, and gain-framed messages were B(SE)=0.30 (0.06), *P*<.001, B(SE)=0.22 (0.06), *P*<.001, and B (SE)=0.18 (0.06), *P*=.002, respectively.

[d]For reference factors (eg, cannot meet), this indicates that data is not applicable. For actual variables (eg, receive simple messages), this indicates that the variable was not included in the model.

[e]Baseline NETP use indicates past 30-day use of NETPs at baseline.

[f]These variables compare receiving 1 message type with its counterpart (gain-framed vs loss-framed, emotional vs rational, and simple vs complex).

Change in Perceived CTP Risk by Type of Message Received

Table 4 presents an observable nonsignificant increase in perceived CTP risk among participants who received emotional messages (*P*=.01), those who received complex messages (*P*=.03), and those who received loss-framed messages (*P*=.01). Among participants receiving complex messages, higher numeracy levels were significantly related to an increase in perceived CTP risk (*P*=.006).

On the other hand, as presented in Table 5, there was no significant increase in perceived CTP risk among participants who received rational (*P*=.20), simple (*P*=.11), or gain-framed messages (*P*=.23). Among participants who received simple messages, numeracy level (*P*=.006) and current CTP use (*P*<.001) were significantly related to higher perceived CTP risk.

Table 4. Change in perceived risk of using CTP[a] among participants receiving emotional messages, those receiving complex messages, and those receiving loss-framed messages.[b]

Characteristics	Emotional (n=314)		Complex (n=324)		Loss-framed (n=321)	
	B (SE)	*P* value	B (SE)	*P* value	B (SE)	*P* value
Time[c]	0.02 (0.07)	.01	0.16 (0.07)	.03	0.18 (0.07)	.01
Crossover group	0.17 (0.07)	.76	−0.09 (0.08)	.24	−0.01 (0.08)	.94
Crossover group by time	−0.05 (0.10)	.64	0.03 (0.10)	.76	0.06 (0.10)	.54
Age	−0.01 (0.01)	.36	−0.01 (0.02)	.66	−0.01 (0.02)	.63
Being female	−0.01 (0.07)	.94	0.05 (0.07)	.45	0.04 (0.07)	.54
Having a child	0.11 (0.11)	.33	0.04 (0.12)	.71	0.19 (0.11)	.09
Basic expenses						
Just meet	0.25 (0.13)	.06	0.12 (0.15)	.43	0.22 (0.13)	.11
Meet adequately	0.27 (0.13)	.04	0.24 (0.15)	.11	0.23 (0.13)	.07
Meet comfortably	0.13 (0.14)	.35	0.11 (0.15)	.47	0.08 (0.13)	.55
Cannot meet (reference)	__[d]	—	—	—	—	—
Education plan						
Associate degree	−0.08 (0.18)	.68	0.19 (0.17)	.27	0.04 (0.18)	.80
Bachelor's degree	−0.002 (0.16)	.99	0.06 (0.16)	.71	−0.002 (0.16)	.99
Master's degree	0.06 (0.16)	.71	0.08 (0.15)	.59	0.01 (0.16)	.96
Doctorate degree	−0.04 (0.16)	.79	0.10 (0.16)	.52	−0.04 (0.16)	.81
Certificate (reference)	—	—	—	—	—	—
Numeracy level	0.02 (0.02)	.28	−0.02 (0.02)	.31	0.01 (0.02)	.52
Baseline CTP use[e]	−0.16 (0.09)	.08	0.07 (0.09)	.45	−0.14 (0.09)	.13
Receive simple messages[f]	0.04 (0.06)	.53	—	—	0.06 (0.06)	.47
Receive gain-framed messages[f]	0.14 (0.07)	.04	0.04 (0.07)	.57	—	—
Receive emotional messages[f]	—	—	0.06 (0.07)	.32	−0.05 (0.06)	.47

[a]CTP: conventional tobacco product.

[b]Three models are presented in this table. Unstandardized coefficients are presented, and the significance level is examined at .007. Multimedia Appendix 5 presents 95% CIs for each coefficient.

[c]The unadjusted time effects among participants receiving emotional, complex, and loss-framed messages were B (SE)=0.16 (0.05), *P*=.001, B (SE)=0.17 (0.05), *P*=.001, and B (SE)=0.21 (0.05), *P*<.001, respectively.

[d]For reference factors (eg, cannot meet), this indicates that data is not applicable. For actual variables (eg, receive simple messages), this indicates that the variable was not included in the model.

[e]Baseline CTP use indicates past 30-day use of CTPs at baseline.

[f]These variables compare receiving 1 message type with its counterpart (gain-framed vs loss-framed, emotional vs rational, and simple vs complex).

Table 5. Change in perceived risk of using CTP[a] among participants receiving rational messages, those receiving simple messages, and those receiving gain-framed messages.[b]

Characteristics	Rational (n=322)		Simple (n=312)		Gain-framed (n=315)	
	B (SE)	*P* value	B (SE)	*P* value	B (SE)	*P* value
Time[c]	0.09 (0.07)	.20	0.11 (0.07)	.12	0.08 (0.07)	.23
Crossover group	−0.10 (0.07)	.19	0.02 (0.07)	.81	−0.06 (0.07)	.41
Crossover group by time	0.12 (0.10)	.22	0.05 (0.10)	.62	0.03 (0.10)	.75
Age	0.01 (0.01)	.31	0.01 (0.01)	.48	0.01 (0.01)	.65
Being female	0.04 (0.06)	.49	−0.02 (0.06)	.81	−0.002 (0.07)	.98
Having a child	0.08 (0.11)	.47	0.18 (0.10)	.07	0.02 (0.11)	.84
Basic expenses						
Just meet	−0.14 (0.12)	.25	0.06 (0.11)	.58	−0.12 (0.13)	.36
Meet adequately	0.01 (0.13)	.96	0.10 (0.11)	.36	0.04 (0.13)	.78
Meet comfortably	−0.13 (0.13)	.31	−0.05 (0.11)	.67	−0.11 (0.14)	.42
Cannot meet (reference)	—[d]	—	—	—	—	—
Education plan						
Associate degree	0.31 (0.16)	.05	−0.001 (0.16)	1.00	0.18 (0.16)	.26
Bachelor's degree	0.28 (0.14)	.05	0.24 (0.14)	.09	0.24 (0.14)	.09
Master's degree	0.29 (0.14)	.04	0.32 (0.14)	.02	0.32 (0.14)	.02
Doctorate degree	0.35 (0.14)	.01	0.24 (0.14)	.08	0.32 (0.14)	.02
Certificate (reference)	—	—	—	—	—	—
Numeracy level	0.01 (0.02)	.65	0.04 (0.01)	.004	0.01 (0.02)	.37
Baseline CTP use[e]	−0.03 (0.08)	.74	−0.30 (0.08)	<.001	−0.05 (0.08)	.51
Receive simple messages[f]	0.09 (0.06)	.11	—	—	0.06 (0.06)	.36
Receive gain-framed messages[f]	−0.03 (0.06)	.64	0.10 (0.06)	.09	—	—
Receive emotional messages[f]	—	—	−0.01 (0.06)	.89	0.09 (0.06)	.37

[a]CTP: conventional tobacco product.

[b]Three models are presented in this table. Unstandardized coefficients are presented with the significance level at .007. Multimedia Appendix 5 presents 95% CIs for each coefficient.

[c]The unadjusted time effects among participants receiving rational, simple, and gain-framed messages were B(SE)=0.16 (0.05), *P*=.002, B(SE)=0.14 (0.05), *P*=.004, and B (SE)=0.10 (0.05), *P*=.03, respectively.

[d]For reference factors (eg, cannot meet), this indicates that data is not applicable. For actual variables (eg, receive simple messages), this indicates that the variable was not included in the model.

[e]Baseline CTP use indicates past 30-day use of CTPs at baseline.

[f]These variables compare receiving 1 message type with its counterpart (gain-framed vs loss-framed, emotional vs rational, and simple vs complex).

Discussion

Principal Findings

Tobacco marketing has successfully crafted messages to promote tobacco use among the general public, particularly among YAs. As a result, there has been evidence of limited public knowledge concerning the harms of tobacco products [82,83], particularly NETPs such as e-cigarettes and hookahs [84]. The use of mobile phones in the United States is nearly ubiquitous. This study expands on previous research by identifying exceptionally successful types of SMS text messages that can be used to correct YAs' perceptions of tobacco risk. Controlling for the type of message received, our results show that YAs had a significant increase in perceived NETP risk. However, when stratifying by message type, we observed a significant increase over time in perceived NETP risk upon exposure to emotional, rational, simple, and gain-framed messages. In addition, YAs generally had an increase in perceived CTP risk. Although not significant, after stratification, we observed an increase in perceived CTP risk upon exposure to emotional, complex, and loss-framed messages.

Previous research on risk perception among YAs supports our results pertaining to emotional and rational messages [85], and it is in line with the ELM of persuasion [61]. In particular, the

harms of NETPs are still unfamiliar to young populations and YAs. To avoid a higher-level effort to process messages regarding NETPs [86], participants receiving simple messages had an increase in perceived NETP risk, whereas participants receiving complex messages did not. This supports the need for simple messages to convey information that requires more effort to understand. Similarly, with a need to engage in peripheral processing, participants who received emotional messages also improved in perceived NETP risk. Interestingly, rational messages also produced a significant increase in perceived NETP risk, indicating the success of central processing of information on NETPs. Conversely, YAs tend to be more familiar with the harms of CTPs [86]. With less need for effortful cognitive information processing, although nonsignificant, participants receiving emotional messages and those receiving complex messages produced an observable increase in perceived CTP risk, whereas participants receiving rational messages and those receiving simple messages did not. Although the ELM may explain these results, it is key for future researchers to further examine this interpretation and measure YAs' need for cognition and familiarity with the products.

However, according to previous research, both gain-framed and loss-framed messages can be effective in increasing tobacco risk perception through different mechanisms [87,88]. In this study, although not significant, our results indicate an observable decrease in perceived CTP risk among YAs exposed to loss-framed messages. Supportive of our findings, previous research has frequently indicated that YAs favor loss-framed messages with health risk themes [88,89]. By highlighting the potential losses that result from tobacco use, it is expected that YAs experience sadness and fear, thereby increasing risk perception [87]. On the other hand, YAs who received gain-framed messages had significantly improved perceived NETP risk levels. This finding agrees with previous research indicating that emphasis on the potential benefits of avoiding tobacco can stimulate a sense of guilt responsible for a decrease in risk perception [83]. In addition, the success of our gain-framed messages can be attributed to their design. In particular, these messages did not directly present the benefits gained as a result of avoiding tobacco (ie, not using tobacco helps one gain certain benefits). Instead, most messages described a lack of loss as being the benefit (ie, not using tobacco helps one avoid negative consequences). This type of message framing takes advantage of a gain-framed design while still describing losses. Regardless of the mechanism at play, our results emphasize the effectiveness of messages that communicate a need to *avoid losses* resulting from NETP use. Future research should consider implementing this message structure to improve the perceived NETP risk.

Limitations

There are some study limitations to be considered. First, this study involved a convenience sample. Nevertheless, the sample is representative of the diverse community college population in terms of demographic characteristics and tobacco use among Texan YAs [90,91]. Second, the trial included a wide variety of tobacco products, making it difficult to attribute the outcomes to messages on specific products. However, the distinction between messages on NETPs (eg, vaping products and hookah) and CTPs (combustible and smokeless products) made it possible to identify successful message types for these 2 common groups of products in the United States. Finally, the loss of participants to follow-up with retention of only 70.1% (446/636) of participants may have made it difficult to capture the significance of some of the observed predictions.

Implications

With our current findings, we cannot conclude that one message type is more effective than another. Nevertheless, this study aimed to identify successful message types individually. Our results suggest that specific types of SMS text messages can be particularly successful. On the basis of our findings, we encourage future researchers to apply emotional, complex, and loss-framed messages when conveying the harm of CTPs. On the other hand, we recommend the use of simple and gain-framed messages to inform about the harms of NETPs. These messages may be emotional or rational.

Our messages can be strategically disseminated within campaigns conducted via social media, smartphone apps, or mass media. Our previous research has posited that YAs are interested in mobile health (mHealth) programs that help them learn about tobacco risks [58], and mHealth programs offer the potential to greatly increase the reach of YAs. Such mHealth programming can present rational, simple, and gain-framed messages for communication of the risk of NETPs. Conversely, emotional, complex, and loss-framed messages can be disseminated to communicate the risk of CTPs.

It is important to note that the appropriateness and impact of the messages are likely to be context-dependent, and the results may have limited transferability. Nevertheless, with the results of this study, researchers and practitioners implementing mHealth programs may take advantage of our tailored messages through larger technology-based programs such as smartphone apps and social media campaigns. If a program were to be designed where individuals could opt in to receive the messages, a separate study might be needed to examine the target populations' needs and preferences with respect to these messages. One promising avenue for future research in this area is the integration of these messages into narratives that can facilitate accurate tobacco risk perception. Several studies have begun to consider the investigation of message framing strategies within narratives, indicating that narratives can be successful with both loss-framed and gain-framed messages [88]. In the next step, we plan to examine how the success of narratives can be improved based on message complexity and emotional appeal. Although mHealth SMS text messaging can efficiently and widely communicate tobacco risk, by integrating narrative-based messages, researchers are likely to improve YAs' engagement through message attention and recall of information.

Acknowledgments

The authors would like to thank the leadership at Houston Community College for supporting their research. The authors would also like to thank all members of the scientific steering committee of the Texas Tobacco Center of Regulatory Science on youth and young adults for their support and valuable recommendations throughout this project. The committee includes Jerome Williams (Prudential Chair in Business, at the Rutgers Business School), Cornelia (Connie) Pechmann (Professor at the Paul Merage School of Business, University of California, Irvine), John Pierce (Professor at the Cancer Prevention Program, Moores Cancer Center, University of California, San Diego, La Jolla, California and the Department of Family Medicine and Public Health, University of California, San Diego), Lisa Hendrickson (Senior Research Scientist at the Stanford Prevention Research Center), and Lois Biener (Senior Research Fellow at the University of Massachusetts Boston). This study was funded by the Tobacco Center of Regulatory Science on Youth and Young Adults (5P50CA180906-02).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description and examples of SMS text messages.
[DOCX File , 14 KB - mhealth_v9i11e25618_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 24721 KB - mhealth_v9i11e25618_app2.pdf]

Multimedia Appendix 3
Associations between population characteristics and primary outcomes.
[DOCX File , 32 KB - mhealth_v9i11e25618_app3.docx]

Multimedia Appendix 4
Demographic characteristics and risk perception ratings.
[DOCX File , 16 KB - mhealth_v9i11e25618_app4.docx]

Multimedia Appendix 5
The 95% CIs for each coefficient and for all tables.
[DOCX File , 74 KB - mhealth_v9i11e25618_app5.docx]

References

1. Prokhorov AV, Khalil GE, Foster DW, Marani SK, Guindani M, Espada JP, et al. Testing the nicotine dependence measure mFTQ for adolescent smokers: a multinational investigation. Am J Addict 2017 Oct;26(7):689-696. [doi: 10.1111/ajad.12583] [Medline: 28708935]
2. Dierker L, Rose J, Selya A, Piasecki TM, Hedeker D, Mermelstein R. Depression and nicotine dependence from adolescence to young adulthood. Addict Behav 2015 Feb;41:124-128 [FREE Full text] [doi: 10.1016/j.addbeh.2014.10.004] [Medline: 25452055]
3. Viswam D, Trotter S, Burge PS, Walters GI. Respiratory failure caused by lipoid pneumonia from vaping e-cigarettes. BMJ Case Rep 2018 Jul 06;2018:bcr2018224350 [FREE Full text] [doi: 10.1136/bcr-2018-224350] [Medline: 29982176]
4. King BA, Jones CM, Baldwin GT, Briss PA. The EVALI and youth vaping epidemics - implications for public health. N Engl J Med 2020 Feb 20;382(8):689-691. [doi: 10.1056/NEJMp1916171] [Medline: 31951683]
5. Klakk H, Kristensen PL, Andersen LB, Froberg K, Møller NC, Grøntved A. Symptoms of depression in young adulthood is associated with unfavorable clinical- and behavioral cardiovascular disease risk factors. Prev Med Rep 2018 May 22;11:209-215. [doi: 10.1016/j.pmedr.2018.05.017] [Medline: 30003016]
6. Carlos WG, Crotty Alexander LE, Gross JE, Dela Cruz CS, Keller JM, Pasnick S, et al. ATS Health Alert-Vaping-associated Pulmonary Illness (VAPI). Am J Respir Crit Care Med 2019 Oct 01;200(7):P15-P16. [doi: 10.1164/rccm.2007P15] [Medline: 31532698]
7. Hooper RW, Garfield JL. An emerging crisis: vaping-associated pulmonary injury. Ann Intern Med 2020 Jan 07;172(1):57-58. [doi: 10.7326/M19-2908] [Medline: 31590182]
8. Leventhal AM, Strong DR, Kirkpatrick MG, Unger JB, Sussman S, Riggs NR, et al. Association of electronic cigarette use with initiation of combustible tobacco product smoking in early adolescence. JAMA 2015 Aug 18;314(7):700-707 [FREE Full text] [doi: 10.1001/jama.2015.8950] [Medline: 26284721]

9. Leventhal AM, Stone MD, Andrabi N, Barrington-Trimis J, Strong DR, Sussman S, et al. Association of e-cigarette vaping and progression to heavier patterns of cigarette smoking. JAMA 2016 Nov 08;316(18):1918-1920 [FREE Full text] [doi: 10.1001/jama.2016.14649] [Medline: 27825000]
10. Goldenson NI, Leventhal AM, Stone MD, McConnell RS, Barrington-Trimis JL. Associations of electronic cigarette nicotine concentration with subsequent cigarette smoking and vaping levels in adolescents. JAMA Pediatr 2017 Dec 01;171(12):1192-1199 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.3209] [Medline: 29059261]
11. Soneji S, Barrington-Trimis JL, Wills TA, Leventhal AM, Unger JB, Gibson LA, et al. Association between initial use of e-cigarettes and subsequent cigarette smoking among adolescents and young adults: a systematic review and meta-analysis. JAMA Pediatr 2017 Jun 26;171(8):788-797. [doi: 10.1001/jamapediatrics.2017.1488] [Medline: 28654986]
12. Schulenberg J, Johnston L, O'Malley P, Bachman J, Miech R, Patrick M. Monitoring the Future National Survey Results on Drug Use, 1975-2019: Volume II, College Students and Adults Ages 19-60. Ann Arbor, MI: Institute for Social Research, The University of Michigan; 2020.
13. Berg CJ, Stratton E, Schauer GL, Lewis M, Wang Y, Windle M, et al. Perceived harm, addictiveness, and social acceptability of tobacco products and marijuana among young adults: marijuana, hookah, and electronic cigarettes win. Subst Use Misuse 2015 Jan;50(1):79-89 [FREE Full text] [doi: 10.3109/10826084.2014.958857] [Medline: 25268294]
14. Tan AS, Mello S, Sanders-Jackson A, Bigman CA. Knowledge about chemicals in e-cigarette secondhand vapor and perceived harms of exposure among a national sample of U.S. Adults. Risk Anal 2017 Jun;37(6):1170-1180 [FREE Full text] [doi: 10.1111/risa.12676] [Medline: 27595498]
15. Gowin M, Cheney MK, Wann TF. Knowledge and beliefs about e-cigarettes in straight-to-work young adults. Nicotine Tob Res 2017 Feb;19(2):208-214. [doi: 10.1093/ntr/ntw195] [Medline: 27613919]
16. Pokhrel P, Lam TH, Pagano I, Kawamoto CT, Herzog TA. Young adult e-cigarette use outcome expectancies: validity of a revised scale and a short scale. Addict Behav 2018 Mar;78:193-199. [doi: 10.1016/j.addbeh.2017.11.019] [Medline: 29195147]
17. Hair E, Rath JM, Pitzer L, Emelle B, Ganz O, Halenar MJ, et al. Trajectories of hookah use: harm perceptions from youth to young adulthood. Am J Health Behav 2017 May 01;41(3):240-247. [doi: 10.5993/AJHB.41.3.3] [Medline: 28376968]
18. O'Brien EK, Nguyen AB, Persoskie A, Hoffman AC. U.S. adults' addiction and harm beliefs about nicotine and low nicotine cigarettes. Prev Med 2017 Mar;96:94-100. [doi: 10.1016/j.ypmed.2016.12.048] [Medline: 28034733]
19. McDonald EA, Ling PM. One of several 'toys' for smoking: young adult experiences with electronic cigarettes in New York City. Tob Control 2015 Nov;24(6):588-593 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051743] [Medline: 25564287]
20. Rodu B, Plurphanswat N, Hughes JR, Fagerström K. Associations of proposed relative-risk warning labels for snus with perceptions and behavioral intentions among tobacco users and nonusers. Nicotine Tob Res 2016 May;18(5):809-816 [FREE Full text] [doi: 10.1093/ntr/ntv168] [Medline: 26253616]
21. Wagoner KG, Cornacchione J, Wiseman KD, Teal R, Moracco KE, Sutfin EL. E-cigarettes, hookah pens and vapes: adolescent and young adult perceptions of electronic nicotine delivery systems. Nicotine Tob Res 2016 Oct;18(10):2006-2012. [doi: 10.1093/ntr/ntw095] [Medline: 27029821]
22. Berry C, Burton S, Howlett E. Are cigarette smokers', e-cigarette users', and dual users' health-risk beliefs and responses to advertising influenced by addiction warnings and product type? Nicotine Tob Res 2017 Oct 01;19(10):1185-1191. [doi: 10.1093/ntr/ntx075] [Medline: 28379568]
23. Kozlowski LT, Sweanor DT. Young or adult users of multiple tobacco/nicotine products urgently need to be informed of meaningful differences in product risks. Addict Behav 2018 Jan;76:376-381 [FREE Full text] [doi: 10.1016/j.addbeh.2017.01.026] [Medline: 28148394]
24. Wong EC, Haardörfer R, Windle M, Berg CJ. Distinct motives for use among polytobacco versus cigarette only users and among single tobacco product users. Nicotine Tob Res 2017 Dec 13;20(1):117-123. [doi: 10.1093/ntr/ntw284] [Medline: 27798088]
25. Noland M, Ickes MJ, Rayens MK, Butler K, Wiggins AT, Hahn EJ. Social influences on use of cigarettes, e-cigarettes, and hookah by college students. J Am Coll Health 2016;64(4):319-328 [FREE Full text] [doi: 10.1080/07448481.2016.1138478] [Medline: 26822236]
26. Mays D, Arrazola RA, Tworek C, Rolle IV, Neff LJ, Portnoy DB. Openness to using non-cigarette tobacco products among U.S. young adults. Am J Prev Med 2016 Apr;50(4):528-534 [FREE Full text] [doi: 10.1016/j.amepre.2015.08.015] [Medline: 26549502]
27. Loukas A, Paddock EM, Li X, Harrell MB, Pasch KE, Perry CL. Electronic nicotine delivery systems marketing and initiation among youth and young adults. Pediatrics 2019 Sep;144(3):e20183601 [FREE Full text] [doi: 10.1542/peds.2018-3601] [Medline: 31451608]
28. Pierce JP, Sargent JD, Portnoy DB, White M, Noble M, Kealey S, et al. Association between receptivity to tobacco advertising and progression to tobacco use in youth and young adults in the path study. JAMA Pediatr 2018 May 1;172(5):444-451. [doi: 10.1001/jamapediatrics.2017.5756] [Medline: 29582078]

29. O'Brien EK, Hoffman L, Navarro MA, Ganz O. Social media use by leading US e-cigarette, cigarette, smokeless tobacco, cigar and hookah brands. Tob Control 2020 Dec;29(e1):e87-e97. [doi: 10.1136/tobaccocontrol-2019-055406] [Medline: 32217772]
30. Padon AA, Maloney EK, Cappella JN. Youth-targeted e-cigarette marketing in the US. Tob Regul Sci 2017 Jan;3(1):95-101 [FREE Full text] [doi: 10.18001/TRS.3.1.9] [Medline: 28083545]
31. Bahreinifar S, Sheon NM, Ling PM. Is snus the same as dip? Smokers' perceptions of new smokeless tobacco advertising. Tob Control 2013 Mar;22(2):84-90 [FREE Full text] [doi: 10.1136/tobaccocontrol-2011-050022] [Medline: 21972063]
32. Carter OB, Donovan R, Jalleh G. Using viral e-mails to distribute tobacco control advertisements: an experimental investigation. J Health Commun 2011 Aug;16(7):698-707. [doi: 10.1080/10810730.2011.551998] [Medline: 21432712]
33. Perez DA, Grunseit AC, Rissel C, Kite J, Cotter T, Dunlop S, et al. Tobacco promotion 'below-the-line': exposure among adolescents and young adults in NSW, Australia. BMC Public Health 2012 Jun 12;12:429 [FREE Full text] [doi: 10.1186/1471-2458-12-429] [Medline: 22691578]
34. Jane LM, Bover MM, Delnevo CD. Tobacco industry direct mail receipt and coupon use among young adult smokers. Prev Med 2015 Feb;71:37-39 [FREE Full text] [doi: 10.1016/j.ypmed.2014.11.030] [Medline: 25511177]
35. Richardson A, Ganz O, Pearson J, Celcis N, Vallone D, Villanti AC. How the industry is marketing menthol cigarettes: the audience, the message and the medium. Tob Control 2015 Nov;24(6):594-600. [doi: 10.1136/tobaccocontrol-2014-051657] [Medline: 25178275]
36. Richardson A, Ganz O, Vallone D. Tobacco on the web: surveillance and characterisation of online tobacco and e-cigarette advertising. Tob Control 2015 Jul;24(4):341-347. [doi: 10.1136/tobaccocontrol-2013-051246] [Medline: 24532710]
37. Carpenter CM, Wayne GF, Pauly JL, Koh HK, Connolly GN. New cigarette brands with flavors that appeal to youth: tobacco marketing strategies. Health Aff (Millwood) 2005;24(6):1601-1610. [doi: 10.1377/hlthaff.24.6.1601] [Medline: 16284034]
38. Ling PM, Glantz SA. Why and how the tobacco industry sells cigarettes to young adults: evidence from industry documents. Am J Public Health 2002 Jun;92(6):908-916. [Medline: 12036776]
39. Rainie L, Perrin A. 10 facts about smartphones as the iPhone turns 10. Pew Research Center. 2017. URL: http://www.pewresearch.org/fact-tank/2017/06/28/10-facts-about-smartphones/ [accessed 2021-11-10]
40. Reynolds R. Camel Snus: we had a feeling you would be stopping. URL: https://snus.tobaccopleasure.com/modules/ [accessed 2021-11-10]
41. Marketing Practices. Philip Morris USA. URL: https://www.philipmorrisusa.com/products/marketing-practices [accessed 2021-11-06]
42. Sears CG, Walker KL, Hart JL, Lee AS, Siu A, Smith C. Clean, cheap, convenient: promotion of electronic cigarettes on YouTube. Tob Prev Cessat 2017 Apr;3:1-5 [FREE Full text] [doi: 10.18332/tpc/69393] [Medline: 28725876]
43. Huang J, Kornfield R, Emery SL. 100 Million Views of Electronic Cigarette YouTube Videos and Counting: Quantification, Content Evaluation, and Engagement Levels of Videos. J Med Internet Res 2016 Mar 18;18(3):e67 [FREE Full text] [doi: 10.2196/jmir.4265] [Medline: 26993213]
44. BinDhim NF, Freeman B, Trevena L. Pro-smoking apps: where, how and who are most at risk. Tob Control 2015 Mar;24(2):159-161. [doi: 10.1136/tobaccocontrol-2013-051189] [Medline: 24046212]
45. Research priorities. US Food and Drug Administration. URL: https://www.fda.gov/TobaccoProducts/PublicHealthScienceResearch/Research/ucm311860.htm [accessed 2021-11-10]
46. Ashley DL, Backinger CL, van Bemmel DM, Neveleff DJ. Tobacco regulatory science: research to inform regulatory action at the Food and Drug Administration's Center for Tobacco Products. Nicotine Tob Res 2014 Aug;16(8):1045-1049. [doi: 10.1093/ntr/ntu038] [Medline: 24638850]
47. Muench F, Weiss RA, Kuerbis A, Morgenstern J. Developing a theory driven text messaging intervention for addiction care with user driven content. Psychol Addict Behav 2013 Mar;27(1):315-321 [FREE Full text] [doi: 10.1037/a0029963] [Medline: 22963375]
48. Scott-Sheldon LAJ, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
49. Obermayer JL, Riley WT, Asif O, Jean-Mary J. College smoking-cessation using cell phone text messaging. J Am Coll Health 2004;53(2):71-78 [FREE Full text] [doi: 10.3200/JACH.53.2.71-78] [Medline: 15495883]
50. Riley W, Obermayer J, Jean-Mary J. Internet and mobile phone text messaging intervention for college smokers. J Am Coll Health 2008;57(2):245-248. [doi: 10.3200/JACH.57.2.245-248] [Medline: 18809542]
51. Sandrick J, Tracy D, Eliasson A, Roth A, Bartel J, Simko M, et al. Effect of a counseling session bolstered by text messaging on self-selected health behaviors in college students: a preliminary randomized controlled trial. JMIR Mhealth Uhealth 2017 May 17;5(5):e67 [FREE Full text] [doi: 10.2196/mhealth.6638] [Medline: 28526664]
52. Head KJ, Noar SM, Iannarino NT, Grant HN. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
53. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]

54. Latimer AE, Krishnan-Sarin S, Cavallo DA, Duhig A, Salovey P, O'Malley SA. Targeted smoking cessation messages for adolescents. J Adolesc Health 2012 Jan;50(1):47-53. [doi: 10.1016/j.jadohealth.2011.04.013] [Medline: 22188833]
55. Toll BA, Salovey P, O'Malley SS, Mazure CM, Latimer A, McKee SA. Message framing for smoking cessation: the interaction of risk perceptions and gender. Nicotine Tob Res 2008 Jan;10(1):195-200 [FREE Full text] [doi: 10.1080/14622200701767803] [Medline: 18188760]
56. Biener L, McCallum-Keeler G, Nyman AL. Adults' response to Massachusetts anti-tobacco television advertisements: impact of viewer and advertisement characteristics. Tob Control 2000 Dec;9(4):401-407 [FREE Full text] [Medline: 11106710]
57. Steward WT, Schneider TR, Pizarro J, Salovey P. Need for cognition moderates responses to framed smoking-cessation messages. J Appl Social Pyschol 2003 Dec;33(12):2439-2464. [doi: 10.1111/j.1559-1816.2003.tb02775.x]
58. Prokhorov AV, Machado TC, Calabro KS, Vanderwater EA, Vidrine DJ, Pasch KP, et al. Developing mobile phone text messages for tobacco risk communication among college students: a mixed methods study. BMC Public Health 2017 Dec 31;17(1):137 [FREE Full text] [doi: 10.1186/s12889-017-4027-z] [Medline: 28143432]
59. Khalil GE, Calabro KS, Crook B, Machado TC, Perry CL, Prokhorov AV. Validation of mobile phone text messages for nicotine and tobacco risk communication among college students: a content analysis. Tob Prev Cessat 2018 Feb;4:1-12. [doi: 10.18332/tpc/84866] [Medline: 29888338]
60. Prokhorov AV, Khalil GE, Calabro KS, Machado TC, Russell S, Czerniak KW, et al. Mobile phone text messaging for tobacco risk communication among young adult community college students: protocol and baseline overview for a randomized controlled trial. JMIR Res Protoc 2018 Oct 15;7(10):e10977 [FREE Full text] [doi: 10.2196/10977] [Medline: 30322833]
61. Petty R, Cacioppo J. The elaboration likelihood model of persuasion. In: Advances in Experimental Social Psychology. New York, NY: Springer; 1986:123-205.
62. O'Keefe D. Elaboration likelihood model. In: The International Encyclopedia of Communication. New York: John Wiley & Sons; 2008:28-27.
63. Chen S, Lee K. The role of personality traits and perceived values in persuasion: an elaboration likelihood model perspective on online shopping. Soc Behav Personal 2008 Nov 01;36(10):1379-1399. [doi: 10.2224/sbp.2008.36.10.1379]
64. Lang A, Yegiyan NS. Understanding the interactive effects of emotional appeal and claim strength in health messages. J Broadcasting Elect Media 2008 Aug 08;52(3):432-447. [doi: 10.1080/08838150802205629]
65. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. In: MacLean LC, Ziemba WT, editors. Handbook of The Fundamentals of Financial Decision Making. Singapore: World Scientific Publishing Co Pte Ltd; 2013:99-127.
66. Kühberger. The influence of framing on risky decisions: a meta-analysis. Organ Behav Hum Decis Process 1998 Jul;75(1):23-55. [Medline: 9719656]
67. O'Keefe DJ, Jensen JD. The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease prevention behaviors: a meta-analytic review. J Health Commun 2007;12(7):623-644. [doi: 10.1080/10810730701615198] [Medline: 17934940]
68. 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]
69. Eysenbach G. CONSORT-EHEALTH: implementation of a checklist for authors and editors to improve reporting of web-based and mobile randomized controlled trials. Stud Health Technol Inform 2013;192:657-661. [Medline: 23920638]
70. 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]
71. Houston community college fact book. Houston Community College. 2017. URL: https://www.hccs.edu/about-hcc/institutional-research/hcc-fact-book/FactBook2018.pdf [accessed 2021-11-10]
72. Johnston LD, Malley PM, Bachman JG, Schulenberg JE, Miech RA. Monitoring the future national survey results on drug use, 1975-2015: Volume ii, college students and adults ages 19-55. Monitoring the Future. 2016. URL: http://monitoringthefuture.org/pubs/monographs/mtf-vol2_2016.pdf [accessed 2021-11-10]
73. Nelson W, Reyna VF, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: theory and practice. Ann Behav Med 2008 Jun;35(3):261-274 [FREE Full text] [doi: 10.1007/s12160-008-9037-8] [Medline: 18677452]
74. Wray RJ, Jupka K, Berman S, Zellin S, Vijaykumar S. Young adults' perceptions about established and emerging tobacco products: results from eight focus groups. Nicotine Tob Res 2012 Feb;14(2):184-190. [doi: 10.1093/ntr/ntr168] [Medline: 22110049]
75. Rath JM, Villanti AC, Abrams DB, Vallone DM. Patterns of tobacco use and dual use in US young adults: the missing link between youth prevention and adult cessation. J Environ Public Health 2012;2012:679134 [FREE Full text] [doi: 10.1155/2012/679134] [Medline: 22666279]
76. Johnston L, O'Malley P, Bachman J, Schulenberg J. Monitoring the Future National Results on Adolescent Drug Use: Overview of Key Findings. Bethesda, MD: National Institute on Drug Abuse; 2009.
77. Loukas A, Batanova M, Fernandez A, Agarwal D. Changes in use of cigarettes and non-cigarette alternative products among college students. Addict Behav 2015 Oct;49:46-51. [doi: 10.1016/j.addbeh.2015.05.005] [Medline: 26046401]

78. Kutner MH, Nachtsheim CJ, Neter J, Li W. Applied linear statistical models. Irwin, NY: McGraw-Hill; 2005.
79. Groenwold RH, Goeman JJ, Cessie SL, Dekkers OM. Multiple testing: when is many too much? Eur J Endocrinol 2021 Mar;184(3):E11-E14. [doi: 10.1530/EJE-20-1375] [Medline: 33300887]
80. Motulsky H. Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking. USA: Oxford University Press; 2010.
81. Feise RJ. Do multiple outcome measures require p-value adjustment? BMC Med Res Methodol 2002 Jun 17;2:8 [FREE Full text] [doi: 10.1186/1471-2288-2-8] [Medline: 12069695]
82. Case K, Crook B, Lazard A, Mackert M. Formative research to identify perceptions of e-cigarettes in college students: implications for future health communication campaigns. J Am Coll Health 2016 Dec;64(5):380-389 [FREE Full text] [doi: 10.1080/07448481.2016.1158180] [Medline: 26979833]
83. Cornacchione J, Wagoner KG, Wiseman KD, Kelley D, Noar SM, Smith MH, et al. Adolescent and young adult perceptions of hookah and little cigars/cigarillos: implications for risk messages. J Health Commun 2016 Jul;21(7):818-825 [FREE Full text] [doi: 10.1080/10810730.2016.1177141] [Medline: 27337629]
84. Wiseman KD, Cornacchione J, Wagoner KG, Noar SM, Moracco KE, Teal R, et al. Adolescents' and young adults' knowledge and beliefs about constituents in novel tobacco products. Nicotine Tob Res 2016 Jul;18(7):1581-1587 [FREE Full text] [doi: 10.1093/ntr/ntw009] [Medline: 26764259]
85. Vidrine JI, Simmons VN, Brandon TH. Construction of smoking-relevant risk perceptions among college students: the influence of need for cognition and message content. J Appl Social Pyschol 2007 Jan;37(1):91-114. [doi: 10.1111/j.0021-9029.2007.00149.x]
86. Copeland AL, Peltier MR, Waldo K. Perceived risk and benefits of e-cigarette use among college students. Addict Behav 2017 Aug;71:31-37. [doi: 10.1016/j.addbeh.2017.02.005] [Medline: 28242533]
87. Liu S, Yang JZ. Incorporating Message Framing into Narrative Persuasion to Curb E-Cigarette Use Among College Students. Risk Anal 2020 Aug;40(8):1677-1690. [doi: 10.1111/risa.13502] [Medline: 32390210]
88. Liu S, Yang JZ. The role of temporal distance perception in narrative vs. non-narrative persuasion related to e-cigarettes. J Health Commun 2020 Jul 16;25(7):543-553. [doi: 10.1080/10810730.2020.1788678] [Medline: 32673189]
89. Kong G, Cavallo DA, Camenga DR, Morean ME, Krishnan-Sarin S. Preference for gain- or loss-framed electronic cigarette prevention messages. Addict Behav 2016 Nov;62:108-113 [FREE Full text] [doi: 10.1016/j.addbeh.2016.06.015] [Medline: 27344117]
90. Perry CL, Creamer MR, Chaffee BW, Unger JB, Sutfin EL, Kong G, et al. Research on youth and young adult tobacco use, 2013-2018, from the Food and Drug Administration-national institutes of health tobacco centers of regulatory science. Nicotine Tob Res 2020 Jun 12;22(7):1063-1076 [FREE Full text] [doi: 10.1093/ntr/ntz059] [Medline: 31127298]
91. Bandiera FC, Loukas A, Li X, Wilkinson AV, Perry CL. Depressive symptoms predict current e-cigarette use among college students in Texas. Nicotine Tob Res 2017 Sep 01;19(9):1102-1106. [doi: 10.1093/ntr/ntx014] [Medline: 28199689]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: CONSORT–Electronic and Mobile Health Applications and Online TeleHealth
CTP: conventional tobacco product
ELM: elaboration likelihood model
mHealth: mobile health
NETP: new and emerging tobacco product
YA: young adult

Edited by L Buis; submitted 13.11.20; peer-reviewed by J Westmaas, A Doyle, E O'Brien, A Lazard, A Galil; comments to author 13.02.21; revised version received 19.04.21; accepted 08.10.21; published 24.11.21.

Please cite as:
Prokhorov AV, Calabro KS, Arya A, Russell S, Czerniak KW, Botello GC, Chen M, Yuan Y, Perez A, Vidrine DJ, Perry CL, Khalil GE
Mobile Text Messaging for Tobacco Risk Communication Among Young Adult Community College Students: Randomized Trial of Project Debunk
JMIR Mhealth Uhealth 2021;9(11):e25618
URL: https://mhealth.jmir.org/2021/11/e25618
doi:10.2196/25618
PMID:34822339

Original Paper

Mechanisms of Smartphone Apps for Cigarette Smoking Cessation: Results of a Serial Mediation Model From the iCanQuit Randomized Trial

Jonathan B Bricker[1,2], PhD; Michael Levin[3], PhD; Raimo Lappalainen[4], PhD; Kristin Mull[1], MS; Brianna Sullivan[1], MS; Margarita Santiago-Torres[1], PhD

[1]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, United States

[2]Department of Psychology, University of Washington, Seattle, WA, United States

[3]Department of Psychology, Utah State University, Logan, UT, United States

[4]Department of Psychology, University of Jyvaskyla, Jyvaskyla, Finland

Corresponding Author:
Jonathan B Bricker, PhD
Division of Public Health Sciences
Fred Hutchinson Cancer Research Center
1100 Fairview Avenue North
M3-B232
Seattle, WA, 98109
United States
Phone: 1 2066675074
Email: jbricker@fredhutch.org

Abstract

Background: Engagement with digital interventions is a well-known predictor of treatment outcomes, but this knowledge has had limited actionable value. Instead, learning why engagement with digital interventions impact treatment outcomes can lead to targeted improvements in their efficacy.

Objective: This study aimed to test a serial mediation model of an Acceptance and Commitment Therapy (ACT) smartphone intervention for smoking cessation.

Methods: In this randomized controlled trial, participants (N=2415) from 50 US states were assigned to the ACT-based smartphone intervention (iCanQuit) or comparison smartphone intervention (QuitGuide). Their engagement with the apps (primary measure: number of logins) was measured during the first 3 months, ACT processes were measured at baseline and 3 months (acceptance of internal cues to smoke, valued living), and smoking cessation was measured at 12 months with 87% follow-up retention.

Results: There was a significant serial mediation effect of iCanQuit on smoking cessation through multiple indicators of intervention engagement (ie, total number of logins, total number of minutes used, and total number of unique days of use) and in turn through increases in mean acceptance of internal cues to smoke from baseline to 3 months. Analyses of the acceptance subscales showed that the mediation was through acceptance of physical sensations and emotions, but not acceptance of thoughts. There was no evidence that the effect of the iCanQuit intervention was mediated through changes in valued living.

Conclusions: In this first study of serial mediators underlying the efficacy of smartphone apps for smoking cessation, our results suggest the effect of the iCanQuit ACT-based smartphone app on smoking cessation was mediated through multiple indicators of engagement and in turn through increases in the acceptance of physical sensations and emotions that cue smoking.

Trial Registration: Clinical Trials.gov NCT02724462; https://clinicaltrials.gov/ct2/show/NCT02724462

(JMIR Mhealth Uhealth 2021;9(11):e32847) doi:10.2196/32847

KEYWORDS

mediation; engagement; digital; mHealth: smartphone; acceptance; smoking; cessation; app; randomized controlled trial; model; intervention

Introduction

Cigarette smoking is a leading cause of premature death and disability [1], attributable to over 1 in 10 deaths worldwide [2]. Barriers to accessing evidence-based smoking cessation treatments include low reimbursement for providers and low demand for in-person treatment [3]. Smartphone apps for smoking cessation have been addressing access barriers by serving as digital interventions with high population-level reach [4]. In the United States, the reach of smartphone apps for smoking cessation has been aided by the fact that as of 2019, 81% of all adults owned smartphones—up from 35% in 2011 [5].

Despite their high population-level reach, very little is known about the potential mediators underlying the efficacy of smartphone apps for smoking cessation [6]. In the broader literature on digital interventions (eg, websites and SMS text messaging) for smoking cessation, we are aware of only 3 randomized controlled trials (RCTs) that reported on their mechanisms of action—with each showing support for the theoretical models guiding their interventions (eg, self-efficacy) [7-9]. Understanding mediators is critical for making future improvements to and guiding optimizations of these behavioral interventions [10]. Intervention components that target specific mechanisms of action can be enhanced, with the goal of creating cost-effective changes to increase intervention efficacy, thereby increasing overall impact. Mediational analysis provides a method to identify potential causal links through which the intervention is efficacious [11].

We recently developed and tested iCanQuit, an Acceptance and Commitment Therapy (ACT)–based smartphone app for smoking cessation [12]. In a large 2-arm RCT, iCanQuit was compared to QuitGuide, a US Clinical Practice Guidelines (USCPG)–based smartphone app. At the 12-month follow-up, iCanQuit was 1.5 times more efficacious than QuitGuide for smoking cessation among 2415 smokers (36% racial/ethnic minority groups) from all 50 US states [12]. The importance of the iCanQuit study is that it is the first full-scale RCT with long-term follow-up to show that a smartphone app was efficacious for smoking cessation [4].

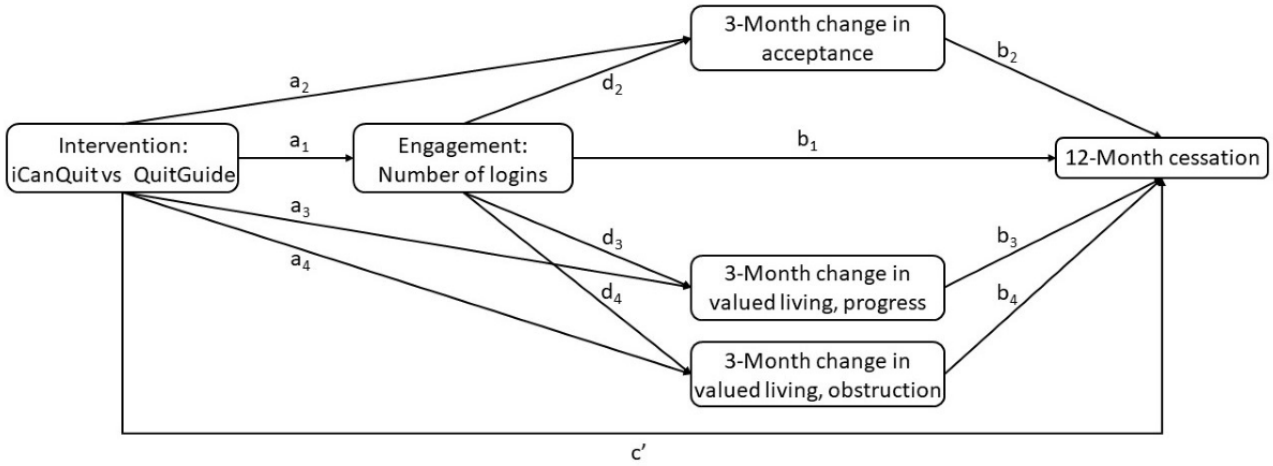
It remains unknown why iCanQuit was efficacious. The iCanQuit intervention targeted 2 core processes of ACT [13]: acceptance and values. Specifically, ACT teaches acceptance of internal cues to smoke (sensations, emotions, and thoughts), which is conceptually distinct from USCPG-based standard approaches that teach avoidance of internal cues to smoke [14,15]. ACT also motivates smokers to quit by appealing to their values, whereas the USCPG-based approaches motivate through reason and logic [14,15]. The iCanQuit app was designed to change the level of enactment of personal values through exercises focusing on valued life domains inspiring a user to quit smoking (eg, family, health, and spirituality) and planning weekly actions to take in those life domains (eg, going on a walk with one's partner).

Acceptance has been identified as a core mediator in ACT-based interventions across a wide variety of content areas [16-18]. For smoking cessation interventions, prior studies have shown that acceptance of internal cues to smoke was a mediator of intervention efficacy [14,19]. For example, we found that in the WebQuit trial of an ACT-based website intervention for smoking cessation, baseline to 3-month increases in acceptance accounted for 80% (P<.001) of the effect of WebQuit.org on the main cessation outcome [14]. There is also evidence indicating that the enactment of values mediates the effects of ACT when applied to various mental health and chronic health conditions [20-23]. These mediation findings are consistent with ACT theory and treatment protocols, including the iCanQuit program, in which there is a strong emphasis on values in addition to acceptance. However, to date, the mediational role of enacting one's values as a way to motivate smokers to quit smoking has not been empirically tested in smoking cessation interventions.

In parallel with studies on psychological mechanisms of action, digital intervention researchers have been studying the role of intervention engagement as a process that predicts treatment outcomes [6,7,24,25]. Our previous study has shown that, in the SmartQuit app that preceded the iCanQuit app, engagement with the intervention and its specific ACT components was predictive of smoking cessation. Participants who completed the program were over 4 times more likely to quit smoking. This app had a tool to track when a user "let a craving pass," defined as noticing a craving and not acting on it by smoking. Usage of this tracking tool predicted a greater likelihood of quitting smoking [26]. Building on this research, the next step is to learn why engagement predicts cessation. By itself, engagement is a limited explanatory variable: engagement describes the user's actions; however, it is unclear how those actions lead to successful treatment outcomes [27].

As shown in Figure 1, we posit that the effect of the intervention (iCanQuit vs QuitGuide) on smoking cessation at the 12-month follow-up may be mediated by engagement (number of logins), which, in turn, impacts 3-month changes in acceptance and valued living. Specifically, the appeal and utility of iCanQuit's content (eg, ACT skills modules) may contribute to higher user engagement as compared to the QuitGuide intervention. This higher engagement may lead to changes in the 2 ACT-based processes targeted in the iCanQuit intervention: (1) higher levels of acceptance of internal smoking cues and (2) enactment of one's values as measured by progress and obstruction of valued living, respectively. Both acceptance of internal cues to smoke and enactment of one's values may then lead to a higher likelihood of quitting smoking. Therefore, this study aimed to test this serial mediational model in the full-scale iCanQuit trial. These results will provide the first known evidence on potential serial mediators of smartphone apps for smoking cessation. While such serial mediational models are useful for developing an in-depth understanding of intervention efficacy, they are rare in smoking cessation research [28-30].

Figure 1. Conceptual model for serial mediation of potential mediators of the iCanQuit intervention.



Methods

Design

Data for this secondary analysis were obtained from all 2415 individuals enrolled in the 2-arm iCanQuit RCT for smoking cessation, with its complete details previously described [12]. In brief, a racially and ethnically diverse sample of 2415 adult daily smokers from all 50 US states were randomized 1:1 to either receive access to an ACT-based smartphone app (iCanQuit) or a USCPG-based smartphone app (QuitGuide) for smoking cessation.

Eligibility Criteria

Eligibility criteria included: (1) being 18 years of age or older, (2) having smoked 5 or more cigarettes per day in the past year, (3) wanting to quit smoking within the subsequent 30 days, (4) if concurrently using any other tobacco products, wanting to quit consuming all tobacco products within 30 days, (5) having an interest in learning skills to quit smoking and being willing to be randomized to either treatment condition, (6) having daily access to their own smartphone, (7) knowing how to download smartphone apps, (8) being willing and able to read in English, (9) having never used QuitGuide and not currently using another smoking cessation treatment, (10) having never participated in our prior studies, (11) no household members having been already enrolled, (12) being willing to complete outcome surveys, and (13) being able to provide contact information for themselves and 2 relatives.

Recruitment, Enrollment, and Follow-up

Adults were recruited nationwide via Facebook ads, a survey sampling company, search engine results, and friends/family referral. Participants completed an encrypted, web-based screening survey and were notified of their eligibility via email. They then clicked on their secured emailed link to the study website, where they provided consent and completed the baseline survey. At each enrollment step, the study was presented as a comparison of 2 smartphone apps for smoking cessation.

Participants were randomized (1:1) to either iCanQuit or QuitGuide using randomly permuted blocks of size 2, 4, and 6, stratified by daily smoking frequency (≤20 vs ≥21), education (≤high school vs ≥some college), race/ethnicity (minority race/ethnicity vs non-Hispanic White), and positive screening for depression (CES-D score ≤15 vs ≥16) [31]. Random assignments were concealed from participants throughout the trial. The random allocation sequence was generated by a database manager and implemented automatically by the study website. Neither research staff nor study participants had access to upcoming randomized assignment into either study arm. In both arms, participants could access their interventions from the moment of randomization and beyond (ie, after the end of the 12-month follow-up period). All participants provided consent online and were compensated to up to US $105 for completing study data collection. All study activities were approved by the institutional review board of the Fred Hutchinson Cancer Research Center.

Interventions

iCanQuit

Participants randomized to the iCanQuit arm received access to download the iCanQuit smartphone app (version 1.2.1). iCanQuit intervenes on the ACT-focused processes of acceptance of internal cues to smoke and enactment of one's values that guide quitting smoking [12]. The acceptance component of the app teaches skills to accept physical sensations, emotions, and thoughts that trigger smoking by distancing from thoughts about smoking, mindfulness skills, and flexible perspective-taking. The values component of the app teaches skills for determining the core life domains that motivate quitting smoking (eg, family, health, and spirituality) and taking repeated small actions within these domains (eg, playing with grandchildren) to develop a smoke-free life. The program is self-paced, and the content is unlocked in a sequential manner across 8 levels. Each of the first 4 levels is made accessible immediately after the prior level is completed, while each of the last 4 is only unlocked upon recording 7 consecutive days without smoking. If a participant lapses, the program encourages (but does not require) them to set a new quit date and return to the first 4 levels for preparation. The program also includes on-demand tools to help in coping with smoking urges

and to track the daily number of cigarettes smoked and urges passed without smoking.

QuitGuide

Participants randomized to the QuitGuide arm received access to download the QuitGuide smartphone app (version 1.2.2). QuitGuide content is delivered in four main sections: (1) "Thinking about quitting," which focuses on motivations to quit by using reason and logic such as identifying reasons to quit and providing information on the health consequences of smoking and quitting; (2) "Preparing to Quit," which helps users develop a plan to quit, identify smoking behaviors, triggers, and reasons for being smoke-free, and social support for quitting; (3) "Quitting," which teaches skills for avoiding cravings to smoke; and (4) "Staying Quit," which presents tips, motivations, and actions to stay smoke-free and skills for coping with slips. No quit smoking medications, coaching, or any other intervention was provided in either intervention arm [12].

Study Measures

Baseline Characteristics and Covariates

Data collected at baseline included age, gender, ethnicity, education, employment, income, marital status, and sexual orientation. Study participants completed validated positive screening tools to assess mental health, including depression [31], panic [32], and posttraumatic stress disorder [33]. Alcohol consumption and heavy drinking were assessed via the Quick Drinking Screen [34]. Smoking behavior variables included nicotine dependence (measured using the Fagerström Test for Nicotine Dependence) [35], number of cigarettes smoked per day, years of smoking, use of e-cigarettes, quit attempts, and relationships with other people who smoke.

As reported in the parent trial paper (and thus not reported in this study), participants were from all 50 US States. The mean age at enrollment was 38.2 (SD 10.9) years. Participants were 70.4% (1700/2415) women and 35.9% (868/2415) reported racial/ethnic minority backgrounds. There were 41.2% (995/2415) with high school or less education. Regarding smoking, 83.1% (2009/2415) had smoked for ≥10 years and 74.7% (1803/2415) smoked more than a half pack (at least 11 cigarettes) per day. There were no significant differences between the 2 arms on any baseline variable (for all, $P>.05$) [12].

Treatment Engagement Mediator: Baseline to 3 Months

Engagement with the assigned app was objectively measured using Google Analytics. The main mediational model's measure of engagement was the number of times each app was opened, consistent with other digital interventions' measures of engagement [7,24,25]. App activity that occurred at least 10 minutes after previous activity was considered a new login. Secondary measures of engagement were the total number of minutes and the unique number of days on which each app was used. To test the proposed mediational model (Figure 1), the first 3 months of utilization data for each participant were used in this study (N=2415).

ACT Theory–Based Mediators: Baseline to 3 Months

Change from baseline to 3 months after randomization in ACT theory–based processes, including acceptance of internal cues to smoke and valued living, were measured using validated tools. Acceptance of internal cues to smoke was measured via the Avoidance and Inflexibility Scale (AIS) [36], using the mean of the three 9-item subscales that assess one's willingness to experience physical sensations, emotions, and thoughts that cue smoking. The items are rated on a 5-point scale from (1) "Not at all" to (5) "Very willing" and averaged, with higher scores indicating greater acceptance. A sample physical sensation item was "How willing are you to notice these bodily sensations without smoking?" and items from the emotions and thoughts subscales were similar, substituting "feelings" or "thoughts" for "bodily sensations." Valued living was measured using the 10-item Valuing Questionnaire (VQ) [37] designed to assess the extent of enactment of personal values. Each item is rated on a 7-point scale ranging from (0) "Not at all true" to (6) "Completely true." Scores were averaged and 2 distinct factors were derived, progress and obstruction, with higher scores indicating either greater progress or greater obstruction toward valued living, respectively. A sample progress item was "I worked toward my goals even if I didn't feel motivated to" and a sample obstruction item was "I was basically on auto-pilot most of the time." Cronbach α (95% CI) values for each of the 3 scales showed good internal consistency: (1) mean acceptance [Cronbach α=.76 (95% CI .75-.77)], (2) valued living, progress subscale [Cronbach α=.88 (95% CI .87-.89)], and (3) valued living, obstruction subscale [Cronbach α=.88 (95% CI .87-.89)].

Smoking Cessation Outcome: 12 Months

The parent trial's primary smoking cessation outcome was specified a priori as self-reported complete-case 30-day point-prevalence abstinence (PPA) at the 12-month follow-up. The secondary smoking cessation outcome for this study was intent-to-treat missing as smoking 30-day PPA at the 12-month follow-up. As reported in the parent trial, for the primary outcome of 30-day PPA at the 12-month follow-up, iCanQuit participants had a 1.49-fold higher odds of quitting smoking as compared to QuitGuide participants (28.2%, 293/1040 abstinent vs 21.1%, 225/1067 abstinent; odds ratio [OR] 1.49, 95% CI 1.22-1.83; $P<.001$). When missing data were coded as smokers, 12-month 30-day PPA results were very similar: 24.1% (293/1214) abstinent for iCanQuit vs 18.7% (225/1201) abstinent for QuitGuide (OR 1.40, 95% CI 1.14-1.71, $P<.001$).

Statistical Analyses

We first compared treatment arms on proposed mediators at 3 months, using a negative binomial model for the number of logins owing to its highly right-skewed distribution and generalized linear models for the remaining mediators (ie, change in mean acceptance, and valued living progress and obstruction subscales). Regression analyses were performed using R (version 4.0.3, The R Foundation) [38] and the "MASS" library for negative binomial regression [39]. Hayes' PROCESS macro (version 3.5) for SAS [40] was used to test serial mediation of the effect of intervention condition on cessation at 12 months through engagement and through changes in acceptance and valued living from baseline to 3 months. Using

the notation in Figure 1, the indirect effect of the intervention on cessation through the number of logins alone was estimated by a_1b_1. Similarly, the indirect effects through change in acceptance and valued living progress and obstruction subscales were estimated by a_2b_2, a_3b_3, and a_4b_4, respectively. The serial mediation effects determined through the number of logins and in turn through change in acceptance and valued living progress and obstruction subscales were estimated by $a_1d_2b_2$, $a_1d_3b_3$, and $a_1d_4b_4$, respectively. Indirect effects were estimated with 5000 bootstrapped samples and were considered statistically significant when 95% CIs did not include zero. Model covariates included the 4 factors used in stratified randomization (ie, education level, heavy smoking [≥21 cigarettes per day], minority race or ethnicity, depression symptoms [20-item Center for Epidemiological Studies-Depression scale score ≥16], and baseline acceptance and valued living scores). This approach, in which the analysis is consistent with the stratified randomization study design, has been recommended to avoid losing power and obtaining incorrect 95% CIs [41,42].

Primary analyses were conducted with complete-case data for all variables in the serial mediation model, which was available for 1846 participants. As reported in the parent trial, the follow-up data retention was 86.7% (n=2093/2415) overall at 3 months (85.9%, 1043/1214 for iCanQuit vs 87.4%, 1050/1201 for QuitGuide [*P*=.28] and 87.2% (n=2107/2415) overall at 12 months (85.7%, 1040/1214 for iCanQuit vs 88.8%, 1067/1201 for QuitGuide [*P*=.02]) [12]. A sensitivity analysis for the serial mediation model was performed using full information maximum likelihood to handle missing data in Mplus [43]. Secondary mediation analyses included all 3 AIS acceptance subscales (ie, willingness to experience physical sensations, emotions, and thoughts that cue smoking), and alternative measures of engagement (ie, total time measured as minutes of app use and the number of unique days of use).

Results

As shown in Table 1, participants randomized to iCanQuit logged into their assigned app for a significantly greater number of times than those randomized to QuitGuide (25.7 vs 7.5 times; *P*<.001). In addition, they had greater baseline to 3-month increases in acceptance of cues to smoke (*P*<.001). However, changes in the valued living subscales of progress and obstruction were not different between the 2 treatment arms (for all, *P*>.05). Table 1 also shows that for every 1-point increase from baseline to 3 months in acceptance of cues to smoke, there was a 6.07-fold higher odds of 12-month smoking cessation (OR 6.07, 95% CI 4.76-7.76, *P*<.001).

Table 1. Differences in mediators between the 2 intervention arms at 3-month follow-up and the effect of each 1-point increase in mediator on 12-month cessation outcomes.

	Relationship between treatment arm and mediator (*a* paths)				*P* value	Relationship between mediator and cessation (*b* paths)	
Mediator	Total (n=1846), mean (SD)	QuitGuide (n=929), mean (SD)	iCanQuit (n=917), mean (SD)	Incidence rate ratio or point estimate (95% CI)		Odds ratio (95% CI)	*P* value
Number of logins	16.5 (32.3)	7.5 (14.0)	25.7 (41.6)	3.46[a] (3.10 to 3.87)	<.001	1.01 (1.01-1.02)	<.001
Change in mean acceptance	0.13 (0.57)	0.06 (0.50)	0.20 (0.62)	0.13[b] (0.09 to 0.18)	<.001	6.07 (4.76-7.76)	<.001
Change in valued living-progress	–0.67 (7.88)	–0.72 (7.75)	–0.62 (8.01)	–0.12[b] (–0.74 to 0.50)	.71	1.04 (1.02-1.05)	<.001
Change in valued living-obstruction	0.43 (8.28)	0.51 (7.79)	0.35 (8.75)	0.15[b] (–0.49 to 0.78)	.65	0.96 (0.95-0.98)	<.001

[a]Incidence rate ratio values.

[b]Point estimate values.

The results of the primary serial mediation model are shown in Table 2 and they show the indirect effects posited by the model rather than individual path coefficients. Baseline to 3-month number of logins (indirect effect a_1b_1=0.09, 95% CI 0.04-0.18, *P*<.001) and change in mean acceptance of internal cues to smoke (indirect effect a_2b_2=0.12, 95% CI 0.04-0.21, *P*<.001) each mediated the effect of intervention condition on smoking cessation at 12 months. There was a significant serial mediation effect of intervention condition on smoking cessation through the number of logins and in turn through the change in mean acceptance (indirect effect $a_1d_2b_2$=0.11, 95% CI 0.07-0.15, *P*<.001). This serial mediation effect corresponds to an OR of 1.11 (95% CI 1.08-1.16). In contrast, none of the pathways through valued living subscales, neither progress nor obstruction, mediated the relationship between intervention condition and cessation. This pattern of results was the same for the missing as smoking cessation outcome.

Table 2. Estimates of indirect effects for all pathways in the serial mediation model.

Mediator	Path	Estimate of indirect effect (95% CI) for complete-case cessation outcome[a]	Estimate of indirect effect (95% CI) for missing as smoking cessation outcome[a]
Number of logins	$a_1 b_1$	0.09 (0.04 to 0.18)[b]	0.10 (0.05 to 0.18)[b]
Change in mean acceptance	$a_2 b_2$	0.12 (0.04 to 0.21)[b]	0.12 (0.04 to 0.20)[b]
Change in valued living, progress subscale	$a_3 b_3$	–0.01 (–0.02 to 0.01)	0.00 (–0.02 to 0.01)
Change in valued living, obstruction subscale	$a_4 b_4$	0.00 (–0.01 to 0.01)	0.00 (–0.01 to 0.01)
Number of logins and change in mean acceptance, in serial	$a_1 d_2 b_2$	0.11 (0.07 to 0.15)[b]	0.10 (0.07 to 0.14)[b]
Number of logins and change in valued living progress, in serial	$a_1 d_3 b_3$	0.00 (–0.001 to 0.01)	0.00 (–0.001 to 0.01)
Number of logins and change in valued living obstruction, in serial	$a_1 d_4 b_4$	0.00 (–0.003 to 0.002)	0.00 (–0.003 to 0.002)

[a]95% CIs that include 0 are nonsignificant. Indirect effect estimate (95% CI) values may be exponentiated to produce estimates on the odds ratio scale.
[b]P<.05.

In secondary analysis models, the pattern of results for the serial mediation model was the same when engagement was measured as the total number of minutes (indirect effect $a_1d_2b_2$=0.09, 95% CI 0.05-0.14, P<.001) or the total number of unique days on which each app was used (indirect effect $a_1d_2b_2$=0.13, 95% CI 0.10-0.17, P<.001). This is consistent with the high correlations between engagement measures, which ranged from 0.72 to 0.91 (results not shown). Results were the same when the mediation model was reanalyzed with full information maximum likelihood (N=2415; data not shown).

The primary mediation model was further elaborated in a sensitivity analysis to determine which acceptance subscales mediated the effect of intervention on smoking cessation at 12 months (Multimedia Appendices 1 and 2). Our results show that change in the mean acceptance of physical sensations (indirect effect a_2b_2=0.03, 95% CI 0.02-0.06, P<.001) and acceptance of emotions (indirect effect a_4b_4=0.09, 95% CI 0.03-0.16, P<.001), but not acceptance of thoughts (indirect effect a_3b_3=0.01, 95% CI –0.02 to 0.04, P>.05), each mediated the effect of the intervention condition on smoking cessation at 12 months. Regarding serial mediation, the effect of the intervention condition on smoking cessation was significantly mediated through the number of logins and in turn through change in the mean acceptance of physical sensations (indirect effect $a_1d_2b_2$=0.03, 95% CI 0.01-0.05, P<.001) and acceptance of emotions (indirect effect $a_1d_4b_4$=0.07, 95% CI 0.04-0.11, P<.001). In contrast, the serial mediation pathway through acceptance of thoughts was not significant (indirect effect $a_1d_3b_3$=0.01, 95% CI –0.01 to 0.03, P>.05). Similar to the primary model, none of the pathways through valued living subscales, neither progress nor obstruction, mediated the relationship between intervention condition and cessation (for all, P>.05).

Discussion

This is the first study of serial mediators underlying the efficacy of smartphone apps for smoking cessation in a nationwide sample of daily smokers. The study tested whether the effect of the iCanQuit (vs QuitGuide) intervention on smoking cessation at the 12-month follow-up was mediated by engagement that in turn impacted 3-month changes in acceptance and valued living. Overall, there was a significant serial mediation effect of iCanQuit on smoking cessation through multiple indicators of engagement (ie, total number of logins, total number of minutes, and total number of unique days or use) and in turn, through change in mean acceptance of internal cues to smoke. Supplementary analysis of the acceptance subscales showed that serial mediation was through acceptance of physical sensations and emotions but not acceptance of thoughts. There was no evidence that the effect of the iCanQuit intervention (vs QuitGuide) was mediated by changes in valued living.

The results significantly advance the understanding of mechanisms underlying interventions for smoking cessation, and digital interventions for smoking cessation in particular. To date, serial mediation models of smoking cessation have been rare. One study found that the effect of telemedicine for smoking cessation on cessation was mediated by providers' support, which, in turn, led to increased self-efficacy and impacted cessation [28]. Another study found that the effect of financial incentives on quitting smoking was mediated only by self-efficacy but not program satisfaction [30]. The unique value of the current study's serial mediation model is in demonstrating how treatment engagement leads to higher cessation outcomes [7,24,25]. Our results suggest that regardless of the measure of engagement, greater treatment participation leads to greater improvements in underlying theoretical processes of behavior change, which in this case was the ACT process of acceptance of internal cues to smoke. This provides empirical support to the clinical premise that greater usage of the mobile app is a key pathway to activating a person's learning of therapeutic processes of change. The serial mediation findings indicate that part of how greater engagement leads to greater likelihood of smoking cessation is through activation of key psychological processes targeted in the intervention. Future research can

examine whether engagement in certain types of clinical content (eg, specific behavior change exercises) have a stronger link to mediating certain therapeutic processes than others, which could provide the empirical guidance to further optimize interventions to increase engagement with behavioral intervention components that most effectively target key psychological processes. This knowledge could inform smartphone intervention designs that coherently connect program engagement, program components, and therapeutic processes to improve treatment outcomes.

The results on acceptance have several important implications for the ACT model of smoking cessation. Eight prior ACT RCTs showed either formal statistical mediation or higher levels of acceptance of internal cues to smoke in the ACT intervention arm [14,15,44-49]. Building on this evidence, our results suggest that acceptance of physical sensations (eg, cravings) and emotions that trigger smoking, but not acceptance of thoughts that trigger smoking, may be important theoretical pathways of smoking cessation. These findings contrast with those of the general ACT therapeutic model, in which acceptance of thoughts, and related changes in how one responds to thoughts, is theorized to be an important therapeutic process for ACT and major component of treatment [13]. If replicated, these findings suggest a potential point for theory refinement in applying the ACT model for smoking cessation.

For intervention design, these findings suggest that future digital ACT-focused smoking cessation interventions should emphasize targeting acceptance of cravings and emotions that cue smoking. This could be accomplished by focusing on intervention exercises that help people (1) identify physical sensations and emotions that trigger smoking behaviors and (2) practice openness and willingness to experience these sensations and emotions. Skills-training in allowing cravings to pass and mindful awareness of cravings and emotions may be especially beneficial. In contrast, these findings suggest that less focus should be on exercises targeting acceptance of thoughts that trigger smoking since this does not appear to mediate treatment effects on smoking cessation.

The results on valued living are novel and have implications for future research. To date, no prior research has examined the role of valued living in smoking cessation. In the broader literature on ACT intervention research, we are only aware of a few pilot studies, all among college students, which showed that the effects of ACT digital interventions for stress, anxiety, and depression were mediated by valued living or meaningfulness [22,23]. Our results on valued living suggest some possibilities. Primarily, valued living or enactment of one's values may not be a mediator of smoking cessation. While it is conceivable that one does not need to work toward broader life goals to quit smoking, this is an unlikely explanation given the central role of motivation in health behavior change overall [50,51]. A more plausible explanation is that the current measure of valued living is not a sensitive measure of valued actions pertinent to smoking cessation. The VQ pertains to one's overall sense of life purpose and goals, whereas a smoking cessation intervention like iCanQuit focuses specifically on valued life domains directly associated with smoking (eg, health) as motivators to take actions toward quitting (eg, setting a quit date). The measure of acceptance was specific to smoking [36], rather than a general construct of acceptance of internal experience [52] and, as observed in this study, the associations between acceptance of smoking cues and smoking cessation were significant. In contrast, the observed associations between valued living, as measured broadly by the VQ, and smoking cessation were minor. Nonetheless, the predictive relationship between valued living and smoking cessation was significant. Thus, another possibility is that while iCanQuit focuses on values specifically in the context of smoking cessation, there may be some benefits to adding a general, less smoking-specific, intervention for valued living. Finally, it is worth noting that prior research has shown mixed evidence for the sensitivity of the VQ in detecting ACT intervention effects, suggesting that there may be limitations in the scale [22,53,54]. Future research can focus on developing a smoking-specific valuing questionnaire with the ultimate goal of testing it in smoking cessation intervention research.

In conclusion, this is the first study of serial mediators underlying the efficacy of smartphone apps for smoking cessation. The effect of the iCanQuit smartphone app on smoking cessation was mediated through multiple indicators of engagement and, in turn, through change in acceptance of physical sensations and emotions. Our results suggest that smoking cessation interventions should focus on increasing treatment engagement with the goal of enhancing the acceptance of cravings and emotions that cue smoking.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Conceptual model for serial mediation analysis including the three acceptance subscales.
[PNG File , 148 KB - mhealth_v9i11e32847_app1.png]

Multimedia Appendix 2
Estimates of indirect effects for pathways in a serial mediation model including the three Avoidance and Inflexibility Scale (AIS) subscales.
[DOCX File , 14 KB - mhealth_v9i11e32847_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 836 KB - mhealth_v9i11e32847_app3.pdf]

References

1. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1659-1724 [FREE Full text] [doi: 10.1016/S0140-6736(16)31679-8] [Medline: 27733284]
2. GBD 2015 Tobacco Collaborators. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. Lancet 2017 May 13;389(10082):1885-1906 [FREE Full text] [doi: 10.1016/S0140-6736(17)30819-X] [Medline: 28390697]
3. Husten CG. A Call for ACTTION: Increasing access to tobacco-use treatment in our nation. Am J Prev Med 2010 Mar;38(3 Suppl):S414-S417. [doi: 10.1016/j.amepre.2009.12.006] [Medline: 20176316]
4. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
5. Mobile Fact Sheet 2019. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-01-02]
6. Browne J, Halverson T, Vilardaga R. Engagement with a digital therapeutic for smoking cessation designed for persons with psychiatric illness fully mediates smoking outcomes in a pilot randomized controlled trial. Transl Behav Med 2021 Sep 15;11(9):1717-1725. [doi: 10.1093/tbm/ibab100] [Medline: 34347865]
7. Graham AL, Papandonatos GD, Cobb CO, Cobb NK, Niaura RS, Abrams DB, et al. Internet and Telephone Treatment for smoking cessation: mediators and moderators of short-term abstinence. Nicotine Tob Res 2015 Mar;17(3):299-308 [FREE Full text] [doi: 10.1093/ntr/ntu144] [Medline: 25156528]
8. Heppner WL, Spears CA, Correa-Fernández V, Castro Y, Li Y, Guo B, et al. Dispositional Mindfulness Predicts Enhanced Smoking Cessation and Smoking Lapse Recovery. Ann Behav Med 2016 Jun;50(3):337-347 [FREE Full text] [doi: 10.1007/s12160-015-9759-3] [Medline: 26743533]
9. Hoeppner BB, Hoeppner SS, Abroms LC. How do text-messaging smoking cessation interventions confer benefit? A multiple mediation analysis of Text2Quit. Addiction 2017 Apr;112(4):673-682 [FREE Full text] [doi: 10.1111/add.13685] [Medline: 27943511]
10. Mackinnon DP, Dwyer JH. Estimating Mediated Effects in Prevention Studies. Eval Rev 2016 Jul 26;17(2):144-158. [doi: 10.1177/0193841x9301700202]
11. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002 Oct;59(10):877-883. [doi: 10.1001/archpsyc.59.10.877] [Medline: 12365874]
12. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of Smartphone Applications for Smoking Cessation: A Randomized Clinical Trial. JAMA Intern Med 2020 Nov 01;180(11):1472-1480 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
13. Hayes SC, Levin ME, Plumb-Vilardaga J, Villatte JL, Pistorello J. Acceptance and commitment therapy and contextual behavioral science: examining the progress of a distinctive model of behavioral and cognitive therapy. Behav Ther 2013 Jun;44(2):180-198 [FREE Full text] [doi: 10.1016/j.beth.2009.08.002] [Medline: 23611068]
14. Bricker J, Wyszynski C, Comstock B, Heffner JL. Pilot randomized controlled trial of web-based acceptance and commitment therapy for smoking cessation. Nicotine Tob Res 2013 Oct;15(10):1756-1764 [FREE Full text] [doi: 10.1093/ntr/ntt056] [Medline: 23703730]
15. Bricker JB, Mull KE, McClure JB, Watson NL, Heffner JL. Improving quit rates of web-delivered interventions for smoking cessation: full-scale randomized trial of WebQuit.org versus Smokefree.gov. Addiction 2018 May;113(5):914-923 [FREE Full text] [doi: 10.1111/add.14127] [Medline: 29235186]
16. Sairanen E, Tolvanen A, Karhunen L, Kolehmainen M, Järvelä-Reijonen E, Lindroos S, et al. Psychological flexibility mediates change in intuitive eating regulation in acceptance and commitment therapy interventions. Public Health Nutr 2017 Jun;20(9):1681-1691. [doi: 10.1017/S1368980017000441] [Medline: 28414018]
17. Østergaard T, Lundgren T, Zettle RD, Landrø NI, Haaland VO. Psychological Flexibility in Depression Relapse Prevention: Processes of Change and Positive Mental Health in Group-Based ACT for Residual Symptoms. Front Psychol 2020;11:528 [FREE Full text] [doi: 10.3389/fpsyg.2020.00528] [Medline: 32292369]
18. Ritzert TR, Berghoff CR, Tifft ED, Forsyth JP. Evaluating ACT Processes in Relation to Outcome in Self-Help Treatment for Anxiety-Related Problems. Behav Modif 2020 Nov;44(6):865-890. [doi: 10.1177/0145445519855616] [Medline: 31220920]
19. Schuck K, Otten R, Kleinjan M, Bricker JB, Engels RCME. Self-efficacy and acceptance of cravings to smoke underlie the effectiveness of quitline counseling for smoking cessation. Drug Alcohol Depend 2014 Sep 01;142:269-276. [doi: 10.1016/j.drugalcdep.2014.06.033] [Medline: 25042212]

20. Lundgren T, Dahl J, Hayes SC. Evaluation of mediators of change in the treatment of epilepsy with acceptance and commitment therapy. J Behav Med 2008 Jun;31(3):225-235. [doi: 10.1007/s10865-008-9151-x] [Medline: 18320301]
21. Ren Z, Zhao C, Bian C, Zhu W, Jiang G, Zhu Z. Mechanisms of the Acceptance and Commitment Therapy: A meta-analytic structural equation model. Acta Psychologica Sinica 2019;51(6):662-676. [doi: 10.3724/sp.j.1041.2019.00662]
22. Levin ME, Haeger JA, Pierce BG, Twohig MP. Web-Based Acceptance and Commitment Therapy for Mental Health Problems in College Students: A Randomized Controlled Trial. Behav Modif 2017 Jan 01;41(1):141-162. [doi: 10.1177/0145445516659645] [Medline: 27440189]
23. Räsänen P, Muotka J, Lappalainen R. Examining mediators of change in wellbeing, stress, and depression in a blended, Internet-based, ACT intervention for university students. Internet Interv 2020 Dec;22:100343 [FREE Full text] [doi: 10.1016/j.invent.2020.100343] [Medline: 32995301]
24. Pechmann C, Delucchi K, Lakon CM, Prochaska JJ. Randomised controlled trial evaluation of Tweet2Quit: a social network quit-smoking intervention. Tob Control 2017 Mar;26(2):188-194 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052768] [Medline: 26928205]
25. Richardson A, Graham AL, Cobb N, Xiao H, Mushro A, Abrams D, et al. Engagement promotes abstinence in a web-based cessation intervention: cohort study. J Med Internet Res 2013 Jan 28;15(1):e14 [FREE Full text] [doi: 10.2196/jmir.2277] [Medline: 23353649]
26. Zeng EY, Heffner JL, Copeland WK, Mull KE, Bricker JB. Get with the program: Adherence to a smartphone app for smoking cessation. Addict Behav 2016 Dec;63:120-124 [FREE Full text] [doi: 10.1016/j.addbeh.2016.07.007] [Medline: 27454354]
27. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring Engagement in eHealth and mHealth Behavior Change Interventions: Viewpoint of Methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
28. Liebmann EP, Preacher KJ, Richter KP, Cupertino AP, Catley D. Identifying pathways to quitting smoking via telemedicine-delivered care. Health Psychol 2019 Jul;38(7):638-647 [FREE Full text] [doi: 10.1037/hea0000740] [Medline: 31021123]
29. Martínez-Vispo C, López-Durán A, Senra C, Rodríguez-Cano R, Fernández Del Río E, Becoña E. Environmental reward and depressive symptoms in the relationship between avoidance and cigarette dependence in treatment-seeking smokers. Psicothema 2020 May;32(2):176-181. [doi: 10.7334/psicothema2019.292] [Medline: 32249742]
30. van den Brand FA, Candel M, Nagelhout G, Winkens B, van Schayck CP. How Financial Incentives Increase Smoking Cessation: A Two-Level Path Analysis. Nicotine Tob Res 2021 Jan 07;23(1):99-106 [FREE Full text] [doi: 10.1093/ntr/ntaa024] [Medline: 31993637]
31. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Appl Psychol Meas 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]
32. Stein MB, Roy-Byrne PP, McQuaid JR, Laffaye C, Russo J, McCahill ME, et al. Development of a brief diagnostic screen for panic disorder in primary care. Psychosom Med 1999;61(3):359-364. [doi: 10.1097/00006842-199905000-00016] [Medline: 10367617]
33. Lang AJ, Wilkins K, Roy-Byrne PP, Golinelli D, Chavira D, Sherbourne C, et al. Abbreviated PTSD Checklist (PCL) as a guide to clinical response. Gen Hosp Psychiatry 2012;34(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2012.02.003] [Medline: 22460001]
34. Roy M, Dum M, Sobell LC, Sobell MB, Simco ER, Manor H, et al. Comparison of the quick drinking screen and the alcohol timeline followback with outpatient alcohol abusers. Subst Use Misuse 2008 Dec;43(14):2116-2123. [doi: 10.1080/10826080802347586] [Medline: 18825590]
35. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
36. Farris SG, Zvolensky MJ, DiBello AM, Schmidt NB. Validation of the Avoidance and Inflexibility Scale (AIS) among treatment-seeking smokers. Psychol Assess 2015 Jun;27(2):467-477 [FREE Full text] [doi: 10.1037/pas0000059] [Medline: 25642937]
37. Smout M, Davies M, Burns N, Christie A. Development of the Valuing Questionnaire (VQ). J Contextual Behav Sci 2014 Jul;3(3):164-172. [doi: 10.1016/j.jcbs.2014.06.001]
38. The R Project for Statistical Computing. The R Foundation. URL: https://www.R-project.org/ [accessed 2021-06-02]
39. Venables W, Ripley B, Venables W. Modern Applied Statistics with S (4th edition). New York, NY: Springer; 2002.
40. Hayes AF. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (2nd edition). New York, NY: Guilford Press; 2017.
41. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. J Clin Epidemiol 1999 Jan;52(1):19-26. [doi: 10.1016/s0895-4356(98)00138-3] [Medline: 9973070]
42. Raab GM, Day S, Sales J. How to select covariates to include in the analysis of a clinical trial. Control Clin Trials 2000 Aug;21(4):330-342. [doi: 10.1016/s0197-2456(00)00061-1] [Medline: 10913808]

43. Muthén L, Muthén B. Mplus Statistical Analysis with Latent Variables User's Guide (8th edition). Los Angeles, CA: Muthén & Muthén; 1998.
44. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 01;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
45. Bricker JB, Bush T, Zbikowski SM, Mercer LD, Heffner JL. Randomized trial of telephone-delivered acceptance and commitment therapy versus cognitive behavioral therapy for smoking cessation: a pilot study. Nicotine Tob Res 2014 Nov;16(11):1446-1454 [FREE Full text] [doi: 10.1093/ntr/ntu102] [Medline: 24935757]
46. Brown RA, Reed KMP, Bloom EL, Minami H, Strong DR, Lejuez CW, et al. Development and preliminary randomized controlled trial of a distress tolerance treatment for smokers with a history of early lapse. Nicotine Tob Res 2013 Dec;15(12):2005-2015 [FREE Full text] [doi: 10.1093/ntr/ntt093] [Medline: 23884317]
47. Gifford EV, Kohlenberg BS, Hayes SC, Antonuccio DO, Piasecki MM, Rasmussen-Hall ML, et al. Acceptance-based treatment for smoking cessation. Behavior Therapy 2004;35(4):689-705. [doi: 10.1016/S0005-7894(04)80015-7]
48. Gifford EV, Kohlenberg BS, Hayes SC, Pierson HM, Piasecki MP, Antonuccio DO, et al. Does acceptance and relationship focused behavior therapy contribute to bupropion outcomes? A randomized controlled trial of functional analytic psychotherapy and acceptance and commitment therapy for smoking cessation. Behav Ther 2011 Dec;42(4):700-715. [doi: 10.1016/j.beth.2011.03.002] [Medline: 22035998]
49. Vilardaga R, Rizo J, Palenski PE, Mannelli P, Oliver JA, Mcclernon FJ. Pilot Randomized Controlled Trial of a Novel Smoking Cessation App Designed for Individuals With Co-Occurring Tobacco Use Disorder and Serious Mental Illness. Nicotine Tob Res 2020 Aug 24;22(9):1533-1542 [FREE Full text] [doi: 10.1093/ntr/ntz202] [Medline: 31667501]
50. Rajani NB, Mastellos N, Filippidis FT. Self-Efficacy and Motivation to Quit of Smokers Seeking to Quit: Quantitative Assessment of Smoking Cessation Mobile Apps. JMIR Mhealth Uhealth 2021 Apr 30;9(4):e25030 [FREE Full text] [doi: 10.2196/25030] [Medline: 33929336]
51. Rajani NB, Mastellos N, Filippidis FT. Impact of Gamification on the Self-Efficacy and Motivation to Quit of Smokers: Observational Study of Two Gamified Smoking Cessation Mobile Apps. JMIR Serious Games 2021 Apr 27;9(2):e27290 [FREE Full text] [doi: 10.2196/27290] [Medline: 33904824]
52. Bond FW, Hayes SC, Baer RA, Carpenter KM, Guenole N, Orcutt HK, et al. Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: a revised measure of psychological inflexibility and experiential avoidance. Behav Ther 2011 Dec;42(4):676-688. [doi: 10.1016/j.beth.2011.03.007] [Medline: 22035996]
53. Bernal-Manrique KN, García-Martín MB, Ruiz FJ. Effect of acceptance and commitment therapy in improving interpersonal skills in adolescents: A randomized waitlist control trial. Journal of Contextual Behavioral Science 2020 Jul;17:86-94. [doi: 10.1016/j.jcbs.2020.06.008]
54. Levin ME, Pierce B, Schoendorff B. The acceptance and commitment therapy matrix mobile app: A pilot randomized trial on health behaviors. J Contextual Behav Sci 2017 Jul;6(3):268-275. [doi: 10.1016/j.jcbs.2017.05.003]

Abbreviations

ACT: Acceptance and Commitment Therapy
AIS: Avoidance and Inflexibility Scale
OR: odds ratio
PPA: point-prevalence abstinence
RCT: randomized controlled trial
USCPG: US Clinical Practice Guidelines
VQ: Valuing Questionnaire

Edited by G Eysenbach; submitted 11.08.21; peer-reviewed by S Hoeppner; comments to author 04.09.21; revised version received 28.09.21; accepted 12.10.21; published 09.11.21.

Please cite as:
Bricker JB, Levin M, Lappalainen R, Mull K, Sullivan B, Santiago-Torres M
Mechanisms of Smartphone Apps for Cigarette Smoking Cessation: Results of a Serial Mediation Model From the iCanQuit Randomized Trial
JMIR Mhealth Uhealth 2021;9(11):e32847
URL: https://mhealth.jmir.org/2021/11/e32847
doi:10.2196/32847
PMID:34751662

Original Paper

Using Co-design in Mobile Health System Development: A Qualitative Study With Experts in Co-design and Mobile Health System Development

Tyler J Noorbergen[1], BIT, BBus, PhD; Marc T P Adam[1], PhD; Timm Teubner[2], PhD; Clare E Collins[3,4], Dip Nutr&Diet, DipClinEpi, BSc, PhD

[1]School of Information and Physical Sciences, College of Engineering, Science and Environment, University of Newcastle, Callaghan, Australia

[2]Technical University Berlin, Berlin, Germany

[3]Hunter Medical Research Institute, New Lambton, Australia

[4]School of Health Sciences, College of Health, Medicine, and Wellbeing, University of Newcastle, Callaghan, Australia

Corresponding Author:
Marc T P Adam, PhD
School of Information and Physical Sciences
College of Engineering, Science and Environment
University of Newcastle
ES214, Es Building
Callaghan, 2308
Australia
Phone: 61 249215197
Email: marc.adam@newcastle.edu.au

Abstract

Background: The proliferation of mobile devices has enabled new ways of delivering health services through mobile health systems. Researchers and practitioners emphasize that the design of such systems is a complex endeavor with various pitfalls, including limited stakeholder involvement in design processes and the lack of integration into existing system landscapes. Co-design is an approach used to address these pitfalls. By recognizing users as experts of their own experience, co-design directly involves users in the design process and provides them an active role in knowledge development, idea generation, and concept development.

Objective: Despite the existence of a rich body of literature on co-design methodologies, limited research exists to guide the co-design of mobile health (mHealth) systems. This study aims to contextualize an existing co-design framework for mHealth applications and construct guidelines to address common challenges of co-designing mHealth systems.

Methods: Tapping into the knowledge and experience of experts in co-design and mHealth systems development, we conducted an exploratory qualitative study consisting of 16 semistructured interviews. Thereby, a constructivist ontological position was adopted while acknowledging the socially constructed nature of reality in mHealth system development. Purposive sampling across web-based platforms (eg, Google Scholar and ResearchGate) and publications by authors with co-design experience in mHealth were used to recruit co-design method experts (n=8) and mHealth system developers (n=8). Data were analyzed using thematic analysis along with our objectives of contextualizing the co-design framework and constructing guidelines for applying co-design to mHealth systems development.

Results: The contextualized framework captures important considerations of the mHealth context, including dedicated prototyping and implementation phases, and an emphasis on immersion in real-world contexts. In addition, 7 guidelines were constructed that directly pertain to mHealth: understanding stakeholder vulnerabilities and diversity, health behavior change, co-design facilitators, immersion in the mHealth ecosystem, postdesign advocates, health-specific evaluation criteria, and usage data and contextual research to understand impact.

Conclusions: System designers encounter unique challenges when engaging in mHealth systems development. The contextualized co-design framework and constructed guidelines have the potential to serve as a shared frame of reference to guide the co-design of mHealth systems and facilitate interdisciplinary collaboration at the nexus of information technology and health research.

(JMIR Mhealth Uhealth 2021;9(11):e27896) doi:10.2196/27896

KEYWORDS

co-design; mHealth; guidelines; qualitative study; mobile phone

Introduction

Background

The proliferation of mobile devices (eg, smartphones and tablets) has enabled new ways of delivering health services via mobile health (*mHealth*) systems [1,2]. Broadly, mHealth can be defined as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices" [3]. The ubiquity and increasing capabilities of these systems have created enormous potential to support individuals in self-managing existing health conditions (eg, diabetes and stroke) and reducing their health risks by supporting healthier lifestyle habits (eg, increasing vegetable intake). The adoption of mHealth systems is steadily growing. In 2018, nearly half of the consumers in health care used mHealth systems compared with one-sixth in 2014. Overall, the global mHealth market is expected to grow from US $28.320 billion in 2018 to US $102.35 billion by 2023 [4].

Researchers have repeatedly emphasized that mHealth systems design and development is a complex endeavor with a range of pitfalls limiting adoption and/or effective usage in practice [5]. This is because the design process commonly entails limited stakeholder involvement [5,6], and solution artifacts lack integration with other health systems or their components [7]. To address these complexities, scholars have suggested co-design for mHealth systems development. Co-design refers to "the creativity of designers and people not trained in design, working together in the design development process" [8]. Research has referred to two main reasons for using co-design: (1) mHealth is a complex environment that requires the involvement of diverse stakeholders (eg, consumers/end users, government, health practitioners, scientists, and software developers) with co-design facilitating necessary collaborations [1,9,10]; (2) using co-design ensures that mHealth systems are underpinned by expert insights and best practices [5,11,12].

Despite repeated calls to use co-design for mHealth systems development [5,6], there is only limited guidance available on how to do so. The existing literature on co-design methodology provides important *general* guidance for the application of co-design frameworks and methods [8,13,14]. However, given the complexities surrounding a person's health and the multitude of stakeholders, there is a need for research that identifies the *specific* challenges system designers face when applying co-design in mHealth and to illustrate ways in which these challenges can be addressed. As such, there is a lack of guidance in the current literature in terms of how one can apply co-design in the mHealth systems context.

Objective

In this paper, we address this research gap by conducting a qualitative study that explores how co-design can be used in mHealth systems development. Specifically, we conducted 16 semistructured interviews to synthesize the theoretical and practical expertise of 8 co-design method experts (CMEs) and 8 mHealth system developers (MSDs) in a rapidly growing application area. Interviews were transcribed and analyzed using thematic analysis [15]. Thereby, the overarching research objectives of this study were (1) to contextualize an existing co-design framework for mHealth applications and (2) to construct guidelines to address common challenges of using co-design in mHealth development.

Theoretical Background and Related Work

Related Work on mHealth Systems Design

mHealth systems have become a growing area for research and practice [16,17]. The two primary application domains that have emerged are (1) disease management and (2) health promotion. First, disease management empowers patients to manage their medical conditions more effectively and independently (eg, controlling blood sugar levels [18,19]). Second, health promotion facilitates better health choices by providing support and encouragement for users to engage in behaviors to lower risk factors and improve health (eg, better diet and smoking cessation). The design of mHealth systems is complex, with a range of pitfalls including limited stakeholder involvement [5], lack of integration with other health systems [7], and disregard of behavior change techniques [5].

Several studies have sought to improve mHealth systems design. McCurdie et al [20] discussed a user-centered design approach for mHealth systems development. The term user-centered design refers to "a design philosophy that places the needs, wants, and limitations of end users at the center of the design process" [21]. However, it should be noted that user-centered design adopts an expert perspective where "trained researchers observe and/or interview largely passive users" [8]. In contrast, in co-design, the user is in the position of being an expert of their own experience and actively plays a "large role in knowledge development, idea generation, and concept development" [8]. Banos et al [22] developed an architecture that showed how specific functionalities and components of mHealth systems could be implemented. Building on a user-centered design, Schnall et al [23] developed a 3-cycle framework (relevance, design, and rigor) to better incorporate end users' preferences. Eckman et al [1] developed an mHealth systems framework that considers design thinking principles, using "a hypothesis-driven method of generating and validating new concepts" [1]. Nahum-Shani et al [24] explored the design of just-in-time adaptive interventions to support users' health behavior change. However, there has been limited focus on how to apply a co-design approach that involves stakeholders within the mHealth context.

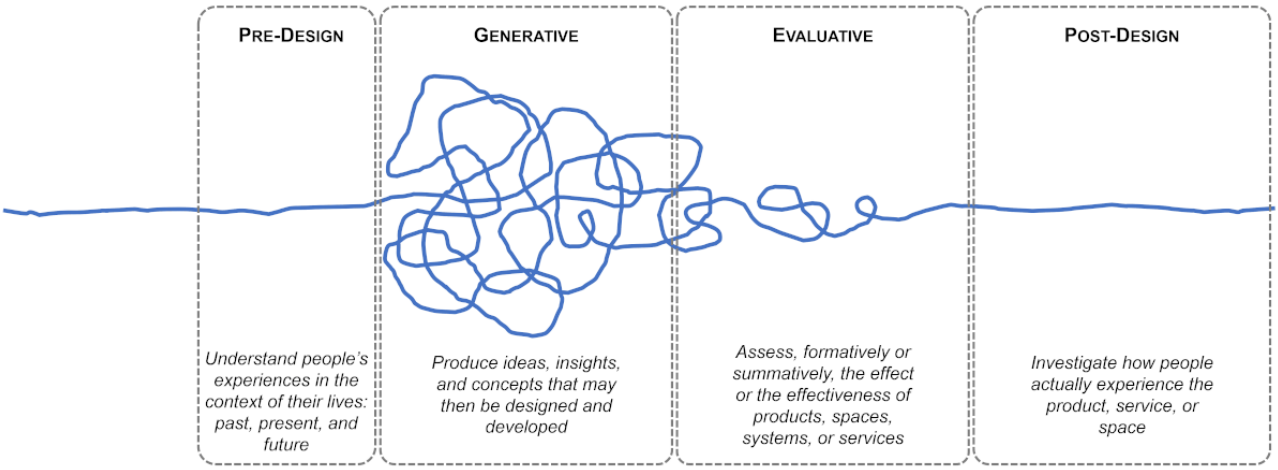
Co-design Frameworks

Researchers have proposed several frameworks to facilitate co-design [8,13,14]. By creating a conceptual structure of the process, these frameworks provide a shared frame of reference for researchers and practitioners engaging in system design. As noted by Sanders and Stappers [13], these frameworks can be

understood as a response to the increased attention to methods: "So many methods, tools and techniques have been introduced that it has become useful to provide frameworks for organizing them" [13]. For instance, the framework by Visser et al [14] structured the co-design process into five phases: preparation, sensitization, sessions, analysis, and communication. Brandt et al [25] described an iterative cycle of making, telling, and enacting. Building on these earlier conceptualizations, the framework by Sanders and Stappers [13] has become one of the most widely recognized co-design resources (538 citations on Google Scholar, February 2021). The framework breaks down the timeline of the co-design process (shown in blue) into 4 interconnected phases (Figure 1, adapted from a study by Sanders and Stappers [13]).

1. The predesign phase is concerned with understanding the surrounding context and people's experiences, exploring knowledge in the user context, establishing goals for future experiences, and sensitizing participants to the problem space [8,13].
2. The generative phase focuses on producing ideas, insights, and concepts that explore the *design space*, with users taking an active role in making through co-creation of conceptual artifacts (eg, journey maps, mock-ups, and storyboards). Although the vision is still fuzzy, these activities test, transform, and refine *ideas, insights, and concepts that may then be designed and developed* [13].
3. The evaluative phase allows users to assess the effects and effectiveness of the devised concepts. The vision of the final artifact becomes more tangible through the evaluation prototypes that allow users *to experience a situation that did not exist before* [13].
4. The postdesign phase captures the notion that once a system is part of a user's lived experiences, it needs to evolve along with their needs, habits, and use patterns. Hence, *the tail end of the postdesign phase [leads] to the front end of another design process* [13].

Figure 1. Co-design framework.



Co-design in mHealth

In recent years, an increasing number of mHealth studies have used a co-design approach. A 2016 review [26] identified early mHealth studies that used co-design with many following an approach similar to the framework by Sanders and Stappers [13]. A 2021 review [27] documented the application contexts (eg, diabetes and nutrition), stakeholders (eg, caregivers, nurses, and specialists), and methods used in co-design mHealth studies, including a mapping to the Sanders and Stappers framework [13]. The methods that have been applied include cultural probes [28-31], storytelling [32-36], and journey maps [31,37,38]. The context of these applications includes both disease management and health promotion. Examples from the disease management context include diabetes [30,39], cancer [40], asthma [41,42], heart failure [43-45], and depression [46]. In health promotion, contexts include nutrition [35,36,47], physical activity [28,35,47], smoking cessation [48,49], and mental health [38,50].

Methods

Overview

In this research, we adopted a constructivist ontological position and acknowledge the socially constructed nature of reality in mHealth systems development [51,52]. Recognizing that *there is no single truth*, constructivist approaches to research generate meaning through a collaborative dialog between researchers and the research participants [51].

Research Participants

We used a purposive sampling method to identify and recruit participants from 2 groups for interviews, namely, CMEs and MSDs. CMEs were recruited on the web using Google, Google Scholar, LinkedIn, Twitter, and ResearchGate to identify experts in co-design (eg, book authors, academics, and consultants). The MSD group was recruited by searching papers and reports by authors with co-design experience in mHealth. Interviewees had to be aged at least 18 years and fluent in English. Individuals were contacted by the first author via email with a study information statement before obtaining written informed

consent. Participation was voluntary and did not involve monetary rewards or other compensation.

Ethics approval was granted by the ethics committee of the University of Newcastle, Australia (H-2019-0064). Data collection and analysis were concurrently performed. The recruitment process continued until data sufficiency was reached (ie, existing categories managed new data without further modifications [53]). The final data set included 16 interviews (8 CMEs and 8 MSDs). On average, CMEs had over 15 years of publication experience (minimum: 2 years; maximum: 25 years) in areas such as co-design or participatory design, cocreation, design thinking, generative design research, and design research methods. On the other hand, MSDs had over 8 years of publication experience on average (minimum: 4 years; maximum: 28 years) in the mHealth literature spanning across multiple areas in disease management (cancer and heart failure) and health promotion (smoking cessation and nutrition). All interviews were audio-recorded (total duration: 14 hours, 15 minutes) and transcribed by the first author. The interview length was between 36 and 72 minutes. Multimedia Appendix 1 provides details on the participants' backgrounds and experiences.

Data Collection

Data were collected between July 2019 and January 2020. Before the interview, the research participants received a two-page information statement via email about the research objectives, scheduled interview duration, and assurance of data anonymization. Individuals who provided written consent to participate were interviewed by the first author at a mutually convenient time using Zoom or Skype videoconferencing as per the interviewee's preference. The interviews were semistructured in nature, with the interviewer using a protocol composed of open-ended questions and probing for additional information when required. The interviews focused on two research objectives: (1) contextualizing an existing co-design framework to the mHealth space and (2) constructing guidelines to address common challenges in this context (see the interview guide in Multimedia Appendix 2). Open-ended questions provided the interviewees with opportunities to speak freely and to guide the discussion in the directions of interest.

Data Analysis

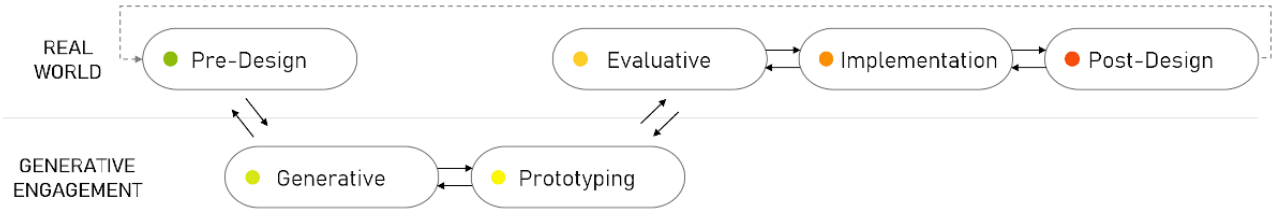
The first author coded the transcripts following the procedure of Braun and Clarke [15], which included (1) familiarization with the data, (2) coding, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) writing up. In step 1, this involved familiarization with the data by repeatedly reading and rereading the transcripts (ie, prolonged engagement). In step 2, the first author performed the initial coding in NVivo. The second author then checked these codes and validated them against the transcripts. Initially, we identified 154 codes from all interviews (eg, power distance and vulnerability). In step 3, the first and second authors clustered nodes into common themes based on coherent patterns. In several discussions between the authors, the identified themes became the foundation of the guidelines. In the results section, data extracts are quoted to support framework contextualization and guideline development. In step 5, the authors further refined the guidelines by eliminating redundant themes and naming the guidelines.

Results

Contextualization of Co-design Framework to mHealth Context

Figure 2 shows the contextualization of the Sanders and Stappers [13] co-design framework for the mHealth setting. It extends the 4 phases of the original framework by dedicated prototyping and implementation phases. A detailed overview of the example quotes for contextualization is provided in Multimedia Appendix 3.

Figure 2. Contextualized co-design framework for mobile health (mHealth).



The first extension was the inclusion of a dedicated *implementation phase*. Interview participants noted that in the context of mHealth, there is a need to separate *implementation* from the evaluative phase. This is because the evaluative phase primarily focuses on testing the *feasibility* of the mHealth system rather than the wider rollout of the system into a complex mHealth ecosystem:

> *You would not naturally do a clinical trial or a randomized control trial in your implementation phase because you first need to be able to test the feasibility.* [MSD8]
>
> *The implementation phase is after we have done the research and probably after we have analyzed the results and come to some kind of conclusions. So, there is a gap then between the generative phase and the implementation phase when we actually do our research. We are checking to make sure that we have got evidence now that would suggest that this is actually going to support people improve their health outcomes. That is your evaluative phase. Let's now go to the implementation phase where we actually deploy it.* [MSD5]

In contrast, a dedicated implementation phase should focus on facilitating the integration of mHealth artifacts into complex systems and stakeholder environments. In other words, it is important to not only consider the design of the technical

mHealth artifact, but everything else around it that is necessary for it to be successfully implemented. This includes important aspects such as documentation, training, and involving key stakeholders in the rollout (see postdesign advocates, guideline 5). The other important consideration discussed in the interviews was the importance of considering implementation right from the beginning of the co-design process:

> *You would want implementation to be on the agenda right from the initial co-design process. You need to have a plan. If you are going to co-design something, you need to have a plan that if it is effective, how could it be brought about, and those discussions or those people involved in that process. [...] Having those people involved from the start is fundamental to the success of implementation, and having plans around that.* [MSD7]

> *You really need to work with the [health system] and that is where that whole implementation phase becomes crucial because even if your thing is beautiful, if it does not have the support to make it work, it will fall down.* [MSD3]

The second extension pertains to a separate prototyping phase before the evaluative phase to acknowledge the complexity of mHealth artifacts and their evaluation requirements (eg, pilot-testing and randomized controlled trials). Including a separate prototyping phase assists in separating generative co-design methods in the generative phase (eg, paper prototyping), from instantiations in which the idea for the solution has become more mature and where (high fidelity) prototyping occurs (ie, hardware and software prototypes). Furthermore, it emphasizes the need for a fully functional prototype at the end of the prototyping phase, suitable for rigorous evaluation in the real-world context as part of the evaluative phase (eg, pilot-testing and randomized controlled trials). In addition, separating the prototyping and evaluative phases can clarify which stakeholders should be involved in which phase and in what capacity they should be involved (eg, app developers in the generative phase may simply observe or consult with the end users, whereas in the prototyping phase, they are developing a prototype):

> *You go from low fidelity [generative], to high fidelity [prototyping], and then to user testing [evaluative]. Generative is like low fidelity brainstorming. Generative design research and making is not user testing. It is different. What you are calling the generative phase is more like wire framing. You increase the fidelity of your prototypes as you go and test along the way.* [MSD3]

> *[Initially,] I would not constrain the end-users with any details about what can and cannot be done. [The mHealth system] would be a magic device and they would act out scenarios without any worry about how this could actually be mocked-up. I would have the developers see and hear that and hopefully then be inspired by it to bring somebody's dream to life [...]. But in a later phase you might hand pick some of the end-users to come and work directly with the developers. Then it's like: 'Well the end-user's dreams are this, but the developer's constraints are these. Can you guys come up with something together?* [CME1]

Finally, it is important to consider the context in which the co-design phases occur. The contextualized framework categorizes the generative and prototyping phases as phases in which *generative engagement* occurs. These phases involve gathering co-design participants from potentially diverse areas (eg, health practitioners and designers) in one place (eg, a co-design workshop in a studio or lab) to engage in generative co-design methods (eg, storyboarding and paper prototyping). However, it is important to note that, especially in the health context, it may not always be possible for end users to gather in the same physical space:

> *Maybe they cannot get there. Maybe socially it is a challenge for them. Maybe you do engage with those people one-on-one, and then bring things together later. So, an interview, or user testing, or even a digital engagement where you are putting something online and getting some feedback.* [CME6]

Hence, for co-design to be accessible to end users, generative engagement does not necessarily need to occur in the presence of all stakeholders or in the same physical space. For example, Smeenk et al [54] described an empathic handover approach in which end users can participate in the early phases of co-design alongside a principal designer who later translates these contributions [54]. On the other hand, there are also co-design phases that require immersion in the *real-world* context in which the mHealth system will eventually be implemented (see also guideline 4). For example, in the predesign phase, interviews or observations may be carried out in a hospital to gain a better understanding of the problem and the stakeholders that need to be involved. Given the focus on the real-world context, this immersion is especially important for the predesign, evaluative, implementation, and postdesign phases:

> *I think a really important part of that was the fact that we were working on-site* [MSD1]

> *We position ourselves in the context by submerging ourselves in all the relevant stakeholders* [MSD8]

Guidelines for Co-Designing in mHealth

On the basis of the thematic analysis of the interviews, we constructed seven guidelines (guidelines 1-7; Textbox 1) to address the challenges in co-designing mHealth systems. Multimedia Appendix 4 provides a detailed overview of example quotes.

Textbox 1. Guidelines 1-7.

Guideline 1

- Carefully consider the unique circumstances of the targeted disease management or health promotion context with respect to its evaluation and integration requirements, stakeholder involvement, and end user vulnerabilities relating to highly personal aspects of a person's health.

Guideline 2

- As early as possible in the co-design process, consult the behavior change literature and/or involve experts in behavior change relevant to the problem context to effectively identify the targeted change in behavior and adequately plan the type and stakeholder involvement of co-design activities.

Guideline 3

- Select and engage co-design facilitators that have an authentic understanding of the intimate problem context (eg, first-hand experience, immersing in problem context, and literature consultation) and operate in an empathetic way to mitigate potential barriers associated with the power distance between mHealth stakeholders.

Guideline 4

- Immerse yourself in the underlying complex health context to identify and understand stakeholders early, include them in defining their involvement in the co-design process along existing health process requirements, recognize the diversity and inherent power distances among stakeholders, and prioritize the needs of the end user.

Guideline 5

- Throughout every phase of co-design, identify potential postdesign advocates from different stakeholder categories who can aid in implementing the mHealth system (eg, training staff in the use of the system) and champion its use in the postdesign phase (eg, providing feedback on system use in practice).

Guideline 6

- In the evaluative phase, ensure that the mHealth system goes through feasibility testing in the real world (pilot-testing and randomized controlled trials) to adequately address ethical considerations in the health context, determine potential risks to the end users caused by the artifact, and clarify whether it accomplishes its intended goals before implementation.

Guideline 7

- In the postdesign phase, collect usage data to observe the mHealth system's impact after it has been implemented and apply contextual co-design methods to understand this impact.

Guideline 1: Understanding Stakeholder Vulnerabilities and Diversity

The interviews emphasized important differences between health promotion and disease management, including (1) that mHealth users have unique vulnerabilities, (2) the diverse array of stakeholders involved, (3) the significance of evaluation, and (4) the actual implementation and translation. Owing to the focus on health outcomes, mHealth typically involves vulnerable user groups (users with health conditions that may create additional barriers to participation, eg, patients). Although this vulnerability may be present in some groups within health promotion (eg, smoking cessation and alcohol reduction), it appears to be most prevalent among the disease management cohort (eg, dementia and autism). This vulnerability creates challenges for (1) recruiting representatives from the target cohort and (2) being mindful of their health vulnerabilities:

> *The first thing that comes to mind [challenge] is getting access to participants. It is really impossible in healthcare. [...] It might be really hard to get people to open up and be honest about their experiences of having a stoma bag [...]. It is just not a subject that ever gets discussed with family members around. [You] can talk to patients one-on-one maybe, but [not] with all their family around them.* [CME2]

With the specific focus of supporting positive health outcomes, the co-design process inherently touches on vulnerable, deeply personal aspects that, in turn, require high levels of trust in the research team and process. Hence, it is vital to select co-design tools and methods that are appropriate in this context and allow vulnerable end users to participate in the best of their capabilities:

> *I knew from experience what it was like to be with someone who has this disease, and that made it easier because I already knew the context, because then you know how to behave. I think for people who are not familiar with that, they must be acquainted with that first.* [CME3]

> *One [challenge] is a lack of trust. [...] It happens with government led and funded projects where people who may have had a lifetime of being let down by organizations and institutions and they may find it difficult to trust that their voice will really be heard and that things will really change because of their participation.* [CME6]

Each specific mHealth context also has its own diversity of stakeholders who need to be identified and involved in the co-design process. There may also be more than one category of end users, such as both the patient and the health practitioner, with differing requirements in terms of evaluation:

> *You have got app designers, [marketing people, health professionals], and you have got the end-users who are trying to grapple with their medical challenges that they have. [...] That would be one of the biggest challenges, getting those people together.* [MSD5]

The last element of this guideline is the integration of mHealth tools into a wider system landscape. One of the largest identified differences between co-design in mHealth and other contexts is that mHealth systems need to integrate into a highly complex health ecosystem involving an array of health processes, systems, and stakeholders:

> *It is not just about the end product, it is about everything that goes with it that we need to test and work out too. So, the instructions that we give to people as to how to use it, how we advertise it, who we train in the facility in terms of helping patients to use it, how we promote it to staff so that they know it is available to their patients as well.* [MSD1]

Guideline 2: Planning for and Assessing Health Behavior Change

The second theme refers to the importance of consulting the behavior change literature and considers directly involving behavior change experts in the co-design process. Overall, the importance of consulting the behavior change literature was mentioned in 9 of the interviews (4 CMEs and 5 MSDs). The importance of behavior change for mHealth systems design relates back to the very nature of the underlying health promotion and disease management contexts, in that the purpose of these systems involves some change in user behavior to address a health goal [55]. For health promotion, this commonly refers to a change in lifestyle behaviors, such as reducing alcohol consumption and quitting smoking [49,56] or improving eating habits [47]. For disease management, examples include regularly performing rehabilitation exercises [41], following a specific medication regime [57], or recording specific aspects of daily activities [39,41]. Interviewees emphasized that because behavior change is not a by-product but is integrally linked to the purpose of the mHealth system, it is vital to explore in the early stages of the co-design process which behaviors are addressed and in what way:

> *You have to engage in the behavior change literature [...]. A health practitioner probably knows that there is behavior change literature to go to, but someone outside that health domain may not know to go to that literature.* [MSD8]

MSD6 elaborated that a key distinguishing factor between mHealth and other contexts is that co-designing mHealth systems is linked to changes in behavior that are often deeply personal to the end users, which are linked to deeply embedded long-term habits (eg, eating, physical activity, and sleep patterns):

> *You are talking about changing behaviors that are there for a reason. They are not just trivial behaviors, they are deeply embedded and they have really unusual reasonings that [...] surprise you. Whereas if you are just designing a booking system or whatever, it is not that emotive.* [MSD6]

Finally, given the focus on mHealth systems to achieve positive health outcomes, it is vital to carefully tailor the co-design activities to the individual circumstances and capabilities of the stakeholders, particularly the end user:

> *Co-design frameworks [are] very focused on picking a series of methods for a workshop, and then saying, 'okay participants, I all want you to do this method using these kinds of materials.' [This] is just completely unfeasible when you have people with only one hand [...]. Co-design [...] for healthcare [...] does have to be approached differently.* [CME2]

Guideline 3: Identifying and Involving Co-design Facilitators

Interviewees emphasized the critical role of facilitators. In mHealth, co-designing involves high stakeholder diversity (eg, app developers, health practitioners, and health insurance providers) while simultaneously addressing highly intimate issues and concerns regarding a person's health (eg, quitting smoking and diabetes self-management). Against this backdrop, the facilitator plays a critical role in involving stakeholders in a trusted, meaningful, and effective way.

Neglecting the role of the facilitator yields a range of risks, including a lack of true involvement (eg, because of power distance between end users and health professionals) and a lack of understanding about end users' lived experiences and perspectives:

> *I think that power balance is particularly interesting in healthcare because it is really hard to say that you do not agree [with] a doctor. [...] They are held up in such high esteem as being experts of the subject matter [...]. So, to then put-up patients in a room [saying] 'co-design with your doctors', it could be really confronting to [say] 'oh I have a different opinion to you and I do not usually get to express it in my experiences with you, but now can I?'* [CME2]

> *We were a little bit disconnected from knowing what it truly means to struggle with [an] addiction that you want to give up and you know is bad for you [...]. All these issues are quite emotional and unless you understand how it really feels I think it is important for whoever is running the workshop [to] have a feel for the topic, a knowledge of what it means.* [MSD6]

To address potential challenges (eg, power distance and lack of empathy), interviewees emphasized that co-design facilitators need an authentic understanding of end users' real-world experiences (eg, through first-hand experience, immersing in problem context, and consulting relevant literature). Facilitators are then able to operate in a more empathetic way, which can help participants feel more comfortable sharing personal experiences regarding their own health. For example, MSD6

stated that facilitators in their smoking cessation project had personal experience with the context. On the basis of ice-breaking exercises, the authentic experiences of the facilitators enabled them to support stakeholders in becoming more comfortable to actively engage with co-design activities. As a result, the facilitators were perceived as more like co-design participants than authority figures. Empathy, or a "soft human touch" (CME4), is a critical skill for a facilitator running co-design workshops to overcome the inherent power distance issue in the mHealth space:

> *I think the more practical power distance issues in sessions can be easily navigated if you just have a bit of a soft human touch to ensure that people do not feel like you are the cocky arrogant researcher, expert, designer, or however you are positioning yourself.* [CME4]

> *There is comfort that comes from people who are like you. [This] is why I saw the two [facilitators] being so successful with the low self-esteem kids because they themselves started the session talking about their problems. The designers running the session were able to talk about their experiences and how they dealt with it and so then they immediately became not the person leading the co-design activity, but a true co-designer.* [MSD6]

Guideline 4: Immersion Into the mHealth Ecosystem

Co-design involves the effective collaboration of system designers with users and other stakeholders. Frequently raised elements include (1) the importance of being immersed in the context where stakeholders are, for optimal problem identification; (2) identification of relevant stakeholders as early as possible to drive their own involvement and contribute to the study design, including ethics approval; and (3) the need for an ongoing relationship with stakeholders in mHealth that recognizes the power distance between stakeholders and prioritizes the needs of end users.

The multiplicity of factors affecting and supporting a person's health renders the environment of mHealth system design inherently complex. Interviewees repeatedly stressed the need for system designers to immerse themselves deeply to effectively identify stakeholders, understand pain points and relationships with one another, and correctly determine the problems they can and cannot address:

> *You do have to be embedded in the space in order to identify it, or you have to be listening to people who are embedded in the space in order to identify it.* [MSD1]

> *We started with a empathize phase, [which] was around interviewing all different types of stakeholders individually to try and understand what their experiences are, what their frustrations are, what their behaviors and pain points are, what they really struggle with.* [MSD2]

Another important aspect relates to involving stakeholders as early as possible because failing to do so is particularly critical, and possibly fatal, in the realm of mHealth. First, because of the array of factors around a person's health, the number of potential stakeholders is high, which requires buffer times for planning, organizing, communicating, and scheduling. Second, the health sector naturally encompasses complex policies and procedures to adhere to privacy regulations and protect and support vulnerable populations. It is therefore vital for stakeholders to become involved sufficiently early to be able to point out procedural constraints in their domains (eg, requirements and time frames for ethics approvals):

> *I would say involve them right from the start. [...] Ascertain to what extent they are going to be able to contribute any of their time [...] and ask them what stage they think they want to be involved [...] and let them drive that process.* [MSD5]

> *The best way to manage the different stakeholders and the management is at different stages [to] highlight the appropriate stakeholders that are necessary and let them know what their voice is and what their purpose is. Basically, letting stakeholders know when their input is important and needed and what the reason for their input is.* [CME8]

Support for positive health outcomes is an ongoing process. It follows that the relationship with stakeholders of mHealth systems design needs to be managed and supported in an ongoing way. Although this holds true for both health promotion and disease management, it is particularly critical in the disease management space. It is also critical to balance the number of participants involved in co-design activities and avoid a potential power imbalance geared toward senior medical practitioners. The resulting power distances between the stakeholders must be carefully considered. After all, the person most affected by the system will be the end users and, hence, it is vital to adequately capture and address their needs:

> *The problem should really be generated in part by the people who are affected when it has something to do with health management. [...] There must be more of an ongoing relationship [with end-users] even if there is not a particular problem yet.* [MSD1]

> *I get really concerned when I see just one or two people with lived experience brought on as kind of the token users to a predominantly professional group and you just think how can those people feel confident and comfortable in that setting, especially in a health context where they are used to being told by the professionals.* [CME6]

Guideline 5: Identifying and Involving Postdesign Advocates

Interviewees repeatedly stressed the challenge of implementing and rolling out mHealth systems. This is linked to the complexity and risk of processes in the health sector and the multitude of stakeholders who need to work together effectively. Against this backdrop, we identified the importance of postdesign advocates as an important theme. CME1 described postdesign advocates as "end-users who are really interested in what you are doing, how you are doing it, and what it could mean for them." MSD7 elaborates that postdesign advocates

are the "people behind it that are going to drive, push, and refer patients or their communities to [the mHealth system]." Identifying postdesign advocates is critical for system designers to support the implementation and postdesign of the mHealth system.

Postdesign advocates need to be stakeholders who are well-connected and respected in the application context. By actively involving them early in the co-design process, their contributions can already be considered in the predesign and generative phases. This establishes a "buy-in" of stakeholders that can later assist in championing the system in the implementation and postdesign phases with the people and communities who are going to use the system. In this way, identifying postdesign advocates can mitigate many challenges in mHealth, including integration into clinical practice and collecting data in the postdesign phase:

> *Implementation of mHealth tools is extraordinarily challenging [...]. There are a lot of barriers of getting things into practice and getting that buy-in from communities [...] can actually aid your implementation because they have already bought in. Because they are invested in it, they are more likely to try and help make it happen.* [MSD7]

Beyond the technical aspects of the designed system, interviewees argued that the integration into the complex processes and the various stakeholders in the health space renders the implementation and rollout of the system exceptionally challenging. Postdesign advocates can be critical to informing and supporting the implementation of the mHealth system in the real world. If health practitioners do not believe that the system will be useful, they will not promote it to their patients and may actively dissuade use:

> *If they were not taught how to use [the app] properly, if they were not given the right support materials, or if it did not get to the right people because the people who did the roll out of it were not briefed well enough around the sorts of people we want it to go to, even if it was really beautifully designed, then it would have failed. So, I am talking about the wraparound services of the thing.* [MSD3]

> *We talk about champions, you have got to have people behind it that are going to drive it and push it. They are going to refer patients or their communities to it, or they are going to support services to use these tools* [MSD7]

Finally, another reason for involving postdesign advocates is that stakeholders in practice are essential to measuring the impact of the mHealth system once it has been implemented in the real world (eg, based on usage data) and being available for follow-up co-design activities in the postimplementation phase (eg, postdesign interviews) to make sense of the usage data:

> *You will find some end-users in this process who are really interested in what you are doing and how you are doing it and what it could mean for them. Those are the kind of people who might become your post-design advocates who would collect this data for you and at a reasonable price because they have a vested interest in seeing how it worked and helping other people manage their lives for example. So, you could build it into the whole process.* [CME1]

Guideline 6: Applying Health-Specific Evaluation Criteria

The evaluation of mHealth systems requires additional considerations compared with other contexts because of the intended and possibly unintended effects of the artifact on people's health. The main elements uncovered from the interviews were as follows: (1) the risks and ethical issues associated with developing solutions in a health care context and (2) the need for feasibility testing in the real world (eg, clinical trials and pilot-testing), before implementation, to ensure that the mHealth system accomplishes what it set out to do and does not pose a risk to the end users.

Interview participants emphasized that because of the focus on people's health, there are additional risks and ethical considerations when co-design mHealth systems for a health care context. For example, MSD1 elaborated:

> *Even though your interruption through technology might end up with things being better, you still have to be very conscious of the fact that there is more at stake if anything goes wrong because I would not want to be involved in a technology that made things more complicated for people who are already in a complicated and stressful situation.* [MSD1]

To navigate these risks, it is important to ensure that an mHealth system goes through feasibility testing such as pilot-testing and randomized controlled trials in the real world so that it can be established that the mHealth system accomplishes its goals and does not pose a risk to the end users:

> *You need a randomized controlled trial of the app first, so that is in the evaluation phase, not the implementation phase, to then prove that it increases patient outcomes and then they might adopt it.* [MSD2]

MSD8 further explained that it is important to understand that there are different levels of feasibility testing that pertain to the quality of the test and the cost to run the test. For example, MSD8 recommended performing pilot-testing before conducting randomized controlled trials for these reasons:

> *We would never as a health researcher or a health clinician move straight into a randomized controlled trial without pilot data first [...]. In terms of costings, randomized control trials are much more expensive to run and they are the gold star or grade one evidence* [MSD8]

MSD1 adds that feasibility testing is not only important for ensuring that the mHealth system works and poses no risk to end users. Owing to the extensive costs of upkeep after implementation, finding problems during feasibility testing is beneficial because they can be fixed by re-entering the earlier co-design phases. Thereby, some problems are more likely to be found because testing is performed in a real-world setting:

> *You need to do it in stages, especially because there is such a massive cost involved in terms of the upkeep of apps [...]. If you can have a prototype, it is not just about testing the prototype, it is also about testing how the prototype works in the real world before you turn it into the end product.* [MSD1]

Guideline 7: Collecting and Analyzing Usage Data to Understand Impact

The final theme focuses on the importance of postdesign. CMEs noted that even though postdesign is important to assess the impact of artifacts in the real world [13], this step is frequently not carried out because of time and cost constraints. However, the postdesign phase is especially important in the mHealth context because impact is driven by changes in health behavior [10], and mHealth systems are intended to be used over extended time spans (eg, diabetes self-management). Hence, mHealth systems must be updated to meet changing user needs:

> *I think post-implementation and the collection of evidence of the impact of that change is absolutely essential because you are talking about people changing their behavior for better health outcomes"* [CME7]

> *All apps need to be updated, and one of the biggest issues with health apps is they are not.* [MSD7]

In addition, interviewees explained that mHealth is in a unique position to measure this impact because of having access to participant usage data from the mHealth system because of their ability to collect usage data. Thereby, it is important that qualitative co-design methods (eg, postdesign interviews) are used to make sense of this participant usage data:

> *You have got all these functionality and metrics that you can get from mHealth that you cannot get anywhere else [such as] Google metrics, Google Analytics, and usage statistics [...]. That is a whole avenue of data that you do not have when you do not have mHealth.* [MSD8]

> *There are ways to get feedback, like usage statistics. Those do not tell you why. Having more qualitative methods to get feedback is really important.* [CME6]

Discussion

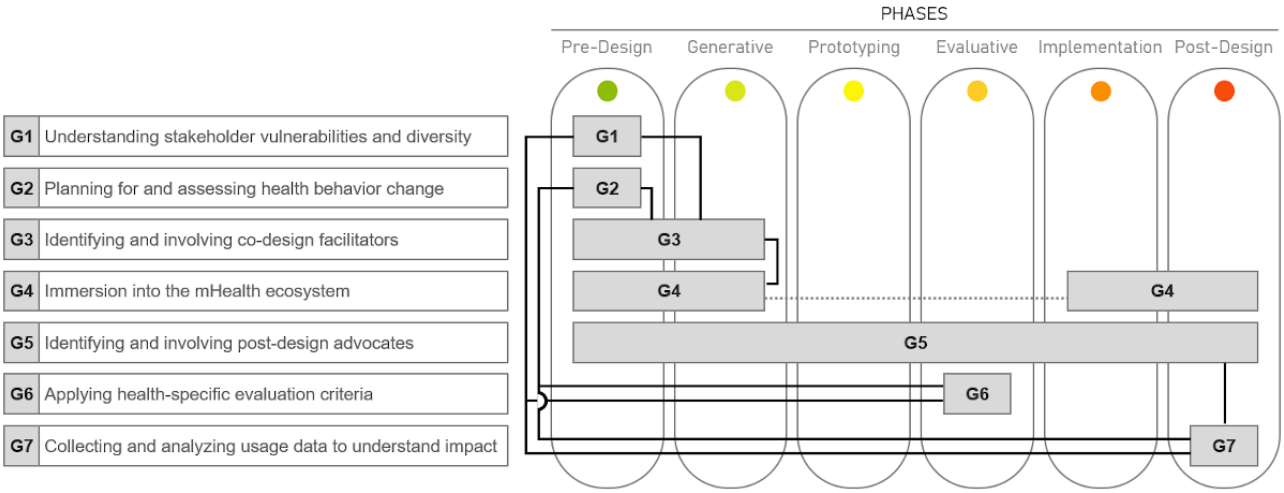
General Discussion

Although extensive research exists on co-design methodology and its general application, limited research has examined the complexities that arise in co-designing mHealth systems. This study aimed to (1) contextualize an existing co-design framework for mHealth and (2) develop guidelines for addressing common challenges. From the 16 interviews conducted with CMEs and MSDs and thematic analysis, we contextualized the co-design framework by Sanders and Stappers [13] to the mHealth context and constructed a set of 7 guidelines.

We identified several important aspects from contextualizing the Sanders and Stappers [13] co-design framework. First, it became apparent that some of the co-design phases should be split up. Although the original framework has an overall generative phase, a separate *prototyping phase* was suggested for mHealth to distinguish between the *generation of early concepts* (eg, low-fidelity prototyping) in the generative phase compared with the *testing of more mature concepts* where maturity is higher (eg, high-fidelity prototyping). Furthermore, a dedicated *implementation phase* distinguishes activities performed during evaluation (eg, pilot-testing and randomized controlled trials) versus implementation (eg, creating documentation, training, and user acceptance). Second, mHealth has its idiosyncrasies regarding the *front-end* of co-design, including the importance of researchers immersing themselves in the complex problem context and diverse stakeholder landscape surrounding a person's health. Furthermore, this diversity of stakeholders can lead to a power distance issue in the *generative phase.* Therefore, it is important to recognize the vulnerability of mHealth end users and their relationships with other stakeholders that could impede participation. The *evaluative phase* is also affected, as mHealth problems are typically riskier compared with other contexts. Thus, pilot-testing and randomized controlled trials were mentioned by interviewees as suitable evaluation methods for mHealth. Finally, the *postdesign phase* plays a specific role in mHealth because of its intended effects on health behavior. However, this cannot be assessed until the system is deployed. Hence, the postdesign phase is necessary to understand this impact on user behavior and to allow for continued monitoring and maintenance.

Addressing the second research objective, seven guidelines were synthesized for applying co-design to mHealth (labeled guideline 1 to guideline 7). As shown in Figure 3, the guidelines pertain to the specific phases of the co-design process. Emphasizing the importance of the front end of co-design, guideline 1 to guideline 4 focus on ensuring that researchers and practitioners establish an intimate understanding of the problem context as early as possible. Interviewees noted that, by following these steps, common challenges such as stakeholder identification, power distance, and lack of trust can be addressed effectively. For instance, by immersing oneself in the mHealth problem context (guideline 4), researchers and practitioners can better understand how end users interface with stakeholders in their health ecosystem, aiding in stakeholder identification. Guideline 5 maps to all phases in the framework as (postdesign) advocates (ie, users championing the system) can be identified in any phase. Interviewees noted that these advocates can help mitigate many issues that can potentially surface in the implementation phase, for instance, by championing the system themselves and by training others. Next, guideline 6 maps to the evaluative phase and emphasizes the importance of health-specific evaluation (eg, pilot-testing and randomized controlled trials) given the high-risk nature of mHealth challenges. Finally, guideline 7 maps to the postdesign phase to ensure that the impact is measured post implementation along with contextual research that informs further system refinements.

Figure 3. Interplay of guidelines and mapping to the co-design phases.



Beyond phase-specific relevance, there is also an important interplay between the guidelines. First, guidelines 1, 2, and 4 are linked to 3 because a deep *understanding* of the problem context is needed to effectively apply guideline 3 (highlighting the importance of the front-end of co-design in mHealth). Without this, many of the challenges identified by the interviewees (eg, power distance, lack of trust, and accessibility of tools and methods) would compromise later co-design phases. Second, there is a link between guideline 1/guideline 2 and guideline 6/guideline 7. Guideline 1 and 2 primarily focus on understanding the problem context and establishing the desired goals of the mHealth system, while guidelines 6 and 7 refer to the evaluation of how well the mHealth system addresses these goals, both pre- and postimplementation. Finally, there is a link between guidelines 5 and 7 since the identified advocates will invariably be needed to understand the impact of the mHealth system in the real world.

There were key differences and similarities between the responses of CMEs and MSDs, with implications for the results presented in this paper. First, the interview guide acknowledged the *differing nature of expertise* between the CME and MSD groups to elucidate the experts' specific domain knowledge (Multimedia Appendix 2). For instance, questions for CMEs primarily referred to co-design in a *general* sense, which tapped into their expertise in co-design applications, processes, phases, methods, and tools. Conversely, questions for MSDs focused on how co-design manifested in their own projects, which led to more *specific* responses about the *contextualization of co-design to mHealth* (challenges they had faced, how they involved mHealth stakeholders, benefit of co-design to their project, etc). The groups expressed similar considerations around the challenges and benefit of co-design, since cost and time constraints are typically common factors in co-design processes, regardless of the context. There was general between-group consensus regarding the aspects that would inform the derivation of guidelines and the contextualization of the framework. Overall, the responses from CMEs tended to refer to general considerations around applying co-design to a complex area, whereas the responses from MSDs were more specific to challenges and best practices based on experience from actual mHealth projects.

Implications

This work has several important implications for researchers and practitioners. First, building on the extensive expertise of CMEs and MSDs who participated in this research, the contextualized framework may provide a shared frame of reference to guide mHealth systems development projects, which are interdisciplinary in nature [5,6]. Rooted in the widely used co-design framework by Sanders and Stappers [13], the contextualized framework brings to light a range of critical considerations that arise in the health context. As a shared frame of reference, the contextualized framework may aid mHealth researchers and practitioners in planning co-design activities and involving stakeholders in all stages of design [1,6]:

> *This is great work. There is definitely work in what you are doing. [...] Any type of framework that helps us to do this on the ground more effectively and in [the mHealth] context, that is the kind of work that we need.* [CME4]

Complementary to the framework, the guidelines point to pitfalls in mHealth systems development along with specific suggestions on how these challenges can be navigated. Multimedia Appendix 5 provides a checklist for co-designing mHealth systems projects according to the 7 guidelines. By facilitating stakeholder engagement and involvement in co-design activities, these guidelines may help researchers and practitioners to ensure that mHealth systems are underpinned by expert insight, reflect the lived experiences of end users, and integrate into the existing system and process landscape [5,11]. In so doing, co-designed mHealth artifacts may enable end users and health professionals to develop a stronger sense of ownership and agency over the outcome. Researchers and practitioners can actively engage postdesign advocates to assist in increasing buy-in from stakeholders, overcoming barriers, and championing the system's implementation and use.

Limitations and Opportunities for Future Research

This study had some limitations. First, it should be noted that most of the interviewees resided in the Oceania region and/or were working within the academic sector. Future research may bring to light potential differences in co-design between industry and academia as well as geographical differences related to

cultural factors (eg, uncertainty avoidance and power distance [58]). Second, while our research builds on the expertise of the interviewed experts beyond the scope of an individual mHealth system, future research is warranted on the usefulness of the contextualized framework and guidelines for the development of actual mHealth systems. Importantly, this evaluation and refinement should also include the perspectives and lived experiences of individuals involved as co-designers in the context of a specific mHealth system. This was beyond the scope of this study, as we considered mHealth system design beyond the scope of a specific mHealth systems development project. Third, because the current focus is on the co-design process as a whole, it was beyond the scope of this study to assess the applicability of specific co-design methods for mHealth (eg, cultural probes and journey maps). Future research needs to investigate the usefulness and boundary conditions of individual co-design methods in mHealth. This would have the potential to illuminate specific activities that can assist in stakeholder engagement and impact determination in the long run.

Conclusions

With the focus of supporting positive health outcomes, researchers and practitioners encounter unique challenges in mHealth systems development. Following a constructivist approach, we interviewed 16 experts in co-design methods and mHealth systems development to contextualize an established co-design framework for the mHealth setting and to construct a set of tangible guidelines to address common challenges in this space. While contextualization emphasizes the need to include dedicated prototyping and implementation phases, the guidelines provide practical insights on how to engage in this process by (1) understanding stakeholder vulnerabilities and diversity, (2) planning for and assessing health behavior change, (3) identifying co-design facilitators, (4) immersing in the mHealth ecosystem, (5) identifying postdesign advocates, (6) applying health-specific evaluation criteria, and (7) analyzing usage data and contextual research to understand impact. We hope that the contextualized framework and guidelines presented in this work will serve as a shared frame of reference to facilitate interdisciplinary collaboration at the nexus of information technology and health research.

Acknowledgments

This research was supported by an Australian Government Research Training Program Scholarship.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Background of interview participants.
[DOCX File , 24 KB - mhealth_v9i11e27896_app1.docx]

Multimedia Appendix 2
Interview guide.
[DOCX File , 23 KB - mhealth_v9i11e27896_app2.docx]

Multimedia Appendix 3
Example quotes for framework contextualization.
[DOCX File , 25 KB - mhealth_v9i11e27896_app3.docx]

Multimedia Appendix 4
Example quotes for the 7 guidelines.
[DOCX File , 31 KB - mhealth_v9i11e27896_app4.docx]

Multimedia Appendix 5
Checklist for system designers.
[DOCX File , 24 KB - mhealth_v9i11e27896_app5.docx]

References

1. Eckman M, Gorski I, Mehta K. Leveraging design thinking to build sustainable mobile health systems. J Med Eng Technol 2016;40(7-8):422-430. [doi: 10.1080/03091902.2016.1218560] [Medline: 27535325]
2. Wowak KD, Adjerid I, Angst CM, Guzman JC. A tutorial on empirical ICT4D research in developing countries: processes, challenges, and lessons. Commun Assoc Inf Syst 2016;38(1):353-374. [doi: 10.17705/1cais.038121]
3. mHealth: new horizons for health through mobile technologies. World Health Organization. 2011. URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2021-06-10]

4. The rise of mHealth apps: a market snapshot. Liquid State. 2018. URL: https://liquid-state.com/mhealth-apps-market-snapshot/ [accessed 2018-03-26]
5. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, et al. Current science on consumer use of mobile health for cardiovascular disease prevention. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
6. Marzano L, Bardill A, Fields B, Herd K, Veale D, Grey N, et al. The application of mHealth to mental health: opportunities and challenges. Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/S2215-0366(15)00268-0] [Medline: 26462228]
7. Winters N, Oliver M, Langer L. Can mobile health training meet the challenge of 'measuring better'? Comp Educ 2017;53(1):115-131. [doi: 10.4324/9780429444494-6]
8. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
9. Medhanyie A, Moser A, Spigt M, Yebyo H, Little A, Dinant G, et al. Mobile health data collection at primary health care in ethiopia: a feasible challenge. J Clin Epidemiol 2015 Jan;68(1):80-86. [doi: 10.1016/j.jclinepi.2014.09.006] [Medline: 25441699]
10. Noorbergen TJ, Adam MT, Attia JR, Cornforth DJ, Minichiello M. Exploring the design of mHealth systems for health behavior change using mobile biosensors. Commun Assoc Inf Syst 2019:944-981. [doi: 10.17705/1cais.04444]
11. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. J Behav Med 2017 Feb;40(1):85-98 [FREE Full text] [doi: 10.1007/s10865-016-9818-7] [Medline: 28058516]
12. Garnett C, Crane D, Michie S, West R, Brown J. Evaluating the effectiveness of a smartphone app to reduce excessive alcohol consumption: protocol for a factorial randomised control trial. BMC Public Health 2016 Jul 08;16(536):1-10 [FREE Full text] [doi: 10.1186/s12889-016-3140-8] [Medline: 27392430]
13. Sanders EB, Stappers PJ. Probes, toolkits and prototypes: three approaches to making in codesigning. CoDesign 2014 Mar 06;10(1):5-14. [doi: 10.1080/15710882.2014.888183]
14. Visser FS, Stappers PJ, van der Lugt R, Sanders EB. Contextmapping: experiences from practice. CoDesign 2005 Apr;1(2):119-149. [doi: 10.1080/15710880500135987]
15. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
16. Danaher BG, Brendryen H, Seeley JR, Tyler MS, Woolley T. From black box to toolbox: outlining device functionality, engagement activities, and the pervasive information architecture of mHealth interventions. Internet Interv 2015 Mar 01;2(1):91-101 [FREE Full text] [doi: 10.1016/j.invent.2015.01.002] [Medline: 25750862]
17. O'Reilly GA, Spruijt-Metz D. Current mHealth technologies for physical activity assessment and promotion. Am J Prev Med 2013 Oct;45(4):501-507 [FREE Full text] [doi: 10.1016/j.amepre.2013.05.012] [Medline: 24050427]
18. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
19. Ho SY, Guo X, Vogel D. Opportunities and challenges in healthcare information systems research: caring for patients with chronic conditions. Commun Assoc Inf Syst 2019:852-873. [doi: 10.17705/1cais.04439]
20. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
21. Jang H, Han SH, Kim JH. User perspectives on blockchain technology: user-centered evaluation and design strategies for dapps. IEEE Access 2020;8:226213-226223. [doi: 10.1109/access.2020.3042822]
22. Banos O, Villalonga C, Garcia R, Saez A, Damas M, Holgado-Terriza JA, et al. Design, implementation and validation of a novel open framework for agile development of mobile health applications. Biomed Eng Online 2015;14 Suppl 2:S6 [FREE Full text] [doi: 10.1186/1475-925X-14-S2-S6] [Medline: 26329639]
23. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
24. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
25. Brandt E, Binder T, Sanders E. Tools and techniques: ways to engage telling, making and enacting. In: Routledge International Handbook of Participatory Design. New York, NY: Routledge; 2012:145-181.
26. Eyles H, Jull A, Dobson R, Firestone R, Whittaker R, Te Morenga L, et al. Co-design of mHealth delivered interventions: a systematic review to assess key methods and processes. Curr Nutr Rep 2016 Jul 4;5(3):160-167. [doi: 10.1007/s13668-016-0165-7]
27. Noorbergen TJ, Adam MT, Roxburgh M, Teubner T. Co-design in mHealth systems development: insights from a systematic literature review. AIS Trans Human-Computer Interact 2021 Jun 30;13(2):175-205. [doi: 10.17705/1thci.00147]

28. Capel T, Schnittert J, Snow S, Vyas D. Exploring motivations of young adults to participate in physical activities. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. 2015 Apr Presented at: CHI Conference on Human Factors in Computing Systems; April 18 - 23, 2015; Seoul, Republic of Korea p. 1409-1414. [doi: 10.1145/2702613.2732800]
29. Danbjørg DB, Villadsen A, Gill E, Rothmann MJ, Clemensen J. Usage of an exercise app in the care for people with osteoarthritis: user-driven exploratory study. JMIR mHealth uHealth 2018 Jan 11;6(1):e11 [FREE Full text] [doi: 10.2196/mhealth.7734] [Medline: 29326092]
30. Kanstrup A. Design concepts for digital diabetes practice: design to explore, share, and camouflage chronic illness. Int J Des 2014;8(3):49-60 [FREE Full text]
31. Wechsler J. HCD mobile health project: post collaboration reflection of researcher and designer. In: Proceedings of the Asia Pacific HCI and UX Design Symposium. 2015 Presented at: APCHIUX '15: Asia Pacific Symposium of HCI and UX Design; December 7 - 10, 2015; Melbourne, Australia p. 16-21. [doi: 10.1145/2846439.2846442]
32. Aljaroodi H, Adam M, Chiong R, Cornforth D, Minichiello M. Empathic avatars in stroke rehabilitation: a co-designed mHealth artifact for stroke survivors. In: Proceedings of the International Conference on Design Science Research in Information System and Technology. 2017 May 23 Presented at: International Conference on Design Science Research in Information Systems and Technology; May 30 - June 1, 2017; Karlsruhe, Germany p. 73-89. [doi: 10.1007/978-3-319-59144-5_5]
33. Cordova D, Bauermeister JA, Fessler K, Delva J, Nelson A, Nurenberg R, Youth Leadership Council. A community-engaged approach to developing an mHealth HIV/STI and drug abuse preventive intervention for primary care: a qualitative study. JMIR mHealth uHealth 2015 Dec 18;3(4):e106 [FREE Full text] [doi: 10.2196/mhealth.4620] [Medline: 26685288]
34. Das A, Bøthun S, Reitan J, Dahl Y. The use of generative techniques in co-design of mHealth technology and healthcare services for COPD patients. In: Proceedings of the International Conference of Design, User Experience, and Usability. 2015 Jul 21 Presented at: International Conference of Design, User Experience, and Usability; August 2-7, 2015; Los Angeles, CA, USA p. 587-595. [doi: 10.1007/978-3-319-20889-3_54]
35. Te Morenga L, Pekepo C, Corrigan C, Matoe L, Mules R, Goodwin D, et al. Co-designing an mHealth tool in the New Zealand Māori community with a "Kaupapa Māori" approach. AlterNative 2018 Jan 23;14(1):90-99. [doi: 10.1177/1177180117753169]
36. Verbiest ME, Corrigan C, Dalhousie S, Firestone R, Funaki T, Goodwin D, et al. Using codesign to develop a culturally tailored, behavior change mHealth intervention for indigenous and other priority communities: a case study in New Zealand. Transl Behav Med 2019 Jul 16;9(4):720-736. [doi: 10.1093/tbm/iby093] [Medline: 30388262]
37. De la Harpe R. Lessons learnt from the participatory design of a mobile care data application in a resource-restricted context. In: Proceedings of the 12th Participatory Design Conference: Exploratory Papers, Workshop Descriptions, Industry Cases. 2012 Presented at: PDC '12: 12th Participatory Design Conference; August 12 - 16, 2012; Roskilde, Denmark p. 125-128. [doi: 10.1145/2348144.2348184]
38. VanHeerwaarden N, Ferguson G, Abi-Jaoude A, Johnson A, Hollenberg E, Chaim G, et al. The optimization of an eHealth solution (Thought Spot) with transition-aged youth in postsecondary settings: participatory design research. J Med Internet Res 2018 Mar 06;20(3):e79 [FREE Full text] [doi: 10.2196/jmir.8102] [Medline: 29510970]
39. Castensøe-Seidenfaden P, Reventlov Husted G, Teilmann G, Hommel E, Olsen BS, Kensing F. Designing a self-management app for young people with type 1 diabetes: methodological challenges, experiences, and recommendations. JMIR mHealth uHealth 2017 Oct 23;5(10):e124 [FREE Full text] [doi: 10.2196/mhealth.8137] [Medline: 29061552]
40. Lipson-Smith R, White F, White A, Serong L, Cooper G, Price-Bell G, et al. Co-design of a consultation audio-recording mobile app for people with cancer: The secondears app. JMIR Form Res 2019 Mar 12;3(1):e11111 [FREE Full text] [doi: 10.2196/11111] [Medline: 30860487]
41. Davis SR, Peters D, Calvo RA, Sawyer SM, Foster JM, Smith L. "Kiss myAsthma": using a participatory design approach to develop a self-management app with young people with asthma. J Asthma 2018 Sep 24;55(9):1018-1027. [doi: 10.1080/02770903.2017.1388391] [Medline: 29064746]
42. Peters D, Davis S, Calvo RA, Sawyer SM, Smith L, Foster JM. Young people's preferences for an asthma self-management app highlight psychological needs: a participatory study. J Med Internet Res 2017 Apr 11;19(4):e113 [FREE Full text] [doi: 10.2196/jmir.6994] [Medline: 28400353]
43. Woods L, Cummings E, Duff J, Walker K. The development and use of personas in a user-centred mHealth design project. In: Proceedings of the 29th Australian Conference on Computer-Human Interaction. 2017 Presented at: OzCHI '17: 29th Australian Conference on Human-Computer Interaction; November 28 - December 1, 2017; Brisbane Queensland Australia p. 560-565. [doi: 10.1145/3152771.3156186]
44. Woods L, Cummings E, Duff J, Walker K. Conceptual design and iterative development of a mHealth app by clinicians, patients and their families. Stud Health Technol Inform 2018;252:170-175. [Medline: 30040701]
45. Woods L, Cummings E, Duff J, Walker K. Design thinking for mHealth application co-design to support heart failure self-management. Stud Health Technol Inform 2017;241:97-102. [Medline: 28809190]
46. Løventoft P, Nørregaard L, Frøkjær E. Designing daybuilder: an experimental app to support people with depression. In: Proceedings of the 12th Participatory Design Conference: Exploratory Papers, Workshop Descriptions, Industry Cases.

2012 Presented at: PDC '12: 12th Participatory Design Conference; August 12 - 16, 2012; Roskilde, Denmark p. 1-4. [doi: 10.1145/2348144.2348146]
47. Dol A, Kulyk O, Velthuijsen H, van Gemert-Pijnen L, van Strien T. Denk je zèlf! Developing a personalised virtual coach for emotional eaters using personas. In: Proceedings of the The Eighth International Conference on eHealth, Telemedicine, and Social Medicine (eTelemed 2016). 2016 Presented at: The Eighth International Conference on eHealth, Telemedicine, and Social Medicine (eTelemed 2016); April 24-28, 2016; Venice, Italy p. 42-47 URL: https://www.researchgate.net/publication/303342099_Denk_je_zelf_Developing_a_Personalised_Virtual_Coach_for_Emotional_Eaters_using_Personas
48. Paay J, Kjeldskov J, Skov M, Lichon L, Rasmussen S. Understanding individual differences for tailored smoking cessation apps. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI '15: CHI Conference on Human Factors in Computing Systems; April 18 - 23, 2015; Seoul, Republic of Korea p. 1699-1708. [doi: 10.1145/2702123.2702321]
49. Paay J, Kjeldskov J, Brinthaparan U, Lichon L, Rasmussen S, Srikandaraja N. QuittyLink: Involving smokers in the design of technology that supports individuals in quitting. In: Proceedings of the Fourth International Conference on Design4Health 2017. 2017 Presented at: Fourth International Conference on Design4Health 2017; December 4-7, 2017; Melbourne, Australia p. 128-132 URL: https://vbn.aau.dk/en/publications/quittylink-involving-smokers-in-the-design-of-technology-that-sup
50. Arslan P, Nam H, Romero M, Perego P, Costa F, Andreoni G, et al. MOHE: Mobile health for moms, kids, adults and elderly. In: Proceedings of the First International Conference on Design Creativity, ICDC 2010. 2010 Presented at: First International Conference on Design Creativity, ICDC 2010; November 29 - December 1, 2010; Kobe, Japan p. 1-6 URL: https://www.academia.edu/27565823/MOHE_Mobile_Health_for_Moms_Kids_Adults_and_Elderly
51. Mann K, MacLeod A. Constructivism: learning theories and approaches to research. In: Cleland J, Durning SJ, editors. Researching Medical Education. New York, United States: John Wiley & Sons, Ltd; 2015:51-65.
52. Wachtler C, Brorsson A, Troein M. Meeting and treating cultural difference in primary care: a qualitative interview study. Fam Pract 2006 Feb;23(1):111-115. [doi: 10.1093/fampra/cmi086] [Medline: 16246851]
53. Varpio L, Ajjawi R, Monrouxe LV, O'Brien BC, Rees C. Shedding the cobra effect: problematising thematic emergence, triangulation, saturation and member checking. Med Educ 2017 Jan;51(1):40-50 [FREE Full text] [doi: 10.1111/medu.13124] [Medline: 27981658]
54. Smeenk W, Sturm J, Eggen B. Empathic handover: how would you feel? Handing over dementia experiences and feelings in empathic co-design. CoDesign 2018;14(4):259-274. [doi: 10.1080/15710882.2017.1301960]
55. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
56. 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]
57. Vilarinho T, Floch J, Stav E. Co-designing a mHealth application for self-management of cystic fibrosis. In: Proceedings of the IFIP Conference on Human-Computer Interaction. 2017 Presented at: IFIP Conference on Human-Computer Interaction; September 25-29 , 2017; Bombay, India p. 3-22. [doi: 10.1007/978-3-319-67687-6_1]
58. Matusitz J, Musambira G. Power distance, uncertainty avoidance, and technology: analyzing hofstede's dimensions and human development indicators. J Technol Hum Serv 2013 Jan;31(1):42-60. [doi: 10.1080/15228835.2012.738561]

Abbreviations

CME: co-design method expert
IT: information technology
mHealth: mobile health
MSD: mHealth system designer

Edited by L Buis; submitted 11.02.21; peer-reviewed by E Sanders, P Cairns; comments to author 07.03.21; revised version received 22.04.21; accepted 03.09.21; published 10.11.21.

Please cite as:
Noorbergen TJ, Adam MTP, Teubner T, Collins CE
Using Co-design in Mobile Health System Development: A Qualitative Study With Experts in Co-design and Mobile Health System Development
JMIR Mhealth Uhealth 2021;9(11):e27896
URL: https://mhealth.jmir.org/2021/11/e27896
doi:10.2196/27896
PMID:34757323

Original Paper

Analysis of Apps With a Medication List Functionality for Older Adults With Heart Failure Using the Mobile App Rating Scale and the IMS Institute for Healthcare Informatics Functionality Score: Evaluation Study

Yohanca Maria Diaz-Skeete[1], BSc, MSci; David McQuaid[2]; Adewale Samuel Akinosun[3]; Idongesit Ekerete[4]; Natacha Carragher[5,6], PhD; Lucia Carragher[1], PhD

[1]NetwellCASALA Advanced Research Centre, Dundalk Institute of Technology, Dundalk, Ireland

[2]CCT College Dublin, Dublin, Ireland

[3]University of Highlands and Islands, Inverness, United Kingdom

[4]Ulster University, Belfast, United Kingdom

[5]National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia

[6]Alcohol, Drugs and Addictive Behaviours, Department of Mental Health and Substance Use, World Health Organization, Geneva, Switzerland

Corresponding Author:
Yohanca Maria Diaz-Skeete, BSc, MSci
NetwellCASALA Advanced Research Centre
Dundalk Institute of Technology
Dublin Road
Marshes Upper
Dundalk, A91 K584
Ireland
Phone: 353 429370497 ext 2592
Email: yohanca.diaz@dkit.ie

Abstract

Background: Managing the care of older adults with heart failure (HF) largely centers on medication management. Because of frequent medication or dosing changes, an app that supports these older adults in keeping an up-to-date list of medications could be advantageous. During the COVID-19 pandemic, HF outpatient consultations are taking place virtually or by telephone. An app with the capability to share a patient's medication list with health care professionals before consultation could support clinical efficiency, for example, by reducing consultation time. However, the influence of apps on maintaining an up-to-date medication history for older adults with HF in Ireland remains largely unexplored.

Objective: The aims of this review are twofold: to review apps with a medication list functionality and to assess the quality of the apps included in the review using the Mobile App Rating Scale (MARS) and the IMS Institute for Healthcare Informatics functionality scale.

Methods: A systematic search of apps was conducted in June 2019 using the Google Play Store and iTunes App Store. The MARS was used independently by 4 researchers to assess the quality of the apps using an Android phone and an iPad. Apps were also evaluated using the IMS Institute for Healthcare Informatics functionality score.

Results: Google Play and iTunes App store searches identified 483 potential apps (292 from Google Play and 191 from iTunes App stores). A total of 6 apps (3 across both stores) met the inclusion criteria. Of the 6 apps, 4 achieved an acceptable MARS score (3/5). The Medisafe app had the highest overall MARS score (4/5), and the Medication List & Medical Records app had the lowest overall score (2.5/5). On average, the apps had 8 functions based on the IMS functionality criteria (range 5-11). A total of 2 apps achieved the maximum score for number of features (11 features) according to the IMS Institute for Healthcare Informatics functionality score, and 2 scored the lowest (5 features). Peer-reviewed publications were identified for 3 of the apps.

Conclusions: The quality of current apps with medication list functionality varies according to their technical aspects. Most of the apps reviewed have an acceptable MARS objective quality (ie, the overall quality of an app). However, subjective quality (ie, satisfaction with the apps) was poor. Only 3 apps are based on scientific evidence and have been tested previously. A total of 2 apps featured all the IMS Institute for Healthcare Informatics functionalities, and half did not provide clear instructions on

how to enter medication data, did not display vital parameter data in an easy-to-understand format, and did not guide users on how or when to take their medication.

(JMIR Mhealth Uhealth 2021;9(11):e30674) doi:10.2196/30674

KEYWORDS

mobile app; mHealth; medication app; heart failure; Mobile App Rating Scale

Introduction

Background

Managing the care of older adults with heart failure (HF) largely centers on symptom and medication management [1]. Medication management in patients with HF is challenging due to frequent medication or dosing changes [2,3] and polypharmacy, as some patients with HF typically take on average 10-25 tablets daily [4]. Polypharmacy is associated with poor adherence to pharmacological therapies, drug interactions, inappropriate drug prescriptions, and other adverse effects [5]. A recent report by the World Health Organization [6] argues that technology can improve patient experiences and medication adherence and enable patients to become active participants in medication reviews. Mobile apps offer the potential to augment care for patients with HF. Apps can potentially support older adults to find information on the medications (ie, drug interactions), track their medication, communicate with health care providers, keep a daily record of their blood pressure and weight measurements, and facilitate an accurate medication history. However, there is a dearth of literature on apps specifically to support medication history. An accurate medication list prevents adverse drug events [7], increases patients' care outcomes, decreases hospitalization and mortality rates [8,9], and supports medication adherence for patients self-managing at home.

Given the complexity of HF self-care, assisting older adults in managing their own care at home is critical to the success of HF management. Emerging evidence suggests that mobile health (mHealth), particularly mobile technologies, can serve as a form of support for patients with HF and may enhance patient-provider collaboration for self-management [1,10]. By their nature, mobile devices, such as phones, are carried by people and, therefore, are always with them, offering opportunities beyond simple remote monitoring to assist with the management of care. In the current context of the COVID-19 pandemic, when the community (and especially older adults) is requested to maintain social distancing, the public health landscape is changing and mHealth has never been so important for treatment [10,11].

For older adults, social isolation and loneliness increase the risk of anxiety, depression symptoms, heart disease, reduction of activities of daily living, morbidity, and mortality [12,13]. Government recommendations to self-isolate during this pandemic have undoubtedly had a detrimental effect on older adults, including those that previously had wide social connections with the community and relatives [14]. Older adults who were previously attending outpatient appointments have seen their access restricted. In Ireland, the Health Service Executive website notes that all outpatient appointments are postponed until further notice [15]. Health care professionals (HCPs) working in outpatient clinics are seeing a reduced number of patients, with most consultations now taking place over the phone, bar medical emergencies. In Ireland, McGlynn [16] drew attention to the sharp decline in cardiac outpatient appointments during March to April 2020 (300,000 appointments) compared with the same period in 2019. Therefore, the need for new models of care in this changed environment to support older adults at home to alleviate their mental and physical burden, as well as provide medical care, is especially timely [10,11].

Across many countries, emerging evidence confirms the important role that mHealth can play in community care, especially during COVID-19. In the United States, there have been 10-fold web-based consultations in a few weeks [17], "...as big a transformation as any ever before in the history of US health care." Similarly, Canada, South Africa, India, and the United Kingdom are conducting health care web-based consultations at an exponential rate [17]. In Ireland, the platform *Attend Anywhere*, endorsed by the Health Service Executive, is now widely available for HCPs to conduct web-based consultations [18]. However, the use of this platform is not even across Irish clinical settings (some HCPs are actively using it and others are not). Many outpatient services are consulting patients over the phone during the COVID-19 pandemic [19,20], except for patients with exacerbated symptoms.

The process of medication review over the phone is difficult and time-consuming [21]. HCPs have to listen attentively to the information the patient is conveying while lacking visual cues (ie, printed medication list or medication blisters provided by pharmacists). Instead, each patient has to spell each medication list over the phone, raising confusion over similarities of the medication name or dose, thus increasing the length of the consultation. An app sharing an up-to-date medication list with HCPs preconsultation could reduce medication and dosing errors, making the consultation process more efficient. A usability study of an app developed for HF self-management was conducted in Australia [22]. A total of 8 participants tested the app over a 14-day period and 6 of them assessed the app using the Mobile App Rating Scale (MARS; widely used to assess the quality of mobile apps). The app was found to be of acceptable quality and beneficial for HF self-management. Interestingly, the medication list feature on the app was considered by the patients to be beneficial; however, none of the patients used it during the 14 days. The authors suggested that participants found it difficult to incorporate the app into their self-care routine [22].

Emerging evidence suggests that apps can support patients by checking drug interactions, tracking medication intake, and facilitating an up-to-date list of medications. However, although

some app interventions have explored HF self-management [23,24], app interventions focusing on older adults with HF maintaining an accurate medication list have been limited and largely unexplored. This paper aims to explore the benefits of apps with a medication list functionality, explore their role in the COVID-19 pandemic context, and assess their quality using the MARS and the IMS Institute for Healthcare Informatics functionality score.

Objectives

The objectives of this review are two-fold: to review apps with a medication list functionality and to evaluate the quality of the apps included in the review with a validated scale. To assess the quality and functionality of the apps, 2 tools were used: the MARS tool and the IMS Institute for Healthcare Informatics functionality score. For the purpose of this paper, an app with a medication list functionality is an app that generates a comprehensive medication history, allowing the patient to email or share the list in real time with HCPs.

Methods

Overview

A systematic search of apps accessible in Ireland was conducted in June 2019 using the Google Play and iTunes App stores. The purpose of this search was to identify apps with a medication list functionality. The search term *medication list* was used to identify apps with a medication list functionality. The term *medication app* was excluded from the search as it identified apps with a different primary purpose (eg, medication alarm, medication tracker, medication reminder, apps providing educational information only, medical decision support systems for clinicians, medication adverse effect, pharmacy locator, and prescription refills). After the initial identification of apps containing a medical list function, the apps were tabulated. If the same app was available on different platforms (iOS or Android), both versions were retained for analysis as apps behave differently depending on the platform, as seen in previous work by Nicholas et al [25]. Inclusion and exclusion criteria (discussed next) were applied to each app to determine whether they should be retained for further analysis.

Apps that met the inclusion criteria were downloaded and evaluated by a team of 4 researchers. The MARS was used to assess the quality of the apps using an Android phone and an iPad. Apps were also evaluated using the IMS Institute for Healthcare Informatics functionality scores.

A Google Scholar search of the apps was conducted to identify apps included in this review that have been evaluated and published in peer-reviewed journals.

Inclusion and Exclusion Criteria

Apps were included if they had a medication list function; if they were updated in the last 2 years; were free of charge, reflecting popular trends in app downloads [26]; were available in English; and had a strict privacy policy written on their website or app store. Although a strict privacy policy is not an ultimate standard, given the sensitive nature of health information, the presence of a transparent privacy policy on medication apps was deemed highly important [27].

Apps were excluded from evaluation if they were a game app, were not available in Ireland, focused solely on a particular medical condition (eg, asthma), were a mobile clinical decision support system, were designed primarily for self-care management of a condition (eg, chronic obstructive pulmonary disease or diabetes), were not available for patients to use at home, and had a barcode scanner that did not recognize medication used in Ireland.

Data Extraction

The following information for each app with a medication list function was downloaded: developer, number of downloads of the app, last update, and description of the app in the app store. The apps were downloaded from the app store, and scientific support was evaluated by investigating their content. Scientific support provides information on the app's evidence of validity, as apps that are not supported by evidence are associated with decrements in quality and safety [28]. For example, issues relating to patient confidentiality, inadequate content present in the app, and malfunctioning clinical decision-making apps could result in negative health outcomes for patients [29]. Apps that fulfilled the inclusion criteria were assessed using the MARS and IMS Institute for Healthcare Informatics functionality scoring criteria [30].

Rating Tools

MARS Description

All apps were subjected to in-depth analysis and evaluation using the MARS. To the best of our knowledge, there are no published studies using the MARS to assess the quality of apps with a medication list functionality. The MARS was developed by a team of researchers at the University of Queensland, Australia, to provide a systematic means of assessing, classifying, and rating the quality of mHealth apps [31].

Within this framework, apps are rated according to 4 objective measures (engagement, functionality, aesthetics, and information quality) and one subjective measure (Table 1). More specifically, engagement involves determining whether the app is fun, interesting, customizable, interactive, and well-targeted to its audience. Functionality assesses whether the app is easy to learn, navigate, and flow logically. The esthetics category evaluates the graphic design, overall visual appeal, color scheme, and stylistic consistency of the app. Information quality involves evaluating whether the app contains high-quality information from a credible source. Subjective quality reflects user satisfaction, app endorsement, and continuity of use (ibid). A complete description of the MARS items and subscales can be found in Multimedia Appendix 1 [31].

Table 1. The Mobile App Rating Scale section scores, overall mean, subjective quality results, and mean total.

App name	Developer	Scores[a]				Overall mean[b]	Subjective quality results[c]
		Engage	Function	Aesthetics	Information		
Dosecast[d]	Montuno Software, LLC	3.3	3.4	3.3	3.1	3.3	2.5
MyTherapy[d]	Smart patient GmbH	3.4	4.0	3.8	3.4	3.7	3.4
Medication List & Medical Records[d]	LSD infotech	2.1	2.9	2.6	2.3	2.5	1.6
Medisafe[d]	MediSafe	3.6	4.5	3.9	4.0	4.0	4.0
MedList Pro[d]	Ramtin Software Solutions	3.2	3.1	2.8	2.7	3.0	1.5
Dosecast[e]	Montuno Software, LLC	3.1	3.2	3.2	2.8	3.1	2.2
My Therapy[e]	Smart patient GmbH	3.4	3.9	3.6	3.2	3.5	2.9
Medisafe[e]	MediSafe Inc	3.6	4.5	3.9	4.0	4.0	3.8
Pill Reminder[e]	Sergio Licea	3.5	4.1	3.9	3.1	3.7	3.6

[a]Engagement: mean 3.24 (SD 0.461); functionality: mean 3.73 (SD 0.602); aesthetics: mean 3.42 (SD 0.506); information: mean 3.17 (SD 0.565).

[b]Overall: mean 3.4 (SD 0.496).

[c]Subjective: mean 2.82 (SD 0.945).

[d]Android platform.

[e]iTunes platform.

The apps were independently reviewed by 4 reviewers using a five-point scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent), as shown in Multimedia Appendix 2. Scores for each category were obtained by calculating the mean of the ratings for each subscale according to the 5 measures described above. The total score for each app was determined using the average of the 4 objective measures. The overall mean app quality total score and the total score for the subjective measure (subjective quality, worth recommending, repeat use of the app, and overall satisfaction) were also calculated.

The reviewers carefully read the MARS instructions, independently reviewed the apps, and provided a rationale for their ratings. Subsequently, they compared the results and reached a consensus on each of the ratings for each of the MARS subscales [31]. Before rating the included apps, each reviewer rated 2 randomly selected apps for training purposes (from those apps that were excluded from the review). The results were discussed to ensure that all reviewers had an understanding of the MARS items and the rating process.

IMS Institute for Healthcare Informatics Functionality Score Description

To complement the MARS quality assessment, another tool was used to independently evaluate app functionalities [30]. This evaluation focused on the scope of functions and the potential role that each functionality plays in supporting self-management for patients with HF.

Unlike MARS, this tool only assesses objective quality and has been used previously to evaluate app capabilities [32,33]. The functionality score consists of 7 functionality criteria and 4 functional subcategories. The complete structure of the IMS Institute for Healthcare Informatics functionality scoring criteria can be found in Multimedia Appendix 3 [30]. If a function was present, it was coded as 1; otherwise, it was coded as 0. Functionality scores ranging from 0 to 11 were generated for each app.

Results

Overview

Google Play and iTunes App stores searches identified 483 potential apps (292 Google Play stores and 191 iTunes App stores), the app selection process for both app stores can be seen in Multimedia Appendices 4 and 5. A total of 6 apps (3 across both stores) met the inclusion criteria. Out of the 6 apps reviewed, 4 achieved an acceptable quality score (MARS score 3/5), one achieved a good quality score (4/5), and one had a poor quality score (2.5/5). The median overall MARS score was 3.5/5, ranging from 2.5/5 to 4/5 (mean 3.4, SD 0.49). As stated earlier, the apps are rated according to 4 objective measures: engagement, functionality, aesthetics, and information. The functionality dimension mean score achieved the highest score (3.7), whereas the mean score for the information and engagement dimensions was the lowest (3.2). The total mean subjective MARS score (2.8/5) was lower than the total mean objective MARS score (3.4/5).

On average, the apps had 8 functions based on the IMS criteria (range 5-11). Two apps achieved the highest IMS functionality criteria score (11 functions), whereas 2 apps achieved the lowest score (5 functions). All functions (n=11) are listed in Table 2.

All apps had collecting, sharing, and recording functionality. However, half of the apps did not provide clear instructions on how to enter medication data, did not display vital parameter data in an easy-to-understand format, and did not guide or provide users with advice on how or when to take their medication. Only 2 apps allowed users to communicate with HCPs, family, and friends in real time.

Table 2. IMS Institute for Healthcare Informatics functionality score results.

IMS functionality scoring criteria	Dosecast (Google Play and iTunes)	My Therapy (Google Play and iTunes)	Medication List & Medical Records (Google Play)	Medisafe (Google Play and iTunes)	Pill Reminder (iTunes)	MedList Pro (Google Play)
Inform		✓[a]		✓		
Instruct		✓		✓	✓	
Record	✓	✓	✓	✓	✓	✓
Collect data	✓	✓	✓	✓	✓	✓
Share data	✓	✓	✓	✓	✓	✓
Evaluate data		✓	✓	✓	✓	
Intervene	✓	✓		✓	✓	✓
Display		✓	✓	✓		
Guide		✓		✓		✓
Remind or alert	✓	✓		✓	✓	✓
Communicate		✓		✓		
Total functions present	5	11	5	11	7	6

[a]✓: Function present in the app.

Quality Assessment Using the MARS

Four out of the six apps achieved acceptable quality, suggesting that most apps are of acceptable quality. None of the apps presented any major technical issues during the review, and all were updated in the last 2 years. Features included in most apps reviewed were medication reminders, medication history logs with the ability to share a medication report with others, vital parameter tracking, and syncing the account to other devices.

Only one app (Medisafe) achieved the highest objective and subjective overall MARS scores. One of the distinctive features of this app was the ability to educate users on how and when to take their medication, drug-drug interaction information, and medication side effects. This information was presented in videos using a clear and concise language and text format. In addition, there is evidence of effectiveness, as the Medisafe app has been previously tested in 2 randomized controlled trials [34,35], as a medication adherence tool using scheduled reminders [36] and as a medication reminder related to patients' intention to use the app [37].

The overall subjective quality dimension was significantly lower (2.8/5) than the overall objective mean of all apps (3.4/5). Only one app (Medication List & Medical Records) achieved a lower objective overall score (2.5/5) than the subjective overall score. The subjective quality represented the opinion of the user on the level of satisfaction with the app, willingness to pay for it, and the extent to which the user will recommend it to others. It appears that the apps reviewed, to a certain extent, are well designed, as most apps achieved an acceptable objective quality. However, the subjective quality was poor, as the reviewers provided lower scores on continuity of use and willingness to recommend the apps to others.

The functionality dimension mean score achieved the highest score (3.7), indicating a trend toward higher responses in a timely manner, intuitiveness and ease of use, and navigation across all apps. However, the mean score for the information and engagement dimensions was the lowest (3.2). The information dimension represents the quality and quantity of information present in the app and the way this information is provided, for example, through the use of different formats, such as videos, text, or graphs. The quality of the information on the apps reviewed varied, as some provided good quality medication information; in others, there was a very poor level of information or none present at all.

None of the apps included in the review performed very well in the engagement dimension (if the app was fun or interesting to the user). Medication apps as a rule are not fun and entertaining; however, most apps allow interactivity (ability to input information and prompting) and customization (sending notifications and setting up medication reminders). Another category in engagement, the target group, evaluated if the app content (ie, visual information, language, and design) was appropriate for the user. Most patients diagnosed with HF are older adults. The apps reviewed were not specifically designed for use by older adults and did not have an age-friendly interface. An example of an age-friendly interface is when an app facilitates older adults to enlarge the font size of the screen if required. Avoiding information overload and providing detailed instructions on how to use the app are also age-friendly interface examples [38]. Three of the apps reviewed (Dosecast, Medication List & Medical Records, and MedList Pro) did not

provide clear instructions on how to use the app or how to input medication. App developers should consult with and take older adults' views, needs, and preferences into consideration to make apps more age-friendly and increase usability [38].

IMS Institute for Healthcare Informatics Functionality Score Evaluation and Implications for Patients With HF

Inform

Only 2 apps (Medisafe and MyTherapy) communicated effectively to users, offering an educational component about medication and the medical condition associated with each medication. In addition, both apps informed users about the medication they are taking and possible drug-to-drug interactions. Due to frequent medication or dosing changes in patients with HF, this is a critical function. Medication adherence and continuous education on the condition and symptom management are vital to reduce rehospitalization, illness progression, and exacerbation of symptoms in patients with HF [39].

Instruct

Some of the apps reviewed (Dosecast, Medication List & Medical Records, and MedList Pro) were not intuitive and did not provide clear instructions on how to enter medication strength, time of the day, route, and setting up medication reminders. In other studies, the level of detailed instructions varied. Providing clear instructions on the use of an app is vital for older adults with HF. Patients with HF are predominantly patients aged ≥ 65 years [40], and old age has been cited as a barrier to app use and uptake [41]. Therefore, app designers should consider the needs of older adults using apps and include basic usability advice and easy-to-understand content [38,42].

Record

All apps allowed users to record their medication and a history of use. Most apps also had the capability to record vital parameters. For patients with HF, blood pressure and weight measurements are useful for monitoring the progress of their illness and identifying when symptoms exacerbate. One of the main goals of HF care is to avoid rehospitalization and major adverse cardiac events [32]. In particular, one app offered the possibility of tracking mental well-being, another key area that needs particular attention. For individuals diagnosed with HF, depression and anxiety are common, leading to rehospitalization, poorer quality of life, and increased morbidity and mortality [43].

Evaluate Data

Four apps (MyTherapy, Medication List & Medical Records, Medisafe, and Pill Reminder) allowed data entered into the app to be analyzed and evaluated by the user, a relative, or an HCP. This functionality allows for information to be shared easily and in a timely manner for HCPs to evaluate it and act accordingly. For example, relatives and clinicians can evaluate whether a person is adhering to medication.

Intervene

Five out of the six apps (except Medication List & Medical Records) had the capability to recommend the user, a relative, or a medical practitioner to intervene based on the data collected. Building from the example provided in "evaluate data," once the health information is evaluated, an intervention can be put in place. Examples of interventions for patients with HF could be to reduce or increase the medication dose, reduce fluid intake, and attend the clinic or emergency department on the day. Other examples are the ability of the app to communicate effectively to provide a positive intervention, that is, reminders to refill their medication or a suggestion to engage in physical activity to achieve their daily activity goal.

Display

Apps with a cluttered or bland display do not engage users and are less likely to offer a positive user experience. In 3 of the apps reviewed (MyTherapy, Medication List & Medical Records, and Medisafe), the data were displayed in a clear and colorful graphical representation format. This function could potentially be effective for patients with HF as they can easily understand health reports, that is, medication, weight, and blood pressure [22]. Consequently, health reports will highlight behavioral changes, for example, to adhere to the medication prescribed, reduce fluid intake, or ring the clinic regarding weight gain.

Guide

Half of the apps (Dosecast, Medication List & Medical Records, and Pill Reminder) did not provide comprehensive guidance or training about the correct administration of medication or advice on the time of day that the medication should be administered, for example, before or after a meal. HF medication management is complex, and continuous education and advice on regular medication use is vital for HF self-management [32].

Reminder or Alert

All apps issued an alert to remind users to take their medication and most allowed the users to tick off their medication once they take it or record it as a missed dose. Most apps also had the capability of reminding users of upcoming medical appointments. Apps with a reminder function improved medication adherence and enhanced complex medication management in patients with HF [32]. One of the apps, Medisafe, alerted users and their relatives of drug-drug interactions and of missed medication doses.

Communicate

This function offers the option to communicate with HCPs or with an online support group in real time. The Medisafe app offers users the possibility to communicate with unlimited Medfriends supporters (relatives, friends, or caregivers). Another app, MyTherapy, provides HCPs with an overview of patients' data to plan for their treatment between visits through the web dashboard function.

Google Scholar Search

A search was conducted to identify the apps included in this review that have been evaluated and published in peer-reviewed

journals. Out of all the apps reviewed, 3 apps were identified in the search: (1) Dosecast, (2) MyTherapy, and (3) Medisafe app.

Discussion

Principal Findings

To our knowledge, this is the first study to assess the quality of apps with a medication list functionality using the MARS and the IMS Institute for Healthcare Informatics functionality scale available to Irish consumers. The most common functionalities found in the apps reviewed were medication reminders, medication history logs, and the ability to share medication reports with others, vital parameter tracking, and syncing the account to other devices. However, half did not provide clear instructions on how to enter medication data, did not display vital parameter data in an easy-to-understand format, and did not guide users on how or when to take their medication.

App users prefer apps that are effective, useful, and easy to use [44]. From the apps reviewed, the Medisafe app achieved the highest objective and subjective overall MARS score and the highest IMS Institute for Healthcare Informatics functionality score. One of the distinctive features of this app is the ability to educate users on how and when to take their medication, drug-drug interaction information, and medication side effects. This information was presented in videos using a clear and concise language and text format. The app was found to be very intuitive and had an age-friendly interface.

The quality and efficacy of health apps is another important factor to be considered by users, as they provide positive user experiences and a higher uptake [45]. Most of the apps included in this review had acceptable quality. However, for users, it is not easy to determine the quality, performance, and trustworthiness of apps. The number of apps available in app stores has been growing exponentially in the last decade [30,46] impacting the user's ability to distinguish app quality and performance. Therefore, a reliable and easy-to-use tool to facilitate this process is warranted [47].

The use of mHealth apps is growing at an exponential rate, but there are questions about their efficacy. One of the methods to check the evidence of efficacy is to conduct a search for peer-reviewed academic evidence. For the purpose of this study, a search was conducted for each app in Google Scholar to identify any published peer-reviewed articles. Three of the reviewed apps were identified in the search. The Medisafe app has been previously tested in 2 randomized controlled trials: (1) a medication adherence study [34] and (2) a medication adherence and blood pressure control study [35]. The MyTherapy app was tested in a study [48] as a medication tracker and reminder, whereas the Dosecast app was included in the IMS Institute for Healthcare Informatics review [30], a MARS review [49], and on a feasibility and acceptability medication adherence experimental trial [50]. However, many widely used apps have not yet been scientifically tested. Therefore, there is a need for more apps to be tested scientifically and the outcomes to be disseminated to inform the mHealth research community [22,51].

The mHealth research community has been actively searching for self-management solutions to support older adults shielding at home during the COVID-19 pandemic. One of their focus areas is to support self-management and medication adherence [52]. Apps with a medication reminder functionality may play a potentially important role in promoting greater self-management of vulnerable older adults living at home. In Ireland, as per March 2021, older adults are shielded, and movement is restricted to a 5 km radius. The monotonous routine makes each day very similar to the previous one, and daily routines, such as taking medication at scheduled times might become easy to forget. All apps reviewed, with the exception of one, have the capability of reminding users to take their medication. Medication reminder apps have been found to be effective and to increase medication adherence [48,49,53-55], even for patients with HF [56]. Another app feature that might be of benefit during the COVID-19 pandemic is the ability of older adults with HF to communicate with their medical team or relatives via the app. As mentioned earlier, in Ireland, the number of in-person HF outpatient consultations has considerably decreased by 2020 [15,16]. As of March 2021, the probability of contracting the virus for health care personnel remains high [57], and patients with HF are considered to be one of the most vulnerable groups [58].

Finally, this paper identified a small number of apps that could be suitable for patients with HF sharing their medication list with HCPs before consultation. Owing to the sensitive nature of health care data and in an era where General Data Protection Regulation (GDPR) legislation considerations are increasingly important, a strict data policy was a criterion for selection. For example, one of the most frequent concerns app users have is how their health data are processed. Under GDPR legislation, app users are encouraged to ask and obtain information in relation to data security and data processing [59]. However, after the introduction of the GDPR in 2018, no specific guidance on privacy has been developed for apps widely available to consumers [60]. Therefore, the need for transparent and easy-to-understand strict privacy policies in health apps should be mandatory if consumers are expected to download and use mHealth apps [60,61] and if clinicians are expected to recommend apps to their patients [62].

Conclusions

The quality of current apps with a medication list functionality varies according to their technical aspects. Most of the reviewed apps have acceptable MARS objective quality. However, the subjective quality or satisfaction with the apps was poor. The objective quality assesses whether an app is interesting, easy to navigate, and overall visual appeal, among other characteristics. Subjective quality reflects user satisfaction, app endorsement, and continuity of use. Only 3 apps are based on scientific evidence and have been tested previously. Two apps featured all the IMS Institute for Healthcare Informatics functionalities and half did not provide clear instructions on how to enter medication data, did not display vital parameter data in an easy-to-understand format, and did not guide users on how or when to take their medication.

To our knowledge, this is the first study to use the MARS to assess the quality of apps with a medication list functionality available in the Irish app stores. The need for an app to support older adults with HF to maintain an accurate medication list is warranted. We recommend that app developers display either in the app or on the website, a transparent and easy-to-understand privacy policy to increase patients' and HCPs' trust and use.

Limitations

One of the limitations of this review is the limited number of apps, as only those with a strict privacy policy written on their website or app store were included. Furthermore, due to the fast pace of app development, it is possible that by the time this paper is published, there may be new apps with a medication list functionality available to Irish consumers.

Acknowledgments

This study was supported by the Eastern Corridor Medical Engineering Centre, a collaborative research project focusing on improving cardiovascular health. The Eastern Corridor Medical Engineering Centre is funded by the European Union's Interreg VA Programme, which is managed by the Special European Union Programmes body.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile App Rating Scale items and subscales.
[DOCX File , 15 KB - mhealth_v9i11e30674_app1.docx]

Multimedia Appendix 2
Mobile App Rating Scale ratings of included apps.
[DOCX File , 16 KB - mhealth_v9i11e30674_app2.docx]

Multimedia Appendix 3
IMS Institute for Healthcare Informatics functionality scoring.
[DOCX File , 13 KB - mhealth_v9i11e30674_app3.docx]

Multimedia Appendix 4
App selection process (Google Play app store).
[DOCX File , 46 KB - mhealth_v9i11e30674_app4.docx]

Multimedia Appendix 5
App selection process (Apple app store).
[DOCX File , 46 KB - mhealth_v9i11e30674_app5.docx]

References

1. Athilingam P, Jenkins B. Mobile phone apps to support heart failure self-care management: integrative review. JMIR Cardio 2018 May 02;2(1):e10057 [FREE Full text] [doi: 10.2196/10057] [Medline: 31758762]
2. Marti CN, Fonarow GC, Anker SD, Yancy C, Vaduganathan M, Greene SJ, et al. Medication dosing for heart failure with reduced ejection fraction - opportunities and challenges. Eur J Heart Fail 2019 Mar;21(3):286-296 [FREE Full text] [doi: 10.1002/ejhf.1351] [Medline: 30537163]
3. Clark AM, Freydberg CN, McAlister FA, Tsuyuki RT, Armstrong PW, Strain LA. Patient and informal caregivers' knowledge of heart failure: necessary but insufficient for effective self-care. Eur J Heart Fail 2009 Jun;11(6):617-621 [FREE Full text] [doi: 10.1093/eurjhf/hfp058] [Medline: 19414477]
4. Clark AM, Davidson P, Currie K, Karimi M, Duncan AS, Thompson DR. Understanding and promoting effective self-care during heart failure. Curr Treat Options Cardiovasc Med 2010 Jan;12(1):1-9. [doi: 10.1007/s11936-009-0053-1] [Medline: 20842477]
5. Butrous H, Hummel SL. Heart failure in older adults. Can J Cardiol 2016 Sep;32(9):1140-1147 [FREE Full text] [doi: 10.1016/j.cjca.2016.05.005] [Medline: 27476982]
6. WHO guideline: recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1 [accessed 2020-04-04]

7. Buning AW, Klopotowska JE, Duyvendak M, Engelen LJ, Arts J. Patient empowerment through provision of a mobile application for medication reconciliation: a proof of concept study. BMJ Innov 2016 Sep 28;2(4):152-157. [doi: 10.1136/bmjinnov-2015-000110]
8. Rose AJ, Fischer SH, Paasche-Orlow MK. Beyond medication reconciliation: the correct medication list. J Am Med Assoc 2017 May 23;317(20):2057-2058. [doi: 10.1001/jama.2017.4628] [Medline: 28426844]
9. Hilmer SN, Gnjidic D, Le Couteur DG. Thinking through the medication list - appropriate prescribing and deprescribing in robust and frail older patients. Aust Fam Physician 2012 Dec;41(12):924-928. [Medline: 23210113]
10. Torous J, Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
11. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for COVID-19. Br Med J 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
12. Santini ZI, Jose PE, Cornwell EY, Koyanagi A, Nielsen L, Hinrichsen C, et al. Social disconnectedness, perceived isolation, and symptoms of depression and anxiety among older Americans (NSHAP): a longitudinal mediation analysis. Lancet Public Health 2020;5(1):62-70. [doi: 10.1016/S2468-2667(19)30230-0]
13. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. PLoS Med 2010 Jul 27;7(7):e1000316 [FREE Full text] [doi: 10.1371/journal.pmed.1000316] [Medline: 20668659]
14. Armitage R, Nellums LB. COVID-19 and the consequences of isolating the elderly. Lancet Public Health 2020 May;5(5):e256 [FREE Full text] [doi: 10.1016/S2468-2667(20)30061-X] [Medline: 32199471]
15. Hospital service disruptions and visiting restrictions (COVID-19). Health Service Executive. URL: https://www2.hse.ie/services/hospital-service-disruptions/hospital-service-disruptions-covid19.html [accessed 2020-04-25]
16. McGlynn M. 'Greatly concerning': Covid-related decline in cardiac outpatient appointments. Irish Examiner. URL: https://www.irishexaminer.com/news/arid-40041818.html [accessed 2020-09-10]
17. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
18. Attend anywhere: healthcare provider video consultations. Health Service Executive. URL: https://healthservice.hse.ie/staff/coronavirus/working-from-home/virtual-health/attend-anywhere-healthcare-provider-video-consultations.html [accessed 2020-04-04]
19. Olwill C, Mc Nally D, Douglas L. Psychiatrist experience of remote consultations by telephone in an outpatient psychiatric department during the COVID-19 pandemic. Ir J Psychol Med 2021 Jun;38(2):132-139 [FREE Full text] [doi: 10.1017/ipm.2020.51] [Medline: 32438945]
20. Usman M, Fahy S. Coping with the COVID-19 crisis: an overview of service adaptation and challenges encountered by a rural Psychiatry of Later Life (POLL) team. Ir J Psychol Med 2020 Jul 02:1-5 [FREE Full text] [doi: 10.1017/ipm.2020.86] [Medline: 32611473]
21. De Bock L, Tommelein E, Baekelandt H, Maes W, Boussery K, Somers A. The introduction of a full medication review process in a local hospital: successes and barriers of a pilot project in the geriatric ward. Pharmacy (Basel) 2018 Feb 28;6(1):21 [FREE Full text] [doi: 10.3390/pharmacy6010021] [Medline: 29495567]
22. Woods LS, Duff J, Roehrer E, Walker K, Cummings E. Patients' experiences of using a consumer mHealth app for self-management of heart failure: mixed-methods study. JMIR Hum Factors 2019 May 02;6(2):e13009 [FREE Full text] [doi: 10.2196/13009] [Medline: 31045504]
23. Athilingam P, Jenkins B, Johansson M, Labrador M. A mobile health intervention to improve self-care in patients with heart failure: pilot randomized control trial. JMIR Cardio 2017 Aug 11;1(2):e3 [FREE Full text] [doi: 10.2196/cardio.7848] [Medline: 31758759]
24. Hägglund E, Lyngå P, Frie F, Ullman B, Persson H, Melin M, et al. Patient-centred home-based management of heart failure. Findings from a randomised clinical trial evaluating a tablet computer for self-care, quality of life and effects on knowledge. Scand Cardiovasc J 2015 Aug;49(4):193-199. [doi: 10.3109/14017431.2015.1035319] [Medline: 25968968]
25. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
26. Azar KM, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
27. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone apps to support self-management of hypertension: review and content analysis. JMIR Mhealth Uhealth 2019 May 28;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]
28. Lui JH, Marcus DK, Barry CT. Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. Prof Psychol Res Pract 2017 Jun;48(3):199-210 [FREE Full text] [doi: 10.1037/pro0000122]
29. Moral-Munoz JA, Esteban-Moreno B, Herrera-Viedma E, Cobo MJ, Pérez IJ. Smartphone applications to perform body balance assessment: a standardized review. J Med Syst 2018 May 29;42(7):119. [doi: 10.1007/s10916-018-0970-1] [Medline: 29845455]

30. Aitken M, Gauntlett C. Patient apps for improved healthcare: from novelty to mainstream. IMS Institute for Healthcare Informatics. Parsippany, NJ: IMS Institute for Healthcare Informatics; 2013. URL: https://silo.tips/download/october-2013-patient-apps-for-improved-healthcare-from-novelty-to-mainstream#modals [accessed 2021-09-08]
31. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
32. Mortara A, Vaira L, Palmieri V, Iacoviello M, Battistoni I, Iacovoni A, et al. Would You Prescribe Mobile Health Apps for Heart Failure Self-care? An Integrated Review of Commercially Available Mobile Technology for Heart Failure Patients. Card Fail Rev 2020 Mar;6:e13 [FREE Full text] [doi: 10.15420/cfr.2019.11] [Medline: 32537246]
33. Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
34. Huang Z, Tan E, Lum E, Sloot P, Boehm BO, Car J. A smartphone app to improve medication adherence in patients with type 2 diabetes in Asia: feasibility randomized controlled trial. JMIR Mhealth Uhealth 2019 Sep 12;7(9):e14914 [FREE Full text] [doi: 10.2196/14914] [Medline: 31516127]
35. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018 Jun 01;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
36. Wade R, Clancey B, Michaeli J. Improvement in antihypertensive and cholesterol-lowering medication persistence using a mobile technology application. Val Health 2016 May;19(3):A306 [FREE Full text] [doi: 10.1016/j.jval.2016.03.658]
37. Park DY, Goering EM, Head KJ, Bartlett Ellis RJ. Implications for training on smartphone medication reminder app use by adults with chronic conditions: pilot study applying the technology acceptance model. JMIR Form Res 2017 Nov 10;1(1):e5 [FREE Full text] [doi: 10.2196/formative.8027] [Medline: 30684397]
38. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: a usability study. JMIR Mhealth Uhealth 2014 Mar 14;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]
39. Gilotra NA, Shpigel A, Okwuosa IS, Tamrat R, Flowers D, Russell SD. Patients commonly believe their heart failure hospitalizations are preventable and identify worsening heart failure, nonadherence, and a knowledge gap as reasons for admission. J Card Fail 2017 Mar;23(3):252-256. [doi: 10.1016/j.cardfail.2016.09.024] [Medline: 27742454]
40. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. Nat Rev Cardiol 2011 Jan;8(1):30-41 [FREE Full text] [doi: 10.1038/nrcardio.2010.165] [Medline: 21060326]
41. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
42. Gao C, Zhou L, Liu Z, Wang H, Bowers B. Mobile application for diabetes self-management in China: Do they fit for older adults? Int J Med Inform 2017 May;101:68-74 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.02.005] [Medline: 28347449]
43. Celano CM, Villegas AC, Albanese AM, Gaggin HK, Huffman JC. Depression and anxiety in heart failure: a review. Harv Rev Psychiatry 2018;26(4):175-184 [FREE Full text] [doi: 10.1097/HRP.0000000000000162] [Medline: 29975336]
44. Gowin M, Cheney M, Gwin S, Wann T. Health and fitness app use in college students: a qualitative study. Am J Health Educ 2015 Jul 06;46(4):223-230. [doi: 10.1080/19325037.2015.1044140]
45. Cheng H, Tutt A, Llewellyn C, Size D, Jones J, Taki S, et al. Content and quality of infant feeding smartphone apps: five-year update on a systematic search and evaluation. JMIR Mhealth Uhealth 2020 May 27;8(5):e17300 [FREE Full text] [doi: 10.2196/17300] [Medline: 32459187]
46. Xu Q, Erman J, Gerber A, Mao Z, Pang J, Venkataraman S. Identifying diverse usage behaviors of smartphone apps. In: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. 2011 Nov 02 Presented at: IMC '11: Internet Measurement Conference; November 2 - 4, 2011; Berlin, Germany p. 329-344. [doi: https://doi.org/10.1145/2068816.2068847]
47. Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? BMC Med 2018 Dec 03;16(1):225 [FREE Full text] [doi: 10.1186/s12916-018-1211-7] [Medline: 30501638]
48. Steinert A, Eicher C, Haesner M, Steinhagen-Thiessen E. Effects of a long-term smartphone-based self-monitoring intervention in patients with lipid metabolism disorders. Assist Technol 2020;32(2):109-116. [doi: 10.1080/10400435.2018.1493710] [Medline: 29944463]
49. Santo K, Singleton A, Chow CK, Redfern J. Evaluating reach, acceptability, utility, and engagement with an app-based intervention to improve medication adherence in patients with coronary heart disease in the medapp-chd study: a mixed-methods evaluation. Med Sci (Basel) 2019 Jun 04;7(6):68 [FREE Full text] [doi: 10.3390/medsci7060068] [Medline: 31167489]
50. Wu YP, Linder LA, Kanokvimankul P, Fowler B, Parsons BG, Macpherson CF, et al. Use of a smartphone application for prompting oral medication adherence among adolescents and young adults with cancer. Oncol Nurs Forum 2018 Jan 01;45(1):69-76 [FREE Full text] [doi: 10.1188/18.ONF.69-76] [Medline: 29251285]

51. Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017 Dec;41(6):625-630. [doi: 10.1111/1753-6405.12688] [Medline: 28749591]
52. Banskota S, Healy M, Goldberg EM. 15 smartphone apps for older adults to use while in isolation during the COVID-19 pandemic. West J Emerg Med 2020 Apr 14;21(3):514-525 [FREE Full text] [doi: 10.5811/westjem.2020.4.47372] [Medline: 32302279]
53. Fallah M, Yasini M. A medication reminder mobile app: does it work for different age ranges. Stud Health Technol Inform 2017;235:68-72. [Medline: 28423757]
54. Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: A randomized study. Am Heart J 2016 Aug;178:85-94 [FREE Full text] [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]
55. Mertens A, Brandl C, Miron-Shatz T, Schlick C, Neumann T, Kribben A, et al. A mobile application improves therapy-adherence rates in elderly patients undergoing rehabilitation: A crossover design study comparing documentation via iPad with paper-based control. Medicine (Baltimore) 2016 Sep;95(36):e4446 [FREE Full text] [doi: 10.1097/MD.0000000000004446] [Medline: 27603339]
56. Goldstein CM, Gathright EC, Dolansky MA, Gunstad J, Sterns A, Redle JD, et al. Randomized controlled feasibility trial of two telemedicine medication reminder systems for older adults with heart failure. J Telemed Telecare 2014 Sep;20(6):293-299 [FREE Full text] [doi: 10.1177/1357633X14541039] [Medline: 24958355]
57. Mira JJ, Carrillo I, Guilabert M, Mula A, Martin-Delgado J, Pérez-Jover MV, SARS-CoV-2 Second Victim Study Group. Acute stress of the healthcare workforce during the COVID-19 pandemic evolution: a cross-sectional study in Spain. BMJ Open 2020 Nov 06;10(11):e042555 [FREE Full text] [doi: 10.1136/bmjopen-2020-042555] [Medline: 33158839]
58. Virani SA, Clarke B, Ducharme A, Ezekowitz JA, Heckman GA, McDonald M, Canadian Heart Failure Society. Optimizing access to heart failure care in Canada during the COVID-19 pandemic. Can J Cardiol 2020 Jul;36(7):1148-1151 [FREE Full text] [doi: 10.1016/j.cjca.2020.05.009] [Medline: 32405146]
59. Muchagata J, Ferreira A. Translating GDPR into the mHealth practice. In: Proceedings of the International Carnahan Conference on Security Technology (ICCST). 2018 Oct Presented at: International Carnahan Conference on Security Technology (ICCST); Oct. 22-25, 2018; Montreal, QC, Canada p. 1-5. [doi: 10.1109/ccst.2018.8585546]
60. Ferretti A, Ronchi E, Vayena E. From principles to practice: benchmarking government guidance on health apps. Lancet Digital Health 2019 Jun;1(2):55-57 [FREE Full text] [doi: 10.1016/S2589-7500(19)30027-5] [Medline: 33323230]
61. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Privacy assessment in mobile health apps: scoping review. JMIR Mhealth Uhealth 2020 Jul 02;8(7):e18868 [FREE Full text] [doi: 10.2196/18868] [Medline: 32459640]
62. Grindrod K, Boersema J, Waked K, Smith V, Yang J, Gebotys C. Locking it down: The privacy and security of mobile medication apps. Can Pharm J (Ott) 2017;150(1):60-66 [FREE Full text] [doi: 10.1177/1715163516680226] [Medline: 28286594]

Abbreviations

GDPR: General Data Protection Regulation
HCP: health care professional
HF: heart failure
MARS: Mobile App Rating Scale
mHealth: mobile health

Edited by L Buis; submitted 24.05.21; peer-reviewed by P Athilingam, H Cheng; comments to author 09.06.21; revised version received 08.07.21; accepted 14.07.21; published 02.11.21.

Please cite as:
Diaz-Skeete YM, McQuaid D, Akinosun AS, Ekerete I, Carragher N, Carragher L
Analysis of Apps With a Medication List Functionality for Older Adults With Heart Failure Using the Mobile App Rating Scale and the IMS Institute for Healthcare Informatics Functionality Score: Evaluation Study
JMIR Mhealth Uhealth 2021;9(11):e30674
URL: https://mhealth.jmir.org/2021/11/e30674
doi:10.2196/30674
PMID:34726613

Review

Mobile Apps Leveraged in the COVID-19 Pandemic in East and South-East Asia: Review and Content Analysis

Bohee Lee[1], PharmD, MPH; Siti Aishah Ibrahim[2], MBChB, MSc; Tiying Zhang[1], BM, MPH

[1]Centre for Population Health Sciences, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

[2]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

Corresponding Author:
Bohee Lee, PharmD, MPH
Centre for Population Health Sciences, Usher Institute
University of Edinburgh
Old Medical School, Teviot Place
Edinburgh, EH8 9AG
United Kingdom
Phone: 44 1316503034
Email: bohee.lee@ed.ac.uk

Abstract

Background: The COVID-19 pandemic increased attention to digital tools to support governmental public health policies in East and South-East Asia. Mobile apps related to the COVID-19 pandemic continue to emerge and evolve with a wide variety of characteristics and functions. However, there is a paucity of studies evaluating such apps in this region, with most of the available studies conducted in the early days of the pandemic.

Objective: This study aimed to examine free apps developed or supported by governments in the East and South-East Asian region and highlight their key characteristics and functions. We also sought to interpret how the release dates of these apps were related to the commencement dates of other COVID-19 public health policies.

Methods: We systematically searched for apps in Apple App Store and Google Play Store and analyzed the contents of eligible apps. Mobile apps released or updated with COVID-19–related functions between March 1 and May 7, 2021, in Singapore, Taiwan, South Korea, China (mainland), Japan, Thailand, Hong Kong, Vietnam, Malaysia, Indonesia, and the Philippines were included. The CoronaNet Research Project database was also examined to determine the timeline of public health policy commencement dates in relation to the release dates of the included apps. We assessed each app's official website, media reports, and literature through content analysis. Descriptive statistics were used to summarize relevant information gathered from the mobile apps using RStudio.

Results: Of the 1943 mobile apps initially identified, 46 were eligible, with almost 70% of the apps being intended for the general public. Most apps were from Vietnam (n=9, 20%), followed by Malaysia, Singapore, and Thailand (n=6 each, 13%). Of note, most apps for quarantine monitoring (n=6, 13%) were mandatory for the target users or a population subset. The most common function was health monitoring (32/46, 70%), followed by raising public health awareness (19/46, 41%) through education and information dissemination. Other functions included monitoring quarantine (12/46, 26%), providing health resources (12/46, 26%). COVID-19 vaccination management functions began to appear in parallel with vaccine rollout (7/46, 15%). Regarding the timing of the introduction of mobile solutions, the majority of mobile apps emerged close to the commencement dates of other public health policies in the early stages of the pandemic between March and April 2020.

Conclusions: In East and South-East Asia, most governments used mobile health apps as adjuncts to public health measures for tracking COVID-19 cases and delivering credible information. In addition, these apps have evolved by expanding their functions for COVID-19 vaccination.

(JMIR Mhealth Uhealth 2021;9(11):e32093) doi:10.2196/32093

KEYWORDS

mobile apps; applications; eHealth; mHealth; mobile health; digital health; telemedicine; telehealth; COVID-19; coronavirus; pandemic; public health; health policy

Introduction

The role of digital technology has reached new heights, with 93% of the world's population having access to mobile broadband networks in 2020 [1]. Today, with more than half of the world's population (approximately 3.8 billion individuals) owning a smartphone, there is enormous potential and increasing opportunity to cost-effectively incorporate mobile apps into pandemic control strategies [2]. Mobile technologies in public health (mHealth), allow individuals to connect with health services, including surveillance, remote monitoring, and health information [3].

mHealth interventions have been continuously evolving in various settings, including resource-limited settings with the surging penetration of smartphones and continuous advancement of relevant technological capabilities [4]. Evidence has shown that mHealth has been used to enable health care providers to reach out to vulnerable individuals, conduct surveillance, and provide treatment, health-related education, and counseling [4-7].

The capabilities of mHealth interventions have grown quickly during the COVID-19 pandemic, but their abundant potential has been constantly predicted by many researchers, even before the pandemic [8]. For instance, a pilot study by Pant Pai et al [9] observed that an unsupervised HIV self-testing strategy using an internet-based mobile app leads to counseling and treatment among patients testing positive in South Africa. A case study in Uganda also highlighted the feasibility of mHealth approaches to implement antimalaria strategies in a transitional country [10].

Since the World Health Organization (WHO) declared COVID-19 a global pandemic in March 2020, the demand for digital tools to reinforce public health measures has drastically increased worldwide [11]. mHealth solutions have been used for early detection, fast screening, patient monitoring, information sharing, education, and treatment management in response to the COVID-19 outbreak [8]. The pandemic has witnessed a rapid proliferation in the application of digital technologies for public health, with many governments around the globe developing mobile apps to reduce the transmission of SARS-CoV-2 [12,13].

Before the advent of COVID-19 vaccines, many governments in East and South-East Asia have gained unprecedented attention for their effective COVID-19 containment and incredibly low death tolls compared to countries in the West [14]. Governments in this region had experienced the consequences of outbreaks such as severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). Therefore, they ensured that their public health systems were better prepared for similar outbreaks by establishing early warning systems and relevant policies [15-18]. Critical medical capacities were augmented while early warning systems and relevant policies were established long before COVID-19 was identified [16,18]. In addition, they actively capitalized on technological solutions to contain the pandemic by leveraging existing regional digital infrastructure through the ASEAN Smart Cities Network (ASCN), a collaborative platform working toward a common goal of smart and sustainable urban development [12,19,20]. These experiences also created a culture of mask-wearing, solidarity, and collective responsibility in the general public [21].

Although a number of systematic reviews had looked at COVID-19–related apps available on a global scale, there is a paucity of studies focusing on mobile apps in this region, which share similar cultural characteristics [13,22,23]. Ming et al [24] found that most apps developed in the United States before May 2020 could trace or map COVID-19 cases and had surveillance features but not educational contents. A recent review by Alanzi [13] examined the functionalities of mobile apps developed by governments in 6 countries including Saudi Arabia, Italy, Singapore, the United Kingdom, the United States, and India as of August 2020. Alanzi [13] found that the most prevalent function was contract tracing, while very few apps had functions for raising public awareness and providing COVID-19–related information. Almalki and Giannicchi [25] assessed mobile apps in a total of 51 countries as of September 2020. They demonstrated that the most common function was basic health information followed by contact tracing, self-assessment, live statistics and the latest news [25]. However, only 5 East and South-East Asian countries (Vietnam, Malaysia, Thailand, Singapore, and South Korea) were included in this assessment.

Given the diverse economic sizes and varying digital adaptation in the East and South-East Asian region [26], it is crucial to know how these governments have developed readiness and abilities to deploy digital technologies integrated with public health measures [14]. In addition, considering the evolving nature of the pandemic, there is a need to examine how COVID-19–related mobile apps are used in the public health context, particularly focusing on this region. Therefore, our review aimed to explore COVID-19–related mobile apps that governments in East and South-East Asia have introduced.

Methods

Search Strategy

This study adopted a systematic search strategy using a modified version of the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines to identify COVID-19–related apps currently freely available in this region and their characteristics and functions [27]. Adjustments were needed because of the different search nature of mobile app stores.

We referred to Bloomberg's Covid Resilience Ranking evaluating the 53 largest economies on their success at containing the virus (March 2021) [28]. This ranking covers a wide range of COVID-19 statuses, from mortality rates and COVID-19 testing to vaccination and lockdown severity, and quality of life during the pandemic [28]. This ranking involved 11 governments in East and South-East Asia as of March 2021: Singapore, Taiwan, Hong Kong, South Korea, China (mainland), Japan, Thailand, Vietnam, Malaysia, the Philippines, and Indonesia. The summarized details of the scores of each selected government based on Bloomberg's Covid Resilience Ranking in March 2021 are presented in Multimedia Appendix 1.

The 2 largest app stores worldwide, iOS-based Apple App Store and Android-based Google Play Store, were searched for potentially relevant mobile apps released or updated from March 1 to May 7, 2021. The following search terms were used: “COVID-19,” “COVID,” “coronavirus,” “corona virus,” “corona,” and “SARS-CoV-2.” To circumvent the regional restriction setting for searching apps, we utilized a website, fnd.ios, to look for apps on Apple App Store and changed the region settings in Google Play Store [29,30]. News articles and media reports were also searched to find further eligible apps that may have been missed. For searching the literature, MEDLINE and Google Scholars were explored by combining 2 search strings, including terms related to mobile apps and COVID-19 such as (“digital health” OR “m-health” OR “mobile health” OR “e-health” OR “mobile apps”) AND (“COVID-19” OR “coronavirus” OR “SARS-CoV-2”). Draft searches were piloted in each database and then finalized. Searches were conducted on May 7, 2021, by 2 reviewers (BL and TZ). To identify and examine the mobile app described in the native language (non-English) of the corresponding government, we searched the app’s official website and news reports to determine whether there was any information provided in English. Google Translator was used if the information about the app was unavailable in either English or the 4 languages spoken by the 3 reviewers (Chinese, Korean, Malaysian, and Japanese).

To evaluate when mobile apps were introduced in relation to other public health policies, we utilized the data set of the CoronaNet Research Project collating governmental public health policies worldwide in the context of COVID-19 [31]. This project comprises a data set providing comprehensive government policies across 195 countries, apprehending 18 broad policy types, including timings of each policy. Any ambiguity was resolved through discussion with a reviewer in the CoronaNet Research Project (CC). We selected national-level policies of 11 governments and validated relevant policies by checking data sources. We narrowed 18 policy types to 6, which were deemed to be associated with the functions of mobile apps such as public awareness measure, COVID-19 testing, quarantine monitoring, health monitoring, vaccination, and health resources [31].

Eligibility Assessment and Selection of Apps

After initial deduplication, 2 authors (SAI, BL) with backgrounds in public health screened mobile apps on the basis of the identified apps’ titles, keywords, and descriptions. Irrelevant apps were excluded during the preliminary screening step. After screening, the 2 reviewers independently assessed the eligibility of mobile apps on the basis of the eligibility criteria. We included apps if they were (1) related to COVID-19, (2) available free of cost with no in-app purchase requirement, (3) released or updated with COVID-19–related functions during the research period, (4) still available to users on the specified search date, (5) developed or supported by governments or authorities, and (6) with full information regarding the app accessible. However, we excluded mobile apps developed by global organizations, nongovernmental organizations, or communities not representing a government or broader regions. Discrepancies were resolved through discussion between 2 reviewers or arbitration by a third reviewer to reach a consensus.

Data Extraction and Synthesis

We used a modified framework of previous studies and the CoronaNet Project database for data extraction [13,31]. This framework covers key characteristics and functions of mobile apps in accordance with coding and policy definitions by the CoronaNet Research Project [31]. Key characteristics include the country of origin, platform availability (Apple App Store and Google Play Store), release date, developer, target users, uptake requirement, and required technology. Key functions were merged into 6 policy types. Definitions of key functions and lists of subordinate functions are described in Textbox 1.

Based on this framework, we developed a data extraction form, and 2 independent reviewers extracted the relevant data. Each app’s official website, relevant media reports, and literature were assessed through content analysis [32]. Through this technique, we identified and quantified relevant keywords indicating key characteristics and functions [32]. At each step, disagreements were resolved by consensus. In case of persistent disagreement, arbitration by the third reviewer settled the discrepancy. Descriptive statistics were used to summarize relevant information gathered from the mobile apps, using RStudio (version 1.3.1056).

Textbox 1. Definitions of key functions and list of subordinate functions of eligible mobile apps.

Public Awareness Measures: Government efforts to disseminate or gather reliable information about COVID-19

- News or government measures
- Up-to-date statistics
- COVID-19 health information
- Health management guidelines
- COVID-19 related services information
- Hotspot/risk area identification

COVID-19 Testing: Government policies to detect COVID-19 cases

- Obtain COVID-19 test
- Report of test results

Quarantine Monitoring: Targets of the policy are obliged to isolate themselves for at least 14 days because there is reason to suspect a person is infected with COVID-19

- Regular health check
- Location tracking

Health Monitoring: Government policies to monitor the health of individuals to limit the spread of COVID-19

- Digital contact tracing
- Digital check-in
- Alert contacts of COVID-19 cases
- Report suspected cases/rule infringement
- Health code/status generator
- Health/travel declaration
- Self-symptom assessment

Vaccination: Government policy made with regards to either the research and development, regulation, production, purchase and distribution of a given COVID-19 vaccine

- Vaccination information
- Vaccination registration/appointment
- Vaccination certificate
- Reporting adverse reactions

Health Resources: Government policies that affect the material (eg, medical equipment, number of hospitals for public health) or human (eg, doctors, nurses) health resources of a country

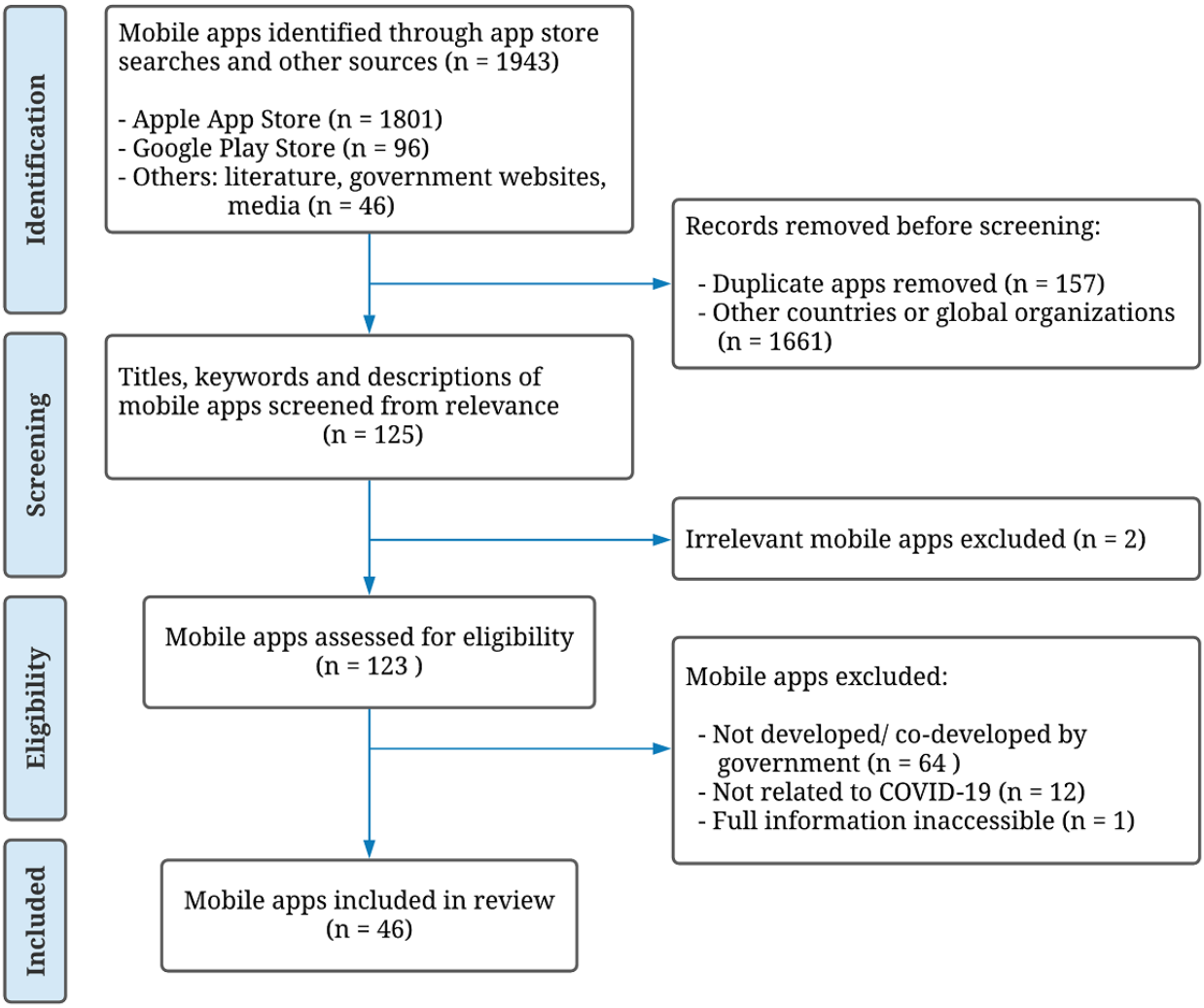
- Virtual medical consultation
- Emergency helpline
- Accessing medical records
- Personal protective equipment distribution

Results

Selected Apps

Figure 1 illustrates an overview of the process involved in selecting the apps for study synthesis. A total of 1943 potential apps were obtained through systematic searches, of which 46 met our eligibility criteria. Although 3 of the apps, namely Alipay, WeChat, and My Health Bank, have pre-existed before March 2020, we included them in the review as they have since been updated to include COVID-19–related services during the pandemic.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the search process.



Characteristics of the Included Apps

All the included apps were free for users to download and use without any in-app purchase requirements. Furthermore, they were official apps developed or supported by the government and maintained by the relevant authority for COVID-19–related service provision. Descriptive analytics related to the characteristics of the apps were summarized and presented in Table 1. Most of the apps (n=9, 20%) were from Vietnam, followed by Malaysia, Singapore, and Thailand (n=6 apps each, 13%). Almost 98% of the apps were available on both iOS and Android platforms through the Apple App Store and Google Play Store.

In total, 24 (52%) apps were mandatory, with a mandate for target users to install them on their smartphones. Target users were mainly a subset of the population only, for example, people living in high-risk areas with stringent pandemic restrictions and confirmed or suspected COVID-19 cases.

Most of these apps (n=32, 70%) were intended for the general public. Six (13%) apps were especially intended for quarantined people: 4 (9%) apps for quarantined residents and 2 (4%) apps for quarantined inbound travelers. Six (13%) apps targeted travelers: domestic and international travelers (n=2, 4%), international travelers including those who required quarantine (n=3, 7%), and outbound travelers (n=1, 2%). Overall, the GPS was the most required technology (n=28, 61%), followed by Bluetooth (n=16, 35%) and the QR scanner (n=16, 35%). Artificial intelligence (AI), the application programming interface (API), and facial-recognition technology were also utilized in 3 (7%) apps. Details of apps with associated characteristics currently available across 11 governments included in this review are described in Multimedia Appendix 2.

Table 1. Overview of the included apps (N=46).

Characteristics	Apps, n (%)
Origin	
China (mainland)	2 (4)
Hong Kong	3 (7)
Indonesia	3 (7)
Japan	4 (9)
Malaysia	6 (13)
Philippines	1 (2.2)
Singapore	6 (13)
South Korea	3 (7)
Taiwan	3 (7)
Thailand	6 (13)
Vietnam	9 (20)
Platform	
iOS (App Store)	45 (98)
Android (Google Play Store)	46 (100)
Uptake requirement	
Mandatory	24 (52)
Voluntary	22 (48)
Target users	
General public	32 (70)
Travelers: domestic and international	2 (4)
Travelers: international	1 (2)
Travelers: requiring quarantine	2 (4)
Travelers: outbound	1 (2)
Foreign workers	1 (2)
Quarantined individuals	4 (9)
Business owners	1 (2)
Vaccinated individuals	2 (4)
Required technology	
GPS	28 (61)
Bluetooth	16 (35)
QR scanner	16 (35)
Others[a]	3 (7)

[a]Other technologies include artificial intelligence (n=1), the application programming interface (API) (n=1), and facial recognition (n=1).

Functions of the Included Apps

Overall, 25 common functions were identified, and they were subsequently organized into 6 overarching domains that characterized the functions of these apps, as shown in Table 2. The functions supported by each app are detailed in Multimedia Appendix 3.

The most common function served by the apps was health monitoring (n=32, 70%). Eleven apps (24%) were used for digital contact tracing by tracking, documenting, and retaining mobile phone users' encounters with other devices using Bluetooth or GPS technologies. Twelve (26%) apps had the function of alerting the contacts of COVID-19 cases. If one of the app users contracted COVID-19, authorities with access to the data could request the infected user to upload the relevant anonymized data for analysis so that others with the same installed app who were in close contact may be alerted for further action. Eleven apps (24%) served the digital check-in function with the same goal for contact tracing: maintaining an

efficient digital log of visitors so that officials could quickly reach out to those who might have been in close contact with a COVID-19 case present in the same events or premises.

The second-most common function associated with the apps was public health awareness (n=19, 41%). More than half of these apps were developed to disseminate the latest news (n=12, 26%) and up-to-date statistics (n=10, 22%). Furthermore, this main function included subordinated functions such as providing health management guidelines (n=9, 20%) and health information and advice about COVID-19 (n=9, 20%) and sharing the location and helpline number of facilities offering services during this pandemic (n=9, 20%). In addition, some apps (n=5, 11%) provided maps of hotspots or high-risk areas with increased COVID-19 transmission to better inform the public of their travel plans.

Table 2. Main functions and subordinate functions of the included apps (N=46).

Main functions and subordinate functions	Apps, n (%)	Apps, %[a]
Public awareness measures	19 (41)	—[b]
News or government measures	12 (26)	7
Up-to-date statistics	10 (22)	6
COVID-19 health information	9 (20)	5
Health management guidelines	9 (20)	5
COVID-19 related services information	9 (20)	5
Hotspot/risk area identification	5 (11)	3
COVID-19 testing	9 (20)	—
Obtain COVID-19 test	4 (9)	2
Report of test results	7 (15)	4
Quarantine monitoring	12 (26)	—
Regular health check	5 (11)	3
Location tracking	10 (22)	6
Health monitoring	32 (70)	—
Digital contact tracing	11 (24)	7
Digital check-in	11 (24)	7
Alert contacts of COVID-19 cases	12 (26)	7
Report suspected cases/rule infringement	5 (11)	3
Health code/status generator	7 (15)	4
Health/travel declaration	7 (15)	4
Self-symptom assessment	8 (17)	5
Vaccination	7 (15)	—
Vaccination information	4 (9)	2
Vaccination registration/appointment	3 (7)	2
Vaccination certificate	4 (9)	2
Reporting adverse reactions	1 (2)	1
Health resources	12 (26)	—
Virtual medical consultation	4 (9)	2
Emergency helpline	7 (15)	4
Accessing medical records	1 (2)	1
Personal protective equipment distribution	4 (9)	2

[a]% values calculated on the basis of the total functions (n=169).

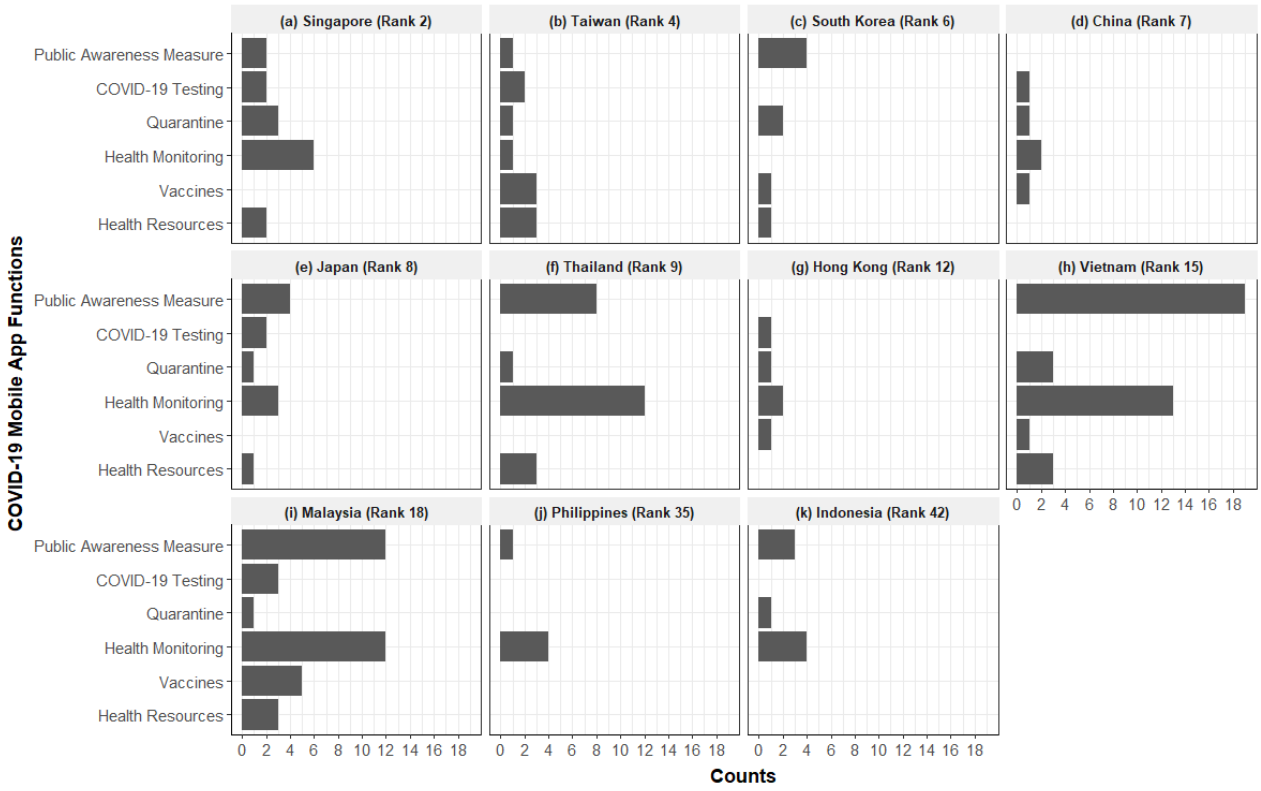
[b]Not applicable.

Seven (15%) apps supported the function for COVID-19 vaccination. Most of these apps provided information regarding COVID-19 vaccines (n=4, 9%) or issued digital proof-of-vaccination (n=4, 9%) to app users who have completed their vaccine doses. Users could also register and make appointments for COVID-19 vaccination (n=3, 7%) via the app.

However, only one of the apps (2%), Taiwan V-watch, allowed users to report vaccination-related adverse reactions.

Figure 2 illustrates the total number of functions served by mobile apps in each government by adding up the number of functions of each app per government. For example, if a government introduced multiple mobile apps having the same functions, the total number of functions will be the sum of each function. Mobile apps in Taiwan and Malaysia had all main functions related to 6 different policy types, and those in Singapore and Japan covered most of the functions except for vaccination. Mobile apps in Thailand, Vietnam, and Malaysia focused on functions for public awareness measures and health monitoring. Among these apps, the MySejahtera app from Malaysia was the most comprehensive app, incorporating public awareness measures, quarantine monitoring, health monitoring, vaccination, and health resources. However, the types of functions served by mobile apps were relatively limited in the Philippines and Indonesia compared to those in the other 9 economies in Bloomberg's Covid Resilience Ranking.

Figure 2. Overview of the key functions of the included government apps.



Relationship Between Government Measures and the Availability of Mobile Apps

Figure 3 shows the timeline of the commencement dates of public health policies and the release dates of mobile apps. Each policy type consists of subtypes, and each point indicates the timepoints of when the policies were implemented. We did not examine the details of each policy.

All governments introduced mobile apps to support COVID-19 mitigation policies. There were no noticeable differences among the included governments with respect to the time of introduction of mobile apps. Furthermore, there was no consistency in the introduction of mobile apps and the initiation of certain types of policies across the governments. Eight governments, namely Singapore, South Korea, China (mainland), Thailand, Hong Kong, Vietnam, Malaysia, and Indonesia, launched their first apps between March and April 2020 (Figure 3).

In 2021, Hong Kong, Taiwan, and South Korea released apps to help track COVID-19 vaccination, registrations, and side effects. Some apps such as WeChat (China [mainland]), MySejahtera (Malaysia), Selangkah (Malaysia), and Bluezone (Vietnam) were updated to include vaccination-related functions.

Figure 3. Governments' COVID-19 policy commencement dates and release dates of the included apps.

Discussion

Principal Findings

This study identified 46 mobile apps developed or supported by 11 governments in East and South-East Asia by using a systematic search method. The most common function was health monitoring. Within the health monitoring function, the most popular function was alerting positive cases, followed by contact tracing and digital check-in. The second-most common function was public awareness measures such as disseminating news or government measures.

Evidence shows that most apps initially focused on disseminating information or monitoring high-risk areas and subsequently had functions for contact tracing [25,33]. As we searched mobile apps cross-sectionally, we did not examine changes in the functions over time. However, most apps in our review had additional functions such as digital check-in, self-assessment of symptoms, virtual medical consultation, COVID-19 testing management, and vaccination-related processes. We noticed that the functions of COVID-19 apps were expanded to cover vaccination-related purposes too. Provision of information and issuance of vaccination certificates were the most frequent functions, followed by vaccination registration or appointment. In other governments, apps not having such functions at the time of our search in May 2021 subsequently integrated the functions in parallel with their nationwide administration of COVID-19 vaccines. In Singapore, test results and vaccination records were added to the pre-existing health information app "HealthHub SG" in February 2021 [34]. In Japan, the COVID-19 vaccination certificate will be available via a QR code using a smartphone in December 2021 [35]. Thus, mobile apps can play an important role in promoting the COVID-19 vaccination programs and increasing their coverage [36].

Since Alanzi [13] reviewed 12 mobile apps in August 2020, we noticed that many mobile apps integrating various functions have emerged. This change might be due to governments' efforts to address users' evolving needs and increase data management efficiency by health authorities [37,38]. Furthermore, some governments such as those of Japan, Malaysia, and Vietnam have developed city-level or state-level apps that provided area-specific information, which supported the local health systems. Given the necessity of crisis management at subnational levels, app-based measures can be promising by promoting regional coordination [39].

Most governments in our review required international travelers to use their apps for health declaration and monitoring. Notably, most quarantine monitoring apps were mandatory for people who required quarantine, mainly international travelers. Compulsory implementation of these apps to other settings or populations would not be simple considering national or regional policies regarding data protection and privacy [40]. Indeed, data security and sharing of data with third parties have been the main reason underlying the reluctance to share information in mobile apps [41,42]. Lack of public trust toward authorities is also a significant reason to refuse privacy trade-off [43,44]. Hence, to maximize the effectiveness of the apps, there must be coordinated legal and ethical governance in place to confer protection against invasion of users' privacy [45].

We examined the timing of the rollout of COVID-19–related mobile apps to assess their relationship with the introduction of other public health measures. All governments included in our review used mobile apps to support the COVID-19 mitigation policies. We found that mobile apps from more successful economies such as Singapore and Malaysia tended to have diverse functions covering various measures. Most apps also first emerged close to the commencement dates of relevant public health policies between March and April 2020. Governments that showed successful performance tended to introduce COVID-19–related apps in the early stages of the pandemic. We did not statistically analyze associations between the timing of introducing apps and epidemiological data. Therefore, further analysis is required.

Although our findings focused on mobile apps, there are various other forms of digital solutions to combat COVID-19. For example, Taiwan did not have a particular mobile app for monitoring quarantine using GPS; however, it initiated the "Entry Quarantine System." This system was achieved by scanning the QR code directly or clicking on its website. Travelers were required to make a web-based health declaration within 2 days before arriving in Taiwan and complete 14-day quarantine at a government facility, a designated hotel, or at home. Thereafter, the "Electronic Fence system" tracks the locations of individuals during their quarantine period using mobile location data to ensure that travelers do not leave their quarantine location [46]. In China (mainland), AI solutions have been used in lung computed tomographic scans, minimizing time and allowing for early diagnosis of COVID-19 cases [47]. Multifaceted digital approaches were utilized, and although they were not substitutes for traditional health care, their integration complemented and enhanced a functioning health system.

It is difficult to determine which mobile app was the most effective in curtailing COVID-19. As of March 24, 2021, Taiwan and Vietnam recorded 0 deaths per 1 million population, 1 in Thailand, 5 in Singapore, 27 in Hong Kong, 3 in China (mainland), 33 in South Korea, 38 in Malaysia, 70 in Japan, 119 in the Philippines, and 146 in Indonesia [28]. Overall, governments introducing mobile apps covering various forms of public health measures showed fewer deaths per million population. However, other factors such as the health system capacity and resources should be considered. For instance, although Malaysia had the most comprehensive COVID-19 apps in our review, Singapore was the top-performing government with the highest COVID-19 resilience in the Asia Pacific region, having the fastest inoculation program and the lowest positive test rate (Multimedia Appendix 1). Future research could therefore consider other domains of public health to assess the performance of COVID-19–related mobile apps.

Our included apps were purposefully selected from governments, which displayed the most cohesive responses to the pandemic as of March 2021. However, the unprecedented infiltration of the highly transmissible delta variant has wrecked the model of COVID-19 containment success exercised in East and South-East Asia. South-East Asia has emerged as the new virus epicenter; the bottom 5 in the latest Bloomberg's Covid Resilience Ranking (August 2021) were all South-East Asian economies [48]. Although these economies showed effective resilience by adapting mobile apps in their public health policies, there are still barriers or blind spots that the current mHealth approaches should overcome.

Practical Implications

This review has several implications for the governments and for public health researchers. Our findings show that governments in East and South-East Asia initiated mobile solutions in the early days of the pandemic, and their COVID-19–related mobile apps were used for various purposes.

Successful performance of mobile apps in both resource-rich and resource-limited settings in this region demonstrated the wide range of applications of these apps and their cost-effectiveness (Multimedia Appendix 3). Although we only compared the timing of the introduction of mobile apps in relation to the commencement dates of other public health policies (Figure 3), we observed how mobile apps are intertwined in the context of public health policies. Governments should consider these mobile solutions in East and South-East Asia to strengthen the current public health system and prepare for subsequent outbreaks.

For public health researchers, there is an enormous potential for such apps, especially in epidemiological research, disease surveillance, and allocation of health resources. Mobile apps can be designed to collect and generate research data to improve our understanding and response to this pandemic.

Limitations and Recommendations for Future Studies

This study has limitations that are important to acknowledge. It is plausible that some apps may have been missed owing to the restrictive setting of several regional app stores. To overcome this issue, we have scoured other sources of information such as current news articles, media reports, and literature to find additional relevant apps. However, it is still likely that some relevant apps were missed as our search terms may not encompass all the available apps, especially those named in the local languages.

Moreover, we did not collect data on the consumer ratings or user feedback of each app. We also neither examined the popularity nor considered the number of app downloads. Although some evidence suggests that contact-tracing apps should be adopted by at least 60%-70% of the population to impact the outbreak transmission rate, much lower app penetration could still be substantial in breaking transmission chains and preventing infection [49-51]. Nevertheless, given that the number of users determines the utility of mobile apps, our findings may not be generalizable to other countries or populations.

We also did not examine the mobile uptake proportion by people from different socioeconomic backgrounds. There is a need to assess how well these mobile apps were accessible by the most deprived individuals, including older individuals, homeless individuals, immigrants, and rural residents [52-54].

Conclusions

In conclusion, our findings added knowledge on the COVID-19–related apps used in 11 governments in East and South-East Asia. The most common function was to monitor public health, followed by disseminating information and health education. Most apps deployed GPS technology, followed by Bluetooth and QR scanner technologies. Most countries in this region adopted mobile apps to support COVID-19 mitigation efforts and introduced them close to the relevant policy commencement dates in the early stages of the pandemic. In addition, some governments, which are relatively successful in suppressing COVID-19, tended to have all-in-one mobile apps or other complementary mobile apps. These apps could play pivotal roles in supporting governments' measures for tracking COVID-19 cases and delivering credible information. Mobile apps catering to the middle-ground strategy of widespread vaccination and reopening of economies can be adopted by the governments to reframe the way of life as we move toward the endemic phase of the COVID-19 pandemic.

Acknowledgments

The authors are grateful for the generous support of Dr Cindy Cheng at CoronaNet Research Project, whose expert guidance assisted in strengthening the manuscript.

Authors' Contributions

BL conceived the study and acquired support from CoronaNet Research Project. SAI and BL developed the study protocol and designed the data extraction rubric. Two authors independently reviewed the mobile apps and extracted, analyzed, and interpreted the data (BL and TZ for app features, BL and SAI for app functions). BL and SAI drafted the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bloomberg's COVID-19 Resilience Ranking (March 2021).
[DOCX File , 19 KB - mhealth_v9i11e32093_app1.docx]

Multimedia Appendix 2
List of included mobile apps and their associated characteristics.
[DOCX File , 35 KB - mhealth_v9i11e32093_app2.docx]

Multimedia Appendix 3
List of included mobile apps with their associated functions.
[DOCX File , 38 KB - mhealth_v9i11e32093_app3.docx]

References

1. Measuring digital development: Facts and figures 2020. International Telecommunication Union. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf [accessed 2021-06-06]
2. Ssali J. Infographic: Half the global population are now using the mobile internet. GSMA Intelligence. 2020. URL: https://tinyurl.com/3tezx9mu [accessed 2021-06-06]
3. mHealth: New horizons for health through mobile technologies. World Health Organization. 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2021-09-05]
4. Wood CS, Thomas MR, Budd J, Mashamba-Thompson TP, Herbst K, Pillay D, et al. Taking connected mobile-health diagnostics of infectious diseases to the field. Nature 2019 Feb;566(7745):467-474 [FREE Full text] [doi: 10.1038/s41586-019-0956-2] [Medline: 30814711]
5. Mackillop L, Loerup L, Bartlett K, Farmer A, Gibson OJ, Hirst JE, et al. Development of a real-time smartphone solution for the management of women with or at high risk of gestational diabetes. J Diabetes Sci Technol 2014 Nov;8(6):1105-1114 [FREE Full text] [doi: 10.1177/1932296814542271] [Medline: 25004915]
6. Estcourt CS, Gibbs J, Sutcliffe LJ, Gkatzidou V, Tickle L, Hone K, et al. The eSexual Health Clinic system for management, prevention, and control of sexually transmitted infections: exploratory studies in people testing for Chlamydia trachomatis. Lancet Public Health 2017 Apr;2(4):e182-e190 [FREE Full text] [doi: 10.1016/S2468-2667(17)30034-8] [Medline: 29253450]
7. Hochstatter KR, Gustafson DH, Landucci G, Pe-Romashko K, Cody O, Maus A, et al. Effect of an mHealth Intervention on Hepatitis C Testing Uptake Among People With Opioid Use Disorder: Randomized Controlled Trial. JMIR Mhealth Uhealth 2021 Feb 22;9(2):e23080 [FREE Full text] [doi: 10.2196/23080] [Medline: 33616545]

8. Asadzadeh A, Kalankesh LR. A scope of mobile health solutions in COVID-19 pandemics. Inform Med Unlocked 2021;23:100558 [FREE Full text] [doi: 10.1016/j.imu.2021.100558] [Medline: 33842688]
9. Pant Pai N, Behlim T, Abrahams L, Vadnais C, Shivkumar S, Pillay S, et al. Will an unsupervised self-testing strategy for HIV work in health care workers of South Africa? A cross sectional pilot feasibility study. PLoS One 2013;8(11):e79772 [FREE Full text] [doi: 10.1371/journal.pone.0079772] [Medline: 24312185]
10. Larocca A, Moro Visconti R, Marconi M. Malaria diagnosis and mapping with m-Health and geographic information systems (GIS): evidence from Uganda. Malar J 2016 Oct 24;15(1):520 [FREE Full text] [doi: 10.1186/s12936-016-1546-5] [Medline: 27776516]
11. Rolling updates on coronavirus disease (COVID-19). World Health Organization. 2020 Jul 31. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen [accessed 2021-06-05]
12. Clavier F, Ghesquiere F. Leveraging digital solutions to fight COVID-19 : Lessons from ASEAN Countries. World Bank. 2021. URL: https://openknowledge.worldbank.org/handle/10986/35126 [accessed 2021-06-21]
13. Alanzi T. A Review of Mobile Applications Available in the App and Google Play Stores Used During the COVID-19 Outbreak. J Multidiscip Healthc 2021;14:45-57 [FREE Full text] [doi: 10.2147/JMDH.S285014] [Medline: 33469298]
14. Chorzempa M, Huang T. Lessons from East Asia and Pacific on taming the pandemic. Peterson institute for international economics. URL: https://www.piie.com/blogs/realtime-economic-issues-watch/lessons-east-asia-and-pacific-taming-pandemic [accessed 2021-07-09]
15. Lim PL. Middle East respiratory syndrome (MERS) in Asia: lessons gleaned from the South Korean outbreak. Trans R Soc Trop Med Hyg 2015 Sep;109(9):541-542 [FREE Full text] [doi: 10.1093/trstmh/trv064] [Medline: 26286944]
16. Chua AQ, Al Knawy B, Grant B, Legido-Quigley H, Lee W, Leung GM, et al. How the lessons of previous epidemics helped successful countries fight covid-19. BMJ 2021 Mar 11;372:n486. [doi: 10.1136/bmj.n486] [Medline: 33707174]
17. Liu Q, Xu K, Wang X, Wang W. From SARS to COVID-19: What lessons have we learned? J Infect Public Health 2020 Nov;13(11):1611-1618 [FREE Full text] [doi: 10.1016/j.jiph.2020.08.001] [Medline: 32888871]
18. An BY, Tang S. Lessons From COVID-19 Responses in East Asia: Institutional Infrastructure and Enduring Policy Instruments. Am Rev Public Adm 2020 Jul 22;50(6-7):790-800. [doi: 10.1177/0275074020943707]
19. Huang Y, Sun M, Sui Y. How digital contact tracing slowed Covid-19 in East Asia. Harvard Business Review. URL: https://hbr.org/2020/04/how-digital-contact-tracing-slowed-covid-19-in-east-asia [accessed 2021-07-05]
20. Tan S, Taeihagh A, Sha K. How Transboundary Learning Occurs: Case Study of the ASEAN Smart Cities Network (ASCN). Sustainability 2021 Jun 07;13(11):6502. [doi: 10.3390/su13116502]
21. Navarro V. Why Asian Countries are Controlling the Pandemic Better Than the United States and Western Europe. Int J Health Serv 2021 Apr;51(2):261-264 [FREE Full text] [doi: 10.1177/0020731421999930] [Medline: 33703963]
22. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A Systematic Review of Smartphone Applications Available for Corona Virus Disease 2019 (COVID19) and the Assessment of their Quality Using the Mobile Application Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
23. Zhang MWB, Chow A, Ho RCM, Smith HE. An Overview of Commercially Available Apps in the Initial Months of the COVID-19 Pandemic. Front Psychiatry 2021;12:557299 [FREE Full text] [doi: 10.3389/fpsyt.2021.557299] [Medline: 33935816]
24. Ming LC, Untong N, Aliudin NA, Osili N, Kifli N, Tan CS, et al. Mobile Health Apps on COVID-19 Launched in the Early Days of the Pandemic: Content Analysis and Review. JMIR Mhealth Uhealth 2020 Sep 16;8(9):e19796 [FREE Full text] [doi: 10.2196/19796] [Medline: 32609622]
25. Almalki M, Giannicchi A. Health Apps for Combating COVID-19: Descriptive Review and Taxonomy. JMIR Mhealth Uhealth 2021 Mar 02;9(3):e24322 [FREE Full text] [doi: 10.2196/24322] [Medline: 33626017]
26. Digital Adoption Index. World Bank Group. URL: https://www.worldbank.org/en/publication/wdr2016/Digital-Adoption-Index [accessed 2021-06-30]
27. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
28. The best and worst places to be as variants outrace vaccinations. Bloomberg. URL: https://www.bloomberg.com/graphics/covid-resilience-ranking/ [accessed 2021-04-10]
29. Google Play. URL: https://play.google.com/store [accessed 2021-05-07]
30. Fnd: Experience the App Store and iTunes Anywhere. URL: https://fnd.io/#/ [accessed 2021-05-07]
31. Cheng C, Barceló J, Hartnett AS, Kubinec R, Messerschmidt L. COVID-19 Government Response Event Dataset (CoronaNet v.1.0). Nat Hum Behav 2020 Jul;4(7):756-768. [doi: 10.1038/s41562-020-0909-7] [Medline: 32576982]
32. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
33. Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M. Features and Functionalities of Smartphone Apps Related to COVID-19: Systematic Search in App Stores and Content Analysis. J Med Internet Res 2020 Aug 25;22(8):e20334 [FREE Full text] [doi: 10.2196/20334] [Medline: 32614777]

34. Health Hub. Ministry of Health Singapore. URL: https://www.healthhub.sg/ [accessed 2021-09-05]
35. About the new corona vaccination certificate for overseas travel. Ministry of Health, Labor and Welfare. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/vaccine_certificate.html [accessed 2021-09-05]
36. Simeoni R, Maccioni G, Giansanti D. The Vaccination Process against the COVID-19: Opportunities, Problems and Support. Healthcare (Basel) 2021 Sep 06;9(9):1165 [FREE Full text] [doi: 10.3390/healthcare9091165] [Medline: 34574939]
37. Steinberg M. LINE as Super App: Platformization in East Asia. Soc Media Soc 2020 Jun 19;6(2):205630512093328. [doi: 10.1177/2056305120933285]
38. How is Covid-19 changing South-east Asia's super apps? Oxford Business Group. URL: https://oxfordbusinessgroup.com/news/how-covid-19-changing-south-east-asia-super-apps [accessed 2021-06-21]
39. OECD policy responses to Coronavirus (COVID-19): The territorial impact of COVID-19: Managing the crisis across levels of government. Organisation for Economic Co-operation and Development. URL: https://www.oecd.org/coronavirus/policy-responses/the-territorial-impact-of-covid-19-managing-the-crisis-across-levels-of-government-d3e314e1/ [accessed 2021-06-04]
40. Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, De Nadai M, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. Sci Adv 2020 Jun;6(23):eabc0764 [FREE Full text] [doi: 10.1126/sciadv.abc0764] [Medline: 32548274]
41. Du L, Raposo VL, Wang M. COVID-19 Contact Tracing Apps: A Technologic Tower of Babel and the Gap for International Pandemic Control. JMIR Mhealth Uhealth 2020 Nov 27;8(11):e23194 [FREE Full text] [doi: 10.2196/23194] [Medline: 33156804]
42. Chan EY, Saqib NU. Privacy concerns can explain unwillingness to download and use contact tracing apps when COVID-19 concerns are high. Comput Human Behav 2021 Jun;119:106718 [FREE Full text] [doi: 10.1016/j.chb.2021.106718] [Medline: 33526957]
43. Cho H, Ippolito D, Yu Y. Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. arXiv Preprint posted online March 24, 2020.
44. Wottrich VM, van Reijmersdal EA, Smit EG. The privacy trade-off for mobile app downloads: The roles of app value, intrusiveness, and privacy concerns. Decis Support Syst 2018 Feb;106:44-52. [doi: 10.1016/j.dss.2017.12.003]
45. Zwitter A, Gstrein OJ. Big data, privacy and COVID-19 – learning from humanitarian expertise in data protection. Int J Humanitarian Action 2020 May 18;5(1). [doi: 10.1186/s41018-020-00072-6]
46. Quarantine system for Entry: Quarantine measures required for incoming passengers 2021. Taiwan Mistry of Health and Welfare. URL: https://hdhq.mohw.gov.tw/Default1?openExternalBrowser=1 [accessed 2021-09-05]
47. Chen B, Marvin S, While A. Containing COVID-19 in China: AI and the robotic restructuring of future cities. Dialogues in Human Geography 2020 Jun 11;10(2):238-241. [doi: 10.1177/2043820620934267]
48. Chang R, Varley K, Munoz M, Tam F, Makol M. The Covid Resilience Ranking: The Best and Worst Places to Be as Delta Wrecks Reopening Plans. Bloomberg. URL: https://www.bloomberg.com/graphics/covid-resilience-ranking/ [accessed 2021-09-05]
49. Digital contact tracing can slow or even stop coronavirus transmission and ease us out of lockdown. Oxford University's Big Data Institute. URL: https://www.research.ox.ac.uk/article/2020-04-16-digital-contact-tracing-can-slow-or-even-stop-coronavirus-transmission-and-ease-us-out-of-lockdown [accessed 2021-06-21]
50. Hart V, Siddarth D, Cantrell B. Outpacing the virus: Digital response to containing the spread of COVID-19 while mitigating privacy risks. Edmond J Safra Center for Ethics. URL: https://ethics.harvard.edu/outpacing-virus [accessed 2021-06-21]
51. O'Neill P. No, coronavirus apps don't need 60% adoption to be effective. MIT Technology Review. URL: https://www.technologyreview.com/2020/06/05/1002775/covid-apps-effective-at-less-than-60-percent-download/ [accessed 2021-06-21]
52. Beaunoyer E, Dupéré S, Guitton MJ. COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. Comput Human Behav 2020 Oct;111:106424 [FREE Full text] [doi: 10.1016/j.chb.2020.106424] [Medline: 32398890]
53. Lee C, Rogers WA, Braunack-Mayer A. Social Justice and Pandemic Influenza Planning: The Role of Communication Strategies. Public Health Ethics 2008 Nov 01;1(3):223-234. [doi: 10.1093/phe/phn031]
54. Van Lancker W, Parolin Z. COVID-19, school closures, and child poverty: a social crisis in the making. Lancet Public Health 2020 May;5(5):e243-e244 [FREE Full text] [doi: 10.1016/S2468-2667(20)30084-0] [Medline: 32275858]

Abbreviations

AI: artificial intelligence
API: application programming interface
ASCN: ASEAN Smart Cities Network
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
WHO: World Health Organization

Edited by L Buis; submitted 14.07.21; peer-reviewed by M Lotto, L Gupta, S Davalbhakta; comments to author 18.08.21; revised version received 07.09.21; accepted 27.10.21; published 11.11.21.

Please cite as:
Lee B, Ibrahim SA, Zhang T
Mobile Apps Leveraged in the COVID-19 Pandemic in East and South-East Asia: Review and Content Analysis
JMIR Mhealth Uhealth 2021;9(11):e32093
URL: https://mhealth.jmir.org/2021/11/e32093
doi:10.2196/32093
PMID:34748515

Original Paper

GPS Mobile Health Intervention Among People Experiencing Homelessness: Pre-Post Study

Leticia R Moczygemba[1], PharmD, PhD; Whitney Thurman[2], RN, PhD; Kyler Tormey[1], BS; Anthony Hudzik[1], BS; Lauren Welton-Arndt[1], BA; Elizabeth Kim[1], BSA

[1]Health Outcomes Division, College of Pharmacy, University of Texas at Austin, Austin, TX, United States

[2]School of Nursing, University of Texas at Austin, Austin, TX, United States

Corresponding Author:
Leticia R Moczygemba, PharmD, PhD
Health Outcomes Division
College of Pharmacy
University of Texas at Austin
1 University Station
Austin, TX, 78712
United States
Phone: 1 5122326880
Email: lrmoczygemba@austin.utexas.edu

Abstract

Background: People experiencing homelessness are at risk for gaps in care after an emergency department (ED) or hospital visit, which leads to increased use, poor health outcomes, and high health care costs. Most people experiencing homelessness have a mobile phone of some type, which makes mobile health (mHealth) interventions a feasible way to connect a person experiencing homelessness with providers.

Objective: This study aims to investigate the accuracy, acceptability, and preliminary outcomes of a GPS-enabled mHealth (GPS-mHealth) intervention designed to alert community health paramedics when people experiencing homelessness are in the ED or hospital.

Methods: This study was a pre-post design with baseline and 4-month postenrollment assessments. People experiencing homelessness, taking at least 2 medications for chronic conditions, scoring at least 10 on the Patient Health Questionnaire-9, and having at least 2 ED or hospital visits in the previous 6 months were eligible. Participants were issued a study smartphone with a GPS app programmed to alert a community health paramedic when a participant entered an ED or hospital. For each alert, community health paramedics followed up via telephone to assess care coordination needs. Participants also received a daily email to assess medication adherence. GPS alerts were compared with ED and hospital data from the local health information exchange (HIE) to assess accuracy. Paired *t* tests compared scores on the Patient Health Questionnaire-9, Medical Outcomes Study Social Support Survey, and Adherence Starts with Knowledge-12 adherence survey at baseline and exit. Semistructured exit interviews examined the perceptions and benefits of the intervention.

Results: In total, 30 participants were enrolled; the mean age was 44.1 (SD 9.7) years. Most participants were male (20/30, 67%), White (17/30, 57%), and not working (19/30, 63%). Only 19% (3/16) of the ED or hospital visit alerts aligned with HIE data, mainly because of patients not having the smartphone with them during the visit, the smartphone being off, and gaps in GPS technology. There was a significant difference in depressive symptoms between baseline (mean 16.9, SD 5.8) and exit (mean 12.7, SD 8.2; t_{19}=2.9; *P*=.009) and a significant difference in adherence barriers between baseline (mean 2.4, SD 1.4) and exit (mean 1.5, SD 1.5; t_{17}=2.47; *P*=.03). Participants agreed that the app was easy to use (mean 4.4/5, SD 1.0, with 5=strongly agree), and the email helped them remember to take their medications (mean 4.6/5, SD 0.6). Qualitative data indicated that unlimited smartphone access allowed participants to meet social needs and maintain contact with case managers, health care providers, family, and friends.

Conclusions: mHealth interventions are acceptable to people experiencing homelessness. HIE data provided more accurate ED and hospital visit information; however, unlimited access to reliable communication provided benefits to participants beyond the study purpose of improving care coordination.

(JMIR Mhealth Uhealth 2021;9(11):e25553) doi:10.2196/25553

KEYWORDS

GPS; mHealth; care coordination; people experiencing homelessness; homelessness; emergency department; health outcomes; health care costs; mobile phone

Introduction

Background

Mobile phone ownership is nearly ubiquitous among American adults, with 96% owning a mobile phone of some type, and most (81%) mobile phones are smartphones [1]. Accordingly, mobile technology is increasingly common in the health care sector. Mobile devices are being used for medical diagnostics [2], disease monitoring [3], smoking cessation [4], and dietary tracking [5]. Smartphone capabilities, including texting and apps, have contributed to improved medication adherence [6], higher attendance at medical appointments [7], and increased vaccination rates [8]. Mobile technology has also been explored as a useful tool to bolster the transmission of information and care coordination during transitions of care [9,10], and studies have demonstrated the potential of mobile technology to improve communication among health care providers and populations at risk for poor outcomes, including people of lower socioeconomic status [11,12].

Recent estimates of mobile phone use among the homeless population indicate that 89% of the people report having and using a mobile phone [13], and researchers have begun to explore the possibility of using mobile technology to improve the health of people experiencing homelessness. For example, Burda et al [14] concluded that mobile phones are a feasible way to monitor and manage medication regimens for people experiencing homelessness with co-occurring disorders. Furthermore, in a survey of people experiencing homelessness, 77% of the respondents were interested in appointment reminders, and most were interested in medication refill reminders (66%) and medication taking reminders (60%) [13]. Despite the accumulating evidence that mobile health (mHealth) interventions among homeless populations are feasible, GPS-enabled mHealth (GPS-mHealth) interventions in this population have remained underexplored. The purpose of this study, therefore, is to investigate the acceptability and preliminary outcomes of a GPS-mHealth intervention designed to improve care coordination in a sample of people experiencing homelessness.

Evidence suggests that the health service experiences of people experiencing homelessness are often interrupted and involve extensive barriers, including unmet physical needs, lack of affordable and available services, and lack of compassion that prevents people experiencing homelessness from accessing appropriate community-based services [15-17]. These barriers lead to disruptions in continuity of care, which is problematic because of evidence that continuity of care—that is, timely, accessible, person-centered, and coordinated care—improves outcomes [18]. Interventions such as case management, respite care, and housing services that target critical transition points have led to decreased acute care use [19] in people experiencing homelessness. Community paramedics have also been used to coordinate care and link high-risk patients to needed health and social services [20], which has led to reduced health care use among diverse populations and improvements in patient outcomes [21]. Despite these multifaceted programs, interventions, and service delivery models intended to improve care coordination among people experiencing homelessness, gaps in services along the continuum of care persist.

Study Premise and Objectives

This study focuses on the significant gap along the continuum of care that begins at the point of an emergency department (ED) visit or hospitalization for people experiencing homelessness. The study intervention was created on the basis of feedback from health care providers and case managers who deliver care to homeless individuals, and the fact that fragmented communication among various health care organizations limits the ability to provide real-time information about ED or hospital visits. When a person experiencing homelessness enters the ED or hospital, they are at high risk of losing contact with community-based health care providers and case managers [22]. This is exacerbated in the people experiencing homelessness living with depression as it is more difficult to manage their chronic conditions, including attending appointments and taking medications as prescribed [23]. The loss of contact between homeless individuals and their community-based care team creates a time of high risk for the individual and represents missed opportunities to provide services and potentially decrease acute health care use. For preventing or minimizing this loss of contact, this study used geofencing to create virtual boundaries that triggered automatic notification of community paramedics if and when a person experiencing homelessness visited an ED or hospital. The use of such geofencing technology in health care has been previously studied in smoking cessation, dietary recommendations, anxiety, and hospitalizations in patients with cardiovascular disease [5,10,24,25]. However, the utility of a GPS-mHealth intervention specifically in transitions of care for people experiencing homelessness has not been previously reported. Therefore, the following research questions guided this study:

1. What is the accuracy of GPS technology in terms of tracking participant visits to the ED or hospital?
2. How do depression symptoms, medication adherence, social support, and experience with and perceptions of GPS and mobile phone technology compare at baseline and exit?
3. What is the number and type of community health paramedic encounters?
4. What concerns do participants express regarding technology or privacy?

Methods

Design and Participants

This study used a pre-post design with assessments at baseline, 1 month, 2 months, 3 months, and 4 months after enrollment to evaluate the acceptability and preliminary outcomes of a GPS-enabled mHealth intervention. Participants were recruited from 2 churches that provided services to people experiencing homelessness. The first serves breakfast at 5:45 AM two mornings each week and is open to anyone in the community. The research staff attended this breakfast once per week for the study duration. Potential participants were referred to study staff for eligibility screening by either the meal program coordinator or the police officer assigned to the downtown Homeless Outreach Service team, whose job function includes attending these twice weekly breakfasts. The second church site doubles as a navigation center for people experiencing homelessness during weekdays. Services at the navigation center include coordinated assessments for housing, assistance with obtaining IDs, and case management. The research staff were on site at the navigation center 2 to 3 days per week for the study duration. Similar to the first site, potential participants were referred by the director of the navigation center or by navigation center volunteers to study staff for study eligibility screening.

Recruitment occurred between October 2018 and April 2019. Community partners assisted with recruitment by distributing flyers to clients and by referring potential participants to research staff. Participants also referred peers who were potentially eligible to the study staff. Potential participants were screened for study eligibility on site at the churches by a member of the research team. The eligibility criteria included (1) being at least 18 years old, (2) currently experiencing homelessness defined as where the person had slept most nights in the past 30 days, (3) score of at least 10 on the Patient Health Questionnaire-9 (PHQ-9), (4) currently prescribed at least 2 medications for chronic medical conditions, (5) diagnosed with at least 1 chronic medical condition, and (6) experienced at least 2 hospitalizations or ED visits in the past 6 months. Exclusion criteria included (1) onset in the past 3 months of depressive symptoms and (2) suicide attempts or suicidal ideation in the past 6 months.

This study was approved by the institutional review board of the university. Individuals interested in participating were screened by research staff, and, if eligible to continue, study details, including the purpose, procedures, risks, and benefits of study participation, were explained. If participants remained interested, informed consent was obtained. None of the participants who were eligible for the study declined to participate after being informed of the study details. After obtaining informed consent, a researcher administered a series of baseline assessments to collect information about demographics, health history, medication adherence, social support, and recent ED visit and hospitalizations. After completion of the surveys, participants were provided with a smartphone activated with a plan for unlimited texting, calling, and data; a hard-plastic smartphone case; and an armband to use for securing the smartphone. Participants were also given US $25 cash for the time spent enrolling in the study and a 31-day unlimited use bus pass to ensure their ability to attend the monthly follow-up assessment. They then received training on the intervention. The training described the expectations of participants, including keeping track of the smartphone, keeping it turned on and charged, attending monthly check-in visits, answering the daily email regarding medication adherence, and responding to community health paramedics or research staff as applicable. The training also included how to use the smartphone, set up voicemail, access email and SMS text messages, and access and use the bus pass. Participants were also informed that one replacement smartphone would be issued if their study smartphone was lost, stolen, or broken during the 4-month study.

GPS-mHealth Intervention

For this study, a mobile app was used to establish and monitor geofences around the 10 EDs located within the city limits where this study took place. The geofences were established using the mobile app so that when a participant entered a local ED or hospital, the research staff and the commander of the community paramedic team would receive an email notification. The email notification sent a secure link to view the participant's name, geofence location, date, and time of entry and exit. On receipt, the commander tasked a community health paramedic member of his team to contact the participant via their smartphone within 2 business days of the geofence entry to follow-up on the visit and any identified health or social needs. The community health paramedic completed an event form documenting the participant-reported reason for the hospital or ED visit, admission, and discharge dates; if the ED or hospital visit was potentially preventable; what intervention may have prevented the ED and hospital visit; and the duration of the community health paramedic visit with the participant.

In addition to the geofencing and care provided by community health paramedics as needed, the intervention had two additional components: (1) monthly in-person meetings and (2) daily adherence reminder emails. In-person meetings occurred between each participant and research staff at enrollment; 1-, 2-, and 3-month follow-up appointments; and at the exit. Monthly follow-up visits (at 1, 2, and 3 months) were scheduled to maintain contact with participants and to identify any issues with the technology. Participants were also asked at these monthly meetings if they had visited the ED or been hospitalized in the past 30 days. At months 1, 2, and 3, participants received US $10 cash and an additional 31-day bus pass. Next, participants received an email every evening at 8 PM asking if they had taken their medications that day. Response options were "yes" or "no," with a follow-up question requesting a short reason why they had not taken their medication if applicable. During the exit interviews, participants responded to a series of questionnaires before engaging in a semistructured interview to assess the overall acceptance of the intervention. Textbox 1 summarizes the interview guidelines. Finally, the local health information exchange (HIE) provided research staff with dates of hospital admissions and ED visits, as applicable, for participants during the 4-month study period.

Textbox 1. Semistructured interview guide.

Questions about the study

1. Please describe your experience with this research study (probe 1: Did you experience benefits from participating in this study? probe 2: Was participating in this study helpful to you? probe 3: Were there any difficulties that you experienced with this study?)
2. Can you share any barriers to study participation that you experienced? (Examples may include keeping the smartphone secure or charged)
3. What strategies did you use to successfully complete the study requirements? (This includes things such as keeping the smartphone charged and operational as well as attendance at monthly check-in visits)
4. What concerns did you have about your visits to the emergency room and hospital being monitored with GPS technology?
5. Can you describe any experiences or interactions you had with community health paramedics?
6. What suggestions do you have for us to improve this intervention for people in the future?

Measurements

Sociodemographic and Health-Related Variables

At baseline, sociodemographic characteristics, including sex, race, highest education obtained, veteran status, and income, were collected. Participants were also asked a series of six questions from the American Community Survey designed to identify individuals who may experience functional limitations [26]. Response options were 1=yes or 0=no. The items were summed for a total score, with higher scores indicating a higher burden of functional limitations. The Cut down, Annoyed, Guilty, and Eye-opener questionnaire, a 4-item screening tool, was used to screen for alcohol use [27]. Response options were 1=yes or 0=no. The items were summed for a total score, and a total score of >2 was considered clinically significant [27]. The single-item screen in which the participant is asked, "how many times in the past year have you used an illicit drug or used a prescription medication for nonmedical reasons?" was used to screen for substance use [28]. Responses ≥1 were considered to be positive.

Health Literacy

Health literacy was measured using the Brief Health Literacy Screening Tool [29], which comprises 4 questions that assess respondents' ability to complete tasks such as filling out medical forms, reading hospital paperwork, and learning about one's medical condition. Each item is worth 1 to 5 points, depending on the response. Scores were summed for a composite score ranging from 4 to 20. Scores of 4-12 indicate limited health literacy, scores of 13-16 indicate marginal health literacy, and scores of 17-20 indicate adequate health literacy [30].

Accuracy of the GPS Technology

The accuracy of the GPS technology was measured in 2 ways. First, when community health paramedics received an alert indicating that a participant had entered a geofence at an area hospital, a community health paramedic attempted to make contact with the participant within 2 business days. If contact was established, the community health paramedic confirmed the visit to the ED and hospital, as indicated by the geofence alert. Second, at the end of the study, the research staff obtained use records from the HIE. These records provided the dates of participants' ED and hospital visits during the study period. Use records for the 25 participants for whom HIE data were collected were triangulated with geofence entry notifications to measure the accuracy of the GPS technology.

Depression

The 9-item PHQ-9 was used to establish participant eligibility and as a baseline measure for depression symptoms. The PHQ-9 is a reliable and valid tool for diagnosing and grading depressive symptom severity [31]. Each item is scored from 0-3 and then summed. Scores of 5, 10, 15, and 20 represent cutoff points for mild, moderate, moderately severe, and severe depression, respectively [31]. To be eligible to participate in this study, individuals were required to score at least 10, indicating moderate depression.

Medication Adherence

Medication adherence was measured using a modified version of the Adherence Starts with Knowledge-12 (ASK-12). The ASK-12 is a brief, 12-item scale with 3 subscales that measure medication behavior, health beliefs, and inconvenience/forgetfulness [32]. For this study, we modified the subset of 5 questions assessing medication behavior into dichotomous yes/no response options to assess medication adherence during the preceding month. The number of yes responses was counted and summed for a medication behavior subscale score. Scores on the full ASK-12, with the modified medication behavior subscale, ranged from 12-40, with higher scores indicating greater barriers to adherence. At baseline, the full scale with the modified behavior subscale was used. At monthly visits and exit, only the modified medication behavior subscale was used as it was unlikely that medication beliefs would change within the short time frame of this study.

Social Support

Social support was measured using the Medical Outcomes Study Social Support Survey, a valid and reliable tool that has been used in multiple groups across various conditions [33]. It includes 19 questions yielding four subscales—emotional/informational support, tangible support, affectionate support, and positive social interaction. Each item is rated using a Likert scale ranging from 1 (none of the time) to 5 (all of the time). The total score was calculated by summing all 19 questions and averaging them. Higher scores represent greater levels of social support.

Experience With and Acceptance of Technology

At baseline, experience with mobile phone technology was measured using a series of questions asking about current mobile phone ownership, mobile phone service, length of time owning a mobile phone, ability to charge the mobile phone, and if the participant had had a mobile phone stolen before. Acceptance of technology was measured at baseline and exit. At baseline, acceptance of technology was measured using a modification of the Technology Acceptance Questionnaire [34]. At baseline, 17 items were used, and at exit, a subset of these items, as well as an additional 8 items, were used. Each item is rated using a Likert scale ranging from 1=strongly disagree to 5=strongly agree. Higher scores indicate greater acceptance of technology. In addition, at exit, participants were asked how often they were able to charge their smartphone with options ranging from "None of the time" to "Always."

Quality of Care Transitions

Self-reported ED and hospital use were assessed at baseline, monthly visits, and exit. If participants indicated that they had visited the ED or hospital within the past month, their experience and perception of patient-centeredness of their care were assessed using the care transitions measure (CTM), a 15-item measure reflecting 4 content domains [35]. The domains include critical understanding, important preferences, management preparation, and care plans [35]. Participants used a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) to rate the quality of various components of a care transition within each domain. Lower scores indicate a poorer quality transition, and higher scores indicate a better transition. The CTM was administered at each monthly visit, during which a participant reported an ED or hospital visit. If someone reported more than 1 visit in the previous month, the CTM was completed only for the most recent visit.

Data Analysis

All statistical analyses were performed using IBM SPSS Statistics for Windows (version 25.0, IBM Corp). Descriptive statistics were used to describe the sociodemographic and health characteristics of the sample and all study measures. Accuracy of the geofence entry notifications was determined by calculating the percentage of notifications that aligned with HIE use data. Paired sample *t* tests were used to compare scores at baseline and exit on the PHQ-9, ASK-12, Medical Outcomes Study, and technology acceptance scales.

Qualitative content analysis was used to identify participants' acceptance of the intervention. All interviews were audio recorded and transcribed verbatim to facilitate coding and analysis. After a thorough reading and deductive coding of 5 representative transcripts by 2 members of the study team (LRM and WT), a consensus meeting was held to discuss and agree upon the codes. Discrepancies were resolved by discussing the context for each phrase being analyzed. After the meeting, a codebook was developed. The remaining interviews were divided between the 2 authors and coded separately. After coding was complete, the study team organized the codes into categories.

Results

Overview

Between October 2018 and April 2019, research staff screened 39 individuals for participation; of the 39 individuals, 32 (82%) met the eligibility criteria, and 30 (77%) were enrolled in the study. The 2 individuals who were eligible to participate but did not enroll did not return for the subsequent enrollment visit in the study after screening. The reasons for ineligibility for the study were not scoring at least 10 on the PHQ-9 (2/39, 5%), not having been to the ED or hospital at least twice in the past 2 months (2/39, 5%), not being prescribed a medication (2/39, 5%), and endorsing suicidal ideation (1/39, 3%). The participant who endorsed suicidal ideation was referred to the public safety officer on site at the community entity for appropriate follow-up and mental health services. Of the 30 participants, 10 (33%) were screened and enrolled at the first church with 2 weekly breakfasts, and the remaining 20 (67%) were screened and enrolled at the navigation center housed in a church.

Of the 30 participants enrolled, 19 (63%) completed the 4-month intervention, with a completion rate of 63%. Of the 11 participants who did not complete the intervention, 6 (55%) were withdrawn from the study after they reported their second smartphone lost or stolen, 2 (18%) notified the research staff that they were moving to a different town, 2 (18%) were lost to follow-up, and 1 (9%) voluntarily withdrew from the study after losing his first smartphone. Of these 11 participants, 4 (36%) completed all but the exit data collection.

Quantitative Results

Participant Demographics and Health-Related Characteristics

Participants comprised 30 people experiencing homelessness. On average, participants were male (20/30, 67%), aged 44.1 years (SD 9.7 years), White (17/30, 57%), never married (17/30, 57%), and not working because of disability or other medical reasons (19/30, 63%). At baseline, participants reported a mean of 2.8 (SD 1.4) chronic conditions, and most (26/30, 87%) experienced multiple chronic conditions. All participants were prescribed at least 2 medications at baseline; 53% (16/30) were prescribed 4 or more medications. Tables 1 and 2 provide a summary of demographic and health-related characteristics.

Table 1. Summary of demographic information (N=30).

Variables	Values
Age (years)	
Mean (SD)	44.1 (9.7)
Median	46
Gender, n (%)	
Male	20 (67)
Female	8 (27)
Transgender female	1 (3)
Other	1 (3)
Self-reported race or ethnicity, n (%)	
White	17 (57)
Black or African American	7 (23)
Hispanic	2 (7)
Native American	1 (3)
Other	3 (10)
Marital status, n (%)	
Married or domestic partnership	4 (13)
Divorced	9 (30)
Single or never married	17 (57)
Children, n (%)	
Yes	18 (60)
Number of children for those with ≥ 1 child	
Mean (SD)	2.9 (1.6)
Median	2
Highest level of education, n (%)	
Less than high school	8 (27)
High school graduate or GED[a]	12 (40)
Trade, technical, or vocational training	4 (13)
Some college	5 (17)
Other	1 (3)
Military veteran, n (%)	
Yes	2 (7)
Employment status[b], n (%)	
Not employed	24 (83)
Employed	5 (17)
Reason if unemployed[b,c], n (%)	
Looking for work	6 (23)
Laid off	2 (8)
Disabled or medical reason	19 (73)
Other	3 (12)
Annual income (US $), n (%)	
0-10,000	27 (90)

Variables	Values
10,001-20,000	3 (10)
Slept most nights[c], n (%)	
On the street	20 (67)
In a shelter	2 (7)
Other	9 (30)
Length of homelessness (years)	
Mean (SD)	8.1 (7.7)
Median	5

[a]GED: general educational development.

[b]Data were missing for some participants.

[c]Respondents may have chosen more than one response.

Table 2. Summary of baseline health information (N=30)[a].

Variables	Values
Number of chronic conditions	
Mean (SD)	2.8 (1.4)
Hypertension, n (%)	19 (63)
Diabetes mellitus, n (%)	5 (17)
High cholesterol, n (%)	9 (30)
Asthma, n (%)	12 (40)
Chronic obstructive pulmonary disease, n (%)	11 (37)
Congestive heart failure, n (%)	2 (7)
Number of prescribed medications, n (%)	
2-3	14 (47)
4-5	12 (40)
≥6	4 (13)
Self-reported number of ED[b] visits or hospitalizations in past 6 months, n (%)	
2	18 (60)
3	6 (20)
4	3 (10)
≥5	3 (10)
Visited ED in past 30 days (self-report), n (%)	
Yes	17 (57)
Visited hospital in past 30 days (self-report), n (%)	
Yes	7 (23)
Functional limitations, n (%)	
Deaf or difficulty hearing (yes)	8 (27)
Blind or difficulty seeing when wearing glasses (yes)	11 (37)
Difficulty walking or climbing stairs (yes)	15 (50)
Difficulty dressing or bathing (yes)	5 (17)
Number of functional limitations, n (%)	
1	8 (27)
2	6 (20)
≥3	6 (20)
CAGE[c] substance abuse screening score, n (%)	
≥2	10 (33)
Drug use in past year, n (%)	
Yes	16 (53)
Health literacy level	
Mean (SD)	13.7 (5.2)
Median	14.5
Limited, n (%)	13 (43)
Marginal, n (%)	5 (17)
Adequate, n (%)	12 (40)

[a]Percentages are out of 30 and more than one response was allowed per respondent.

[b]ED: emergency department.

[c]CAGE: Cut down, Annoyed, Guilty, and Eye-opener.

Accuracy of the GPS Technology

Accuracy of the GPS technology was calculated for the 25 participants who completed a release of information form, giving permission for the research team to access data in the HIE. During the 4-month study period, HIE use data indicated that these participants made 16 hospital or ED visits. Community health paramedics received 14 total geofence entry notifications during the study period; of these 14 notifications, 2 (14%) were from participants without a release of information for the HIE. Thus, community health paramedics received 12 geofence entry notifications for the 25 participants from whom HIE data were available. However, only 3 of the geofence entry notifications were consistent with the HIE use data for an overall accuracy rate of 19%.

Of the 16 ED and hospital visits reported by the HIE data for which community health paramedics did not receive geofence entry notifications, 4 (25%) occurred during the first month of the intervention, a time during which the research staff identified a technical issue with the mobile app and geofence entries were not being received. Of these 16 visits, 3 (19%) occurred in the window of time during which the participant was without the study-assigned smartphone as the smartphone had been stolen or misplaced but not yet replaced. It is unclear why the remaining 43% (6/14) ED and hospital visits reported by the HIE data did not result in a geofence notification entry.

Community Health Paramedic Interventions

Community health paramedics successfully reached participants to conduct follow-up and provide care coordination assistance after 79% (11/14) of geofence notifications. Of these 11 contacts, 10 (91%) lasted ≤10 minutes, and 1 (9%) contact lasted between 11 and 20 minutes. Of these 11 contacts, 3 (27%) participants reported having accompanied a friend or family member to the ED and were not seen themselves, and 1 (9%) participant reported having visited the hospital campus for a scheduled medical visit. Thus, 36% (4/11) of these geofence notifications were classified as false positives. Of the remaining 7 contacts, 3 (43%) aligned with the HIE notification data. Of the remaining 4 contacts, 3 (75%) did not align with the HIE data, as the participants did not have a release of information form on the file. It is unclear why the remaining contact did not register with the HIE.

Community health paramedics determined that 43% (3/7) of the ED visits were emergent and likely unavoidable. Reasons for the emergent ED visits included chronic pulmonary obstructive disease exacerbation, a physical altercation at a local shelter, and uncontrolled epigastric pain. Reasons for the remaining ED visits were skin irritation because of scabies infection, shoulder pain, and 2 visits for gastrointestinal illness. Community health paramedics judged each of these 4 visits to be due to ambulatory care–sensitive conditions that could have been appropriately managed in the outpatient setting.

Depression

There was a significant difference in depressive symptoms between baseline (mean 16.9, SD 5.8) and exit (mean 12.7, SD 8.2; t_{19}=2.892; P=.009), indicating fewer depressive symptoms at the 4-month exit.

Medication Adherence

At baseline, scores on the ASK-12 ranged from 14-30 (mean 20.5, SD 4.4). Among those who completed the 4-month intervention, there was a significant difference in medication behavior between baseline (mean 2.4, SD 1.4) and exit (mean 1.5, SD 1.5; t_{17}=2.47; P=.03), indicating that at the 4-month exit visit, there were fewer barriers to taking medications.

Social Support

There was no significant difference in social support between baseline (mean 3.2, SD 1.1) and exit (mean 2.9, SD 1.3; t_{18}=1.25; P=.23).

Experience With and Acceptance of Technology

At baseline, 50% (15/30) of participants reported having a mobile phone. Of these 15 patients, 12 (80%) had current wireless service (4/12, 33% participants had pay as you go service plans; 3/12, 25% had prepaid plans, 3/12, 25% had month-to-month contracts; and 2/12, 17% had free minutes through government-funded plans). Of the 15 participants with mobile phones, 13 (87%) reported that their mobile phones could support both SMS text messaging and mobile apps. At the exit interview, participants agreed that the smartphone app was easy to use (mean 4.4, SD 1.0), that they had the knowledge to use the smartphone app (mean 4.6, SD 0.5), and that they planned to continue using both a smartphone (mean 4.5, SD 0.6) and GPS technology (mean 4.4, SD 0.5). The acceptance of technology questionnaire indicated that participants had a high level of agreement at baseline and exit with items such as having the resources and knowledge to use smartphone technology and being comfortable with the health care team being alerted about ED or hospital use. There was a significant increase in agreement level from baseline (mean 3.9, SD 0.8) to exit (mean 4.4, SD 0.5) for the item, “My friends would encourage me to use this Smartphone app.” Participants’ agreement level increased for several other items, such as having the knowledge and resources to use GPS technology from baseline to exit, but not significantly. Table 3 summarizes the participants’ technology acceptance at baseline and the 4-month exit interview.

Table 3. Perceptions of acceptance of technology at baseline and 4-month exit interview.

Item	Baseline, mean (SD)	Exit, mean (SD)	*P* value
I have the resources necessary to use smartphone technology	4.4 (0.5)	4.4 (1.0)	.86
I have the knowledge necessary to use smartphone technology	4.6 (0.4)	4.7 (0.5)	.28
I can get help from others when I have difficulties using smartphone technology	4.2 (0.7)	4.5 (0.8)	.28
I find GPS technology useful in my daily life	4.2 (0.8)	4.4 (1.0)	.39
I find GPS technology easy to use	4.1 (0.8)	4.3 (1.1)	.33
I have the resources necessary to use GPS technology	3.8 (1.2)	4.2 (1.3)	.29
I have the knowledge necessary to use GPS technology	4.1 (1.0)	4.6 (0.5)	.06
I am comfortable with my health data being stored online	3.6 (1.3)	4.1 (0.9)	.13
I believe my health information will be protected on a smartphone	3.7 (1.2)	3.6 (1.1)	.88
I am comfortable with my health care team being alerted when I go to the emergency department or hospital	4.8 (0.4)	4.7 (0.5)	.33
I think using GPS is a good way to notify my health care team when I visit the emergency department or hospital	4.5 (0.6)	4.7 (0.5)	.27
I think using this smartphone app can help me improve my overall health	4.3 (0.6)	4.5 (0.5)	.16
My friends would encourage me to use this smartphone app	3.9 (0.8)	4.4 (0.5)	.04
My family members would encourage me to use this smartphone app	4.0 (1.0)	4.1 (1.0)	.85

Quality of Care Transitions

At baseline, 57% (17/30) of participants self-reported at least one ED or hospital visit in the previous 30 days and completed the CTM-15. At months 1, 2, 3, and exit, 33% (10/30), 13% (4/30), 13% (4/30), and 17% (5/30) of participants, respectively, self-reported at least 1 ED or hospital visit in the previous 30 days and completed the CTM-15 for their most recent visit. The mean score for the critical understanding and management preparation domains was 4.1, indicating that participants generally agreed that they left the hospital or ED understanding how to manage medications and their health. The mean score for the preferences important domain was 4.0 (SD 0.1), which means that participants agreed that hospital staff took their preferences for health care needs into account when planning for discharge. The lowest level of agreement was with the care plan domain (mean 3.8, SD 0.0). Table 4 provides a summary of the scores for each item and domain.

Table 4. Summary of perceptions of the quality-of-care transitions using the care transitions measure (N=40 hospital or ED[a] visits).

Domains and items	Mean (SD)[b]
Critical understanding	
When I left the hospital or ED, I clearly understood the purpose for taking each of my medications.	4.1 (0.8)
When I left the hospital or ED, I clearly understood how to take each of my medications, including how much I should take and when.	4.3 (0.7)
When I left the hospital or ED, I clearly understood the possible side effects of each of my medications.	4.0 (1.1)
When I left the hospital or ED, I had a good understanding of the things I was responsible for in managing my health.	4.2 (0.8)
When I left the hospital or ED, I was confident that I knew what to do to manage my health.	3.9 (1.0)
When I left the hospital or ED, I was confident I could actually do the things I needed to do to take care of my health.	3.8 (1.1)
Domain overall mean	4.1 (0.2)
Preferences important	
Before I left the hospital or ED, the staff and I agreed about clear health goals for me and how those would be reached.	3.9 (1.2)
The hospital staff took my preferences into account in deciding what my health care needs would be when I left the hospital or ED.	4.1 (1.0)
The hospital staff took my preferences into account in deciding where my health care needs would be met when I left the hospital or ED.	4.0 (1.1)
Domain overall mean	4.0 (0.1)
Management preparation	
When I left the hospital or ED, I had all the information I needed to be able to take care of myself.	4.0 (0.9)
When I left the hospital or ED, I clearly understood how to manage my health.	3.9 (0.9)
When I left the hospital or ED, I clearly understood the warning signs and symptoms I should watch for to monitor my health condition.	4.2 (0.8)
When I left the hospital or ED, I had a good understanding of my health condition and what makes it better or worse.	4.1 (0.9)
Domain overall mean	4.1 (0.1)
Care plan	
When I left the hospital or ED, I had a readable and easily understood written list of appointments I needed to complete within the next several weeks.	3.8 (1.1)
When I left the hospital or ED, I had a readable and easily understood written plan that described how all of my health care needs were going to be met.	3.8 (1.2)
Domain overall mean	3.8 (0.0)

[a]ED: emergency department.

[b]Participants indicated their level of agreement with each item using a Likert scale from 1=strongly disagree to 5=strongly agree.

Qualitative Findings

Of the 30 participants, 17 (57%) completed an exit interview. During data analysis, the first 2 authors of this study organized the codes into the following categories: (1) benefits of study participation, (2) challenges to study participation, (3) perceptions of GPS technology, and (4) suggestions for improvement.

Overall, participants reported positive experiences with study participation. They also identified several benefits, defined as any real or perceived aid or assistance from participating in the research study or having access to the unlimited use of a smartphone. Benefits included self-management support, improved social connections, and improved well-being. An example of how study participation provided self-management support is demonstrated by this quote:

> *[...] there was a time when I [...] would be confused as to whether or not I took my medicine. Sometimes I would go days without even thinking about it, you know? But now, I am confident knowing that every morning you know "Bam!", you know it's [daily email] right there and I had my medication and had taken it. There was never any more confusion.*

Social connections were facilitated by the ability to call friends and family, to stay up to date on current events by reading the news on the internet, and to use social media sites. Several participants described using the smartphone to reconnect with the family from out of state. One participant put it succinctly as follows:

> *[...] being homeless, you can be very bored sometimes with nothing to do. And, [with the phone] I had something to do. You can read the news and find out*

> *what's going on in the world. Or, you know, keep in touch with my friends with email.*

Participants also described improved well-being as they did not have to worry about paying for their smartphone, were able to travel to appointments because of bus pass on the smartphone, and felt more secure in their environments with the ability to contact the police or emergency medical services in the case of an emergency. An example of how study participation improved well-being is demonstrated by the following quote:

> *It was a godsend. It really was, I mean because I didn't have to worry about a lot of things. I could make phone calls when I needed to. It just took a lot of burden off me, knowing that I had a bus pass. I had a phone I could use you know if I got in trouble or something or was in a bad situation.*

Challenges to study participation were defined as circumstances in which participants had to navigate to access, use, and benefit from services and resources, including the research study itself. Challenges included differential treatment because of homelessness, difficulty with technology, and keeping the smartphone secure. For example, differential treatment resulted in participants having difficulty keeping their smartphones charged as business owners do not allow people experiencing homelessness to spend time charging smartphones in their establishments. Some participants also described trouble with technology, such as short battery life and slow internet service. Finally, keeping the smartphone secure required constant vigilance on the part of participants, and even with creative solutions for safekeeping, many experienced theft or damage to their smartphones. One participant expressed his desire for smartphones to be replaced up to 4 times, saying as follows:

> *[...] the fact is, anything can happen out here. Like you know, I was charging my phone at Starbucks. I fell asleep, and when I woke up, my phone was stolen. Got my second phone…but I forgot to put the case back on and water hits it and its out.*

Perceptions of the GPS technology were uniformly positive, as each participant who completed the exit interview denied having concerns about the community health paramedics or research staff knowing when they visited the ED or hospital. One participant clearly articulated this by saying the following:

> *...you know, that kind of thing right now is the least of my concerns. If you're sleeping in an alley or somewhere else, you're not really worried about somebody knowing that you've been to the hospital, or at least I'm not.*

Suggestions for improvement included two main subcategories: helping to complete the study requirements and tailoring the intervention. Participants suggested more teaching about using the smartphone and its functions and providing portable battery chargers to help overcome some of the technical challenges to study completion that participants faced. Participants also suggested sending daily messages via text instead of email and indicated that personalized and tailored medication reminders for their individual medication regimens would be helpful.

Discussion

Principal Findings

The results of this study contribute to a small but growing body of literature documenting the utility of mHealth interventions among people experiencing homelessness. First, our findings suggest that GPS technology is not a reliable method for tracking visits to the ED or hospital among people experiencing homelessness. The geofence notifications aligned with objective HIE use data only 18.8% of the time, indicating that the community health paramedics were unable to connect with participants to provide follow-up assistance with care coordination after most participant ED and hospital visits. This finding was surprising given recent evidence that a smartphone app used by 12 patients with low income had 75% accuracy in detecting real-time ED or hospital use over a 3-month period [36]. It is likely that the results of this study are inconsistent with this prior evidence because of variations in the real-world use of smartphones among a population without consistent access to electricity. Specifically, a strategy that participants used to preserve the smartphone battery was to power the smartphone off when it was not in use. As geofence technology relies on real-time transmission of data, it is likely that one reason entry notifications were not received was as the smartphone was turned off when the geofence entry occurred.

Despite findings that GPS technology is not reliable for real-time ED or hospital use data, overall, participants expressed positive views of GPS technology. Participants embraced the idea of GPS being used by health care and other service providers to locate them if needed and described feeling more secure with the knowledge they could be found. This is similar to findings by Liss et al [9], who found that high-risk primary care patients were willing to use GPS technology to facilitate care coordination. Findings by Liss et al [9] also align with prior work by Moczygemba et al [37], which indicate that clinicians and care managers are particularly interested in using mHealth for care coordination among high-risk patients and patients experiencing homelessness [9,13]. This is particularly important as community health paramedics indicated that 57% (4/7) of ED or hospital visits were likely nonemergent visits that could have been addressed in the outpatient setting. Collectively, these findings suggest the need for app development and refinement as the GPS location tracking apps that are currently in the market do not have face validity or the specific functionality needed for use in the health care setting.

There was a significant decrease in depression symptoms from baseline to exit, which aligns with the qualitative findings where participants reported improved well-being and an overall positive experience with the intervention at the study exit. In contrast, a 1-month, pre-post study of homeless young adults (aged 18-24 years) who participated in a remote mental health intervention, which included SMS text messaging, did not find a difference in depression symptoms [38]. This may be as it takes longer than 1 month to see a difference in depression symptoms, although this finding warrants further study. The results also indicate an improvement in medication adherence as measured by the ASK-12. These findings support the findings

of Morawski et al [39], in which the use of a smartphone app resulted in improved medication adherence among patients with hypertension. Participants in this study viewed the daily email question regarding medication adherence as a helpful reminder that supported adherence. The data also suggest that participants used their smartphones as a self-management support tool by downloading specific medication adherence apps or by setting alarms to help with medication management. This use of the smartphone as a tool is also evidenced by overall high scores regarding acceptance of technology at baseline and exit.

Although there was no significant difference in social support between baseline and exit, the qualitative data suggest that the smartphone had an impact on participants' social connections. Prior evidence clearly indicates that social support can have a protective influence on multiple health outcomes among people experiencing homelessness [40] and that mobile phones are critical for individuals experiencing homelessness to maintain social connectedness to family and friends [41]. Thus, measuring social support in future studies investigating mHealth interventions among people experiencing homelessness is important for ascertaining a holistic picture of the benefits of smartphone technology among the homeless population.

Overall, the participants rated care transitions from the ED or hospital to the community fairly high. However, the results suggest that specific aspects of transitions could be improved. For example, in the critical understanding domain, two items related to understanding what and how to manage health on discharge and one item related to medication side effects scored lower than the remaining domain items. Future studies could investigate adapting the mHealth intervention to provide targeted follow-up post-ED or hospital discharge as well as specific guidance related to medication side effects to maximize adherence and optimize outcomes. Furthermore, the care plan domain scored the lowest among the four domains. This further supports the need to adapt the intervention to provide two-way communication between people experiencing homelessness and service providers to ensure that needed follow-up care is received in a timely and accessible manner.

Study Limitations

The findings of this study should be interpreted with caution because of the study's limitations. Participants were recruited from one city in a large, southern state using convenience sampling; therefore, the generalizability of the findings is unknown. Furthermore, participants were recruited from community sites, which may have biased the results. The pre-post design is subject to bias, and as study participants were selected on the basis of their PHQ-9 score, it is possible that regression to the mean occurred for the depression symptom outcome. There were also baseline differences in PHQ-9 scores between the groups that did and did not complete the study (t_{21}=–2.17; P=.02) with the group that did not complete the study having a higher mean score at baseline than the group that did finish. The small sample size, although sufficient for answering this study's research questions, may further limit the generalizability of the findings.

Future Directions

The findings from this study point to several directions for future research. First, based on participants' responses to the daily email medication adherence message and their stated preferences for SMS text messages, a subsequent study tested an expanded SMS text messaging intervention. That study also included testing the use of remote location services preinstalled on the smartphone to locate participants during business hours. The findings also suggest that in addition to unlimited access to a smartphone, access to unlimited transportation can facilitate the ability of people experiencing homelessness to self-manage chronic illness. Thus, future research could investigate the impact of providing accessible transportation on health outcomes and use. Finally, because of the shortcomings of GPS technology in communicating real-time health care use information for people experiencing homelessness and as there is an operational HIE in the local area, future research investigating care coordination should incorporate the HIE to ensure transmission of objective use data. Qualitative findings also suggest that mHealth interventions, particularly unlimited access to a smartphone and bus pass transportation, have numerous benefits for well-being and the ability of people experiencing homelessness to meet social needs. These concepts need to be explored quantitatively in future studies. Furthermore, coupling access to a smartphone and transportation with health care programs should be pursued at a policy level for local programs [42,43].

Conclusions

mHealth interventions are acceptable to people experiencing homelessness and positively affected depression symptoms and medication adherence. Objective data from the HIE provided more accurate ED and hospital use information compared with alerts relying on predefined geofences. Despite this, participants favorably viewed GPS technology, warranting further exploration of GPS technology as a tool for facilitating care coordination among people experiencing homelessness.

Acknowledgments

The authors gratefully acknowledge the partners who made this study possible. This includes the Sunrise Community Navigation Center, First United Methodist Church of Austin, Austin-Travis County Paramedics, and Integrated Care Collaborative. The authors also thank Azova, a digital platform for this study.

Conflicts of Interest

None declared.

References

1. Demographics of mobile device ownership adoption in the United States. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-05-20]
2. Kwon L, Long K, Wan Y, Yu H, Cunningham B. Medical diagnostics with mobile devices: comparison of intrinsic and extrinsic sensing. Biotechnol Adv 2016;34(3):291-304. [doi: 10.1016/j.biotechadv.2016.02.010] [Medline: 26952640]
3. Logan AG, McIsaac WJ, Tisler A, Irvine MJ, Saunders A, Dunai A, et al. Mobile phone-based remote patient monitoring system for management of hypertension in diabetic patients. Am J Hypertens 2007 Sep;20(9):942-948. [doi: 10.1016/j.amjhyper.2007.03.020] [Medline: 17765133]
4. Vidrine D, Arduino R, Lazev A, Gritz E. A randomized trial of a proactive cellular telephone intervention for smokers living with HIV/AIDS. AIDS 2006 Jan 09;20(2):253-260. [doi: 10.1097/01.aids.0000198094.23691.58] [Medline: 16511419]
5. Dorsch MP, An LC, Hummel SL. A novel just-in-time contextual mobile app intervention to reduce sodium intake in hypertension: protocol and rationale for a randomized controlled trial (LowSalt4Life Trial). JMIR Res Protoc 2018 Dec 07;7(12):e11282 [FREE Full text] [doi: 10.2196/11282] [Medline: 30530462]
6. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
7. Guy R, Hocking J, Wand H, Stott S, Ali H, Kaldor J. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. Health Serv Res 2012 Apr;47(2):614-632 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01342.x] [Medline: 22091980]
8. Stockwell MS, Kharbanda EO, Martinez RA, Vargas CY, Vawdrey DK, Camargo S. Effect of a text messaging intervention on influenza vaccination in an urban, low-income pediatric and adolescent population: a randomized controlled trial. JAMA 2012 Apr 25;307(16):1702-1708. [doi: 10.1001/jama.2012.502] [Medline: 22535855]
9. Liss D, Serrano E, Wakeman J, Nowicki C, Buchanan DR, Cesan A, et al. "The Doctor Needs to Know": acceptability of smartphone location tracking for care coordination. JMIR Mhealth Uhealth 2018 May 04;6(5):e112 [FREE Full text] [doi: 10.2196/mhealth.9726] [Medline: 29728349]
10. Nguyen KT, Olgin JE, Pletcher MJ, Ng M, Kaye L, Moturu S, et al. Smartphone-based geofencing to ascertain hospitalizations. Circ Cardiovasc Qual Outcomes 2017 Mar;10(3):e003326 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.116.003326] [Medline: 28325751]
11. 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]
12. Patrick K, Griswold WG, Raab F, Intille SS. Health and the mobile phone. Am J Prev Med 2008 Aug;35(2):177-181 [FREE Full text] [doi: 10.1016/j.amepre.2008.05.001] [Medline: 18550322]
13. Moczygemba LR, Cox LS, Marks SA, Robinson MA, Goode JR, Jafari N. Homeless patients' perceptions about using cell phones to manage medications and attend appointments. Int J Pharm Pract 2017 Jun;25(3):220-230 [FREE Full text] [doi: 10.1111/ijpp.12321] [Medline: 27896909]
14. Burda C, Haack M, Duarte A, Alemi F. Medication adherence among homeless patients: a pilot study of cell phone effectiveness. J Am Acad Nurse Pract 2012 Nov;24(11):675-681. [doi: 10.1111/j.1745-7599.2012.00756.x] [Medline: 23088699]
15. Gelberg L, Browner CH, Lejano E, Arangua L. Access to women's health care: a qualitative study of barriers perceived by homeless women. Women Health 2004;40(2):87-100. [doi: 10.1300/J013v40n02_06] [Medline: 15778140]
16. Khandor E, Mason K, Chambers C, Rossiter K, Cowan L, Hwang SW. Access to primary health care among homeless adults in Toronto, Canada: results from the Street Health survey. Open Med 2011;5(2):e94-103 [FREE Full text] [Medline: 21915240]
17. Skosireva A, O'Campo P, Zerger S, Chambers C, Gapka S, Stergiopoulos V. Different faces of discrimination: perceived discrimination among homeless adults with mental illness in healthcare settings. BMC Health Serv Res 2014 Sep 07;14:376 [FREE Full text] [doi: 10.1186/1472-6963-14-376] [Medline: 25196184]
18. Lamanna D, Stergiopoulos V, Durbin J, O'Campo P, Poremski D, Tepper J. Promoting continuity of care for homeless adults with unmet health needs: the role of brief interventions. Health Soc Care Community 2018 Jan;26(1):56-64. [doi: 10.1111/hsc.12461] [Medline: 28569397]
19. Tomita A, Herman DB. The impact of critical time intervention in reducing psychiatric rehospitalization after hospital discharge. Psychiatr Serv 2012 Sep 01;63(9):935-937 [FREE Full text] [doi: 10.1176/appi.ps.201100468] [Medline: 22810163]
20. Brubaker A, Sasser M. Incorporating Community Health Paramedics into Your Homeless Healthcare Program. 2018 Healthier Texas Summit. URL: https://healthiertexassummit.com/wp-content/uploads/2018/10/Advancing-Homeless-Healthcare-with-Community-Health-Paramedics.pdf [accessed 2021-09-09]
21. Thurman WA, Moczygemba LR, Tormey K, Hudzik A, Welton-Arndt L, Okoh C. A scoping review of community paramedicine: evidence and implications for interprofessional practice. J Interprof Care 2021;35(2):229-239. [doi: 10.1080/13561820.2020.1732312] [Medline: 32233898]

22. Davis MM, Devoe M, Kansagara D, Nicolaidis C, Englander H. "Did I do as best as the system would let me?" Healthcare professional views on hospital to home care transitions. J Gen Intern Med 2012 Dec;27(12):1649-1656 [FREE Full text] [doi: 10.1007/s11606-012-2169-3] [Medline: 22829294]
23. Alafaireet P, Houghton H, Petroski G, Gong Y, Savage G. Toward determining the structure of psychiatric visit nonadherence. J Ambul Care Manage 2010;33(2):108-116. [doi: 10.1097/JAC.0b013e3181d91649] [Medline: 20228633]
24. Naughton F, Hopewell S, Lathia N, Schalbroeck R, Brown C, Mascolo C, et al. A context-sensing mobile phone app (Q Sense) for smoking cessation: a mixed-methods study. JMIR Mhealth Uhealth 2016 Sep 16;4(3):e106 [FREE Full text] [doi: 10.2196/mhealth.5787] [Medline: 27637405]
25. Pramana G, Parmanto B, Lomas J, Lindhiem O, Kendall PC, Silk J. Using mobile health gamification to facilitate cognitive behavioral therapy skills practice in child anxiety treatment: open clinical trial. JMIR Serious Games 2018 May 10;6(2):e9 [FREE Full text] [doi: 10.2196/games.8902] [Medline: 29748165]
26. American Community Survey (ACS). The United States Census Bureau. URL: https://www.census.gov/programs-surveys/acs [accessed 2020-07-27]
27. Ewing JA. Detecting alcoholism. The CAGE questionnaire. JAMA 1984 Oct 12;252(14):1905-1907. [doi: 10.1001/jama.252.14.1905] [Medline: 6471323]
28. Smith PC, Schmidt SM, Allensworth-Davies D, Saitz R. A single-question screening test for drug use in primary care. Arch Intern Med 2010 Jul 12;170(13):1155-1160 [FREE Full text] [doi: 10.1001/archinternmed.2010.140] [Medline: 20625025]
29. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]
30. Haun J, Luther S, Dodd V, Donaldson P. Measurement variation across health literacy assessments: implications for assessment selection in research and practice. J Health Commun 2012;17 Suppl 3:141-159. [doi: 10.1080/10810730.2012.712615] [Medline: 23030567]
31. 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] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
32. Matza LS, Park J, Coyne KS, Skinner EP, Malley KG, Wolever RQ. Derivation and validation of the ASK-12 adherence barrier survey. Ann Pharmacother 2009 Oct;43(10):1621-1630. [doi: 10.1345/aph.1M174] [Medline: 19776298]
33. Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med 1991;32(6):705-714. [doi: 10.1016/0277-9536(91)90150-b] [Medline: 2035047]
34. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
35. Coleman EA, Mahoney E, Parry C. Assessing the quality of preparation for posthospital care from the patient's perspective: the care transitions measure. Med Care 2005 Mar;43(3):246-255. [doi: 10.1097/00005650-200503000-00007] [Medline: 15725981]
36. Liss DT, Brown T, Wakeman J, Dunn S, Cesan A, Guzman A, et al. Development of a smartphone app for regional care coordination among high-risk, low-income patients. Telemed J E Health 2020 Nov;26(11):1391-1399. [doi: 10.1089/tmj.2019.0176] [Medline: 32040386]
37. Moczygemba L, Marks S, Cox L, Robinson M. Homeless care providers' perspectives about medication-related problems. 2013 American Pharmacists Association Annual Meeting. Los Angeles, CA 2013.
38. Schueller SM, Glover AC, Rufa AK, Dowdle CL, Gross GD, Karnik NS, et al. A mobile phone-based intervention to improve mental health among homeless young adults: pilot feasibility trial. JMIR Mhealth Uhealth 2019 Jul 02;7(7):e12347 [FREE Full text] [doi: 10.2196/12347] [Medline: 31267980]
39. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018 Jun 01;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
40. Hwang SW, Kirst MJ, Chiu S, Tolomiczenko G, Kiss A, Cowan L, et al. Multidimensional social support and the health of homeless individuals. J Urban Health 2009 Sep;86(5):791-803 [FREE Full text] [doi: 10.1007/s11524-009-9388-x] [Medline: 19629703]
41. Eyrich-Garg KM. Mobile phone technology: a new paradigm for the prevention, treatment, and research of the non-sheltered "street" homeless? J Urban Health 2010 May;87(3):365-380 [FREE Full text] [doi: 10.1007/s11524-010-9456-2] [Medline: 20397058]
42. Thurman W, Moczygemba LR, Hilbelink M. Smartphones and transportation as tools to meet the health and social needs of people experiencing homelessness. In: Proceedings of the 2020 Academy Health National Health Policy Conference. 2020 Presented at: 2020 Academy Health National Health Policy Conference; Feb 11, 2020; Washington, DC.
43. Raven MC, Kaplan LM, Rosenberg M, Tieu L, Guzman D, Kushel M. Mobile phone, computer, and internet use among older homeless adults: results from the HOPE HOME cohort study. JMIR Mhealth Uhealth 2018 Dec 10;6(12):e10049 [FREE Full text] [doi: 10.2196/10049] [Medline: 30530464]

Abbreviations

ASK-12: Adherence Starts with Knowledge-12
CTM: care transitions measure
ED: emergency department
HIE: health information exchange
mHealth: mobile health
PHQ-9: Patient Health Questionnaire-9

Edited by L Buis; submitted 06.11.20; peer-reviewed by T Powell-Wiley, E Green; comments to author 12.03.21; revised version received 27.05.21; accepted 06.08.21; published 03.11.21.

Please cite as:
Moczygemba LR, Thurman W, Tormey K, Hudzik A, Welton-Arndt L, Kim E
GPS Mobile Health Intervention Among People Experiencing Homelessness: Pre-Post Study
JMIR Mhealth Uhealth 2021;9(11):e25553
URL: https://mhealth.jmir.org/2021/11/e25553
doi:10.2196/25553
PMID:34730550

Original Paper

Gait and Axial Spondyloarthritis: Comparative Gait Analysis Study Using Foot-Worn Inertial Sensors

Julie Soulard[1,2], PT, PhD; Jacques Vaillant[1], PT, PhD; Athan Baillet[3], MD, Prof Dr, PhD; Philippe Gaudin[3], MD, Prof Dr, PhD; Nicolas Vuillerme[1,4,5], PhD

[1]University Grenoble Alpes, AGEIS, La Tronche, France

[2]Grenoble Alpes University Hospital, Grenoble, France

[3]University Grenoble Alpes, CNRS, Grenoble Alpes University Hospital, Grenoble INP, TIMC-IMAG UMR5525, Grenoble, France

[4]Institut Universitaire de France, Paris, France

[5]LabCom Telecom4Health, Orange Labs & Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP-UGA, Grenoble, France

Corresponding Author:
Julie Soulard, PT, PhD
University Grenoble Alpes
AGEIS
Faculty of Medicine, Jean Roget Building
Place du commandant Nal
La Tronche, 38700
France
Phone: 33 476637104
Email: juliesoulard.physio@gmail.com

Abstract

Background: Axial spondyloarthritis (axSpA) can lead to spinal mobility restrictions associated with restricted lower limb ranges of motion, thoracic kyphosis, spinopelvic ankylosis, or decrease in muscle strength. It is well known that these factors can have consequences on spatiotemporal gait parameters during walking. However, no study has assessed spatiotemporal gait parameters in patients with axSpA. Divergent results have been obtained in the studies assessing spatiotemporal gait parameters in ankylosing spondylitis, a subgroup of axSpA, which could be partly explained by self-reported pain intensity scores at time of assessment. Inertial measurement units (IMUs) are increasingly popular and may facilitate gait assessment in clinical practice.

Objective: This study compared spatiotemporal gait parameters assessed with foot-worn IMUs in patients with axSpA and matched healthy individuals without and with pain intensity score as a covariate.

Methods: A total of 30 patients with axSpA and 30 age- and sex-matched healthy controls performed a 10-m walk test at comfortable speed. Various spatiotemporal gait parameters were computed from foot-worn inertial sensors including gait speed in ms^{-1} (mean walking velocity), cadence in steps/minute (number of steps in a minute), stride length in m (distance between 2 consecutive footprints of the same foot on the ground), swing time in percentage (portion of the cycle during which the foot is in the air), stance time in percentage (portion of the cycle during which part of the foot touches the ground), and double support time in percentage (portion of the cycle where both feet touch the ground).

Results: Age, height, and weight were not significantly different between groups. Self-reported pain intensity was significantly higher in patients with axSpA than healthy controls (P<.001). Independent sample t tests indicated that patients with axSpA presented lower gait speed (P<.001) and cadence (P=.004), shorter stride length (P<.001) and swing time (P<.001), and longer double support time (P<.001) and stance time (P<.001) than healthy controls. When using pain intensity as a covariate, spatiotemporal gait parameters were still significant with patients with axSpA exhibiting lower gait speed (P<.001), shorter stride length (P=.001) and swing time (P<.001), and longer double support time (P<.001) and stance time (P<.001) than matched healthy controls. Interestingly, there were no longer statistically significant between-group differences observed for the cadence (P=.17).

Conclusions: Gait was significantly altered in patients with axSpA with reduced speed, cadence, stride length, and swing time and increased double support and stance time. Taken together, these changes in spatiotemporal gait parameters could be interpreted as the adoption of a so-called cautious gait pattern in patients with axSpA. Among factors that may influence gait in patients with

axSpA, patient self-reported pain intensity could play a role. Finally, IMUs allowed computation of spatiotemporal gait parameters and are usable to assess gait in patients with axSpA in clinical routine.

Trial Registration: ClinicalTrials.gov NCT03761212; https://clinicaltrials.gov/ct2/show/NCT03761212

International Registered Report Identifier (IRRID): RR2-10.1007/s00296-019-04396-4

(JMIR Mhealth Uhealth 2021;9(11):e27087) doi:10.2196/27087

KEYWORDS

ankylosing spondylitis; spondyloarthritis; gait; locomotion; pain; mobility; spatiotemporal; digital health; sensors

Introduction

The generic term spondyloarthritis (SpA) regroups part of chronic inflammatory diseases with common clinical, genetic, and pathophysiological features [1,2]. Diagnosis of SpA is based on the Assessment of Spondyloarthritis International Society (ASAS) criteria [3]. Two groups of SpA are defined: axial SpA (axSpA) with main manifestations being on spinal and sacroiliac joints, and peripheral SpA with main manifestations being arthritis, enthesitis or dactylitis [1,3]. In this study, we will focus on axSpA which is divided into its radiographic (ankylosing spondylitis, AS) and its nonradiographic (nr-axSpA) forms [1,2]. Note that patients with axSpA could represent 0.13% to 1.4% of the world population [4].

Clinical manifestation of axSpA includes chronic inflammatory back pain and morning stiffness [5]. As a consequence of inflammation, structural damage can occur and lead to spinal mobility restrictions [5] associated with restricted lower limb ranges of motion [6], thoracic kyphosis [7], spinopelvic ankylosis [7], decrease in muscle strength [8], and sarcopenia [8]. It is well known that factors such as limited range of motion [9], reduced muscle strength [10], sarcopenia [11], thoracic kyphosis, and spinopelvic alignment [12] can have consequences on spatiotemporal gait parameters during walking.

It is interesting to note that a recent review concluded that no published work has investigated spatiotemporal gait parameters in patients with axSpA [13]. This is not the case for patients with AS, a subgroup of axSpA [3]. A recent review [14] reported that 21 articles assessed gait in AS. Interestingly, only 4 of them (19%) used a healthy control group for comparison of spatiotemporal gait parameters [15-18]. What is more, results of these 4 studies are rather mixed and have reached somewhat inconsistent results and raised unanswered questions [15-18]. Some studies, indeed, reported gait impairment in patients with AS who presented with lower gait speed [18] and lower stride length [16,18] than healthy controls. Other studies reported gait speed [15,17], stride length [15,17], cadence [16,17], swing time, and stance time percentages [18] of patients with AS similar to those of healthy controls. How can we explain these observed differences? It is possible that the relatively small sample size of these studies (from n=10 [17] to n=18 [18] in each group) represented an obstacle to the identification of any significant group differences. Note that this limitation is that of the authors themselves (“However, further study should be performed on a larger sample subjects” [15] and “the sample size was limited” [18]). It is also possible that self-reported pain intensity at the time of assessment played a role in these divergent results. On the one hand, it is recognized that low back pain is one of the main symptoms of axSpA [5] and inflammatory back pain is a central criteria for disease diagnosis [3]. On the other hand, it is also well established that low back pain could significantly affect spatiotemporal gait parameters during walking [19-21]. For instance, previous studies have reported significant differences in spatiotemporal gait parameters between patients with low back pain and healthy matched controls [19,20]. Patients with low back pain presented lower gait speed [19,20] and cadence [20] and shorter stride length [19,20] than healthy matched controls during walking. It is important to mention that self-reported pain intensity at time of evaluation was not reported in all studies on gait and AS. In particular, only studies from Mangone et al [17] and Zhang et al [17] have reported this parameter. Regardless of this, a careful examination and comparative analysis of these two published works [17,18] nevertheless has drawn our attention to more specifically take into consideration the possible impact of pain on spatiotemporal gait parameters during walking. To support this view, let us first consider the work of Mangone et al [17]. Analysis of spatiotemporal gait showed no significant between-group difference for gait speed (AS: 0.94 [SD 0.2] ms^{-1} vs healthy controls: 0.96 [SD 0.2] ms^{-1}, *P*=.78) and stride length (AS: 1.09 [SD 0.1] m vs healthy controls: 1.14 [SD 0.2] m, *P*=.40) [17]. Concomitantly, no between-group difference was observed for self-reported pain intensity reported at time of evaluation assessed with the visual analog scale (VAS-AS: 1.0 [SD 1.3] versus healthy controls: 0.7 [SD 1.1]) [17]. Worthy of note also are the very low self-reported pain intensity scores of close to 0. A value of 0 on the VAS is considered as no pain while a value above 3 is considered as moderate pain [22]. In other words, participants of Mangone et al study [17] could thus be considered as pain-free participants.

Unlike the findings of Mangone et al [17], analysis of data from Zhang et al [18] revealed between-group significant difference in spatiotemporal gait. Lower gait speed (AS: 1.15 [SD 0.21] ms^{-1} vs healthy controls: 1.25 [SD 0.09] ms^{-1}, *P*=.009) and shorter stride length (stride length/height: AS: 0.70 [SD 0.97] m/m vs healthy controls: 0.76 [SD 0.42] m/m, *P*=.002) were observed in patients with AS (n=18) than in healthy controls (n=18) [18]. Meanwhile, pain intensity scores reported with the VAS in patients with AS only [18] were 3.89 [SD 1.64]. This value is above 3 and hence considered as moderate pain [22]. This self-reported pain intensity score is 3 times higher than that reported by patients with AS involved in the study of Mangone et al [17]. Although self-reported pain intensity was not collected in healthy controls, it is probable that the value for healthy controls would have been close to 0 like in Mangone

et al [17] study. Taken together, the two studies have prompted us to consider that inconsistent and inconclusive results on gait in AS could stem from self-reported pain intensity at the time of the evaluation per se. Moreover, inertial measurement units (IMUs) are becoming helpful to assess gait in different populations [23,24]. IMUs allow computation of spatiotemporal gait parameters in clinical practice that are reliable in patients with axSpA. While previous studies demonstrated the advantages of using IMUs in axSpA to assess spinal mobility [25] or level of physical activity [26], no study assessed gait parameters using IMUs in patients with axSpA.

Overall, because of the lack of published works available on gait in patients with axSpA [13] and considering the divergent results obtained in the studies that have assessed spatiotemporal gait parameters in AS [14], which could be partly explained by self-reported pain intensity scores, this study was designed to compare spatiotemporal gait parameters in patients with axSpA and matched healthy individuals without and with pain intensity score as a covariate.

Methods

Study Design

The Function, Locomotion, Measurement, Inflammation (FOLOMI) study was approved by local ethics committee (CPP Ile De France 1, RCB: 2017-A03468-45) and registered with ClinicalTrials.gov [NCT03761212] and followed the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) checklist. All participants of the study signed informed consent.

Participants

The sample size of this study was calculated using difference between patients with AS and healthy controls in stride length in the Zebouni et al [16] study with a standard deviation of 0.12, expected difference of 0.14, significance level of 0.05, and power of 80%. The sample size was estimated at 12 in each group using a sample size calculator [27,28]. It was increased to 30 to allow the use of parametric tests.

Data for this cross-sectional study are a subset of individuals recruited in the FOLOMI prospective study that has been described in a previous publication [29]. The first 30 patients with axSpA included in FOLOMI study and 30 age- and sex-matched healthy controls were studied in this work. Inclusion and noninclusion criteria of the FOLOMI study are detailed below for patients with axSpA and for healthy controls in Textboxes 1 and 2.

Textbox 1. Patients with axSpA. axSpA: axial spondyloarthritis

Inclusion criteria:

- aged 18 to 65 years at time of their first evaluation
- axSpA (based on ASAS criteria [3] or AS (based on modified New York Criteria [30])
- able to walk 180 m without technical help
- with stable treatment for 3 months
- with a public health insurance (French social security)

Exclusion criteria:

- musculoskeletal, cardiorespiratory or neurologic disease that could affect gait
- hip or knee arthroplasty done or planned in the following 18 months
- not able to speak French
- desire of pregnancy in the following 18 months
- adults protected by laws (Article L1121-5)

Textbox 2. Healthy controls.

Inclusion criteria:

- aged 18 to 65 years at time of evaluation
- able to walk 180 m without technical help
- with a health insurance

Exclusion criteria:

- musculoskeletal, cardiorespiratory or neurologic disease that could affect gait
- hip or knee arthroplasty done
- not able to speak French

Clinical Characteristics of the Participants

Age, sex, weight, height, self-reported pain intensity at time of evaluation, and pain location were collected for both patients with axSpA and healthy controls by the same observer (JS) [29]. Self-reported pain intensity at time of evaluation score was assessed with the VAS, a horizontal line of 10 cm in length, anchored by word descriptors with no pain on the left side and the worst imaginable pain on the right side [22]. Participants were asked to mark the point corresponding to their current pain. Participants were asked to localized their pain using a pain areas figure [31].

For patients with axSpA only, disease clinical characteristics including treatment, disease duration, and morning stiffness and self-assessment questionnaires including the Bath Ankylosing Spondylitis Functional Index (BASFI) [32] and the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) [33] were also collected [29].

Experimental Protocol

Participants performed gait assessments described in a previous publication [29]. In this study, data from the 10-meter walk test (10MWT) in single-task condition only were considered. Participants performed a 10MWT on a 14-meter walkway at comfortable walking speed [34] in single-task condition (3 trials). Gait assessments were performed by the same examiner (JS). Participants wore walking shoes, with 2 inertial measurement units with triaxial accelerometers and gyroscopes (Physilog5, Gait Up), placed above both feet (behind the base of the fifth metatarsal) [35]. The two first and last steps were removed from the analysis [36,37], and at least 16 steps were included in the analysis. For patients with axSpA, regarding the possible consequences of morning stiffness on functional limitations [38], gait assessment was performed at least 2 hours from the end of morning stiffness.

Spatiotemporal Gait Outcomes

After checking for nonsignificant differences between left and right feet, the following spatiotemporal gait parameters were computed using Gait Analysis Software (version 5.3.0, Gait Up) with the mean of right and left foot values for each trial:

- Speed (ms^{-1}): mean walking stride velocity of forward walking
- Cadence (steps/minute): number of steps in a minute
- Stride length (m): distance between two consecutive footprints on the ground, from the heel of a foot to the heel of the same foot, one cycle after
- Swing time (%): portion of the cycle during which the foot is in the air and does not touch the ground
- Stance time (%): portion of the cycle during which part of the foot touches the ground
- Double support time (%): portion of the cycle where both feet touch the ground

The mean between trial 2 and 3 was calculated for each spatiotemporal gait parameter as it has recently been shown to be the more reliable to assess spatiotemporal gait parameters when performing a 10MWT at comfortable speed [35].

Data Analysis

Data analysis were performed using SPSS (version 20, IBM Corp) and Excel (Microsoft Corp). Independent samples *t* tests were used to compare patients with axSpA and healthy controls in terms of age, gender, height, weight, self-reported pain intensity scores, and spatiotemporal gait parameters.

In the interest of further discerning differences that could exist as a function of group versus changes in self-reported pain intensity scores, the spatiotemporal gait parameters were further analyzed between groups using 1-way analyses of covariance (ANCOVAs) with the addition of pain intensity score as a covariate. Statistical threshold for all analyses was set at P=.05. Effect size (Cohen d and partial η^2) and 95% confidence intervals were also calculated.

Results

Demographic and Clinical Assessments

Demographic and clinical assessments for patients with axSpA and healthy controls are shown in Table 1. When comparing patients with axSpA and healthy controls, there were no significant differences for age, height, or weight, but patients with axSpA had higher self-reported pain intensity (P<.001; Table 1). In healthy controls, pain was located at the low back (1/30, 3%), knees (1/30, 3%), or shoulders (1/30, 3%). In patients with axSpA, pain was located at the low back (17/30, 57%), bottom or sacroiliac joints (12/30, 40%), thoracic back (11/30, 37%), cervical back (14/30, 47%), sternum or ribs (1/30, 3%), hips (6/30, 20%), knees (9/30, 30%), ankle or feet (3/30, 10%), shoulders (6/30, 20%), elbows (4/30, 13%), or hands (5/30, 17%).

Table 1 also presents pharmacological treatments and disease characteristics for patients with axSpA. Most patients with axSpA included in this study had anti-TNF treatment (21/30, 70%), low disease activity with BASDAI <4 (BASDAI: 3.04 [SD 1.90]), and low impact of axSpA on physical function (BASFI: 2.86 [SD 2.04]).

Table 1. Patients with axSpA and healthy controls clinical characteristics (n=60).

Clinical characteristics	Healthy controls (n=30)	Patients with axSpA[a] (n=30)	Independent *t* test		
			t	*P* value	95% CI
Demographics					
Age (years), mean (SD)	45.70 (10.60)	45.37 (10.54)	0.121	.90	–5.13 to 5.79
Gender (male), n (%)	20 (67)	20 (67)	—[b]	—	—
Weight (kg), mean (SD)	70.25 (10.27)	74.15 (12.94)	–1.294	.20	–9.94 to 2.13
Height (cm), mean (SD)	174.47 (7.48)	170.77 (7.82)	1.873	.07	–0.25 to 7.65
Self-reported pain intensity scores at time of evaluation, mean (SD)	0.20 (0.66)	3.12 (2.38)	–6.463	<.001	–3.82 to –2.02
Pharmacological treatment, n (%)					
Anti-TNF[c]	0 (0)	21 (70)	—	—	—
Anti-IL-17A[d]	0 (0)	2 (7)	—	—	—
DMARDs[e]	0 (0)	3 (10)	—	—	—
NSAIDs[f]	0 (0)	7 (23)	—	—	—
Pain relief	0 (0)	7 (23)	—	—	—
No treatment	30 (100)	3 (10)	—	—	—
Disease, mean (SD)					
Disease duration from diagnosis (years)	—	11.77 (10.11)	—	—	—
BASDAI[g]	—	3.04 (1.90)	—	—	—
BASFI[h]	—	2.86 (2.04)	—	—	—
Morning stiffness duration (min)	—	28.17 (33.71)	—	—	—

[a]axSpA: axial spondyloarthritis.

[b]Not applicable.

[c]Anti-TNF: antitumor necrosis factor.

[d]Anti-IL-17A: anti-interleukine-17A.

[e]DMARD: disease modifying antirheumatic drug.

[f]NSAID: nonsteroidal anti-inflammatory agent.

[g]BASDAI: Bath Ankylosing Spondylitis Disease Activity Index.

[h]BASFI: Bath Ankylosing Spondylitis Functional Index.

Spatiotemporal Gait Parameters

Spatiotemporal gait parameters for patients with axSpA and healthy controls are shown in Table 2. Independent sample *t* tests without covariate indicated that patients with axSpA presented lower gait speed (*P*<.001) and cadence (*P*=.004), shorter stride length (*P*<.001) and swing time (*P*<.001), and longer double support time *P*<.001) and stance time (*P*<.001) than matched healthy controls (Table 2).

ANCOVA comparisons of spatiotemporal gait parameters between groups revealed that a significant effect of group was found (*F*: 3.434, *P*=.004, partial η^2: 0.320). Results for each spatiotemporal gait parameter can be found in Table 2. When using self-reported pain intensity score as a covariate, spatiotemporal gait parameters were still significant with patients with axSpA exhibiting lower gait speed (*P*<.001), shorter stride length (*P*=.001) and swing time (*P*<.001), and longer double support time (*P*<.001) and stance time (*P*<.001) than matched healthy controls except for cadence which was not significant (*P*=.17; Table 2).

Table 2. Spatiotemporal gait parameters obtained in patients with axSpA and healthy controls in single-task condition with t test and ANCOVA results when taking self-reported pain intensity as a covariate.

Spatiotemporal gait parameters	Healthy controls (n=30), mean (SD)	Patients with axSpA[a] (n=30), mean (SD)	Independent *t* test				ANCOVA[b]		
			t	*P* value	Cohen *d*	95% CI	*F*	*P* value	Partial η^2
Speed (ms^{-1})	1.50 (0.16)	1.27 (0.17)	5.528	<.001	1.17	0.15 to 0.32	15.268	<.001	0.211
Cadence (steps/min)	113.89 (6.35)	108.41 (7.85)	2.97	.004	0.72	1.79 to 9.17	1.922	.17	0.033
Stride length (m)	1.56 (0.14)	1.38 (0.15)	4.679	<.001	1.04	0.10 to 0.25	13.508	.001	0.192
Double support time (%)	19.43 (3.42)	22.99 (2.50)	–4.609	<.001	–1.03	–5.11 to –2.01	13.948	<.001	0.197
Swing time (%)	39.84 (1.77)	38.20 (1.19)	4.201	<.001	0.96	0.86 to 2.41	14.011	<.001	0.197
Stance time (%)	60.16 (1.77)	61.80 (1.19)	–4.201	<.001	–0.96	–2.41 to –0.86	14.011	<.001	0.197

[a]axSpA: axial spondyloarthritis.

[b]ANCOVA: one-way analysis of covariance using pain as covariate.

Discussion

Principal Findings

Only a few studies have assessed gait in the broader spectra of axSpA [13,26,39]. What is more, these studies have used clinical measurements of gait (ie, 6-min walk test [26] or 6-meter maximum velocity test [39]) without a healthy control group for comparison. Inconsistent results were found in patients with AS regarding spatiotemporal gait parameters [15-18], which may be explained by the rather small sample sizes of these studies and by self-reported pain intensity scores reported by the patients at the time of the evaluation.

This study was hence specifically designed to evaluate and compare spatiotemporal gait in 30 patients with axSpA and 30 matched healthy controls without and with pain intensity score as a covariate.

We found that patients with axSpA walked with reduced speed, cadence, stride length, and swing time and increased double support and stance time and that pain could per se partly explain this gait behavior. These results are in line with those recently reported by Zhang et al [18]. However, it should be noted that we further broaden the range of patients by including patients with axSpA, including AS and nr-axSpA, while Zhang et al [18] assessed gait in patients with AS and with hip involvement only. To our knowledge, this is the first study comparing spatiotemporal gait parameters in the broad range of patients with axSpA and matched healthy individuals [13]. Zhang et al [18] used a 3D motion-capture system, which is hardly accessible to clinical routine, while we used IMUs positioned on the feet, allowing computation of spatiotemporal gait parameters in clinical practice or in an ecological environment [23,40]. Finally, contrary to the Zhang et al [18] study, we included pain as a covariate to examine whether and to what extent self-reported pain intensity score could explain the gait differences observed between patients with axSpA and healthy controls.

Our results first showed a significant decrease of gait speed (control: 1.50 [SD 0.16] vs axSpA: 1.27 [SD 0.17] m/s, Δ=–16.6%, *P*<.001) of patients with axSpA as compared to matched healthy controls. This statistically significant difference is accompanied by a Cohen *d* effect size of 1.17, hence suggesting that the between groups difference for the gait speed is large (*d*>0.8) [41]. In the absence of published work on gait in patients with axSpA [13] and although the included population was broader (axSpA vs AS), we were inclined to compare our results with those obtained in patients with AS. With this in mind, our result is in line with that reported in patients with AS by Zhang et al [18], who compared 18 patients with AS to 18 healthy matched controls (control: 1.25 [SD 0.09] vs AS: 1.15 [SD 0.21] m/s, Δ=–8.3%, *P*=.009). Gait speed of patients and healthy controls measured in this study was slightly higher than that reported in Zhang et al [18] (this study axSpA: 1.27 [SD 0.17] vs Zhang et al [18] AS: 1.15 [SD 0.21] m/s, Δ=–9.9%; this study control: 1.50 [SD 0.16] vs Zhang et al [18] control: 1.25 [SD 0.09] m/s, Δ=–16.6%). If gait were assessed along 10 meters in both studies, Zhang et al [18] included gait initiation, steady-state walking, and gait termination in the analysis. In our study, the acceleration and deceleration phases achieved during gait initiation and termination were not included. We used a 14-meter walkway [42-44] and removed the two first and the two last steps of the trials [36,37], as previously proposed in other studies that have assessed spatiotemporal gait parameters during walking [45-47]. When compared to other studies on AS, our result on gait speed does not corroborate those of Del Din et al [15] (12 AS vs 12 controls, control: 1.12 [SD 0.25] vs AS: 1.05 [SD 0.23] m/s, Δ=–6.45%, *P*=.33) and Mangone et al [17] (17 AS vs 10 controls, control: 0.96 [SD 0.2] vs AS: 0.94 [SD 0.2] m/s, Δ=–2.1%, *P*=.78), who did not report any significant between-group differences for the gait speed.

Our results further showed a significantly shorter stride length in patients with axSpA than in matched healthy controls (control: 1.56 [SD 0.14] vs axSpA: 1.38 [SD 0.15] m, Δ=–12.2%, *P*<.001) with a large Cohen *d* effect size of 1.04. This result is in agreement with the decrease in stride length of patients with AS observed in two previous studies by Zebouni et al [16] (12 AS vs 11 controls, control: 0.72 [SD 0.13] vs AS: 0.58 [SD 0.11] m, Δ=–21.5% , *P*<.05) and Zhang et al [18] (stride length/height: control: 0.76 [SD 0.42], AS: 0.70 [SD 0.97],

Δ=–8.2%, *P*=.002). However, our result is not in line with two other studies on AS by Del Din et al [15] and Mangone et al [17], who did not report any significant differences in stride length between AS and controls (control: 1.29 [SD 0.30] vs AS: 0.98 [SD 0.58] m, Δ=–27.3%, *P*=.27 [15]; control: 1.14 [SD 0.2] vs AS: 1.09 [SD 0.1] m, Δ=–4.48%, *P*=.40 [17]).

Our results further revealed a significant reduction of cadence in patients with axSpA as compared to matched healthy controls (control: 113.89 [SD 6.35] vs axSpA: 108.41 [SD 7.85] steps/min, Δ=–4.9%, *P*=.004) with a medium effect size (Cohen *d*: 0.72). This result does not support the previous findings of Zhang et al [18], Zebouni et al [16], or Mangone et al [17], as no significant difference of cadence between patients with AS and healthy controls was observed (control: 0.94 [SD 0.04] vs AS: 0.95 [SD 0.09] /s, Δ=1.06%, *P*=.601 [18]; control: 103.2 [SD 6.6] vs AS: 102.6 [SD 9] steps/min, Δ=–0.58%, *P*=nonsignificant [16]; control: 101.4 [SD 8.7] vs AS: 102.4 [SD 13.3] steps/min, Δ=0.98%, *P*=.65 [17]).

In addition to these three routinely used spatiotemporal gait parameters, we further computed temporal distribution of gait cycle phases using swing time, stance time, and double support time percentages. The distribution of swing and stance period are temporal indicators of gait pattern [48] and often used as objectives in gait rehabilitation [49]. Indeed, the percentage times spent on swing and stance phases are determined by various factors including balance [50] and push-off force generation responsible for step asymmetry in chronic hemiparesis [51] and are associated with gait speed [52]. Only one study on patients with AS assessed these two temporal parameters [18]. Our results showed shorter swing time percentages (control: 39.84% [SD 1.77%] vs axSpA: 38.20% [SD 1.19%] of gait cycle, Δ=–4.2%, *P*<.001, Cohen *d*: 0.96) and longer stance time percentages (control: 60.16% [SD 1.77%] vs axSpA: 61.8% [SD 1.19%] of gait cycle, Δ=2.69%; *P*<.001, Cohen *d*: –0.96) in patients with axSpa than matched healthy controls. Once again, our results are not in agreement with the existing literature as no significant difference with healthy controls of swing period was found by Zhang et al [18] (right: control: 38.61% [SD 1.55%] vs AS: 38.29% [SD 2.62%] of gait cycle, Δ=–0.83%, *P*=.64; left: control: 38.49% [SD 1.66%] vs AS: 38.12% [SD 3.95%] of gait cycle, Δ=–0.97%, *P*=.57).

Our results further showed longer double support time percentages in patients with axSpa than matched healthy controls (control: 19.43% [SD 3.42%] vs axSpA: 22.99% [SD 2.5%] of gait cycle, Δ=16.8%, *P*<.001, Cohen *d*:–1.03). Note that the Cohen *d* effect size for double support time can be considered as large (>0.8). Interestingly, double support time percentage values obtained in this study cannot be compared to other studies as this parameter has never been assessed in AS [14].

To conclude, both the results of this study and those published elsewhere revealed a remarkable lack of consensus in the academic literature on gait and AS, although the low number of published studies and various methodologies make comparisons rather difficult. What explanation could we have for these differences?

Note that the demographic and clinical characteristics of patients and healthy controls (age, weight, and height) and disease duration of patients with axSpA involved in this study (age: 45 years, disease duration: 11.77 years) were comparable to those reported in previous studies (age between 38 and 49.4 years [15-18]; disease duration between 9.3 and 15 years [15-18]) and hence may not account for the observed divergent results.

We further assessed if divergent results previously reported on gait in AS [14] could be partly explained by self-reported pain intensity score at the time of the evaluation per se. The second statistical analysis presented in this study showed that when adjusting for self-reported pain intensity, patients with axSpA still presented lower gait speed, shorter stride length and swing time, and longer double support time. Interestingly, our results also revealed that there were no longer statistically significant between-group differences observed for the cadence. Taken together, these results suggest that differences between groups on cadence observed in this study could thus stem from self-reported pain intensity at the time of the evaluation per se and could explain why previous studies in AS did not find significant differences in cadence [16,17] and reported low pain intensity in patients [17]. In a complementary way, results also suggest that differences between groups on the other spatiotemporal gait parameters observed in this study could not stem from self-reported pain intensity at the time of the evaluation per se. In other words, conclusions should be made with caution with respect to the influence of pain. Whether or not self-reported pain intensity per se could play a role in gait impairment observed in patients with axSpA still remains an open, unresolved question.

To synthesize the findings, patients with axSpA presented lower gait speed and cadence, shorter stride length and swing time, and longer double support time and stance time than matched healthy controls during walking. Taken together and looked into as a whole, these changes in spatiotemporal gait parameters could be interpreted as the adoption of a more conservative or less destabilizing gait in patients with axSpA (Figure 1). These results represent the characteristically so-called cautious gait pattern commonly observed in older persons [53] but also in individuals with gait disorders (eg, patients with cerebellar ataxia [54], with sensory ataxia [54], adults with obesity [55-57], and with low back pain [19-21]). This typical characteristic of cautious gait has already been observed in patients with AS [15,16]. However, these studies found that stride length was significantly shortened [16] or found only “a trend towards reduction” in gait speed or stride length [15] in patients with AS as compared to controls. Overall, it has been emphasized that individuals compensate for their balance disorders and/or gait by being more cautious during walking. Hence, adopting a more conservative gait pattern, characterized in particular by a slow gait speed, shortened stride/step length, reduced cadence, and an increased time spent in double limb support could be viewed as an adaptation to ensure or increase stability and maintain a safe gait [53,58].

Figure 1. Illustration of a healthy gait and a cautious gait pattern characterized by reduced gait speed and cadence, shortened stride length, and increased double support time.

It is important to note that the differences in spatiotemporal gait parameters obtained between patients with axSpA and healthy controls were outside the standard error of measurement and minimal detectable change (MDC), the minimum value for which a difference can be considered as real [59]. MDC adapted to our group (〖MDC〗 _group= 〖MDC〗 _(individual)÷√n [59,60]) was 0.01 for speed and stride length, between 0.74 and 0.92 for cadence, between 0.24 and 0.53 for double support, and between 0.25 and 0.28 for swing time and stance time [35]. Accordingly, the significant between-group differences observed for these 6 spatiotemporal gait parameters cannot be considered as a measurement error. All in all, our findings showed that patients with axSpA adopted a cautious gait pattern in a similar fashion as the previously mentioned populations [20,53-57].

Taken together, the results of this study are promising for clinical application of gait analysis. We demonstrated that

assessing gait in patients with AS using foot-worn inertial sensors is feasible in clinical settings. Spatiotemporal gait parameters (such as stride length or cadence) are the most used parameters in clinical gait analysis and are easy to understand by both clinicians [23] and patients. IMUs, by allowing rapid and easy-to-perform computation of spatiotemporal gait parameters at a low cost and without limitation of the testing environment, are gaining interests for clinicians [23,61]. The 10MWT used in this study is also routinely used by physiotherapists or medical doctors to evaluate gait in clinical and rehabilitation settings. In addition to the time taken to complete this test [62,63], foot-worn inertial sensors enabled the quantitative gait patterns analysis of patients with axSpa with the computation of spatiotemporal gait parameters that were presented in an intuitive and comprehensible manner. We believe that integrating quantitative gait analysis with wearable IMU systems for clinical assessments could be advantageous for clinicians to better understand movement-related disorders for better functional diagnosis, guidance of treatment planning, monitoring of disease progress, and tracking of recovery [64]. In the near future, we can expect that mobile phone–based gait assessment apps will be used to monitor gait in daily life [65] and permit clinicians to remotely monitor patients' conditions [66,67].

Limitations

Some limitations of the study should be acknowledged. First, patients included in the study were aged between 18 and 65 years with a pathology evolving with age and an increase of stiffness and limitations. Assessments of older patients could be interesting to capture gait alterations associated with disease evolution. Second, although self-reported pain intensity measured at time of assessment was significantly higher in patients with axSpA than healthy controls, levels of pain were quite low (3.12 [SD 2.38]). Patients included in this study were stable (ie, with stable treatment for at least 3 months at time of inclusion) and may not represent the whole population of patients with axSpA [68]. Further studies are thus necessary to explore gait in the broad disease of axSpA. Patient-reported pain intensity is commonly measured with the single VAS. However, VAS alone may not capture all features of pain [69,70] and may be not sufficient to assess pain in patients with axSpA [71]. Finally, additional research is required to determine whether factors other than pain may influence gait in patients with axSpA.

Conclusions

To our knowledge, this is the first study comparing spatiotemporal gait parameters in a broad range of patients with axSpA and matched healthy individuals. Our results provide a comprehensive overview of the alterations of gait in patients with axSpA with reduced speed, cadence, stride length, and swing time and increased double support and stance. When all these changes in spatiotemporal gait parameters are taken together and looked into as a whole, it is possible to consider that patients with axSpA adopt a so-called cautious gait pattern. It is the first study to include pain intensity as a covariate to explain spatiotemporal gait parameters in patients with AS or axSpA. Although not a definitive finding, our results suggest that among factors that may influence gait in patients with axSpA, patient self-reported pain intensity could play a role and hence should be addressed when assessing gait in this population.

Acknowledgments

The authors would like to thank all the participants of the study and the FOLOMI (Function, Locomotion, Measurement, Inflammation) team for conceiving and designing the FOLOMI study (JS, NV, PG, AB, JV, Laurent Grange, Jean-Luc Cracowski, and Robert Juvin). We also would like to thank Thomas Carlin for his help on data analysis. This work is supported by grant PHRIP-16-0528 from the French Ministry of Health and Solidarity with the Nursing and Paramedical Hospital Research Program year 2016 and grants ANR-10-AIRT-05 and ANR-15-IDEX-02 from the French National Research Agency in the framework of the Investissements d'avenir program. The sponsors had no involvement in the design of the study; the collection, analysis, and interpretation of data; or in the writing of the manuscript. This study is part of the PhD thesis of the first author, JS (University Grenoble Alpes, Autonomy, Gerontology, E-health, Imaging and Stroke (AGEIS), France and Grenoble Alps University Hospital, Grenoble, France). This work further forms part of a broader translational and interdisciplinary research program GaitAlps (NV). Research data are available on request to the corresponding author.

Authors' Contributions

The FOLOMI (Function, Locomotion, Measurement, Inflammation) team conceptualized and designed the study [29]. JV and NV supervised the project. JS was responsible for the acquisition and analysis of the data. JS, JV, and NV interpreted the results. JS drafted the first version of the manuscript. JV and NV revised the article critically for important intellectual content. All authors read and approved the final version and agreed to be accountable for all aspects related to the accuracy or integrity of the work.

Conflicts of Interest

None declared.

References

1. Proft F, Poddubnyy D. Ankylosing spondylitis and axial spondyloarthritis: recent insights and impact of new classification criteria. Ther Adv Musculoskelet Dis 2018 Jun;10(5-6):129-139 [FREE Full text] [doi: 10.1177/1759720X18773726] [Medline: 29942364]
2. Raychaudhuri SP, Deodhar A. The classification and diagnostic criteria of ankylosing spondylitis. J Autoimmun 2014 Feb;48-49:128-133. [doi: 10.1016/j.jaut.2014.01.015] [Medline: 24534717]
3. Rudwaleit M, van der Heijde D, Landewé R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis 2009 Jun;68(6):777-783. [doi: 10.1136/ard.2009.108233] [Medline: 19297344]
4. Bohn R, Cooney M, Deodhar A, Curtis J, Golembesky A. Incidence and prevalence of axial spondyloarthritis: methodologic challenges and gaps in the literature. Clin Exp Rheumatol 2018;36(2):263-274 [FREE Full text] [Medline: 29148402]
5. Sieper J, Poddubnyy D. Axial spondyloarthritis. Lancet 2017 Jul 01;390(10089):73-84. [doi: 10.1016/S0140-6736(16)31591-4] [Medline: 28110981]
6. Bot S, Caspers M, Van Royen BJ, Toussaint H, Kingma I. Biomechanical analysis of posture in patients with spinal kyphosis due to ankylosing spondylitis: a pilot study. Rheumatology (Oxford) 1999 May;38(5):441-443. [doi: 10.1093/rheumatology/38.5.441] [Medline: 10371282]
7. Sato T, Yonezawa I, Inoue H, Tada K, Kobayashi S, Hayashi E, Juntendo AS Research Group. Relationship between characteristics of spinopelvic alignment and quality of life in Japanese patients with ankylosing spondylitis: a cross-sectional study. BMC Musculoskelet Disord 2020 Jan 18;21(1):41 [FREE Full text] [doi: 10.1186/s12891-020-3040-z] [Medline: 31954409]
8. Sahin N, Ozcan E, Baskent A, Karan A, Kasikcioglu E. Muscular kinetics and fatigue evaluation of knee using by isokinetic dynamometer in patients with ankylosing spondylitis. Acta Reumatol Port 2011;36(3):252-259 [FREE Full text] [Medline: 22113600]
9. Ro D, Lee J, Lee J, Park J, Han H, Lee M. Effects of knee osteoarthritis on hip and ankle gait mechanics. Adv Orthop 2019;2019:9757369 [FREE Full text] [doi: 10.1155/2019/9757369] [Medline: 31019809]
10. Suh MJ, Kim BR, Kim SR, Han EY, Nam KW, Lee SY, et al. Bilateral quadriceps muscle strength and pain correlate with gait speed and gait endurance early after unilateral total knee arthroplasty: a cross-sectional study. Am J Phys Med Rehabil 2019 Oct;98(10):897-905. [doi: 10.1097/PHM.0000000000001222] [Medline: 31094710]
11. Tanishima S, Hagino H, Matsumoto H, Tanimura C, Nagashima H. Association between sarcopenia and low back pain in local residents prospective cohort study from the GAINA study. BMC Musculoskelet Disord 2017 Nov 15;18(1):452 [FREE Full text] [doi: 10.1186/s12891-017-1807-7] [Medline: 29141602]
12. Otayek J, Bizdikian AJ, Yared F, Saad E, Bakouny Z, Massaad A, et al. Influence of spino-pelvic and postural alignment parameters on gait kinematics. Gait Posture 2020 Feb;76:318-326. [doi: 10.1016/j.gaitpost.2019.12.029] [Medline: 31891899]
13. Soulard J, Vaillant J, Vuillerme N. Gait in patients with axial spondyloarthritis: a systematic review of the literature. Curr Rheumatol Rev 2021 Sep 20:1. [doi: 10.2174/1573397117666210921114949] [Medline: 34548003]
14. Soulard J, Vaillant J, Agier C, Vuillerme N. Gait characteristics in patients with ankylosing spondylitis: a systematic review. Clin Exp Rheumatol 2021;39(1):173-186 [FREE Full text] [Medline: 33025884]
15. Del DS, Carraro E, Sawacha Z, Guiotto A, Bonaldo L, Masiero S, et al. Impaired gait in ankylosing spondylitis. Med Biol Eng Comput 2011 Jul;49(7):801-809. [doi: 10.1007/s11517-010-0731-x] [Medline: 21229328]
16. Zebouni L, Helliwell PS, Howe A, Wright V. Gait analysis in ankylosing spondylitis. Ann Rheum Dis 1992 Jul 01;51(7):898-899 [FREE Full text] [doi: 10.1136/ard.51.7.898] [Medline: 1632666]
17. Mangone M, Scettri P, Paoloni M, Procaccianti R, Spadaro A, Santilli V. Pelvis-shoulder coordination during level walking in patients with ankylosing spondylitis. Gait Posture 2011 May;34(1):1-5. [doi: 10.1016/j.gaitpost.2011.02.002] [Medline: 21377366]
18. Zhang G, Li J, Xia Z, Xu W. The gait deviations of ankylosing spondylitis with hip involvement. Clin Rheumatol 2019 Apr 4;38(4):1163-1175. [doi: 10.1007/s10067-018-4401-y] [Medline: 30607653]
19. Demirel A, Onan D, Oz M, Ozel Aslıyuce Y, Ulger O. Moderate disability has negative effect on spatiotemporal parameters in patients with chronic low back pain. Gait Posture 2020 Jun;79:251-255. [doi: 10.1016/j.gaitpost.2020.05.015] [Medline: 32460134]
20. Bonab M, Colak TK, Toktas ZO, Konya D. Assessment of spatiotemporal gait parameters in patients with lumbar disc herniation and patients with chronic mechanical low back pain. Turk Neurosurg 2020;30(2):277-284 [FREE Full text] [doi: 10.5137/1019-5149.JTN.27499-19.2] [Medline: 32091127]
21. Hicks GE, Sions JM, Coyle PC, Pohlig RT. Altered spatiotemporal characteristics of gait in older adults with chronic low back pain. Gait Posture 2017 Jun;55:172-176 [FREE Full text] [doi: 10.1016/j.gaitpost.2017.04.027] [Medline: 28458149]
22. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. J Clin Nurs 2005 Aug;14(7):798-804 [FREE Full text] [doi: 10.1111/j.1365-2702.2005.01121.x] [Medline: 16000093]
23. Iosa M, Picerno P, Paolucci S, Morone G. Wearable inertial sensors for human movement analysis. Expert Rev Med Devices 2016 Jul;13(7):641-659. [doi: 10.1080/17434440.2016.1198694] [Medline: 27309490]
24. Chen S, Lach J, Lo B, Yang G. Toward pervasive gait analysis with wearable sensors: a systematic review. IEEE J Biomed Health Inform 2016 Nov;20(6):1521-1537. [doi: 10.1109/JBHI.2016.2608720] [Medline: 28113185]

25. Aranda-Valera I, Cuesta-Vargas A, Garrido-Castro J, Gardiner P, López-Medina C, Machado P, et al. Measuring spinal mobility using an inertial measurement unit system: a validation study in axial spondyloarthritis. Diagnostics (Basel) 2020 Jun 24;10(6):426 [FREE Full text] [doi: 10.3390/diagnostics10060426] [Medline: 32599741]
26. Coulter EH, McDonald MT, Cameron S, Siebert S, Paul L. Physical activity and sedentary behaviour and their associations with clinical measures in axial spondyloarthritis. Rheumatol Int 2020 Mar;40(3):375-381 [FREE Full text] [doi: 10.1007/s00296-019-04494-3] [Medline: 31848736]
27. Lwanga S, Lemeshow S. Sample size determination in health studies : a practical manual. Geneva: World Health Organization; 1991. URL: https://apps.who.int/iris/handle/10665/40062 [accessed 2021-09-18]
28. Sample size calculator. URL: https://wnarifin.github.io/ssc/ss2mean.html [accessed 2021-09-18]
29. Soulard J, Vuillerme N, Gaudin P, Grange L, Baillet A, Cracowski J, et al. Gait as predictor of physical function in axial spondyloarthritis: the prospective longitudinal FOLOMI (Function, Locomotion, Measurement, Inflammation) study protocol. Rheumatol Int 2019 Oct;39(10):1681-1688. [doi: 10.1007/s00296-019-04396-4] [Medline: 31392500]
30. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum 1984 Apr;27(4):361-368. [doi: 10.1002/art.1780270401] [Medline: 6231933]
31. Escalante A, Lichtenstein MJ, White K, Rios N, Hazuda HP. A method for scoring the pain map of the McGill Pain Questionnaire for use in epidemiologic studies. Aging (Milano) 1995 Oct;7(5):358-366. [doi: 10.1007/BF03324346] [Medline: 8719602]
32. Calin A, Garrett S, Whitelock H, Kennedy L, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994 Dec;21(12):2281-2285. [Medline: 7699629]
33. El Miedany Y, Youssef S, Mehanna A, Shebrya N, Abu Gamra S, El Gaafary M. Defining disease status in ankylosing spondylitis: validation and cross-cultural adaptation of the Arabic Bath Ankylosing Spondylitis Functional Index (BASFI), the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), and the Bath Ankylosing Spondylitis Global score (BASG). Clin Rheumatol 2008 May 5;27(5):605-612. [doi: 10.1007/s10067-007-0755-2] [Medline: 18180979]
34. Graham JE, Ostir GV, Fisher SR, Ottenbacher KJ. Assessing walking speed in clinical research: a systematic review. J Eval Clin Pract 2008 Aug;14(4):552-562 [FREE Full text] [doi: 10.1111/j.1365-2753.2007.00917.x] [Medline: 18462283]
35. Soulard J, Vaillant J, Balaguier R, Baillet A, Gaudin P, Vuillerme N. Foot-worn inertial sensors are reliable to assess spatiotemporal gait parameters in axial spondyloarthritis under single and dual task walking in axial spondyloarthritis. Sensors (Basel) 2020 Nov 12;20(22):6453 [FREE Full text] [doi: 10.3390/s20226453] [Medline: 33198119]
36. Truong P, Lee J, Kwon A, Jeong G. Stride counting in human walking and walking distance estimation using insole sensors. Sensors (Basel) 2016 Jun 04;16(6):823 [FREE Full text] [doi: 10.3390/s16060823] [Medline: 27271634]
37. Anwary A, Yu H, Vassallo M. An automatic gait feature extraction method for identifying gait asymmetry using wearable sensors. Sensors (Basel) 2018 Feb 24;18(2):676 [FREE Full text] [doi: 10.3390/s18020676] [Medline: 29495299]
38. Song Y, Wang C, Chen H. Functional limitation and associated factors in outpatients with ankylosing spondylitis in Southwest China. Clin Rheumatol 2017 Apr;36(4):871-877. [doi: 10.1007/s10067-017-3563-3] [Medline: 28168660]
39. Eppeland S, Diamantopoulos AP, Soldal D, Haugeberg G. Short term in-patient rehabilitation in axial spondyloarthritis: the results of a 2-week program performed in daily clinical practice. BMC Res Notes 2013 May 07;6(1):185 [FREE Full text] [doi: 10.1186/1756-0500-6-185] [Medline: 23651526]
40. Kavanagh JJ, Menz HB. Accelerometry: a technique for quantifying movement patterns during walking. Gait Posture 2008 Jul;28(1):1-15. [doi: 10.1016/j.gaitpost.2007.10.010] [Medline: 18178436]
41. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edition, reprint. New York: Psychology Press; 2009.
42. Cheng DK, Nelson M, Brooks D, Salbach NM. Validation of stroke-specific protocols for the 10-meter walk test and 6-minute walk test conducted using 15-meter and 30-meter walkways. Top Stroke Rehabil 2020 May;27(4):251-261. [doi: 10.1080/10749357.2019.1691815] [Medline: 31752634]
43. Chan WL, Pin TW. Reliability, validity and minimal detectable change of 2-minute walk test, 6-minute walk test and 10-meter walk test in frail older adults with dementia. Exp Gerontol 2019 Jan;115:9-18. [doi: 10.1016/j.exger.2018.11.001] [Medline: 30423359]
44. Nascimento LR, Caetano LCG, Freitas DCMA, Morais TM, Polese JC, Teixeira-Salmela LF. Different instructions during the ten-meter walking test determined significant increases in maximum gait speed in individuals with chronic hemiparesis. Rev Bras Fisioter 2012 Apr;16(2):122-127. [doi: 10.1590/s1413-35552012005000008] [Medline: 22378478]
45. Graser JV, Letsch C, van Hedel HJA. Reliability of timed walking tests and temporo-spatial gait parameters in youths with neurological gait disorders. BMC Neurol 2016 Jan 31;16(1):15 [FREE Full text] [doi: 10.1186/s12883-016-0538-y] [Medline: 26830919]
46. Sorsdahl AB, Moe-Nilssen R, Strand LI. Test-retest reliability of spatial and temporal gait parameters in children with cerebral palsy as measured by an electronic walkway. Gait Posture 2008 Jan;27(1):43-50. [doi: 10.1016/j.gaitpost.2007.01.001] [Medline: 17300940]

47. Pérez-Sanpablo AI, Quinzaños-Fresnedo J, Loera-Cruz R, Quiñones-Uriostegui I, Rodriguez-Reyes G, Pérez-Zavala R. Validation of the instrumented evaluation of spatio-temporal gait parameters in patients with motor incomplete spinal cord injury. Spinal Cord 2017 Feb 28;55(7):699-704. [doi: 10.1038/sc.2017.4]
48. Mariani B, Rouhani H, Crevoisier X, Aminian K. Quantitative estimation of foot-flat and stance phase of gait using foot-worn inertial sensors. Gait Posture 2013 Feb;37(2):229-234. [doi: 10.1016/j.gaitpost.2012.07.012] [Medline: 22877845]
49. Jayakaran P, DeSouza L, Cossar J, Gilhooly K. Influence of a walking aid on temporal and spatial parameters of gait in healthy adults. PM R 2014 Sep;6(9):796-801. [doi: 10.1016/j.pmrj.2014.02.005] [Medline: 24534098]
50. Lopes PG, Lopes JAF, Brito CM, Alfieri FM, Rizzo Battistella L. Relationships of balance, gait performance, and functional outcome in chronic stroke patients: a comparison of left and right lesions. Biomed Res Int 2015;2015:716042-716048 [FREE Full text] [doi: 10.1155/2015/716042] [Medline: 26583129]
51. Balasubramanian CK, Bowden MG, Neptune RR, Kautz SA. Relationship between step length asymmetry and walking performance in subjects with chronic hemiparesis. Arch Phys Med Rehabil 2007 Jan;88(1):43-49. [doi: 10.1016/j.apmr.2006.10.004] [Medline: 17207674]
52. Wang Y, Mukaino M, Ohtsuka K, Otaka Y, Tanikawa H, Matsuda F, et al. Gait characteristics of post-stroke hemiparetic patients with different walking speeds. Int J Rehabil Res 2020 Mar;43(1):69-75 [FREE Full text] [doi: 10.1097/MRR.0000000000000391] [Medline: 31855899]
53. Herssens N, Verbecque E, Hallemans A, Vereeck L, Van Rompaey V, Saeys W. Do spatiotemporal parameters and gait variability differ across the lifespan of healthy adults? A systematic review. Gait Posture 2018 Jul;64:181-190. [doi: 10.1016/j.gaitpost.2018.06.012] [Medline: 29929161]
54. Pirker W, Katzenschlager R. Gait disorders in adults and the elderly: a clinical guide. Wien Klin Wochenschr 2017 Feb;129(3-4):81-95 [FREE Full text] [doi: 10.1007/s00508-016-1096-4] [Medline: 27770207]
55. Lim J, Kim J, Seo K, van Emmerik REA, Lee S. The effects of mobile texting and walking speed on gait characteristics of normal weight and obese adults. Motor Control 2020 Sep 10:1-17. [doi: 10.1123/mc.2020-0006] [Medline: 32916659]
56. Maktouf W, Durand S, Boyas S, Pouliquen C, Beaune B. Interactions among obesity and age-related effects on the gait pattern and muscle activity across the ankle joint. Exp Gerontol 2020 Oct 15;140:111054. [doi: 10.1016/j.exger.2020.111054] [Medline: 32791335]
57. McGraw B, McClenaghan BA, Williams HG, Dickerson J, Ward DS. Gait and postural stability in obese and nonobese prepubertal boys. Arch Phys Med Rehabil 2000 Apr;81(4):484-489. [doi: 10.1053/mr.2000.3782] [Medline: 10768540]
58. Cromwell RL, Newton RA. Relationship between balance and gait stability in healthy older adults. J Aging Phys Act 2004 Jan;12(1):90-100. [doi: 10.1123/japa.12.1.90] [Medline: 15211023]
59. Washabaugh EP, Kalyanaraman T, Adamczyk PG, Claflin ES, Krishnan C. Validity and repeatability of inertial measurement units for measuring gait parameters. Gait Posture 2017 Jun;55:87-93 [FREE Full text] [doi: 10.1016/j.gaitpost.2017.04.013] [Medline: 28433867]
60. Lu W, Wang C, Lin J, Sheu C, Hsieh C. The minimal detectable change of the simplified stroke rehabilitation assessment of movement measure. J Rehabil Med 2008 Aug;40(8):615-619 [FREE Full text] [doi: 10.2340/16501977-0230] [Medline: 19020694]
61. Muro-de-la-Herran A, Garcia-Zapirain B, Mendez-Zorrilla A. Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. Sensors (Basel) 2014 Feb 19;14(2):3362-3394 [FREE Full text] [doi: 10.3390/s140203362] [Medline: 24556672]
62. Amatachaya S, Kwanmongkolthong M, Thongjumroon A, Boonpew N, Amatachaya P, Saensook W, et al. Influence of timing protocols and distance covered on the outcomes of the 10-meter walk test. Physiother Theory Pract 2020 Dec;36(12):1348-1353. [doi: 10.1080/09593985.2019.1570577] [Medline: 30704332]
63. Lindholm B, Nilsson MH, Hansson O, Hagell P. The clinical significance of 10-m walk test standardizations in Parkinson's disease. J Neurol 2018 Aug;265(8):1829-1835 [FREE Full text] [doi: 10.1007/s00415-018-8921-9] [Medline: 29876762]
64. Baker R, Esquenazi A, Benedetti MG, Desloovere K. Gait analysis: clinical facts. Eur J Phys Rehabil Med 2016 Aug;52(4):560-574 [FREE Full text] [Medline: 27618499]
65. Lee S, Walker RM, Kim Y, Lee H. Measurement of human walking movements by using a mobile health app: motion sensor data analysis. JMIR Mhealth Uhealth 2021 Mar 05;9(3):e24194 [FREE Full text] [doi: 10.2196/24194] [Medline: 33666557]
66. Pratap A, Grant D, Vegesna A, Tummalacherla M, Cohan S, Deshpande C, et al. Evaluating the utility of smartphone-based sensor assessments in persons with multiple sclerosis in the real-world using an app (elevateMS): observational, prospective pilot digital health study. JMIR Mhealth Uhealth 2020 Oct 27;8(10):e22108 [FREE Full text] [doi: 10.2196/22108] [Medline: 33107827]
67. Muntaner-Mas A, Martinez-Nicolas A, Quesada A, Cadenas-Sanchez C, Ortega FB. Smartphone app (2kmFIT-App) for measuring cardiorespiratory fitness: validity and reliability study. JMIR Mhealth Uhealth 2021 Jan 08;9(1):e14864 [FREE Full text] [doi: 10.2196/14864] [Medline: 33416503]
68. Madsen OR. Stability of fatigue, pain, patient global assessment and the Bath Ankylosing Spondylitis Functional Index (BASFI) in spondyloarthropathy patients with stable disease according to the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI). Rheumatol Int 2018 Mar 3;38(3):425-432. [doi: 10.1007/s00296-017-3920-1] [Medline: 29299630]

69. Kaciroti N, DosSantos MF, Moura B, Bellile EL, Nascimento TD, Maslowski E, et al. Sensory-discriminative three-dimensional body pain mobile app measures versus traditional pain measurement with a visual analog scale: validation study. JMIR Mhealth Uhealth 2020 Aug 19;8(8):e17754 [FREE Full text] [doi: 10.2196/17754] [Medline: 32124732]
70. Pathak A, Sharma S, Jensen MP. The utility and validity of pain intensity rating scales for use in developing countries. Pain Rep 2018;3(5):e672 [FREE Full text] [doi: 10.1097/PR9.0000000000000672] [Medline: 30534623]
71. Duruöz MT, Acer Kasman S, Şahin N, Sezer, Bodur H, Ketenci A, Baklacıoğlu. Validity, reliability, and factor structure of the Istanbul Low Back Pain Disability Index in axial spondyloarthritis. Mod Rheumatol 2020 Mar 03:1-7. [doi: 10.1080/14397595.2020.1733172] [Medline: 32075459]

Abbreviations

10MWT: 10-meter walk test
AGEIS: Autonomy, Gerontology, eHealth, Imaging, and Stroke
ANCOVA: one-way analysis of covariance
AS: ankylosing spondylitis
ASAS: Assessment of Spondyloarthritis International Society
axSpA: axial spondyloarthritis
BASDAI: Bath Ankylosing Spondylitis Disease Activity Index
BASFI: Bath Ankylosing Spondylitis Functional Index
FOLOMI study: Function, Locomotion, Measurement, Inflammation study
IMU: inertial measurement unit
nr-axSpA: nonradiographic axial spondyloarthritis
SpA: spondyloarthritis
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials
VAS: visual analog scale

Edited by L Buis; submitted 11.01.21; peer-reviewed by V Shah, Y Man; comments to author 10.05.21; revised version received 18.05.21; accepted 23.07.21; published 09.11.21.

Please cite as:
Soulard J, Vaillant J, Baillet A, Gaudin P, Vuillerme N
Gait and Axial Spondyloarthritis: Comparative Gait Analysis Study Using Foot-Worn Inertial Sensors
JMIR Mhealth Uhealth 2021;9(11):e27087
URL: https://mhealth.jmir.org/2021/11/e27087
doi:10.2196/27087
PMID:34751663

Original Paper

Experiences of Wearable Technology by Persons with Knee Osteoarthritis Participating in a Physical Activity Counseling Intervention: Qualitative Study Using a Relational Ethics Lens

Jenny Leese[1,2,3,4], PhD; Graham MacDonald[1], MA; Catherine L Backman[1,2], OT, PhD; Anne Townsend[2,5], PhD; Laura Nimmon[1], PhD; Linda C Li[1,2], PT, PhD

[1]Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada

[2]Arthritis Research Canada, Vancouver, BC, Canada

[3]Ottawa Hospital Research Institute, Ottawa, ON, Canada

[4]Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[5]Faculty of Health and Medicine, Lancaster University, Lancaster, United Kingdom

Corresponding Author:
Linda C Li, PT, PhD
Faculty of Medicine
The University of British Columbia
Vancouver, BC
Canada
Phone: 1 604 207 4020
Email: lli@arthritisresearch.ca

Abstract

Background: Current evidence indicates physical activity wearables could support persons with knee osteoarthritis (OA) to be more physically active. However, recent evidence also identifies some persons with arthritis experience guilt or worry while using a wearable if they are not as active as they feel they should be. Questions remain around how persons with knee OA experience benefits or downsides using a wearable in their everyday lives. Better understanding is needed if wearables are to be incorporated in arthritis self-management in ethically aware ways.

Objective: Using an ethics lens, we aimed to describe a range of experiences from persons with knee OA who used a wearable during a physical activity counseling intervention study.

Methods: This is a secondary analysis of qualitative interviews nested within a randomized controlled trial. Guided by phenomenography, we explored the experiences of persons with knee OA following participation in a physical activity counseling intervention that involved using a Fitbit Flex and biweekly phone calls with a study physiotherapist (PT) in an 8-week period. Benefits or downsides experienced in participants' relationships with themselves or the study PT when using the wearable were identified using a relational ethics lens.

Results: Interviews with 21 participants (12 females and 9 males) aged 40 to 82 years were analyzed. Education levels ranged from high school graduates (4/21, 19%) to bachelor's degrees or above (11/21, 52%). We identified 3 categories of description: (1) participants experienced their wearable as a motivating or nagging influence to be more active, depending on how freely they were able to make autonomous choices about physical activity in their everyday lives; (2) some participants felt a sense of accomplishment from seeing progress in their wearable data, which fueled their motivation; (3) for some participants, sharing wearable data helped to build mutual trust in their relationship with the study PT. However, they also expressed there was potential for sharing wearable data to undermine this trust, particularly if this data was inaccurate.

Conclusions: Findings provide an early glimpse into positive and negative emotional impacts of using a wearable that can be experienced by participants with knee OA when participating in a randomized controlled trial to support physical activity. To our knowledge, this is the first qualitative study that uses a relational ethics lens to explore how persons with arthritis experienced changes in their relationship with a health professional when using a wearable during research participation.

(JMIR Mhealth Uhealth 2021;9(11):e30332) doi:10.2196/30332

KEYWORDS

relational ethics; physical activity; wearable; arthritis; qualitative

Introduction

Affecting an estimated 302 million people worldwide, osteoarthritis (OA) is one of the most common forms of arthritis, and it is a leading cause of disability among older adults [1-5]. Evidence-based practice guidelines recommend physical activity as a key component of optimal OA self-management due to its beneficial effects on pain, mobility, and quality of life [6-8]. However, most people with OA do not meet these recommendations, and supporting persons with knee OA to be more active remains problematic [9,10].

The use of consumer-available, activity-monitoring wearable devices offers a promising strategy to increase physical activity among persons with knee OA. Indeed, literature exists to indicate that wearable technology-enabled interventions can significantly increase moderate-to-vigorous physical activity (MVPA) among adults with knee OA [11-15]. These findings build on previous research to suggest that these interventions may significantly improve MVPA participation if the devices are integrated as part of a multifaceted intervention involving counseling with a health professional [16,17].

Evidence exists, however, to indicate that using digital health technologies (including wearables) in the practice of self-monitoring and self-management may be experienced positively or negatively by persons with chronic illness in the context of their everyday lives [18,19]. In a recent synthesis of qualitative evidence, Leese et al [20] provided an early glimpse of ethical issues identified in the perspectives of persons with arthritis on the use of wearables to support physical activity participation in their everyday lives. These ethical issues were expressed by persons with arthritis as benefits and downsides in their relationships with themselves (ie, their self-perception) and their health professionals. It was found, for example, that persons with OA expressed a general opinion that communication with their health professionals could be enhanced or challenged through the use of a physical activity wearable in their everyday self-management, depending on the quality of their existing relationship [21-23]. However, empirical evidence on the experiences of persons with OA participating in physical activity interventions involving a health professional and using a wearable device is minimal [20]. Furthermore, there is little knowledge on the impact of using a wearable in patient-health professional relations, which are laden with power dynamics [24]. Leese et al [20] also identified that while some persons with chronic illness, including OA, spent more time in physical activity and felt more confident about managing their health while using a wearable, others experienced guilt or worry when they were not as active as they felt they should be while using wearables [25].

An in-depth understanding of a fuller spectrum of experiences of persons with OA is needed if wearable-enabled physical activity counseling programs to support arthritis self-management are to be implemented in ways that appropriately consider ethical issues, such as benefits and downsides encountered by patients. Thus, this study aimed to examine a range of experiences encountered by persons with OA who participated in a study of a wearable-enabled physical activity counseling intervention, with particular attention paid to any influences on participants' relationships with themselves and with the physiotherapist (PT).

Methods

Study Design

This study was a qualitative secondary analysis of semistructured interview data with participants of a mixed-methods study named MONITOR-OA [11]. The original study involved a 6-month proof-of-concept randomized controlled trial (RCT) to examine the efficacy of a technology-enabled counseling intervention for increasing MVPA among persons with knee OA and a qualitative component to describe the participants' experiences of the intervention [11]. Fundamental qualitative description and conventional content analysis methods guided the qualitative component of the original study [26,27]. As a complete data set was already present, our secondary analysis used phenomenographic analytical methods and introduced a relational ethics lens to shape interpretations of the original interview data, in line with Varpio et al's [28] description of a theory-informing inductive data analysis study design [29,30]. Relational ethics is a broad theoretical lens that continues to evolve from critiques of a strong individualistic perspective that dominated traditional bioethics discourse [31-34]. These critiques highlight the complex ways in which persons develop within (and are inherently shaped by) relationships (personal and institutional, past and present) that are an integral part of one's life. They expand on traditional bioethics principles by locating ethical issues in the context of everyday relational settings [31]. Phenomenography is concerned with relations between a person and a specified aspect of the world as it appears to them [29]. Phenomenographic analytic methods and a relational ethics lens thus offer appropriate theoretical grounding to explore the particular focus of our study on everyday ethical issues experienced by participants in their relationships with themselves and the PT while using a physical activity wearable. Our approach rests on assumptions that reality is socially and experientially constructed, and to understand these realities, researchers need to explore the meanings constructed by individuals or groups.

The RCT component of the original study took place between November 2015 and June 2017 in Vancouver, Canada. Participants attended a 1.5-hour session, where they received (1) 15-minute group education about physical activity, (2) a Fitbit Flex (Google LLC), and (3) individual counseling with a study PT who was trained in motivational interviewing [35]. The individual counseling session followed the brief action planning approach, whereby PTs guided participants to identify activity goals, develop an action plan, and identify barriers and solutions [36].

Participants were asked to wear the Fitbit wristband 24 hours a day except during water-based activities or when recharging the battery. The physical activity data were wirelessly synchronized with Fitbit's online Dashboard that could be viewed only by the participants and their study PT. During the intervention period, the PT reviewed the participants' physical activity data on the Dashboard and reviewed their activity goals during 4 biweekly phone calls. Participants could also contact the PT via email in-between the scheduled calls. After the 8-week intervention concluded, all participants were invited to take part in an interview about their experiences.

The research protocol was approved by the University of British Columbia Behavioral Research Ethics Board (H14-01762) and was published in clinicaltrials.gov (NCT02315664). Informed consent to use interview data for the secondary analysis was obtained from all participants at the time of the original study.

Participants

Further details of the original RCT have been described elsewhere [11]. Briefly, individuals were eligible if they were adults living in Vancouver, Canada, with a physician-confirmed diagnosis of knee OA or who passed two criteria for early OA: (1) aged 50 years or older and (2) experienced knee pain during the previous year lasting more than 28 separate or consecutive days [37]. Individuals were excluded if they (1) had been diagnosed with inflammatory arthritis, connective tissue diseases, fibromyalgia, or gout; (2) used antirheumatic drugs or gout medications; (3) had previously undergone knee arthroplasty; (4) had suffered an acute knee injury in the past 6 months; (5) had a body mass index of 40kg/m^2 or higher; (6) had received a steroid injection or a hyaluronate injection in the last 6 months; and (7) were using medications which impaired physical activity tolerance (eg, beta-blockers), or had an inappropriate level of risk for increasing their physical activity. Participants were also excluded if they did not have access to a computer in their home or did not have a personal email address.

The original RCT had 61 participants, of which 56 completed the in-depth interview after the intervention. As this study aimed to examine the range of experiences among participants, sampling was mostly theoretical to maximize variation across demographic characteristics (eg, age, sex, and education). Our analysis focused on a purposive subsample of 21 of these interview participants (13 females and 8 males), ranging in age from 40 to 82 years (Table 1; participants chose their pseudonyms). Participants came from a variety of household compositions, with education levels ranging from high school graduate (4/21, 19%) to a bachelor's degree or above (11/21, 52%), and annual household incomes ranging from under CAD $12,000 (US $9,750; 1/21, 5%) to over CAD $100,000 (US $81,240; 4/21, 19%).

Table 1. Participants' sociodemographic characteristics.

Participant pseudonym	Age at consent	Sex	Education	Annual household income, CAD[a]	Marital status	Living status	Other conditions
Martha	82	Female	Bachelor's degree or higher	$60,001-$80,000	Widowed	Alone	Circulation problems; cancer
Lenny	68	Female	Trades certificate, vocational school diploma, apprenticeship	$40,001-$60,000	Widowed	Alone	None reported
Anne	61	Female	Bachelor's degree or higher	$60,001-$80,000	Separated/divorced	Alone	Allergies; breathing problems; osteoporosis; osteopenia
Marco	64	Male	Bachelor's degree or higher	Over $100,000	Married/common law	Significant other	Allergies
Don	63	Male	Bachelor's degree or higher	Prefer not to answer	Married/common law	Significant other	High blood pressure; allergies
Bruce	58	Male	Grade 11 to 13 (including GED[b])	Over $100,000	Married/common law	Significant other	High blood pressure
Darius	64	Male	Nonuniversity certificate below Bachelor's level	$60,001-$80,000	Married/common law	Significant other	Allergies; kidney, bladder, or urinary problems
Minnekhada	58	Male	Grade 11 to 13	$60,001-$80,000	Married/common law	Significant other	None reported
Gavin	61	Male	Bachelor's degree or above	Over $100,000	Married/common law	Significant other	None reported
Joe	71	Male	Bachelor's degree or above	$24,001-$40,000	Never married	Alone	None reported
Hazel	63	Female	Bachelor's degree or above	$60,001-$80,000	Separated/divorced	With relatives or others	Digestive system problems; allergies; breathing problems; psoriasis; mental health or emotional problems
Tony	77	Male	Trades certificate, vocational school diploma, apprenticeship	$12,001-$24,000	Married/common law	Significant other	Cancer
Yoda	40	Female	Bachelor's degree or above	Prefer not to answer	Married/common law	With children	None reported
Zed	56	Female	Nonuniversity certificate below Bachelor's level	Prefer not to answer	Separated/Divorced	With children	Allergies; diabetes; breathing problems
Denny	61	Female	Non-university certificate below Bachelor's level	$80,001-$100,000	Married/common law	With children	None reported
Logan Kale	41	Female	Grade 11 to 13	$24,001-$40,000	Married/common law	With children	Fibromyalgia
Olivia	69	Female	Bachelor's degree or above	$12,001-$24,000	Separated/divorced	Alone	Diabetes; cancer
Jane	53	Female	Grade 11 to 13	Prefer not to answer	Separated/divorced	With relatives or others	Cerebrovascular problems; headaches
Daenerys	57	Female	Bachelor's degree or above	Over $100,000	Married/common law	Alone	Digestive system problems; allergies; kidney, bladder, or urinary problems

Participant pseudonym	Age at consent	Sex	Education	Annual household income, CAD[a]	Marital status	Living status	Other conditions
Sansa	57	Female	Bachelor's degree or above	$80,001-$100,000	Married/common law	With children	None reported
Biker	61	Female	Non-university certificate below Bachelor's level	$60,001-$80,000	Married/common law	Significant other	Skin problems

[a]A currency exchange rate of CAD $1=US $0.80 is applicable.

[b]GED: general education diploma.

Interviews

The first author (JL) conducted 13 of the 21 in-depth one-on-one interviews, and the second author (GM) conducted the remaining 8 interviews. JL (cisgender woman) and GM (cisgender man) were PhD Candidates with approximately 10 years of experience each in qualitative interviewing. All interviews followed a semistructured interview guide including the following questions: Can you tell me about any physical activity you do? Since taking part in this study, have you noticed any changes in your physical activity? How easy or difficult has it been using the Fitbit? What about your experience with the study PT? Prompts and probes were also used as reminders during active listening, and considerable attention was given to the interviewee articulating their experience in their own words. Interviews (lasting approximately 60-90 mins) were audio-recorded and transcribed verbatim.

Data Analysis

Phenomenographic analytical methods were used to describe the different (positive and negative) ways in which persons with knee OA experienced the physical activity counseling intervention and the various meanings they attributed to this phenomenon in their world [29]. Although the roots of phenomenographic research methods lie in learning studies, the methods have been applied to a variety of other issues inside and outside the field of education [38-41]. We use this analytic approach for two reasons. First, the MONITOR-OA RCT provided a complete data set of interview transcripts in which participants' experiences using a wearable in a physical activity intervention was a prominent issue. Second, this study focused on exploring differences among the collective experience that persons with OA had when participating in the intervention. Both reasons aligned with phenomenography, justifying it as an appropriate approach for analysis.

We analyzed the data through a relational ethics lens, following Heaton's [42,43] definition of supra-analysis, which is one type of secondary analysis involving examining pre-existing data from a new theoretical perspective, thereby transcending the aims and focus of the original research. We reused our self-collected data to examine it from a new theoretical perspective than the primary study, specifically a relational ethics perspective. Relational ethics places the principles of bioethics that are traditionally used to guide moral practice in health care (eg, autonomy, nonmaleficence, beneficence, and justice) within the context of close-up relationships [30,34,44]. It assumes that all relationships are moral and attends to the commitment to ethical action in one's relationship to oneself and the other in every situation or encounter on an ordinary everyday basis [34,44]. Issues of relational autonomy, such as engagement and partnership, have been previously identified in perspectives of persons with arthritis on their use of wearable devices to self-monitor physical activity [20]. Building on this earlier research, our study focused on participants' experiences of the benefits or drawbacks of using a wearable, with a particular interest in any impacts experienced in their relationships with themselves and with the study PT. Therefore, relational ethics is a suitable conceptual lens through which to continue the exploration of ethical issues identified from the perspectives of persons with arthritis.

Our analysis was informed by the 7 steps of phenomenographic analysis as described by Sjöström and Dahlgren [41]: familiarization, compilation, condensation, grouping, comparison, naming the categories, and contrastive comparison (Textbox 1). Data were reviewed carefully for themes that attend to morality in everyday relationships, including (but not limited to) self-control, engagement, and partnership [30,34,44]. Regular meetings were held with the research team, some of whom had experience in phenomenographic analysis, to enhance rigor and check, test, and probe preliminary findings [45].

Informed by calls to critically consider the relevance of notions of saturation within our study's context, we did not consider saturation to be meaningful to our research objective and methodological orientation because we did not seek to saturate theoretical categories, themes, or data [46-49]. Our sampling and analysis ceased when the research team reached an agreement that findings were sufficiently varied to describe a range of participants' experiences relevant to our research objective. Based on the previous experience of phenomenographers, 15 to 25 interviews are typically preferred in a phenomenographical study [50]. While we acknowledge that what is determined as an appropriate sample size for one qualitative study is not necessarily an appropriate sample size for another qualitative study [51], the combination of team agreement regarding content and the number of interviews compared to past phenomenographic studies supported our decision.

Textbox 1. Data analysis process.

Familiarization: Our analytic method began with JL reading and re-reading all of the transcripts. GM also read and re-read a varied sample of 4 of these transcripts.

Compilation: Next, JL and GM independently identified excerpts of data in these 4 transcripts that were found to be of interest for the question being investigated.

Condensation: JL and GM met regularly to discuss and narrow down a selection of data excerpts from the 4 transcripts that they identified as relevant.

Grouping: JL and GM sorted quotes within this selected data pool into piles based on similarity to create a preliminary set of categories of description. At the end of this phase, the following preliminary categories were identified: feeling accountable, changing awareness/perceptions, increasing physical activity, feeling better, having an objective measure, participant-physio relationship, reaching physical activity goals, relating to the Fitbit, emotional impacts of using Fitbit, and making progress.

Comparison: Next, using QSR International NVivo software (version 12) to organize data, JL tested the preliminary categories against the remaining transcripts. She sorted and re-sorted data extracts from the remaining transcripts. This process entailed refining or collapsing the preliminary categories. Preliminary categories were also compared and differentiated from one another in terms of differences (making criterion attributes for each category explicit). All team members met to discuss the categories as they were developed.

Naming: 3 final categories were defined to emphasize their essence were agreed upon through discussion with all team members.

Contrastive Comparison: Informed by a discussion with all team members, JL described the unique character of each final category as well as the resemblances between them. She also checked the final categories against the original transcript data.

Results

Overview

Phenomenographic analysis provided a rich interpretation of how persons with knee OA experienced benefits and drawbacks in using a Fitbit during their participation in a physical activity counseling intervention study. Our analysis revealed 3 main categories: (1) making choices about physical activity with or without a wearable, (2) emotional dimensions of adding awareness about physical activity, and (3) reviewing wearable data with the study PT: issues of accountability and trust. Key quotes are presented in Textboxes 2-4 to illustrate each category, and supplementary supporting quotes can be found in Multimedia Appendix 1.

Making Choices About Physical Activity With or Without a Wearable

Participants described how their Fitbit had influenced their choices to increase their walking to reach their daily step goal (Textbox 2, quotes 1-4). Some attributed human-like traits to their Fitbit when describing its influence on their activity choices (Textbox 2, quotes 1-3). For example, Hazel described her Fitbit as “a little person on my wrist…a little friend” that was a welcome source of friendly and gentle encouragement to meet her step goals. For her, Fitbit was “happy when I do 10,000 steps…just like a friend supporting me…There’s a gentle persuasion.” Others expressed how they felt pressured by their Fitbit to meet their activity goals. Martha, for example, felt ambivalently toward her Fitbit’s influence in her decision to walk more, reflecting that “it probably gives me some incentive to walk a little further just to placate the Fitbit…it nags you…I’m fine with that…it’s probably a good thing to have something that makes you get up and go.” Denny felt “forced to kind of do some more activities” by her Fitbit in the evenings when she felt too tired. Sansa indicated she felt a sense of responsibility to meet her activity goals and viewed the Fitbit as “there to keep me accountable.”

Some participants described certain days they did not engage with their Fitbit; for example, on days they were experiencing “too much pain” or on “some days I don’t care” to be more active. Zed found that Fitbit’s influence did “start me going” to meet step goals “in the beginning” of the study, and its influence “wanes over time” once his increase in walking became part of his habitual routine. Others explained that wearing their Fitbit did not add any value as they were making choices to meet their activity goals regardless (Textbox 2, quote 5). As a busy mother, Yoda found that Fitbit was generally irrelevant as meeting her activity goals was not her main priority (something for which she suggested “shows perhaps an inherent bad attitude”). She recounted, “I wasn’t invested that I absolutely had to do, come hell or high water, these steps so I’d be like yeah, I just didn’t walk very much today.”

Textbox 2. Supporting quotes for "Making choices about physical activity with or without a wearable."

1_Hazel: the Fitbit…It's like a little person on my wrist…it's a little friend…and I tap into it and go, "I'm almost to my 10,000 but I could do a bit more"…throughout my day I can refer to it and I think it's on my side. It's happy when I do 10,000 steps…just like a friend supporting me, encouraging me…whether I achieve it or not it's still there with me…Whether I do it or not, it's up to me. There's a gentle persuasion…a Fitbit helps me feel less alone…it's huge for me because you know most of my life I've not had a lot of support and so support is huge, just huge.

2_Sansa: when I go on there and I see that I haven't reached my six or whatever, I'll in the evening go for you know a five-minute walk or a 10-minute walk or you know try to do that after dinner. So it does give me a little bit more kind of accountability or kind of check where I'm at. I think it's kind of like I have a partner in crime, it's just kind of there to keep me accountable.

3_Martha: It probably gives me some incentive to walk a little further just to placate the Fitbit…I forget to check but I think it does give you an incentive to get out and do something because it's there and it nags you…I'm fine with that…it's probably a good thing to have something that makes you get up and go.

4_Denny: I finally get home say around seven in the evening…just kind of want to eat and then just do nothing…I know you're supposed to move [laughs]…But sometimes I'm just too tired and in the evenings, I'm forced to kind of do some more activities…having the Fitbit it does make me feel I need to move more…I have definitely gone for more walks.

5_Joe: I max out in my activities, so I don't need this monitoring as a way of positive feedback or gratification to give me incentive. I personally don't need that…I'm actually walking and all of that. If somebody says, "Oh you should dance more in the evening," I say, "Well no I can't dance anymore." I can only go so far then I drop dead right? So I'm maxed out. I can't add much more here…So when this six months is over, I'll take the Fitbit and throw it away because it has no relevance to my life…I know what I'm doing and I don't care what this little machine tells me.

Emotional Dimensions of Adding Awareness About Physical Activity

Some participants highlighted there was an emotional dimension to the heightened awareness of their physical activity levels they experienced through using the wearable (Textbox 3, quotes 1-3). Some described how this added awareness about their activity levels prompted feelings of accomplishment and gratification, which fueled their motivation. For example, Daenerys recalled, "It lets me feel as though I'm accomplishing something every day…I feel pretty happy…and then it's kind of fun to see how much more I can do." Olivia also felt "instant gratification" as she could see improvement in Fitbit's feedback on her step goals.

While Logan Kale felt "happiness, accomplishment" when she reached her activity goal most days, she also found that on days when she was "in too much pain" to reach her step goal, she would "try not to beat myself up over it" and "kind of get stuck on that hamster wheel of negative thoughts and have to zip it." Biker commented that she did not "feel bad" when Fitbit data showed she had not met their step goal because "I like to exercise and so being active is not an issue for me…actually I kind of find it a kind of cozy feeling thinking 'I haven't completed everything. There's still more to do.'" She expected, however, that others who were less active than her "might feel sad" if they did not meet a physical activity goal.

Textbox 3. Supporting quotes for "Emotional dimensions of adding awareness about physical activity."

1_Daenerys: I will always keep the Fitbit and always have one I think because it lets me feel as though I'm accomplishing something every day, like I have it set to a pretty low number of steps every day. It's set to 3000 but when I look at my results I can be over 3000. When it goes off during the day, I feel pretty happy about, "Okay, I've accomplished that much today," and then it's kind of fun to see how much more I can do.

2_Logan Kale: when I would be close to my steps, if I would see…you just tap it [the Fitbit] and then you'll know if you're close or not and I would just make that extra effort to meet that mark. Like instead of driving to work, I would walk to work. On most days it was easy…just seeing that number of how you're so close. "Got to get over that hump." Once a goal you've set and, you know, when you reach that goal, you feel good about it. It's just happiness, accomplishment. The other days I'm just in too much pain. I'm like, "I'm not walking home." I want to be that person going on those hikes. I don't want to be that person just sitting there. You want to always try to do better the next week but then if it doesn't happen, I try not to beat myself up over it anymore because, you know, the next day could be better. I'm like, "Okay well you didn't do well this week. What's the problem? You shouldn't be doing that. You shouldn't be doing this. You should be doing that," and I just kind of get stuck on that hamster wheel of negative thoughts and have to zip it.

3_Biker: [responding to *you mentioned that there were some goals that you didn't meet…*] I don't feel bad at all. [Laughs] I just kind of go, "Oh that's life" because I know that I'm keeping really active so yeah it's not a problem to me. I guess if I thought, "Gee I'm not very active and I'm not meeting any of my goals" then I might feel sad about it but because I know that I like to exercise and so being active is not an issue for me. So the goals I set are kind of like…in a perfect world this is what I would like to do but the world isn't perfect and it's okay. You know, I'm working, I'm certainly getting tons of weekly exercise in and so if I don't meet some aspect of it, it'll be okay. You know also I guess I can also look at it and kind of go, "If I really, really wanted to do Yoga, I could put a DVD on and do some at home" but again, I like the Yoga for the social aspect so that's not much of an incentive to do it on my own in my living room…actually I kind of find it a kind of cozy feeling thinking, "I haven't completed everything. There's still more to do." So I'm not taking it kind of like I'm a failure, I'm taking it more as, "Oh there's still more to do and you can keep on growing, keep on improving" so.

Reviewing Wearable Data With the Study PT: Issues of Accountability and Trust

Participants described how their choices to be physically active had been influenced because the study PT had access to review wearable data on their physical activity (Textbox 4, quotes 1-2). For example, Denny described how she had "tried a little bit harder [laughs] maybe for a few weeks" to reach her activity goals because she knew the study PT had access to her wearable data and found this to be "very, very encouraging." Darius also

indicated that he felt a sense of accountability to meet the physical activity goals he had agreed with the study PT, as he recalled "a little bit of pressure…I need to watch to keep my promise."

For one participant, accountability issues intertwined with trust issues in his relationship with the study PT when reviewing his wearable data together (Textbox 4, quotes 3-4). Recounting how his wearable data had not always accurately tracked his step count because it sometimes had "slipped into sleep mode," Gavin described a phone call he had with the study PT in which he explained that his subjective account of his physical activity should be trusted as more credible than his wearable data, and he should not be held accountable, because "it wasn't my fault I didn't make my 10,000." When he felt his Fitbit had tracked his step count more accurately, Gavin described this data as "some form of proof" that could be shared with the study PT "for her to gauge" his activity. He also commented that "if [Fitbit data] wasn't there… then 'oh yeah I climbed Mt. Everest this weekend'…I could've been making up anything about that," emphasizing that, when accurate, Fitbit data may give a more reliable account of his physical activity than himself.

Textbox 4. Supporting quotes for "Reviewing wearable data with the study PT: Issues of accountability and trust."

1_Denny:…you have a knowledgeable person telling you, you're doing the right thing type thing…I tried a little bit harder [laughs] maybe for a few weeks…because you actually have another person kind of monitoring you and you also…you want to try…it was very, very encouraging…I think it's really good for her to be able to see and for me to know that somebody is monitoring me. I think maybe that makes me [laughs] take a few more steps maybe.

2_Darius: It's just a little bit of pressure of keeping the, the steps they...Because I, I need to watch to keep my promise, you know, as far as I can…she [the physiotherapist] called me like every two weeks. I think her purpose is to motivate me for keeping my, my promise to keep activities.

3_Gavin: I do have a grump with the Fitbit over the times where it's gone into sleep mode so many times…Before I know it I've lost 3000 steps…you go like, "I get a lot more steps today than that's showing me. I know it's slipped into the sleep mode activity and it's not"…I figured I should have my entire European boot badge by now so it's not fair…I knew I was getting a phone call from the physio and I said, "Oh yeah but it wasn't my fault I didn't make my 10,000. This stupid band didn't log on properly." Oh she said, "Oh yeah, it happens."

4_Gavin: I think [Fitbit data] at least gave a better conversational point in terms of seeing how you were doing, a reference check for the physio when we're checking in. If it wasn't there it would really…there would be nothing for her to gauge because then, "What did you do?" "Oh yeah I climbed Mt. Everest this weekend and am feeling really good. Kilimanjaro is tomorrow. Not bad."… we would've been talking in a fairy land because I could've been making up anything about that…"

Discussion

Principal Findings

Our findings provide novel insight into different ways in which persons with knee OA experienced their use of a wearable positively or negatively during their research participation. Firstly, the contradictions in the data were fascinating. Participants experienced their wearable positively as a motivating influence and more negatively as a nagging reminder to be more active. From a relational ethics perspective, these findings shed light on how participants' experience of their wearable impacts their autonomy positively or negatively.

Autonomy is a central notion in modern health care ethics that is understood as the capacity to direct one's own choices freely and intentionally [31]. A relational approach builds on this traditional understanding of autonomy by focusing attention on relationships and interdependencies that may support or impair a person's capacity for autonomy [52,53]. The relational autonomy perspective highlighted that many participants experienced the relationship with their wearable as a support to their autonomy. The device helped them take more control in making choices to be more active. However, some participants who described their capacity to make autonomous choices about their physical activity as impaired (eg, due to tiredness) experienced tension and ambivalence in the relationship with their wearable, as they described feeling pressured or forced to be more active at times when they did not entirely wish to move. These findings align well with previous research suggesting that using a wearable may be experienced as autonomy-enhancing or autonomy-undermining by persons with arthritis, regardless of whether they used the wearable to support physical activity during research participation or as part of their everyday self-management [20]. Leese et al [23] have reported some persons with arthritis experienced their wearable as a motivating or autonomy-enhancing support in their everyday self-management when used as a "nice reinforcement" to an already physically active lifestyle.

Findings in our study also indicated positive and negative experiences of participants using a wearable, depending on how freely they were able to direct their own choices to be active given their specific situation or set of circumstances. They resonate with a relational ethics approach that recognizes exercising autonomy requires relevant capabilities, which are dynamically shaped by a person's situation or set of circumstances [54]. They also resonate with calls from some health professionals for academic literature to facilitate the positive development of self-tracking technology in self-management by reflecting on context-relativity (rather than focusing on "ideal" situations) [55]. We posit that further research is thus warranted to build a greater understanding of the everyday contexts in which the use of a wearable may be experienced as autonomy-enhancing or autonomy-undermining by persons self-managing chronic illness.

At this early stage of the potential integration of wearables into arthritis self-management, our findings can contribute to ongoing conversations in clinical practice. They suggest health professionals may wish to carefully consider a person's capability to make autonomous choices about their physical activity if using a wearable in their everyday self-management. They, therefore, align well with research suggesting how health

professionals may tailor their support for arthritis self-management in ways that take a person's capacities to engage in physical activity within a 24-hour day context into account [56,57]. Further research could guide forms of support that health professionals may offer to persons with arthritis using a wearable who struggle to freely direct their own choices about physical activity in their everyday self-management. Without this support, using a wearable in everyday self-management may be experienced as a nagging or autonomy-undermining influence by some persons with arthritis, adding to a struggle to feel in control of their choices to be more active. Evidence exists to indicate that feeling this sense of control in self-management is at least one of the mechanisms responsible for improvements in health behaviors and health status [58-60].

Secondly, our findings build on previous research indicating that persons with arthritis experienced an enhanced awareness of their activity levels when using wearable technology [25,61]. The insights emphasized how a heightened awareness impacted participants' emotions positively and negatively. Some participants experienced feelings of accomplishment when their wearable data illustrated that they had reached their physical activity goal, which often fueled their motivation to do more. One participant, however, experienced negative thoughts on days that his wearable data indicated he had not reached his physical activity goal. Mercer et al [25] also found that persons with chronic illness (including arthritis) experienced negative feelings in using a wearable during a 15-day research period as they were concerned they were not sufficiently active. From a relational ethics perspective, these experiences speak to a theme of embodiment, emphasizing the importance of complex emotions or feelings in a commitment to ethical action in one's relationship with oneself and others [34,62]. It remains unknown how others with arthritis may be emotionally impacted if using a wearable in a "real-world" context, outside of research participation. Our findings, therefore, raise questions about how the use of a wearable device within a counseling program may impact the emotional wellbeing of individuals, particularly as they evaluate their capacity and progress in managing their health.

Thirdly, findings indicate how participants experienced issues of accountability and trust differently when reviewing their wearable data with the study PT. Some participants were motivated by a sense of accountability to "keep my promise" to meet the activity goal agreed with the study PT. For some of these participants, wearable data served as "some form of proof" that the study PT could trust to "gauge" whether they had met this goal. One participant, however, expressed how his wearable data was not to be trusted at times and recalled reaching a mutually satisfactory interpretation of this data as inaccurate with the study PT. These different experiences illustrate how sharing wearable data helped build mutual trust and engagement between participants and the study PT. At the same time, this finding also indicates how sharing wearable data threatened to undermine mutual trust and engagement within this relationship.

The fundamental role of building trust in interpersonal relationships in health care has been emphasized elsewhere through a relational ethics lens [62-64]. A relational ethics lens can also be explored here through the relational theme of engagement. Genuine engagement is understood by Bergum [34] to be "located in the shared moment when people have found a way to look at something together, freely accepting or declining the interpretation that each other offers, until they reach a meaning they both affirm." Our findings support the general opinion previously expressed by persons with OA that sharing their wearable data may enhance their communication with health professionals in everyday self-management under specific conditions (eg, if there was a good rapport already established in the relationship or if a health professional would welcome the wearable data being shared) [20]. They also raise questions about the sense of accountability experienced when sharing wearable data with a health professional, in terms of how far this may be a burden to persons with arthritis in their everyday self-management [65,66]. Further research is therefore needed to gain a better understanding of the relational conditions in which persons with arthritis may experience issues of accountability, trust, and engagement positively or negatively when sharing wearable data with their health professionals in everyday self-management.

Limitations

There were limitations to this study. As a secondary analysis, the data were not created with the relational ethics lens in mind, and therefore potentially important experiences might not be fully elicited in the interviews. Nonetheless, transcripts were purposively selected to offer sufficient variation in participants' experiences relevant to our objective. A phenomenographic approach also allowed us to identify overarching meanings that crossed transcripts and were implicitly presented by the collective group. To better examine the transferability of findings, further research is needed to explore the experiences of a more diverse sample using a wearable in the context of their everyday self-management of arthritis outside of research participation. It may be that persons of diverse genders or cultural backgrounds, for example, encounter different experiences. Our subsample, however, is varied and represents a typical OA group in terms of age and sex.

Conclusions

Our findings provide insight into different ways in which persons with OA experienced their use of a wearable during participation in a physical activity counseling intervention study positively or negatively. Drawing on a relational ethics lens, we identified how issues of relational autonomy, embodiment, accountability, trust, and genuine engagement were present in these experiences. These issues have implications for learning how to develop and implement wearable-enabled physical activity programs to support arthritis self-management in ways that seriously factor in ethical considerations. We present these salient ethical issues for further discussion and to guide future empirical investigation of the use of wearables in arthritis self-management.

Acknowledgments

JL was supported by a Canadian Institutes of Health Research Doctoral Research Award. GM was supported by a Doctoral Salary Award from The Arthritis Society. LCL is supported by the Harold Robinson/Arthritis Society Chair in Arthritic Diseases award, the Canada Research Chair Program, and the Michael Smith Foundation for Health Research Scholar Award.

Authors' Contributions

JL, AT, CLB, LN, and LCL contributed to the design and planning of the study. JL and GM contributed to the analysis, and all authors contributed to the interpretation of data. JL drafted the manuscript, and all authors critically reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary supporting quotes.
[PDF File (Adobe PDF File), 46 KB - mhealth_v9i11e30332_app1.pdf]

References

1. Bombardier C, Hawker G, Mosher D. The impact of arthritis in Canada: Today and over the next 30 years. Toronto, ON: Arthritis Alliance of Canada; 2011. URL: https://www.arthritisalliance.ca/images/PDF/eng/Initiatives/20111022_2200_impact_of_arthritis.pdf [accessed 2021-10-13]
2. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: Estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014 Jul;73(7):1323-1330. [doi: 10.1136/annrheumdis-2013-204763] [Medline: 24553908]
3. Cisternas MG, Murphy L, Sacks JJ, Solomon DH, Pasta DJ, Helmick CG. Alternative methods for defining osteoarthritis and the impact on estimating prevalence in a US population-based survey. Arthritis Care Res 2016 Apr 27;68(5):574-580. [doi: 10.1002/acr.22721]
4. Murphy L, Schwartz T, Helmick C, Renner J, Tudor G, Koch G, et al. Lifetime risk of symptomatic knee osteoarthritis. Arthritis Rheum 2008 Sep 15;59(9):1207-1213 [FREE Full text] [doi: 10.1002/art.24021] [Medline: 18759314]
5. GBD 2015 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet 2016 Oct 8;388(10053):1603-1658. [doi: 10.1016/S0140-6736(16)31460-X]
6. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Rheumatol 2020 Feb;72(2):220-233. [doi: 10.1002/art.41142]
7. Bannuru R, Osani M, Vaysbrot E, Arden N, Bennell K, Bierma-Zeinstra S, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019 Nov;27(11):1578-1589. [doi: 10.1016/j.joca.2019.06.011]
8. Rausch Osthoff A, Niedermann K, Braun J, Adams J, Brodin N, Dagfinrud H, et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. Ann Rheum Dis 2018 Sep;77(9):1251-1260. [doi: 10.1136/annrheumdis-2018-213585] [Medline: 29997112]
9. Wallis J, Webster K, Levinger P, Taylor N. What proportion of people with hip and knee osteoarthritis meet physical activity guidelines? A systematic review and meta-analysis. Osteoarthritis Cartilage 2013 Nov;21(11):1648-1659. [doi: 10.1016/j.joca.2013.08.003]
10. Jakiela J, Waugh E, White D. Walk at least 10 minutes a day for adults with knee osteoarthritis: Recommendation for minimal activity during the COVID-19 pandemic. J Rheumatol 2021 Feb;48(2):157-159. [FREE Full text] [doi: 10.3899/jrheum.200914] [Medline: 32801138]
11. Li LC, Sayre EC, Xie H, Falck RS, Best JR, Liu-Ambrose T, et al. Efficacy of a community-based technology-enabled physical activity counseling program for people with knee osteoarthritis: Proof-of-concept Study. J Med Internet Res 2018 Apr 30;20(4):e159. [doi: 10.2196/jmir.8514] [Medline: 29712630]
12. Conn VS, Hafdahl AR, Minor MA, Nielsen PJ. Physical activity interventions among adults with arthritis: meta-analysis of outcomes. Semin Arthritis Rheum 2008 Apr;37(5):307-316. [doi: 10.1016/j.semarthrit.2007.07.006] [Medline: 17888500]
13. Li LC, Feehan LM, Xie H, Lu N, Shaw CD, Gromala D, et al. Effects of a 12-week multifaceted wearable-based program for people with knee osteoarthritis: Randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 03;8(7):e19116. [FREE Full text] [doi: 10.2196/19116] [Medline: 32618578]

14. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: Systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819. [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
15. Chaudhry UAR, Wahlich C, Fortescue R, Cook DG, Knightly R, Harris T. The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults. Int J Behav Nutr Phys Act 2020 Oct 09;17(1):129 [FREE Full text] [doi: 10.1186/s12966-020-01020-8] [Medline: 33036635]
16. 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]
17. Davergne T, Pallot A, Dechartres A, Fautrel B, Gossec L. Use of wearable activity trackers to improve physical activity behavior in patients with rheumatic and musculoskeletal diseases: A systematic review and meta-analysis. Arthritis Care Res 2019 Apr 23;71(6):758-767. [doi: 10.1002/acr.23752]
18. Lupton D. The digitally engaged patient: Self-monitoring and self-care in the digital health era. Soc Theory Health 2013 Jun 19;11(3):256-270. [doi: 10.1057/sth.2013.10]
19. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: Promises and barriers. PLoS Med 2016 Feb;13(2):e1001953. [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
20. Leese J, Geldmanc J, Zhu S, Macdonald G, Pourrahmat M, Townsend A, et al. The perspectives of persons with arthritis on the use of wearable technology to self-monitor physical activity: A qualitative evidence synthesis. Arthritis Care Res (Hoboken) 2021 Feb 28 (forthcoming). [doi: 10.1002/acr.24585] [Medline: 33644994]
21. Papi E, Belsi A, McGregor AH. A knee monitoring device and the preferences of patients living with osteoarthritis: a qualitative study. BMJ Open 2015 Sep 07;5(9):e007980. [doi: 10.1136/bmjopen-2015-007980]
22. Belsi A, Papi E, McGregor AH. Impact of wearable technology on psychosocial factors of osteoarthritis management: a qualitative study. BMJ Open 2016 Feb 03;6(2):e010064. [doi: 10.1136/bmjopen-2015-010064]
23. Leese J, Macdonald GG, Tran BC, Wong R, Backman CL, Townsend AF, et al. Using physical activity trackers in arthritis self-management: A qualitative study of patient and rehabilitation professional perspectives. Arthritis Care Res 2019 Jan 29;71(2):227-236. [doi: 10.1002/acr.23780]
24. Nimmon L, Stenfors-Hayes T. The "Handling" of power in the physician-patient encounter: perceptions from experienced physicians. BMC Med Educ 2016 Apr 18;16:114. [FREE Full text] [doi: 10.1186/s12909-016-0634-0] [Medline: 27091146]
25. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: A mixed-methods evaluation. JMIR mHealth uHealth 2016 Jan 27;4(1):e7. [doi: 10.2196/mhealth.4225]
26. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
27. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g]
28. Varpio L, Paradis E, Uijtdehaage S, Young M. The distinctions between theory, theoretical framework, and conceptual framework. Acad Med 2020 Jul;95(7):989-994. [doi: 10.1097/ACM.0000000000003075] [Medline: 31725464]
29. Marton F. Phenomenography—A research approach to investigating different understandings of reality. Journal of Thought 1986;21(3):49 [FREE Full text]
30. Bergum V, Dossetor J. Relational Engagement. In: Relational Ethics: The Full Meaning of Respect. Hagerstown, MD: University Publishing Group; 2005:103-135.
31. Beauchamp T, Childress J. Principles of biomedical ethics. 6th ed. New York: Oxford University Press; 2009.
32. Nedelsky J. Reconceiving autonomy. Yale J Law Fem 1989;1(1):7. [doi: 10.1093/acprof:oso/9780195147964.003.0004]
33. Donchin A. Reworking autonomy: Toward a feminist perspective. New York, USA: Camb Q Healthc Ethics; 1995:44.
34. Bergum V. Relational ethics for health care. Toronto, ON: Pearson; 2013:127.
35. Miller W, Rollnick S. Motivational interviewing: Helping people change. New York: Guildford Press; 2013.
36. Gutnick D, Reims K, Davis C, Gainforth H, Jay M, Cole S. Brief action planning to facilitate behaviour change and support patient self-management. Journal of Clinical Outcomes Management 2014;21:17-29 [FREE Full text]
37. Marra CA, Cibere J, Tsuyuki RT, Soon JA, Esdaile JM, Gastonguay L, et al. Improving osteoarthritis detection in the community: Pharmacist identification of new, diagnostically confirmed osteoarthritis. Arthritis Rheum 2007 Oct 15;57(7):1238-1244. [doi: 10.1002/art.23019]
38. Stenfors-Hayes T, Hult H, Dahlgren M. A phenomenographic approach to research in medical education: A phenomenographic approach to research. Med Edu 2013;47(3):261-270. [doi: 10.1111/medu.12101]
39. McDonald HN, Dietrich T, Townsend A, Li LC, Cox S, Backman CL. Exploring occupational disruption among women after onset of rheumatoid arthritis. Arthritis Care Res (Hoboken) 2012 Feb;64(2):197-205. [FREE Full text] [doi: 10.1002/acr.20668] [Medline: 22006430]
40. Eurenius E, Biguet G, Stenström CH. Attitudes toward physical activity among people with rheumatoid arthritis. Physiother Theory Pract 2009 Jul 10;19(1):53-62. [doi: 10.1080/09593980307967]

41. Sjostrom B, Dahlgren LO. Applying phenomenography in nursing research. J Adv Nurs 2002 Nov;40(3):339-345. [doi: 10.1046/j.1365-2648.2002.02375.x]
42. Heaton J. SAGE research methods core. In: Reworking qualitative data. London: Thousand Oaks; 2004.
43. Heaton J. Secondary analysis of qualitative data: An overview. Hist Soc Res 2008;33(3):33-45.
44. Austin W. Relational ethics in forensic psychiatric settings. J Psychosoc Nurs Ment Health Serv 2001 Sep;39(9):12-17. [doi: 10.3928/0279-3695-20010901-04]
45. Sandbergh J. Are phenomenographic results reliable? High Educ Res Dev 2006 Nov;16(2):203-212. [doi: 10.1080/0729436970160207]
46. Leese J, Li L, Nimmon L, Townsend A, Backman C. Moving beyond "Until Saturation Was Reached": Critically examining how saturation is used and reported in qualitative research. Arthritis Care Res (Hoboken) 2021 Sep;73(9):1225-1227. [doi: 10.1002/acr.24600] [Medline: 33756068]
47. O'Reilly M, Parker N. 'Unsatisfactory Saturation': a critical exploration of the notion of saturated sample sizes in qualitative research. Qual Res 2012 May 17;13(2):190-197. [doi: 10.1177/1468794112446106]
48. Morse JM. The significance of saturation. Qual Health Res 1995 May 01;5(2):147-149. [doi: 10.1177/104973239500500201]
49. Bowen GA. Naturalistic inquiry and the saturation concept: a research note. Qual Res 2008 Feb;8(1):137-152. [doi: 10.1177/1468794107085301]
50. Dunkin R. Using phenomenography to study organisational change. In: Bowen JA, Walsh E, editors. Phenomenography. Melbourne: RMIT University Press; 2000:94-103.
51. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822x05279903]
52. MacDonald C. Nurse autonomy as relational. Nurs Ethics 2016 Aug 18;9(2):194-201. [doi: 10.1191/0969733002ne498oa]
53. Entwistle VA, Carter SM, Cribb A, McCaffery K. Supporting patient autonomy: The importance of clinician-patient relationships. J Gen Intern Med 2010 Mar 6;25(7):741-745. [doi: 10.1007/s11606-010-1292-2]
54. Entwistle VA, Watt IS. Treating patients as persons: a capabilities approach to support delivery of person-centered care. Am J Bioeth 2013;13(8):29-39. [FREE Full text] [doi: 10.1080/15265161.2013.802060] [Medline: 23862598]
55. Gabriels K, Moerenhout T. Exploring entertainment medicine and professionalization of self-care: Interview study among doctors on the potential effects of digital self-tracking. J Med Internet Res 2018 Jan 12;20(1):e10. [FREE Full text] [doi: 10.2196/jmir.8040] [Medline: 29330140]
56. Li LC, Feehan LM, Hoens AM. Rethinking physical activity promotion during the COVID-19 pandemic: Focus on a 24-hour day. J Rheumatol 2021 Aug 15;48(8):1205-1207. [FREE Full text] [doi: 10.3899/jrheum.201595] [Medline: 33589559]
57. Feehan L, Lu N, Xie H, Li L. Twenty-four-hour activity and sleep profiles for adults living with arthritis: Habits matter. Arthritis Care Res (Hoboken) 2020 Dec;72(12):1678-1686. [doi: 10.1002/acr.24424] [Medline: 33025679]
58. Lorig K, Gonzalez V, Ritter P. Community-based Spanish language arthritis education program: A randomized trial. Med Care 1999 Sep;37(9):957-963. [doi: 10.1097/00005650-199909000-00011] [Medline: 10493473]
59. Lenker S, Lorig K, Gallagher D. Reasons for the lack of association between changes in health behavior and improved health status: An exploratory study. Patient Edu Couns 1984 Jan;6(2):69-72. [doi: 10.1016/0738-3991(84)90036-3]
60. Bandura A. Self-efficacy: The exercise of control. New York: W. H. Freeman; 1997.
61. Pellegrini CA, Ledford G, Chang RW, Cameron KA. Understanding barriers and facilitators to healthy eating and physical activity from patients either before and after knee arthroplasty. Disabil Rehabil 2017 May 05;40(17):2004-2010. [doi: 10.1080/09638288.2017.1323026]
62. Kunyk D, Austin W. Nursing under the influence. Nurs Ethics 2011 Jun 06;19(3):380-389. [doi: 10.1177/0969733011406767]
63. Ellis C. Telling secrets, revealing lives. Qual Inq 2016 Jun 29;13(1):3-29. [doi: 10.1177/1077800406294947]
64. Meloni F, Vanthuyne K, Rousseau C. Towards a relational ethics: Rethinking ethics, agency and dependency in research with children and youth. Anthropol Theory 2015 Mar 16;15(1):106-123. [doi: 10.1177/1463499614565945]
65. Redman BK. Responsibility for control; ethics of patient preparation for self-management of chronic disease. Bioethics 2007 Jun;21(5):243-250. [doi: 10.1111/j.1467-8519.2007.00550.x] [Medline: 17845469]
66. Redman BK. The ethics of self-management preparation for chronic illness. Nurs Ethics 2016 Aug 18;12(4):360-369. [doi: 10.1191/0969733005ne801oa]

Abbreviations

MVPA: moderate-to-vigorous physical activity
OA: osteoarthritis
PT: physiotherapist
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 11.06.21; peer-reviewed by K Stampe; comments to author 05.07.21; revised version received 27.08.21; accepted 19.09.21; published 12.11.21.

Please cite as:
Leese J, MacDonald G, Backman CL, Townsend A, Nimmon L, Li LC
Experiences of Wearable Technology by Persons with Knee Osteoarthritis Participating in a Physical Activity Counseling Intervention: Qualitative Study Using a Relational Ethics Lens
JMIR Mhealth Uhealth 2021;9(11):e30332
URL: https://mhealth.jmir.org/2021/11/e30332
doi:10.2196/30332
PMID:34766912

Original Paper

Feasibility, Usability, and Effectiveness of a Machine Learning–Based Physical Activity Chatbot: Quasi-Experimental Study

Quyen G To[1], PhD; Chelsea Green[1], BSc; Corneel Vandelanotte[1], PhD

Physical Activity Research Group, Appleton Institute, Central Queensland University, Rockhampton, Australia

Corresponding Author:
Quyen G To, PhD
Physical Activity Research Group
Appleton Institute
Central Queensland University
554-700 Yaamba Rd, Norman Gardens
Rockhampton, 4701
Australia
Phone: 61 7 4930 6456
Email: q.to@cqu.edu.au

Abstract

Background: Behavioral eHealth and mobile health interventions have been moderately successful in increasing physical activity, although opportunities for further improvement remain to be discussed. Chatbots equipped with natural language processing can interact and engage with users and help continuously monitor physical activity by using data from wearable sensors and smartphones. However, a limited number of studies have evaluated the effectiveness of chatbot interventions on physical activity.

Objective: This study aims to investigate the feasibility, usability, and effectiveness of a machine learning–based physical activity chatbot.

Methods: A quasi-experimental design without a control group was conducted with outcomes evaluated at baseline and 6 weeks. Participants wore a Fitbit Flex 1 (Fitbit LLC) and connected to the chatbot via the Messenger app. The chatbot provided daily updates on the physical activity level for self-monitoring, sent out daily motivational messages in relation to goal achievement, and automatically adjusted the daily goals based on physical activity levels in the last 7 days. When requested by the participants, the chatbot also provided sources of information on the benefits of physical activity, sent general motivational messages, and checked participants' activity history (ie, the step counts/min that were achieved on any day). Information about usability and acceptability was self-reported. The main outcomes were daily step counts recorded by the Fitbit and self-reported physical activity.

Results: Among 116 participants, 95 (81.9%) were female, 85 (73.3%) were in a relationship, 101 (87.1%) were White, and 82 (70.7%) were full-time workers. Their average age was 49.1 (SD 9.3) years with an average BMI of 32.5 (SD 8.0) kg/m2. Most experienced technical issues were due to an unexpected change in Facebook policy (93/113, 82.3%). Most of the participants scored the usability of the chatbot (101/113, 89.4%) and the Fitbit (99/113, 87.6%) as at least "OK." About one-third (40/113, 35.4%) would continue to use the chatbot in the future, and 53.1% (60/113) agreed that the chatbot helped them become more active. On average, 6.7 (SD 7.0) messages/week were sent to the chatbot and 5.1 (SD 7.4) min/day were spent using the chatbot. At follow-up, participants recorded more steps (increase of 627, 95% CI 219-1035 steps/day) and total physical activity (increase of 154.2 min/week; 3.58 times higher at follow-up; 95% CI 2.28-5.63). Participants were also more likely to meet the physical activity guidelines (odds ratio 6.37, 95% CI 3.31-12.27) at follow-up.

Conclusions: The machine learning–based physical activity chatbot was able to significantly increase participants' physical activity and was moderately accepted by the participants. However, the Facebook policy change undermined the chatbot functionality and indicated the need to use independent platforms for chatbot deployment to ensure successful delivery of this type of intervention.

(JMIR Mhealth Uhealth 2021;9(11):e28577) doi:10.2196/28577

KEYWORDS

conversational agent; virtual coach; intervention; exercise; acceptability; mobile phone

Introduction

Background

It has been established that physical activity reduces the risk of mortality and many health conditions such as cardiovascular diseases, type 2 diabetes, and cancer [1]. However, less than half of Australian adults meet the physical activity guidelines of at least 150 minutes of vigorous-moderate-intensity physical activity per week [2]. It was estimated that physical inactivity accounted for 53.8 billion in health care costs and an additional 13.7 billion in productivity losses worldwide in 2013 [3]. Therefore, interventions to increase physical activity are needed [4]. To date, many of these interventions have been delivered face-to-face and are expensive [5]. Therefore, there is a need for low-cost interventions targeting large populations.

With the advancement of mobile technology, people can access the internet almost everywhere and at any time. It is estimated in 2019 that 4.48 billion people are active internet users, 4.07 billion are unique mobile internet users, and 3.66 billion are active mobile social media users [6]. This indicates that mobile health (mHealth) has the potential to offer a great platform for behavior change interventions that can reach a large number of people at a low cost. In the last decade, many eHealth and mHealth interventions targeting physical activity have been examined [7-10], many of which use email, SMS, and websites as delivery tools. Overall, these interventions have been able to produce moderate effect sizes in increasing physical activity [7-10]. As such, there is still room to further increase the effectiveness of behavioral eHealth and mHealth interventions. One often cited problem in this area is the low levels of engagement and interaction with eHealth and mHealth interventions [11]. As there is evidence that the more participants use the interventions, the more effective the interventions tend to be [5,12], an important aim is to design eHealth and mHealth interventions that will lead to higher levels of engagement.

The use of chatbots is a potential innovative avenue for achieving higher levels of engagement. A chatbot or conversational agent is a computer program that can interact with users [13]. Equipped with natural language processing capability, a modern chatbot can effectively engage in conversations with users [14]. Chatbots can help save human resources while providing instant responses to requests. In particular, chatbots can also help users monitor participants' progress by continuously evaluating physical activity data from wearable sensors and smartphones. Applying machine learning algorithms can also enable chatbots to provide personalized activity recommendations to a specific user. Chatbots can be embedded into different platforms, such as websites, apps, messaging programs or other social media to reach large numbers of people easily and conveniently. As such, chatbots have been adopted across many industries, such as finance, e-commerce, and health care [14-16].

Recent reviews indicate that health behavior change interventions using chatbots have mostly focused on mental health [17,18]. Among the few studies that used chatbots to promote physical activity and healthy diet [19], only 2 evaluated increases in physical activity [20,21]. However, the study conducted in Switzerland was designed to test differences in daily step goals among 3 groups (cash incentives vs charity incentives vs no incentives) rather than the effectiveness of the chatbot [21]. Only one study in Australia evaluated the effectiveness of a chatbot in improving diet and physical activity [20]. Although this study focused more on diet than physical activity, it did show a large increase in physical activity (approximately 110 moderate-to-vigorous physical activity min/week). However, the chatbot evaluated in this study did not provide automatic daily updates that remind the participants about their physical activity goals and did not automatically adjust participants' goals based on their current physical activity level.

Objectives

Given the lack of studies on the effectiveness of physical activity chatbots, the aim of this study is to investigate the feasibility, usability, and effectiveness of an interactive machine learning–based physical activity chatbot that uses natural language processing and adaptive goal setting.

Methods

Study Design and Participants

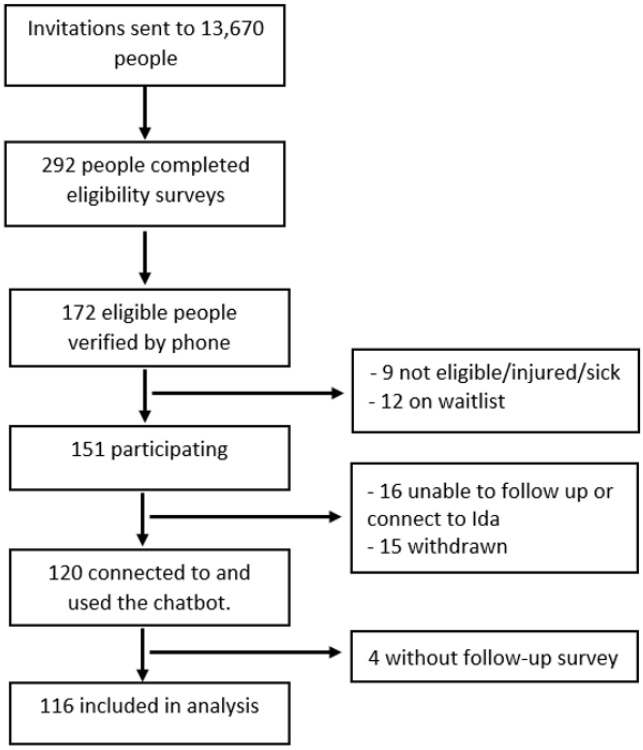
A quasi-experimental design without a control group was conducted with outcomes evaluated at 2 time points—baseline and 6 weeks after participants started to use the chatbot. Prospective participants were recruited from a list of people who had previously used the 10,000 Steps program [22]. To be eligible, potential participants had to be inactive (<20 min/day of moderate-to-vigorous physical activity), live in Australia, have internet access and a smartphone, aged at least 18 years, self-reported motivation to improve physical activity (targeting those in need of support to become more active), not already participating in another physical activity program, not already owning and used a physical activity tracking device (eg, pedometer, Fitbit [Fitbit LLC], and Garmin) within the last 12 months, and able to safely increase their activity levels. Those who were interested in the study and clicked on the link attached to the invitation emails were directed to a web-based survey. Prospective participants were provided with a participant information sheet and contact details of the research team and then asked to answer a series of screening questions to assess eligibility. If eligible, they completed a web-based consent form and baseline survey questions. After 6 weeks, the participants were asked to complete a follow-up web-based survey to assess changes over time.

Owing to an unexpected Facebook policy change (we used the Messenger app to host the chatbot, which is owned by Facebook) that blocked the chatbot from sending out new messages to

participants who did not respond to the previous message within 24 hours, we were forced to stop the study at that point. As the recruitment was rolling, 48 participants had already completed the study when the Facebook policy change was implemented. For those who were still engaged in the study at that time, a follow-up survey was sent to them immediately at the time of implementation of this policy, resulting in a shorter intervention period.

Invitation emails were sent to 13,670 email addresses registered in the 10,000 Steps program database between September and November 2020 (Figure 1). A total of 2.14% (292/13, 670) of people completed the eligibility survey during the recruitment period, with 58.9% (172/292) people deemed eligible. Eligible people were contacted by phone for verification. This resulted in 9 people being excluded from the study because they did not meet the eligibility criteria upon verification, were no longer interested, or had an illness or injury that prevented them from taking part. When recruitment closed, 12 people were placed on a waitlist. Another 16 people were excluded because they were unable to be recontacted or to connect to the chatbot, and 15 withdrew because of an illness or personal issues. As a result, 120 participants were enrolled at baseline. However, only 116 completed the follow-up surveys.

Figure 1. Participant flowchart.



This study was approved by the Human Research Ethics Committee of the Central Queensland University (application #0000022181). This study was retrospectively registered on the Australian New Zealand Clinical Trials Registry (ACTRN12621000345886).

Procedures

Participants who agreed to participate, provided written consent, and completed the baseline survey were mailed a package including a Fitbit Flex 1 activity tracker (with instructions on how to use it), a participant information sheet, and instructions on how to download the Fitbit app on their smartphone and how to create a Fitbit account. Follow-up phone calls were conducted to ensure that participants received the package and were able to install the Fitbit app and use the Fitbit device.

Participants wore their Fitbit for 7 days to collect their baseline physical activity data before connecting to the chatbot. To connect to the chatbot, participants were instructed to download and open the Messenger app on their smartphones and complete the secure verification process (only study participants were able to connect with the chatbot). Once verified, the participants started to receive daily messages and were able to interact with the chatbot. Participants were asked to engage with the chatbot (intervention) for a period of 6 weeks.

Follow-up surveys were sent to the participants via email. Four reminders (a combination of text messages, email, and phone calls), each of which was 3 days apart, were sent to ask participants to complete the follow-up survey. A research assistant was available during the intervention period to assist participants with any technical issues that the participants may encounter.

Intervention

The chatbot, named *Ida*, was created and technically managed by an Australian company called SmartAI. However, the natural language processing capability was powered by Dialogflow (Google Inc), an advanced Google machine learning platform for creating conversational artificial intelligence applications. The Facebook Messenger app was selected for the deployment of the chatbot because of its popularity. The Fitbit Flex 1 device

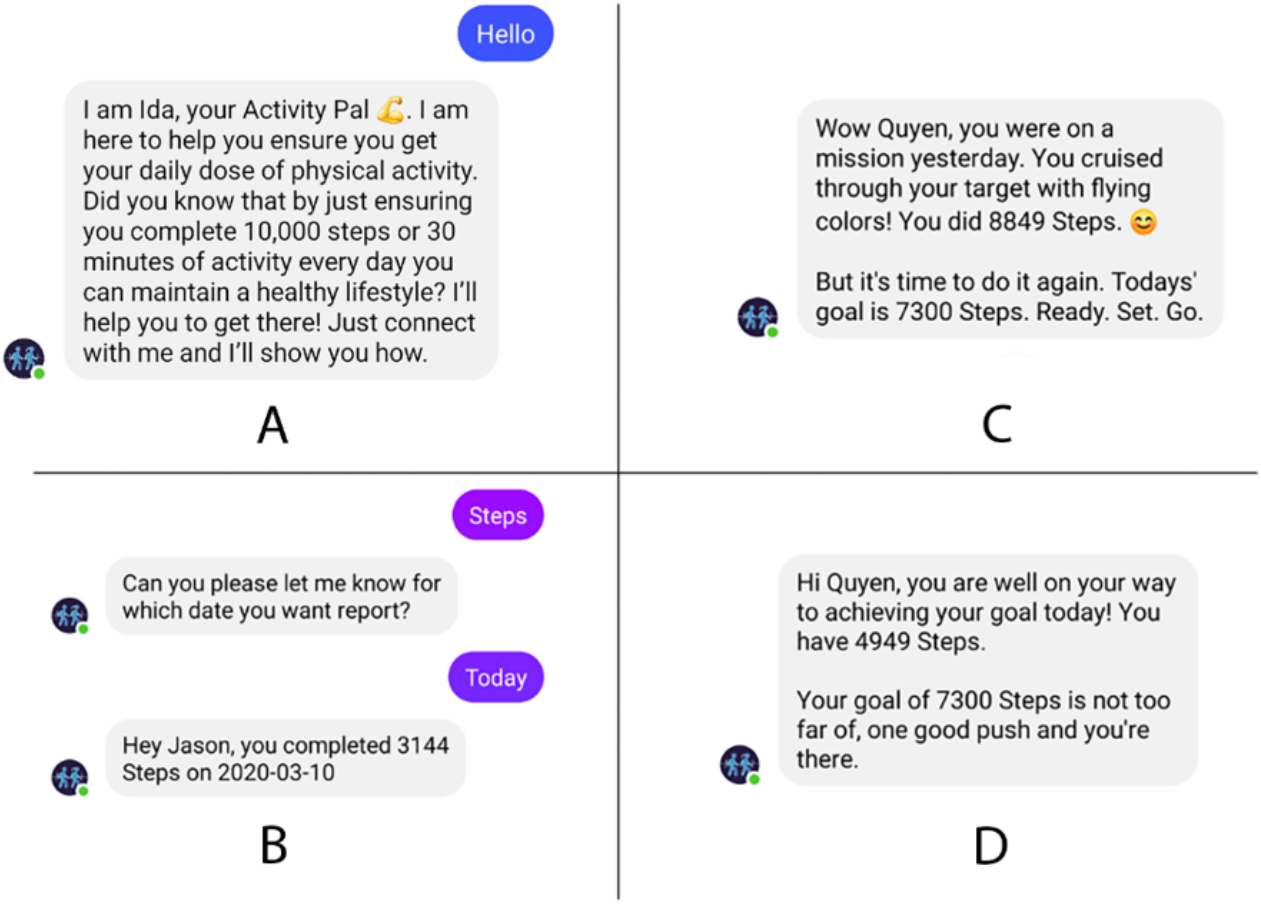
was used to measure the participants' daily physical activity. Fitbit activity data were synced from the Fitbit platform to the chatbot platform and used by the chatbot to monitor the participants' progress over time.

The intervention was designed using the COM-B model. The COM-B model forms the core of the Behavior Change Wheel, a behavioral system focusing on 3 components: capability, opportunity, and motivation [23]. As explained below, the messages delivered by the chatbot aimed to increase participants' motivation (ie, motivational messages), capability (ie, through ongoing adaptive feedback on goal achievement), and opportunity (ie, educational content and activity reminders throughout the day helped participants become more aware of physical activity opportunities). The chatbot supported participants through 2 groups of actions: proactive and reactive.

Proactive actions include the following: (1) Providing an update on participants' physical activity level achieved the previous day and informing them of the goal they needed to achieve on the current day. This message was sent early in the morning at the time selected by each participant. (2) Sending out 1 or 2 additional messages later in the day to encourage participants trying to achieve their daily goal or indicate they were doing great and had already achieved the goal when the message was sent. The number of messages and times was selected by each participant. (3) Automatically adjusting the daily activity goals based on the average physical activity level achieved during the 7 previous days. The type of goal (step counts or minutes) and the amount per day (eg, 8000 steps/day or 35 min/day) that the participant wanted to achieve by the end of the study was also chosen by each participant. The goal was automatically adjusted to increase by 500 steps/day or 5 minutes of moderate-vigorous physical activity/day if the participant, on average, met their current goal over the last 7 days [24,25]. If not, the same goal was used. We used a combination of moderate and vigorous physical activity assessed by Fitbit to calculate the physical activity min/week. The information needed to personalize the chatbot was collected during the verification phone calls and added to the participants' profile page on the chatbot platform, which was only accessible to the research team.

Reactive actions, which occurred when the participants sent a request for information to the chatbot, include (1) providing sources of information on the benefits of physical activity, (2) sending general motivational messages to encourage participants to become more active, and (3) checking participants' activity history (ie, the step counts or minutes that were achieved on any day) as requested. Examples of these messages are shown in Figure 2.

Figure 2. Message examples: (A) introduction, (B) request on step counts, (C) message upon reaching the goal, (D) message encouraging the participant to try reaching the goal.



Measures

Demographic characteristics were self-reported at baseline. Age, height, weight, years of schooling, and average daily work time (hours) were used as continuous variables; categorical variables included gender (male or female), marital status (not in a relationship or in a relationship), ethnicity (White or other), living area (major city, regional, or remote area), work status (full-time or other), and annual household income (≥Aus $130,000 [US $94,900], Aus $78,000 to <Aus $130,000 [US $56,940- $94900], or <Aus $78,000 [US $56,940]). Weight was also self-reported at follow-up. BMI was calculated as weight (kg)/height (m^2) and was analyzed as a secondary outcome.

Physical activity was objectively measured using the Fitbit Flex 1. Although this device records both step counts and physical activity minutes, only step counts were used in the analysis. This is because the Fitbit only recorded the minutes if a user was active for at least 10 minutes, whereas all steps were counted regardless of whether they occurred during bouts of activity (10 minutes) or not.

Self-reported physical activity was assessed at baseline (before receiving a Fitbit) and follow-up using the Active Australia Survey [26]. These questions asked about minutes participants spent on walking, moderate and vigorous physical activity per day, and the number of days they spent engaging in these activities in the last week. The total amount of time spent engaging in walking and moderate and vigorous physical activity in a week was calculated by adding the above times (with vigorous physical activity time doubled as per scoring instructions) [26]. Participants who spent at least 150 minutes of moderate-to-vigorous physical activity per week were categorized as meeting the Australian physical activity guidelines.

Usability and acceptability were assessed at follow-up using the System Usability Scale (SUS) [27] and other self-reported questions. The SUS includes 10 questions with 5 response options from *strongly agree* to *strongly disagree*. As recommended, the original scores for each question were converted to new scores, summed, and multiplied by 2.5, to generate an SUS score between 0 and 100 [27]. We used the cutoffs suggested by Bangor et al [28] to classify the SUS scores into 4 groups: excellent (85.58-100), good (72.75-85.57), OK (52.01-72.74), and poor (0-52.00). Other self-reported questions asked about the usefulness of the chatbot, willingness to use the chatbot in the future and recommend it to others, whether a participant experienced any technical issues, Fitbit wear time, and frequency of using the Fitbit app.

Power and Data Analysis

Posthoc power calculation was conducted for Fitbit step counts using the following parameters: difference in means, SDs, and correlation between step counts at 2 time points. The posthoc power for this study was 81.3%.

Fitbit data were cleaned and processed using the Python v3.7 (Python Software Foundation). As step counts of <1000 indicate that the Fitbit was not worn all day [29,30], these counts were removed. A 7-day moving average for the daily mean steps was generated and used to show changes in steps over the study period. Average step counts were also calculated for weeks with at least 4 days of valid data; however, only data in the first week as baseline data (before the date that participants connected to the chatbot) and the last week of participation as follow-up data were used in the outcome analysis. As participation duration was different among the participants, doing this allowed the analysis to be performed for all participants together.

SAS v9.4 (SAS Institute) was used for the analysis. Baseline characteristics were compared among those participating <4 weeks, 4 to <6 weeks, and ≥6 weeks using Fisher exact tests for categorical variables and Welch analysis of variance for continuous variables, except for daily work time and total physical activity minutes, which were tested using Kruskal–Wallis tests. As a robustness check, the analysis was performed separately for 2 samples, a full sample and a subsample (excluding those using the chatbot <4 weeks). This ensures that the results reflect the effectiveness of the intervention for those with sufficient exposure to the chatbot.

Generalized linear mixed models were used to identify changes in the outcomes. Normal distribution and identity link were used for BMI and Fitbit step counts. As total physical activity minutes were highly skewed, PROC TRANSREG was used to conduct the Box-Cox transformation analysis, and as a result, a fourth root transformation was applied. Generalized linear mixed models with normal distribution and log link were used for the transformed total physical activity minutes. Estimates were converted back into ratios for interpretative purposes. Empirical estimators were used to obtain the robust SEs. Binary distribution and logit link were used to determine the outcome of meeting physical activity guidelines. For each outcome, 2 models were run to generate crude estimates and estimates adjusted for sample characteristics including age, gender, marital status, years of schooling, ethnicity, household income, living area, work status, and daily work time. Differences in BMI, step counts, and total physical activity minutes between the follow-up and baseline were reported with a 95% CI. Odds ratios (ORs) and 95% CIs were reported for meeting the physical activity guidelines. All *P* values were 2-sided and considered significant if <.05.

Results

Baseline Characteristics

Table 1 presents the baseline characteristics of the sample. Most of the participants were female (95/116, 81.9%), in a relationship (85/116, 73.3%), White (101/116, 87.1%), and full-time workers (82/116, 70.7%). The participants had an average age of 49.1 (SD 9.3) years, with an average BMI of 32.5 (SD 8.0) kg/m^2, and 81.9% (95/116) of the participants were either overweight or obese. The average step count was <6000 (SD 2391) steps/day. Only 13.8% (16/116) of the participants met the physical activity guideline. There were no significant differences in these characteristics among those with different participation durations (*P*>.05). Among the 116 participants, 17 participated in <4 weeks, 51 between 4 and <6 weeks, and 48 at least 6 weeks.

Table 1. Characteristics at baseline by participation duration (N=116).

	All (N=116)	<4 weeks (n=17)	4-<6 weeks (n=51)	At least 6 weeks (n=48)	*P* value
Gender, n (%)					.49
Male	21 (18.1)	2 (11.8)	12 (23.5)	7 (14.6)	
Female	95 (81.9)	15 (88.2)	39 (76.5)	41 (85.4)	
Marital status, n (%)					.17
Not in a relationship	31 (26.7)	7 (41.2)	15 (29.4)	9 (18.8)	
In a relationship	85 (73.3)	10 (58.8)	36 (70.6)	39 (81.3)	
Ethnicity, n (%)					.27
White	101 (87.1)	13 (76.5)	44 (86.3)	44 (91.7)	
Others	15 (12.9)	4 (23.5)	7 (13.7)	4 (8.3)	
Living areas, n (%)					.72
Major city	56 (48.3)	9 (52.9)	26 (51)	21 (43.8)	
Regional or remote areas	60 (51.7)	8 (47.1)	25 (49)	27 (56.3)	
Work status, n (%)					.53
Full-time	82 (70.7)	14 (82.4)	34 (66.7)	34 (70.8)	
Others	34 (29.3)	3 (17.7)	17 (33.3)	14 (29.2)	
Annual household income in Aus $ (US $), n (%)					.85
≥130,000 (≥94,900)	35 (30.2)	6 (35.3)	14 (27.5)	15 (31.3)	
78,000 to <130,000 (56,940-94,900)	39 (33.6)	5 (29.4)	20 (39.2)	14 (29.2)	
<78,000 (<56,940)	42 (36.2)	6 (35.3)	17 (33.3)	19 (39.6)	
Average age (years), mean (SD)	49.1 (9.3)	48.9 (11.0)	50.3 (9.0)	48.1 (9.0)	.48
Average height (cm), mean (SD)	167.4 (8.9)	163.7 (11.6)	169.4 (9.2)	166.7 (7.0)	.11
Average weight (kg), mean (SD)	91.3 (24.7)	94.3 (18.1)	91.3 (27.8)	90.3 (23.5)	.77
Average BMI (kg/m^2), mean (SD)	32.5 (8.0)	35.3 (6.7)	31.7 (8.6)	32.4 (7.7)	.19
Average years of schooling, mean (SD)	15.8 (3.5)	15.5 (3.8)	15.6 (3.4)	16.0 (3.6)	.78
Average daily work time (h/day), mean (SD)	8.0 (2.0)[a]	8.0 (1.9)	7.8 (1.8)[b]	8.0 (2.2)[c]	.97[d]
Average step counts/day, mean (SD)	5933 (2391)[e]	5761 (2076)[f]	6466 (2800)[g]	5428 (1895)[h]	.12
Average total physical activity (min/week), mean (SD)	86.5 (137.5)	72.4 (58.0)	91.4 (143.8)	86.3 (151.8)	.80[i]
Met physical activity recommendation, n (%)					.12
No	100 (86.2)	16 (94.1)	40 (78.4)	44 (91.7)	
Yes	16 (13.8)	1 (5.9)	11 (21.6)	4 (8.3)	

[a]n=106.
[b]n=45.
[c]n=44.
[d]Fisher Exact tests or Welch analysis of variance was used unless indicated otherwise.
[e]n=108.
[f]n=14.
[g]n=48.
[h]n=46.
[i]Kruskal–Wallis test was used.

Usability and Acceptability

Table 2 shows data on process evaluation on the implementation of the intervention. The average usability score for the chatbot was 61.6 (9.7) with majority of the participants scoring the chatbot as *OK* (89/113, 78.8%) or *good* (12/113, 10.6%). Although less than half would recommend the chatbot to others (49/113, 43.4%) and about one-third (40/113, 35.4%) would

continue to use the chatbot in the future, more than half (60/113, 53.1%) agreed that the chatbot helped them become more active. About one-third thought the chatbot was quite or very useful in helping them increase confidence for engaging in regular physical activity, and in helping them stay motivated to participate in physical activity. About one-quarter thought the chatbot was useful in helping them overcome barriers, increase support they receive, and plan for physical activity during the study period. Most of the participants (106/113, 93.8%) read the messages that the chatbot sent out, and about half of the participants sent messages to the chatbot at least once a day. On average, the participants sent 6.7 messages to the chatbot per week and spent 5.1 minutes with the chatbot per day. About one-quarter liked very much the messages that the chatbot sent out. However, only 43.4% (49/113) thought that the chatbot understood their messages most of the time. Most participants experienced technical issues (93/113, 82.3%) and stopped receiving the chatbot messages at any time during the study (95/113, 84.1%).

Table 2. Usability and acceptability of the chatbot and Fitbit.

	Value
System Usability Scale score for the chatbot (n=113), n (%)	
Good	12 (10.6)
OK	89 (78.8)
Poor	12 (10.6)
Would recommend chatbot to others (n=113), n (%)	
Strongly agree or agree	49 (43.4)
Neutral	36 (31.9)
Strongly disagree or disagree	28 (24.7)
Would continue to use the chatbot in future (n=113), n (%)	
Strongly agree or agree	40 (35.4)
Neutral	32 (28.3)
Strongly disagree or disagree	41 (36.3)
The chatbot helped me to be more active (n=113), n (%)	
Strongly agree or agree	60 (53.1)
Neutral	27 (23.9)
Strongly disagree or disagree	26 (23)
Usefulness—the chatbot helps me to increase confidence for physical activity participation (n=113), n (%)	
Not at all useful or a little useful	52 (46)
Somewhat useful	27 (23.9)
Quite useful or very useful	34 (30.1)
Usefulness—the chatbot helps me to overcome barriers to physical activity participation (n=113), n (%)	
Not at all useful or a little useful	60 (53.1)
Somewhat useful	24 (21.2)
Quite useful or very useful	29 (25.7)
Usefulness—the chatbot increased support for being active (n=113), n (%)	
Not at all useful or a little useful	59 (52.2)
Somewhat useful	26 (23)
Quite useful or very useful	28 (24.8)
Usefulness—the chatbot helped me plan to be active (n=113), n (%)	
Not at all useful or a little useful	63 (55.7)
Somewhat useful	20 (17.7)
Quite useful or very useful	30 (26.6)
Usefulness—the chatbot helped me to stay motivated (n=113), n (%)	
Not at all useful or a little useful	47 (41.6)
Somewhat useful	26 (23)
Quite useful or very useful	40 (35.4)
Read the chatbot messages (n=113), n (%)	
Always	71 (62.8)
Most of the time	35 (31)
Sometimes or rarely	7 (6.2)
Frequency of sending messages to chatbot (n=77), n (%)	
Several times a day	30 (26.6)

	Value
Once a day	28 (24.8)
Less than once a day	19 (48.6)
Average messages/week sent to the chatbot (n=113), mean (SD)	6.7 (7.0)
Average time/day spent with the chatbot (minutes; n=113), mean (SD)	5.1 (7.4)
Liked the chatbot messages (n=113), n (%)	
Very much	26 (23)
Average	42 (37.2)
A little or not at all	45 (39.8)
Understood the chatbot messages (n=113), n (%)	
Always or most of the time	49 (43.3)
Sometimes	34 (30.1)
Rarely or never	30 (26.5)
Technical issues during the study (n=113), n (%)	
Yes	93 (82.3)
No	20 (17.7)
Chatbot stopped sending motivational messages or updates at any time (n=113), n (%)	
Yes	95 (84.1)
No	18 (15.9)
System Usability Scale for Fitbit Flex 1 (n=112), n (%)	
Good	22 (19.6)
OK	77 (68.8)
Poor	13 (11.6)
Average weeks of wearing the Fitbit (n=112), mean (SD)	5.4 (1.1)
Average day/week of wearing the Fitbit (n=112), mean (SD)	6.7 (0.9)
Average h/day of wearing the Fitbit (n=112), mean (SD)	19.5 (5.5)
Frequency of using the Fitbit app (n=112), n (%)	
<1/day	19 (17)
Once a day	27 (24.1)
At least twice a day	66 (58.9)

The average usability for Fitbit was 64.0 (SD 11.1) with majority scoring the Fitbit usability as *OK* (77/113, 68.1%) or *Good* (22/113, 19.5%). More than half of the participants used the Fitbit app at least twice a day. On average, the participants wore the Fitbit for 5.4 weeks, 6.7 days per week, and 19.5 h/day.

Effectiveness of the Intervention

Figure 3 shows changes in the average number of Fitbit steps per day over the study period. The average steps increased throughout the study. Data for the first 7 days of Fitbit use, before receiving access to the chatbot, did not show an increase in mean step (Figure 4).

Figure 3. Change in mean daily step over time.

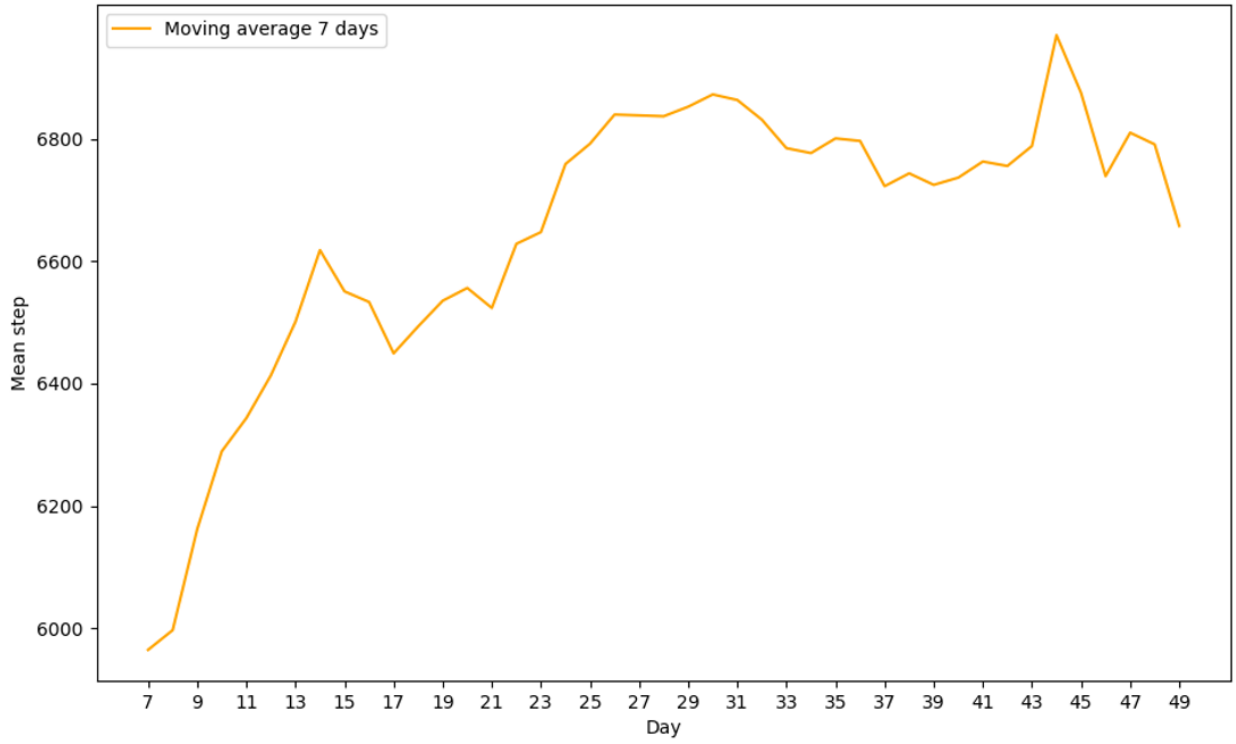


Figure 4. Change in mean daily step over time.

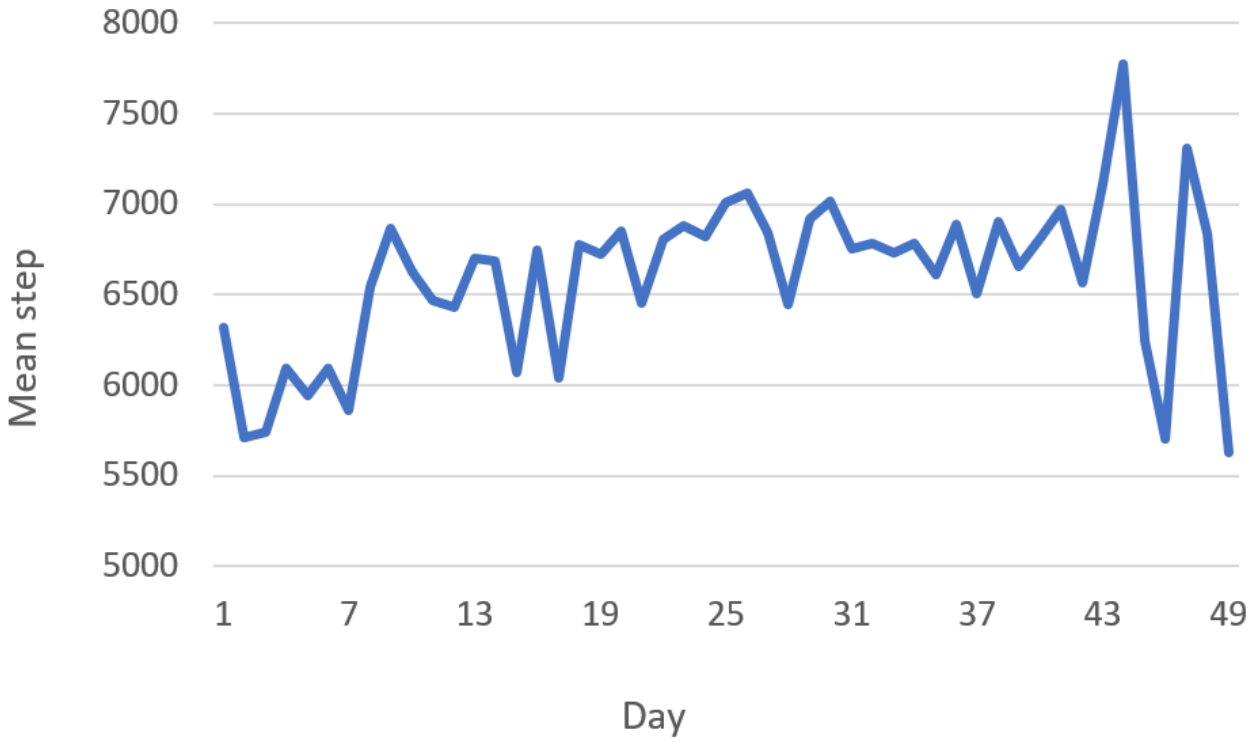


Table 3 shows improvement in the outcomes between follow-up and baseline. For both samples, BMI was improved but was not statistically significant at follow-up compared with baseline. On average, participants recorded significantly more steps at follow-up compared with baseline in the full sample (increase of 627, 95% CI 219-1035 steps/day) and in the subsample that excludes those with <4 weeks of exposure to the chatbot (increase of 564, 95% CI 120-1009 steps/day). Similarly, the total physical activity minutes at follow-up were 3.58 (95% CI 2.28-5.63) times higher in the full sample and 4.17 (95% CI 2.55-6.80) times higher in the subsample than at baseline, representing an increase 154.2 and 176.6 min/week, respectively. Participants were also more likely to meet the physical activity guideline at follow-up compared with baseline in the full sample (OR 6.37, 95% CI 3.31-12.27) and in the subsample (OR 6.41, 95% CI 3.14-13.09).

Table 3. Differences in the outcomes between follow-up and baseline.

	Baseline		Follow-up		Crude estimate (95% CI)	Adjusted estimate (95% CI)[a]
	Participants, n	Value	Participants, n	Value		
Full sample (N=116)						
BMI (kg/m^2), mean (SD)	116	32.5 (8.0)	116	32.4 (8.0)	−0.08 (−0.34 to 0.17)	−0.13 (−0.37 to 0.11)
Step counts/day, mean (SD)	108	5933 (2391)	102	6570 (2326)	633[b] (242 to 1024)	627[b] (219 to 1035)
Total physical activity (min/week), mean (SD)[c]	116	86.5 (137.5)	116	240.7 (233.6)	4.04[d] (2.59 to 6.29)	3.58[d] (2.28 to 5.63)
Meeting physical activity guidelines[e], n (%)						
No	116	100 (86.2)	116	54 (46.6)	1.0	1.0
Yes	116	16 (13.8)	116	62 (53.5)	7.18[d] (3.89 to 13.24)	6.37[d] (3.31 to 12.27)
Subsample (n=99); excludes those with <4 week of chatbot use						
BMI (kg/m^2), mean (SD)	99	32.0 (8.1)	99	31.9 (8.2)	−0.08 (−0.37 to 0.21)	−0.13 (−0.4 to 0.14)
Step counts/day, mean (SD)	94	5958 (2444)	89	6530 (2297)	576[b] (153 to 998)	564[f] (120 to 1009)
Total physical activity (min/week), mean (SD)[c]	99	88.9 (147.0)	99	265.5 (240.5)	4.69[d] (2.92 to 7.55)	4.17[d] (2.55 to 6.80)
Meeting physical activity guidelines[e], n (%)						
No	99	84 (84.9)	99	43 (43.4)	1.0	1.0
Yes	99	15 (15.1)	99	56 (56.6)	7.29[d] (3.77 to 14.12)	6.41[d] (3.14 to 13.09)

[a]Adjusted for age: gender, marital status, years of schooling, ethnicity, household income, living area, work status, and work duration.

[b]P<.01.

[c]Estimates were converted back to ratios.

[d]P<.001.

[e]Estimates are odds ratios.

[f]P<.05.

Discussion

Principal Findings

This study examined the feasibility, usability, and effectiveness of a physical activity chatbot with natural language processing capability and adaptive goal setting delivered via the Facebook Messenger app. Significant improvements in both the step count and self-reported physical activity were observed. These findings are consistent with those from another Australian study examining a combined diet and physical activity chatbot using natural language processing [20]. The effect on self-reported physical activity in this study (approximately 160 min/week) was similar to an increase of approximately 110 min/week of physical activity measured by accelerometers after a 12-week intervention in another study [20]. However, the effect measured by the Fitbits appeared to be smaller, as the average step count only increased by approximately 600 steps/day or 4200 steps/week, which roughly corresponds to an increase of 42 minutes of physical activity per week (assuming it takes approximately 10 minutes to take 1000 steps) [20]. The differences between self-reported and objectively measured physical activity are likely due to recall and social desirability bias [31]. It is also worth noting that an increase of 1000 steps/day can help reduce all-cause mortality risk between 6% and 36%, although an increase of approximately 600 steps may not have a similar effect [32]. In addition, the technical issues experienced in this study might have reduced the effectiveness of the intervention and explained the differences between the 2 studies. Moreover, it is possible that a longer study duration would have resulted in higher effectiveness, and that 6 weeks was not long enough to demonstrate the full potential of the chatbot. Those in the subsample that excluded those who had <4 weeks of exposure to the chatbot, had a higher increase in the total physical activity minutes compared with the full sample, which indicates that more exposure might have resulted in better outcomes. Finally, we also found a study that used a chatbot to promote stair climbing, and it also reported a significant increase in physical activity after 12 weeks of

intervention, although it is not clear how physical activity was measured and what the effect size was [33].

The findings showed that the participants liked the chatbot with some even asking for continuing to use it after they completed the 6-week trial. Nevertheless, usability for both the chatbot and the Fitbit was rated as *OK* for majority of the participants. For most usability and acceptability indicators, less than half of the participants provided answers in favor of the chatbot. This level of usability (SUS scores of 61.6) is comparable with the 2 psychological therapy chatbots with SUS scores of 63.6 and 57.0 [34] but lower than that of asthma management (SUS score of about 83) [35] and depression prevention chatbots (SUS score of approximately 75.7) [36]. A study by Nadarzynski et al [37] also showed higher acceptability, with 67% of participants who would like to use a health chatbot. The lower acceptability in this study is likely due to technical issues that most participants experienced during the study. Some technical issues related to the use of the Fitbit, including malfunctioning Fitbits and broken bands and cables, were expected and dealt with by the research team. Most of the other technical issues (eg, the chatbot stopped sending daily notifications and difficulties connecting to the chatbot) were fixed by the management company. However, one issue beyond the control of the management company that resulted in the end of the study was that Facebook changed its policy to block the chatbot sending out messages to participants who did not respond to the chatbot within 24 hours. Facebook implemented this new policy to prevent chatbots from spamming its Messenger app users, however, inadvertently disabled the functionality of the chatbot, which sent messages wanted by our participants. Despite our efforts, it was not possible to contact Facebook to explain and reverse the situation. The Facebook policy change offers the most likely explanation for the discrepancy between high engagement and low usability scores; that is, most participants used the chatbot until the end of the implementation, but at the time of the Facebook policy change, they experienced serious technical issues undermining the usability of the chatbot.

Previous studies have also shown higher usability of Fitbit use [38-40] compared with this study. A possible explanation might be that because of budgetary reasons, the research team was forced to use Fitbit Flex 1, which is an old model in this study. Apart from being outdated in terms of user expectations (newer models with better functionality dominate the market), the long shelf life of the Fitbits meant that battery and connectivity problems were more prevalent than normal. We recommend the use of newer and higher-quality activity trackers to increase the feasibility of future chatbot-based physical activity interventions. Furthermore, we recommend that future chatbots be hosted on flexible messaging platforms that can be contacted for assistance in dealing with similar issues should they arise. However, the disadvantage of using such platforms is that people may be less familiar with the platform and more reluctant to use them, as few of their friends and family are likely to use those messaging services. In addition, rather than relying on an external technical company to develop and host the chatbot, it would be better if the research team is capable of doing this by itself, so that upgrading chatbot functionality and responding to potential technical problems is faster and more efficient.

The results also showed that BMI did not significantly improve at follow-up. This finding is not surprising, as our study did not target weight loss and therefore, no direct activity related to weight loss or weight maintenance was delivered. This is different from the other Australian chatbot-based physical activity interventions, which showed a significant decrease in weight at week 12 [20]. However, that study also included a large dietary component and allowed more time (12 weeks) for weight loss to occur [20]. As this study was not designed to evaluate the effect of each component (diet and physical activity) separately, it is impossible to determine whether the improvement in weight was due to increases in physical activity. Furthermore, current evidence regarding the effects of physical activity and exercise on weight loss is not strong [41].

Strengths and Limitations

This study has several strengths: (1) both objective and subjective measures of physical activity were used to obtain accurate and complementary data on the effectiveness of the intervention [42] and (2) a high retention rate means that selection bias due to loss to follow-up was likely minimal. However, this study also has limitations. First, as the study was designed as a quasi-experiment without a control group, it was not possible to control for unknown confounders. It is also likely that the increase in steps occurred just by using the Fitbit [43]. However, it is worth noting that step counts in the baseline week before the participants started using the chatbot did not increase (Figure 4). Second, the sample (majority were women and White with high BMI) was not representative of the broader Australian population, so generalizability of the findings may be limited, although external validity was not the main focus of the intervention. Third, the technical issues caused by Facebook's policy changes, which were beyond the research team's control, were likely responsible for a reduction in chatbots' usability, acceptability, and effectiveness. Finally, the short duration of the intervention (6 weeks) is a limitation, and the effects of chatbot interventions with a longer duration need to be examined.

Conclusions

The machine learning–based physical activity chatbot was able to significantly increase participants' physical activity and was moderately accepted by the participants. However, a Facebook policy change undermined the chatbot functionality and indicated the need to use independent platforms for chatbot deployment so that this type of intervention could be successfully delivered.

Future studies with stronger designs, such as randomized controlled trials, in which the effect of the activity trackers can be isolated, are needed to confirm these findings. Research is also required to determine whether chatbot-based interventions could be effective for broader populations. Furthermore, technology to develop and evaluate more comprehensive chatbot interventions already exists. In addition to natural language processing, Fitbit integration and adaptive goal setting, it is possible to use deep reinforcement learning with feedback loops and integrate more real-time data sources (eg, GPS and weather data) to enable chatbots to personally tailor and continuously adapt cues to action to ensure the timing, frequency, context,

and content are optimally suited for each participant. It is important that such comprehensive physical activity chatbots should be developed and evaluated in future studies.

Acknowledgments

This study was supported by a Future Leader Fellowship (ID 100427) awarded to CV by the National Heart Foundation of Australia, and a postdoctoral fellowship and a commencement grant (ID RSH/5409) awarded to QGT by Central Queensland University.

Conflicts of Interest

None declared.

References

1. Physical Activity Guidelines Advisory Committee Scientific Report. Physical Activity Guidelines Advisory Committee. 2018. URL: https://health.gov/sites/default/files/2019-09/PAG_Advisory_Committee_Report.pdf [accessed 2021-11-18]
2. Impact of physical inactivity as a risk factor for chronic impact of physical inactivity as a risk factor for chronic conditions: Australian burden of disease study. Australian Institute of Health and Welfare. 2017. URL: https://www.aihw.gov.au/reports/burden-of-disease/impact-of-physical-inactivity-chronic-conditions/summary [accessed 2021-11-08]
3. Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van Mechelen W, Lancet Physical Activity Series 2 Executive Committee. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet 2016 Sep 24;388(10051):1311-1324. [doi: 10.1016/S0140-6736(16)30383-X] [Medline: 27475266]
4. Sweet SN, Fortier MS. Improving physical activity and dietary behaviours with single or multiple health behaviour interventions? A synthesis of meta-analyses and reviews. Int J Environ Res Public Health 2010 Apr 16;7(4):1720-1743 [FREE Full text] [doi: 10.3390/ijerph7041720] [Medline: 20617056]
5. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-28.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
6. Global digital population as of October 2019 (in millions). Statista. 2019. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2019-12-11]
7. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
8. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
9. Schembre SM, Liao Y, Robertson MC, Dunton GF, Kerr J, Haffey ME, et al. Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. J Med Internet Res 2018 Mar 22;20(3):e106 [FREE Full text] [doi: 10.2196/jmir.8701] [Medline: 29567638]
10. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
11. Karekla M, Kasinopoulos O, Neto DD, Ebert DD, Van Daele T, Nordgreen T, et al. Best practices and recommendations for digital interventions to improve engagement and adherence in chronic illness sufferers. Eur Psychol 2019 Jan;24(1):49-67. [doi: 10.1027/1016-9040/a000349]
12. Short CE, Rebar AL, Plotnikoff RC, Vandelanotte C. Designing engaging online behaviour change interventions: a proposed model of user engagement. Eur Health Psychol 2015;17(1):32-38 [FREE Full text]
13. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
14. AI Multiple. 2021. URL: https://blog.aimultiple.com/top-chatbot-success/ [accessed 2021-11-08]
15. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
16. HealthTap Launches Dr. A.I. Business Wire. 2017. URL: https://www.businesswire.com/news/home/20170110005266/en/HealthTap-Launches-Dr.-A.I [accessed 2021-11-08]
17. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]

18. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346 [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]
19. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: viewpoint. J Med Internet Res 2020 Sep 30;22(9):e22845 [FREE Full text] [doi: 10.2196/22845] [Medline: 32996892]
20. Maher CA, Davis CR, Curtis RG, Short CE, Murphy KJ. A physical activity and diet program delivered by artificially intelligent virtual health coach: proof-of-concept study. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e17558 [FREE Full text] [doi: 10.2196/17558] [Medline: 32673246]
21. Kramer J, Künzler F, Mishra V, Smith S, Kotz D, Scholz U, et al. Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
22. Vandelanotte C, Van Itallie A, Brown W, Mummery W, Duncan M. Every Step Counts: Understanding the Success of Implementing the 10,000 Steps Project. In: Maeder A, Champion S, Moores C, editors. Information Technology Based Methods for Health Behaviours (Part of "Global Telehealth 2019: Studies in Health Technology and Informatics" series). Amsterdam: IOS Press Books; 2020:15-30.
23. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
24. Mitchell M, White L, Lau E, Leahey T, Adams MA, Faulkner G. Evaluating the carrot rewards app, a population-level incentive-based intervention promoting step counts across two Canadian provinces: quasi-experimental study. JMIR Mhealth Uhealth 2018 Sep 20;6(9):e178 [FREE Full text] [doi: 10.2196/mhealth.9912] [Medline: 30148712]
25. Feig EH, Harnedy LE, Celano CM, Huffman JC. Increase in daily steps during the early phase of a physical activity intervention for type 2 diabetes as a predictor of intervention outcome. Int J Behav Med 2021 Dec;28(6):834-839. [doi: 10.1007/s12529-021-09966-0] [Medline: 33575971]
26. Australian Institute of Health and Welfare. The Active Australia Survey: A Guide and Manual for Implementation, Analysis and Reporting. Cranberra: Australian Institute of Health and Welfare; 2003.
27. Brooke J. SUS-A quick and dirty usability scale. In: Usability Evaluation in Industry. Boca Raton, Florida: CRC Press; 1996.
28. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Human Comput Interac 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
29. Chomistek A, Yuan C, Matthews C, Troiano R, Bowles H, Rood J, et al. Physical activity assessment with the ActiGraph GT3X and doubly labeled water. Med Sci Sports Exerc 2017 Sep;49(9):1935-1944 [FREE Full text] [doi: 10.1249/MSS.0000000000001299] [Medline: 28419028]
30. Mitchell M, White L, Lau E, Leahey T, Adams MA, Faulkner G. Evaluating the carrot rewards app, a population-level incentive-based intervention promoting step counts across two canadian provinces: quasi-experimental study. JMIR Mhealth Uhealth 2018 Sep 20;6(9):e178 [FREE Full text] [doi: 10.2196/mhealth.9912] [Medline: 30148712]
31. Vandelanotte C, Short CE, Plotnikoff RC, Rebar A, Alley S, Schoeppe S, et al. Are web-based personally tailored physical activity videos more effective than personally tailored text-based interventions? Results from the three-arm randomised controlled TaylorActive trial. Br J Sports Med 2021 Mar 03;55(6):336-343. [doi: 10.1136/bjsports-2020-102521] [Medline: 33144346]
32. Hall KS, Hyde ET, Bassett DR, Carlson SA, Carnethon MR, Ekelund U, et al. Systematic review of the prospective association of daily step counts with risk of mortality, cardiovascular disease, and dysglycemia. Int J Behav Nutr Phys Act 2020 Jun 20;17(1):78 [FREE Full text] [doi: 10.1186/s12966-020-00978-9] [Medline: 32563261]
33. Piao M, Ryu H, Lee H, Kim J. Use of the healthy lifestyle coaching chatbot app to promote stair-climbing habits among office workers: exploratory randomized controlled trial. JMIR Mhealth Uhealth 2020 May 19;8(5):e15085 [FREE Full text] [doi: 10.2196/15085] [Medline: 32427114]
34. Bennion MR, Hardy GE, Moore RK, Kellett S, Millings A. Usability, acceptability, and effectiveness of web-based conversational agents to facilitate problem solving in older adults: controlled study. J Med Internet Res 2020 May 27;22(5):e16794 [FREE Full text] [doi: 10.2196/16794] [Medline: 32384055]
35. Kadariya D, Venkataramanan R, Yup HY, Kalra M, Thirunarayanan K, Sheth A. kBot: knowledge-enabled personalized chatbot for self-management of asthma in pediatric population. In: Proceedings of the 2019 IEEE International Conference on Smart Computing (SMARTCOMP). 2019 Presented at: 2019 IEEE International Conference on Smart Computing (SMARTCOMP); Jun 12-15, 2019; Washington, DC, USA. [doi: 10.1109/SMARTCOMP.2019.00043]
36. Bresó A, Martínez-Miranda J, Botella C, Baños RM, García-Gómez JM. Usability and acceptability assessment of an empathic virtual agent to prevent major depression. Expert Syst J Knowl Eng 2016 May 25;33(4):297-312 [FREE Full text] [doi: 10.1111/exsy.12151]
37. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]

38. Vandelanotte C, Duncan MJ, Maher CA, Schoeppe S, Rebar AL, Power DA, et al. The effectiveness of a web-based computer-tailored physical activity intervention using Fitbit activity trackers: randomized trial. J Med Internet Res 2018 Dec 18;20(12):e11321 [FREE Full text] [doi: 10.2196/11321] [Medline: 30563808]
39. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. Wearable activity tracker use among Australian adolescents: usability and acceptability study. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e86 [FREE Full text] [doi: 10.2196/mhealth.9199] [Medline: 29643054]
40. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
41. Cox CE. Role of physical activity for weight loss and weight maintenance. Diabetes Spectr 2017 Aug;30(3):157-160 [FREE Full text] [doi: 10.2337/ds17-0013] [Medline: 28848307]
42. Skender S, Ose J, Chang-Claude J, Paskow M, Brühmann B, Siegel EM, et al. Accelerometry and physical activity questionnaires - a systematic review. BMC Public Health 2016 Jun 16;16:515 [FREE Full text] [doi: 10.1186/s12889-016-3172-0] [Medline: 27306667]
43. Chaudhry UA, Wahlich C, Fortescue R, Cook DG, Knightly R, Harris T. The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults. Int J Behav Nutr Phys Act 2020 Oct 09;17(1):129 [FREE Full text] [doi: 10.1186/s12966-020-01020-8] [Medline: 33036635]

Abbreviations

mHealth: mobile health
OR: odds ratio
SUS: System Usability Scale

Edited by L Buis; submitted 07.03.21; peer-reviewed by C Mauch, Z Yin; comments to author 25.03.21; revised version received 25.04.21; accepted 03.09.21; published 26.11.21.

Please cite as:
To QG, Green C, Vandelanotte C
Feasibility, Usability, and Effectiveness of a Machine Learning–Based Physical Activity Chatbot: Quasi-Experimental Study
JMIR Mhealth Uhealth 2021;9(11):e28577
URL: https://mhealth.jmir.org/2021/11/e28577
doi:10.2196/28577
PMID:34842552

Original Paper

Predicting Participant Compliance With Fitness Tracker Wearing and Ecological Momentary Assessment Protocols in Information Workers: Observational Study

Gonzalo J Martinez[1], MSc; Stephen M Mattingly[1], PhD; Pablo Robles-Granda[2], PhD; Koustuv Saha[3,4], PhD; Anusha Sirigiri[5], PhD; Jessica Young[6], MSc; Nitesh Chawla[1], PhD; Munmun De Choudhury[4], PhD; Sidney D'Mello[7], PhD; Gloria Mark[8], PhD; Aaron Striegel[1], PhD

[1]Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, United States
[2]Thomas M Siebel Center for Computer Science, University of Illinois Urbana-Champaign, Urbana, IL, United States
[3]Microsoft Research, Montreal, QC, Canada
[4]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States
[5]Indian School of Business Gachibowli, Hyderabad Telangana, India
[6]Center for Research Computing, University of Notre Dame, Notre Dame, IN, United States
[7]Institute of Cognitive Science, University of Colorado Boulder, Boulder, CO, United States
[8]Informatics Department, University of California, Irvine, Irvine, CA, United States

Corresponding Author:
Gonzalo J Martinez, MSc
Computer Science and Engineering
University of Notre Dame
400 Main Building
Notre Dame, IN, 46556
United States
Phone: 1 574 631 5000
Email: gmarti11@nd.edu

Abstract

Background: Studies that use ecological momentary assessments (EMAs) or wearable sensors to track numerous attributes, such as physical activity, sleep, and heart rate, can benefit from reductions in missing data. Maximizing compliance is one method of reducing missing data to increase the return on the heavy investment of time and money into large-scale studies.

Objective: This paper aims to identify the extent to which compliance can be prospectively predicted from individual attributes and initial compliance.

Methods: We instrumented 757 information workers with fitness trackers for 1 year and conducted EMAs in the first 56 days of study participation as part of an observational study. Their compliance with the EMA and fitness tracker wearing protocols was analyzed. Overall, 31 individual characteristics (eg, demographics and personalities) and behavioral variables (eg, early compliance and study portal use) were considered, and 14 variables were selected to create beta regression models for predicting compliance with EMAs 56 days out and wearable compliance 1 year out. We surveyed study participation and correlated the results with compliance.

Results: Our modeling indicates that 16% and 25% of the variance in EMA compliance and wearable compliance, respectively, could be explained through a survey of demographics and personality in a held-out sample. The likelihood of higher EMA and wearable compliance was associated with being older (EMA: odds ratio [OR] 1.02, 95% CI 1.00-1.03; wearable: OR 1.02, 95% CI 1.01-1.04), speaking English as a first language (EMA: OR 1.38, 95% CI 1.05-1.80; wearable: OR 1.39, 95% CI 1.05-1.85), having had a wearable before joining the study (EMA: OR 1.25, 95% CI 1.04-1.51; wearable: OR 1.50, 95% CI 1.23-1.83), and exhibiting conscientiousness (EMA: OR 1.25, 95% CI 1.04-1.51; wearable: OR 1.34, 95% CI 1.14-1.58). Compliance was negatively associated with exhibiting extraversion (EMA: OR 0.74, 95% CI 0.64-0.85; wearable: OR 0.67, 95% CI 0.57-0.78) and having a supervisory role (EMA: OR 0.65, 95% CI 0.54-0.79; wearable: OR 0.66, 95% CI 0.54-0.81). Furthermore, higher wearable compliance was negatively associated with agreeableness (OR 0.68, 95% CI 0.56-0.83) and neuroticism (OR 0.85, 95% CI 0.73-0.98). Compliance in the second week of the study could help explain more variance; 62% and 66% of the variance in

EMA compliance and wearable compliance, respectively, was explained. Finally, compliance correlated with participants' self-reflection on the ease of participation, usefulness of our compliance portal, timely resolution of issues, and compensation adequacy, suggesting that these are avenues for improving compliance.

Conclusions: We recommend conducting an initial 2-week pilot to measure trait-like compliance and identify participants at risk of long-term noncompliance, performing oversampling based on participants' individual characteristics to avoid introducing bias in the sample when excluding data based on noncompliance, using an issue tracking portal, and providing special care in troubleshooting to help participants maintain compliance.

(JMIR Mhealth Uhealth 2021;9(11):e22218) doi:10.2196/22218

KEYWORDS

adherence; compliance; wearables; smartphones; research design; ecological momentary assessment; mobile sensing; mobile phone

Introduction

Background

In the past decade, an increasing variety of sensors have been used as research tools, such as smartphones [1-9], wearables [9-19], ecological momentary assessments (EMAs) [20-23], social media [24,25], and other sensing modalities [26-30]. However, the effectiveness of these studies depends on the completeness of the data generated, which further relies on participant compliance. When designing an observational study, the most important factor for ensuring quality data is compliance (sometimes referred to as adherence [31]).

Typically, compliance decreases throughout the life of a study [1]. Compliance has been found to be as low as 16% by the end of an almost year-long study [32], and it varies considerably from 80% to between 10% and 20% during certain periods in the study [33]. This pattern persists in shorter studies as well. For example, in the study of wearable compliance by Evenson et al [34], it was found that only 78% of >15,000 participants completed at least 21% of the possible data collection. This can considerably reduce the sample size available for analysis, should continuous measurements be necessary. For instance, despite enrolling 646 participants, Wang et al [35] only analyzed data from 159 participants because of the lack of compliance. Finally, other studies have found compliance to be related to participant characteristics [12,13,31,36], which increases the odds of introducing bias when excluding participants from analysis because of noncompliance [37]. Therefore, we posit that the ability to identify the compliance of individuals early and resolve issues that can affect study participation would be invaluable to the research community. Prior works have found associations between certain participant characteristics that could be used to predict compliance early in a study, although the evidence is conflicting.

Early works in the field by Schüz et al [22] and Courvoisier et al [38] suggested that compliance with EMAs was not associated with participant characteristics. Specifically, Schüz et al [22] administered random prompts for EMAs in a 6-day study of 119 smokers and found no association between EMA compliance and smoking habits, race, sex, education level, or marital status. Similarly, Courvoisier et al [38] found no associations between a phone call–based EMA monitoring protocol and sex, age, education level, linguistic region, life satisfaction, or personality. However, more recent studies have found conflicting evidence. Dzubur et al [39] examined EMA responses and reported approximately 80% compliance, but compliance was influenced by the participant-level factors of income and ethnicity; lower-income or Hispanic mothers were less likely to respond to surveys. Finally, in a meta-analysis on factors that contribute to EMA compliance administered from smartphones or with wearables for those aged <18 years, Wen et al [36] reported a general response rate of approximately 78% to surveys, which varied based on clinical status; those without disorders had lower response rates with more prompts (≥6 times: 75%; 2-3 times: 92%), whereas those with a clinical status responded more often to increased prompts (≥6 times: 89%; 2-3 times: 74%). These papers suggest that participant-level factors such as income, ethnicity, and clinical status can interact with compliance in longitudinal studies using EMAs.

Several studies using wearable accelerometers have found associations between compliance and various participant characteristics, such as income, age, smoking, and having tertiary education [12,13,34]. A 4-day study involving 3601 participants [13] found that higher compliance, defined as wearing time, was associated with being older, not smoking, and having a full-time job, tertiary education, and high self-reported health, whereas no associations were found with income level or sex. As noted earlier, the study by Evenson et al [34] defined compliance as wearing an accelerometer for 10 hours a day, 3 days out of 6 days, in a study of 15,153 participants in a Hispanic community. The study reported higher compliance for those participants who are married or partnered; those with higher household income; those who are male, older, and employed or retired; those not born in the United States; those preferring Spanish over English; and those having a lower BMI. Similarly, a repeated measures study of adolescent females that deployed accelerometers 7 months apart found that physical activity level and race were associated with compliance [12]. The same study found that compliance was trait-like; higher compliance in the first session was associated with higher compliance in subsequent redeployments of accelerometers 7 months apart.

Recent studies that included smartwatches, fitness trackers, or smartphones have also found participant characteristics to be related to compliance [31-33,40]. Harari et al [33] analyzed 3 student population samples to understand participants' motivations for self-tracking using passive sensing and active

logging in relation to compliance. The study correlated participant characteristics with behaviors that motivated self-tracking and found that more agreeable, younger, and extroverted participants had increased productivity and health behaviors that were positively correlated with compliance. In addition, the study found that neuroticism, openness, and being female were correlated with well-being and daily activities motivation to self-track, which were in turn positively correlated with compliance. No significant correlations were found with conscientiousness and well-being measures. On the other hand, in a 4-year study of 698 college students where compliance was studied as a binary variable, it was found that extraversion and openness negatively correlated with compliance, whereas conscientiousness and agreeableness positively correlated with compliance, and neuroticism did not significantly correlate with compliance [31]. In addition, the study reported that the first month of compliance correlated with whole study compliance, a result in line with Rowlands et al [12].

Jeong et al [18] studied how individuals perceive their study participation in a sample of 50 students using Apple Watch and found that limitations of the devices themselves or personal reasons could get in the way of study participation and cause noncompliance. The study found that when participants needed to charge the devices during the night or at least once a day, compliance decreased. Similarly, participants would forget to wear the smartwatch depending on certain situations, such as staying at home for the weekend or going out with friends. However, the reported patterns of wearing behavior do not necessarily match those of modern wearables where battery life lasts multiple days or those of a working population [41].

Objective

In summary, there is extensive literature showing that participant characteristics and compliance are related, albeit with conflicting results. However, there remain several issues before these findings could be used in practice. For instance, several works define compliance as a binary variable with two outcomes depending on a specific definition that is not universally applicable, for example, in ≥80% [31], 10 hours a day for 3 out of 7 days [34], or >16 hours a day for 7 days [12]. The use of such specific and inconsistent definitions of compliance makes it challenging to apply the findings from other studies with different thresholds to meet the requirements of a new study. In addition, most existing works do not provide any metric of model fit or error that provides guidance on the predictive power of the models, with the exceptions of Lee et al [13] reporting R^2=0.03 and Hermsen et al [32] reporting R^2=0.099. Existing works rely only on a training set of data, that is, no testing set was used to report predictive performance. This means that we cannot know beforehand if participant characteristics could be effective in predicting compliance before a study starts or early on. Given these limitations, the objective of our paper is to address the following research questions (RQs) using a generalizable definition of compliance and considering personal characteristics that are commonly or easily assessed in other studies:

- RQ1: To what extent can personal characteristics measured before the start of a study predict long-term compliance?
- RQ2: How does early assessment of compliance (ie, during the first 2 weeks of study participation) predict future compliance?
- RQ3: Are participants' perceptions of study participation, feedback, and issue reporting correlated with compliance?

Methods

We used our Tesserae [42] study—a 1-year, large-scale, multimodal study of working professionals with a rich set of psychological and health-related data to extract meaningful variables that explain variations in compliance.

Participants and Recruitment

Tesserae recruited 757 participants from cognitively demanding professions (eg, information workers) to participate in a 1-year study exploring the extent to which widely available sensing streams could predict various individual difference variables and job performance dimensions. As such, the study was observational in nature and did not implement interventions beyond interacting with participants to resolve participation issues. Individuals were drawn from throughout the United States. Participants were enrolled both in person and remotely from January 2018 to July 2018. The study concluded data gathering in mid-April 2019. A timeline of the study can be found in Figure 1. Participants were divided into 2 sets: blinded and nonblinded (Table 1). Responses to initial and daily surveys from the blinded set were withheld from researchers by the study sponsor until the end of phase 1 of the multimodal objective sensing to assess individuals with context [43] program in May 2020. Researchers had full access to participant data in the nonblinded set.

Figure 1. Timeline of study participation.

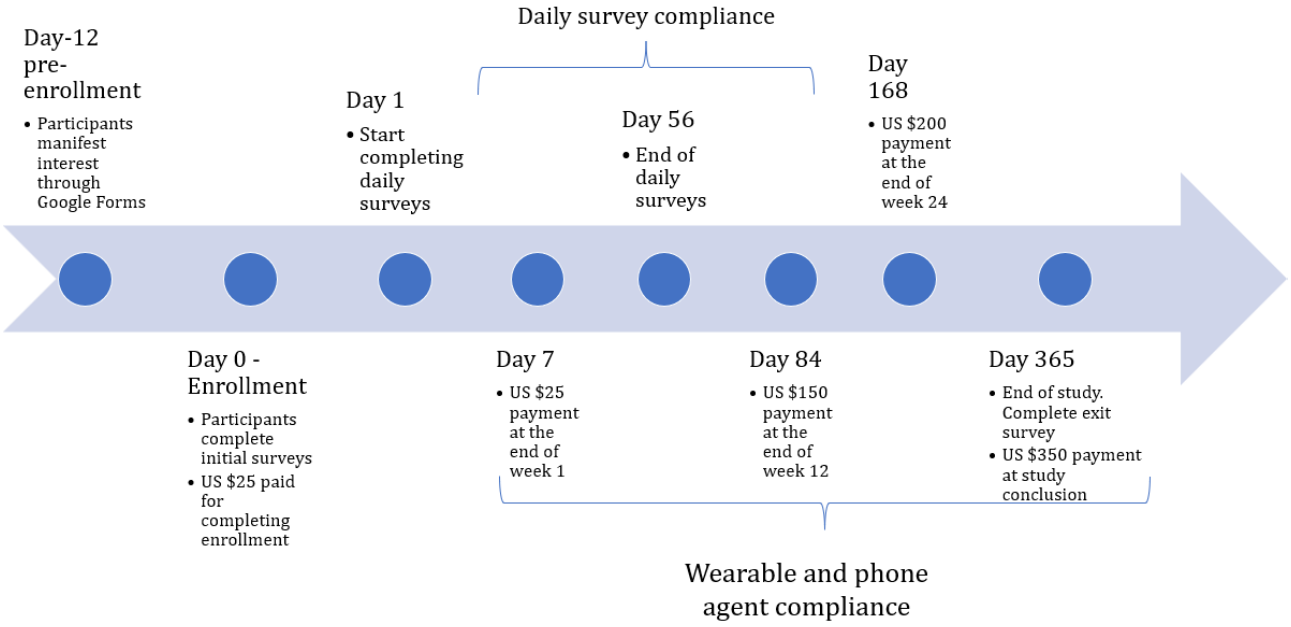


Table 1. Cohort distribution in the blinded and nonblinded samples after data preprocessing. A chi-square test of independence showed that the cohort distribution is different between the blinded and nonblinded samples ($\chi^2(4)=129.53$; $P<.001$).

Cohort	Blinded, n (%)	Nonblinded, n (%)
Large multinational technology services firm	42 (28)	249 (41.7)
Large midwestern technology or engineering firm	34 (22.7)	144 (24.1)
Midwestern software engineering firm	5 (3.3)	21 (3.5)
Midwestern university	57 (38)	32 (5.4)
Various other companies	12 (8)	151 (25.3)
Total	150 (100)	597 (100)

Procedures

Individuals participating in the study were provided with a wearable (Garmin vivoSmart 3); a smartphone agent (phone app) derived from StudentLife [2]; and a set of Bluetooth beacons to demarcate home, work, and proximity to others in the study and were requested to provide read access to social media (Facebook and Twitter) [44]. An initial set of psychological and health-related surveys were collected (Multimedia Appendix 1 [45-57]) at enrollment in the study. In addition, short, daily versions of many of the aforementioned surveys were administered, as well as context, current stress level, and current anxiety level assessments. Daily surveys were administered via the Qualtrics Experience Management platform, prompted by an SMS text message, designed to be answered in <2 minutes, and with a 4-hour window for completion. Users received daily survey prompts at either 8 AM, noon, or 4 PM during the initial 8 weeks (56 days) of a participant's year in the study. Participant compliance and troubleshooting were provided through a user-facing web-based portal and managed by the study personnel. In the case of data from participants belonging to the blinded set, researchers did not send the surveys or receive the responses directly. These surveys were administered by the study sponsor and stored apart from nonblinded data, allowing researchers the use of nonblinded data for exploratory analyses while still having a separate sample to test out-of-sample performance.

The compliance rate for the study was tracked for 3 of the major sensing streams: daily surveys, wearables, and phone agents, which was not analyzed (see the *Measures* section). Bluetooth beacon compliance was not tracked as an individual could be complying without being in the range of a beacon for the study. Conversely, social media compliance was not tracked, as it was an opt-in sensor and only required one-time authorization.

Participants were compensated based on their compliance. Participants could continuously review their compliance and report issues through a dynamic web-based portal. On the basis of cohort requirements, participants were either paid a stipend or entered into a weekly lottery. For stipend participants, those with average compliance of at least 80% across all streams could receive up to US $750 at the completion of the study, broken up as shown in Figure 1. Lottery participants received a ticket per day for each compliant (>80%) stream (wearable, phone agent, and daily survey). A US $250 weekly lottery was held for every 25 participants.

Experimenters could increase compliance in one or more streams at their discretion because of changes in circumstances that precluded compliance for a limited time. For example,

participants whose wearables broke and who reported it to researchers received full compliance for the wearable stream until they received a replacement wearable. In the duration of the study, 325 wearable device replacements were issued [41]. Other examples include international travel, which prohibited SMS text message receipt of the daily survey; a damaged or replaced cell phone; or change of carrier that affected phone agent compliance and the receipt of daily surveys. If the participants did not inform the researchers regarding such problems related to their sensing streams, they received zero compliance in that period.

Periodic reminders were sent every week via email to participants exhibiting noncompliance (missing recent data or cumulative noncompliance). Participants could decide to stop participating at any point in the study. In addition, participants exhibiting continued noncompliance without response over multiple months (≥3) were considered ineligible for subsequent rewards and were excluded from the study. In total, 107 participants were considered to have dropped out. Nevertheless, although these participants were considered to have dropped from the Tesserae study, they were still considered in the analyses in this work.

Measures

Compliance

Daily survey compliance was defined as the response rate, that is, the number of surveys responded to over the number of surveys sent. Not receiving a prompt because of the phone being turned off or being out of coverage was not differentiated from receiving and not responding to a prompt; both were considered noncompliant. Wearable compliance was computed in 30-minute nonoverlapping windows, whereby an individual was considered to have been compliant if any wearable data from any stream of the fitness tracker (eg, heart rate, step count, or physical activity) was recorded within that window. However, considering the heart rate or the combination of all sensor streams with the Garmin vivoSmart 3 leads to differences in the calculation of compliance of <1% [41]. Therefore, wearable compliance is the number of 30-minute windows with any data over the number of 30-minute windows in the study, with 48 windows a day for up to 365 days.

Participants were requested to wear the device 24-7 to capture their sleep and daily activities. The roughly 5-day battery life and rapid charge rate allowed a brief charging window each day to not exhaust the wearable battery [41]. Given that the first 2 weeks of compliance would be used to predict long-term compliance (RQ2), we discarded these 2 weeks from the dependent variables and calculated them starting at week 3. We referred to these simply as *wearable compliance* and *survey compliance*, whereas the variables used as predictors were referred to as *wearable compliance in week 2* and *survey compliance in week 2* (see Figure 2 and Table 2 for compliance distributions and correlations).

Figure 2. Distribution of compliance in the nonblinded set in red (n=597) and the blinded set in blue (n=150). The superposition of both sets is in purple.

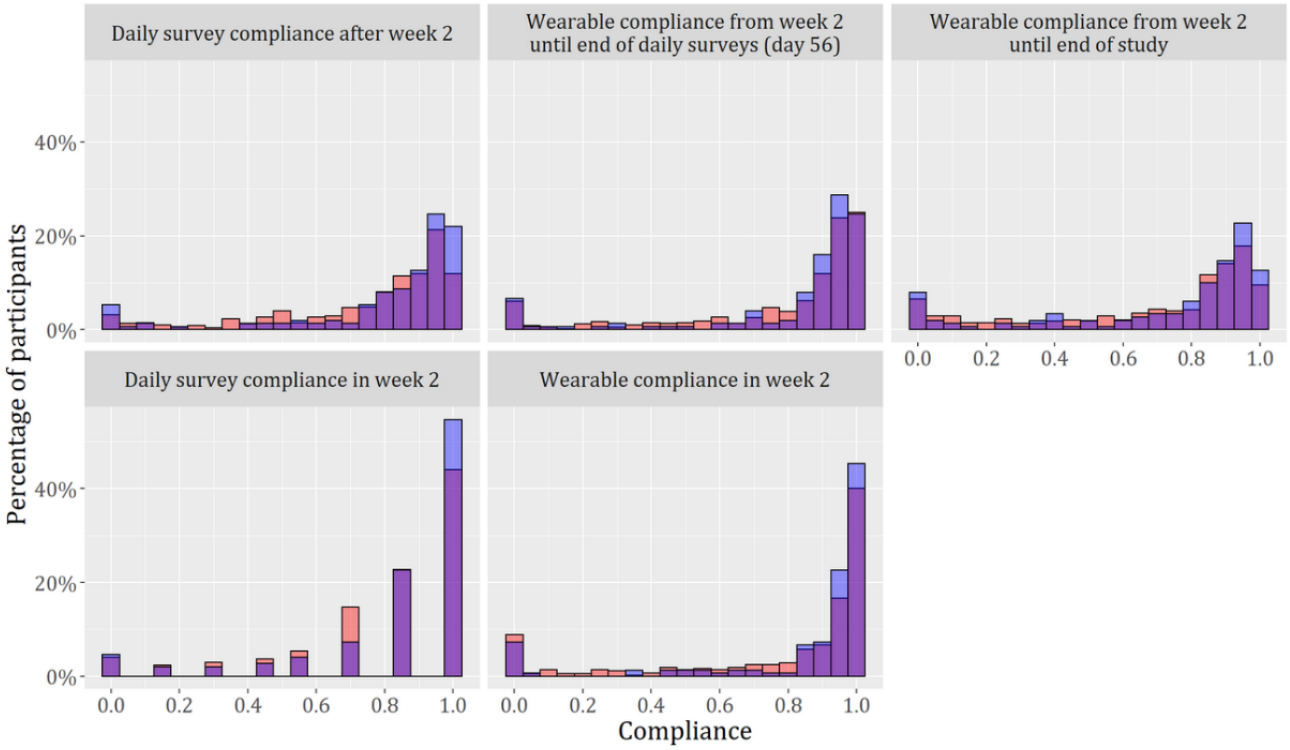


Table 2. Pearson correlation (*r*) of compliance in the nonblinded (below the diagonal) and blinded sets (above the diagonal).

	Daily survey compliance after week 2	Wearable compliance after week 2	Daily survey compliance in week 2	Daily wearable compliance in week 2
Daily survey compliance after week 2	1	0.79	0.81	0.69
Wearable compliance after week 2	0.73	1	0.60	0.72
Daily survey compliance in week 2	0.69	0.51	1	0.58
Daily wearable compliance in week 2	0.56	0.61	0.54	1

Similar to wearable compliance, phone agent compliance was calculated in half-hour time windows. However, the phone agent was intended to run without any user input or action apart from the initial installation and updates. This presented several challenges for researchers, as participants used 112 different models of mobile devices from 14 manufacturers throughout the study. The main challenges were issues of high battery use, failure to run continuously in the background, and an initial lack of feedback to participants regarding whether data were being collected. As a result, missing data more likely reflected the researchers' technical ability to keep the app running on all devices rather than user characteristics or behaviors. Therefore, we decided *not* to consider phone agent compliance rates.

Demographics, Psychological Traits, and Health-Related Characteristics

The shared Tesserae demographics, psychological, and behavioral data set comprised 31 variables that were collected during the study and could be used as predictors. These variables can be categorized as demographics, personality [45], anxiety [46], affect [47], health [48-52], cognitive ability [53], job performance [54-57], and behavior characteristics collected through the study website [42], such as log-ins and issue tickets submitted. Psychometrically validated inventories were used to collect all survey-based measures, except for demographics (Table 3).

Table 3. Demographic questions asked at the onset of study participation.

Item	Question	Options
Age	How old are you?	• Any number
Sex	Are you male or female?	• Male • Female
Household income level	In which of these groups did your total household income (from all sources) fall in 2016?	• <US $25,000 • US $25,000 to US $49,999 • US $50,000 to US $74,999 • US $75,000 to US $99,999 • US $100,000 to US $124,999 • US $125,000 to US $150,000 • >US $150,000
Supervise	Think about your main job. Do you supervise or manage anyone in this job?	• Yes • No
English as native language	Is English your native language?	• Yes • No
Education level	What is your highest level of education?	• Some high school (or equivalent) • High school degree (or equivalent) • Some college • College degree • Some graduate school • Master's degree • Doctoral degree, such as a PhD, MD, or JD
Had a wearable	Do you currently use a wearable like a Fitbit or other fitness device?	• Yes • No

Data Exclusion and Preprocessing

Most of the predictors were obtained from the initial survey conducted during enrollment. Variables in the blinded and nonblinded samples were treated in the same manner. The variables for household income and education were relabeled to address class imbalance. Classes of <US $25,000 and from US $25,000 to US 49,999 were merged into one class (<US $50,000), whereas the classes US $100,000 to US $124,999 and US $125,000 to US $150,000 were merged into a single class as well (US $100,000 to US $150,000). In the case of education, the classes were coalesced into *no college degree*,

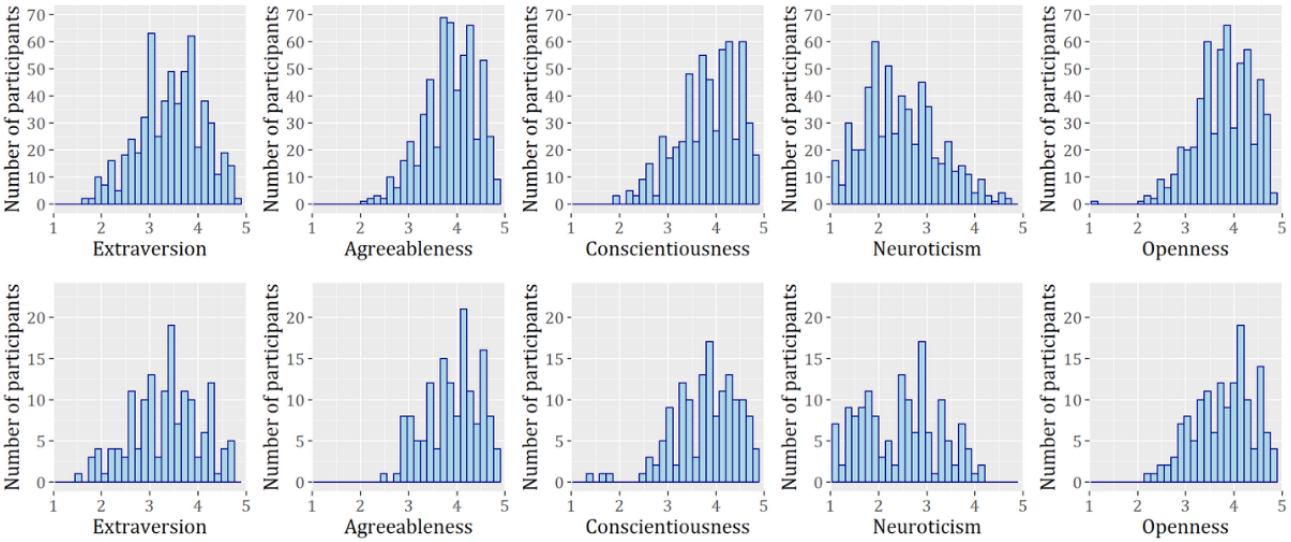
college degree, and *graduate degree*. A total of 32 respondents did not answer the wearable ownership question. To avoid discarding these observations, we generated an additional *unknown* category. All other categorical variables <5 missing observations. Missing survey data were excluded. After data preprocessing, our nonblinded sample contained 597 participants, and our blinded sample contained 150 participants. The participants' demographics are available in Table 4. The distribution of participants' personality variables are available in Figure 3.

Table 4. Demographics of the blinded set and nonblinded set participants. Results of chi-square tests of independence are shown in the table.

Demographics	Nonblinded (n=597)	Blinded (n=150)	Chi-square (*df*)	*P* value
Age (years)				
Values, mean (SD)	34.33 (9.37)	37.18 (10.83)	N/A[a]	N/A
Values, range	21-68	20-63	N/A	N/A
Sex, n (%)				
Male	346 (58)	93 (62)	0.7 (1)	.42
Female	251 (42)	57 (38)	0.7 (1)	.42
Income (US $), n (%)				
<49,999	46 (7.7)	7 (4.7)	4.8 (4)	.31
50,000-74,999	126 (21.1)	34 (22.7)	4.8 (4)	.31
75,000-99,999	129 (21.6)	33 (22.0)	4.8 (4)	.31
100,000-150,000	162 (27.1)	50 (33.3)	4.8 (4)	.31
150,000	134 (22.4)	26 (17.3)	4.8 (4)	.31
Education, n (%)				
No college degree	43 (7.2)	9 (6)	0.3 (2)	.86
College degree	326 (54.6)	84 (56)	0.3 (2)	.86
Graduate degree	228 (38.2)	57 (38)	0.3 (2)	.86
Supervisor role, n (%)				
Nonsupervisor	313 (52.4)	88 (58.7)	1.6 (1)	.20
Supervisor	284 (47.6)	62 (41.3)	1.6 (1)	.20
English as first language, n (%)				
No	86 (14.4)	8 (5.3)	8.2 (1)	.004
Yes	511 (85.6)	142 (94.7)	8.2 (1)	.004
Had a wearable, n (%)				
No	277 (46.4)	63 (42)	1.2 (2)	.56
Yes	292 (48.9)	78 (52)	1.2 (2)	.56
Unknown	28 (4.7)	9 (6)	1.2 (2)	.56

[a]N/A: not applicable.

Figure 3. Distribution of personalities in the nonblinded set (top) and blinded set (bottom). Independent sample *t* tests found no significant differences between the two sets across personalities (all *P* values >.09).



Analyses

Addressing RQ1 and RQ2

Exploratory Analysis and Variable Selection

We conducted exploratory analyses on the nonblinded set to determine whether variables that were not studied before in the context of compliance or without enough supporting theory behind them are related to compliance, that is, job performance, anxiety, sleep quality, affect, smoking and alcohol use, and physical activity. We compiled the results of our exploratory analysis, show the distribution of other variables in Multimedia Appendix 1, and show a model including all available variables in Multimedia Appendix 2 [12,13,22,52,58-67] along with the interpretation of these variables and their possible relationship with compliance. As these variables, as well as statistics from our study portal use, are uncommon and specific to our study, we conducted our main analyses with a reduced set of variables (14/31, 45%) comprising demographics, personality, and early compliance. We considered our reduced set to be generalizable to more studies (eg, age as opposed to log-ins to our study portal) and to be of interest based on related work. Table 5 shows the means and SDs of the selected variables.

Table 5. Means and SDs of continuous variables in the models. Independent sample t tests (2-tailed) show no evidence of differences in the means between the nonblinded and blinded data sets in all variables tested, except survey compliance and wearable compliance in week 2.

Variable	Nonblinded, mean (SD)	Blinded, mean (SD)	*t* test (*df*)	*P* value
Age (years)	34.33 (9.37)	37.18 (10.83)	—[a]	—
Extraversion	3.45 (0.67)	3.34 (0.73)	1.60 (218)	.11
Agreeableness	3.88 (0.56)	3.93 (0.58)	−0.99 (225)	.33
Conscientiousness	3.89 (0.66)	3.85 (0.68)	0.69 (224)	.49
Neuroticism	2.46 (0.78)	2.46 (0.83)	0.03 (220)	.98
Openness	3.82 (0.60)	3.82 (0.63)	−0.01 (222)	>.99
Number of log-ins in week 2	0.40 (0.94)	0.34 (0.67)	0.82 (312)	.41
Number of issues in week 2	0.00 (0.01)	0.00 (0.00)	1.32 (292)	.19
Survey compliance	0.75 (0.26)	0.80 (0.26)	−2.18 (229)	.03
Wearable compliance	0.68 (0.32)	0.73 (0.31)	−1.54 (234)	.12
Survey compliance in week 2	0.80 (0.27)	0.84 (0.27)	−1.50 (229)	.13
Wearable compliance in week 2	0.78 (0.32)	0.85 (0.28)	−2.45 (256)	.01

[a]The samples of age in each set were not normal (Shapiro-Wilk test: *P*<.001) and appeared as not drawn from the same distribution (Kolmogorov-Smirnov test: *P*=.01). Therefore, the differences in age were not assessed with *t* tests.

Among the demographic measures collected (Table 3), 2 measures are unique to our data set: previous wearable ownership and having a supervising role. Previous wearable ownership is of special interest because an association with compliance could point to a useful characteristic that could be applied to other populations as well. This variable can contribute in contrasting ways to wearable compliance. On the one hand, one can expect this variable to positively affect wearable

compliance, as those with prior wearable experience may be familiar with wearable requirements and capabilities. On the other hand, prior work found that those who did not own a smartwatch and received one in a study could feel motivated to be more diligent and have higher wearable compliance [18]. In addition, we included *supervisory role* in examining compliance. A supervisory role may indicate a busy life or schedule that impairs the ability to participate successfully in the study. However, this variable only applies to a working population.

Model Creation

The response variables in our models are, by definition, ratios with possible values constrained to the interval from 0 to 1 (inclusive). Given our intention of having interpretable models and that our response variables are proportions (windows that contained data over total windows and response rate), we decided to create the models using the beta regression model proposed by Ferrari et al [68], which was specifically designed to model rates and proportions using the beta distribution. We relied on the flexibility of the beta distribution to take different shapes and to represent probabilities and proportions and assumed that the beta distribution would be able to represent distributions that are not normally distributed, such as those in Figure 2 (lack of normality further confirmed through Shapiro-Wilk [69] test $P<.001$). However, the beta regression model cannot handle values of exactly 0 or 1. Thus, we transformed the dependent variables using the following equation [70], with y being the response variable:

For example, the responses of 0 and 1 were transformed to 0.0008 and 0.9992, respectively.

Using 14 variables as predictors, we created hierarchical beta regression models that addressed RQ1 and RQ2. Specifically, to address RQ1, we created model 1s (s=survey as dependent variable) and model 1w (w=wearable as dependent variable), including demographics and personality measures as predictors that can be assessed before starting the study and formally enrolling participants. To address RQ2, we created 4 different models. Models 2s/2w added survey compliance in week 2, which entailed a greater effort to collect than a single survey; however, it did not require giving a wearable to participants while allowing the capture of trait-like compliance. Models 3s/3w added in daily wearable compliance in week 2, which implies that to be able to measure all the predictors, participants would have to have a wearable device for 2 weeks. Note that early compliance variables are entered in the model as percentages to obtain the OR when there is a 1% change in early compliance.

Model Evaluation

As the models contained a different number of variables and R^2 can be inflated because of overfitting of the data, we computed the Akaike information criterion and conducted likelihood ratio tests to compare models trained in the nonblinded data set. We computed R^2, root mean square error, and mean absolute error (MAE) with 5-fold cross-validation to ensure that models trained in a reduced set of the same data have a good fit, have low prediction error, and are robust and unchanging, given more or less information. Finally, we assessed the out-of-sample performance of the model on the blinded data set that was not used (or seen) during modeling.

Given the multiple comparisons involved in our models, all the *P* values of the predictors in each model presented in the results were adjusted using the false discovery rate [71] correction in the stats package [72]. We interpreted the results when the adjusted *P* values in the models were <.10. Multicollinearity was assessed using the generalized variance inflation factors [73]. Visual inspection of diagnostic plots, such as residuals versus indices of observations, Cook distance plot, and residuals versus linear predictor, was conducted following the recommendations of Ferrari et al [68].

Addressing RQ3

A total of 623 individuals completed an *optional* assessment of participation in the study (see Table 6 for inventory and responses). The 5-point Likert scale items were scored from 1 to 5, and nonresponses were dropped. Responses were correlated with compliance in the study. Note that although the items asked in this survey relate to study participation, they were not used for modeling compliance, given that they were asked at the end of the study (RQ1 and RQ2).

Table 6. Exit survey questions related to compliance (N=623 participants answered the exit survey in Tesserae).

Question and item scale and options	Percentage of responses, n (%)[a]	Values, mean (SD)	Total responses, n
How difficult was it to participate in our study?			
1–Extremely easy	215 (37.3)	1.90 (0.89)	576
2–Somewhat easy	251 (43.6)	1.90 (0.89)	576
3–Neither easy nor difficult	71 (12.32)	1.90 (0.89)	576
4–Somewhat difficult	35 (6.1)	1.90 (0.89)	576
5–Extremely difficult	4 (0.7)	1.90 (0.89)	576
What were the biggest difficulties in maintaining compliance? (Select all that apply)[b]			
Technical issues (eg, device broke and did not work)	280 (50.5)	N/A[c]	555
Personal reasons (eg, friends, family, sickness, injury, and travel)	203 (36.6)	N/A	555
Work reasons (eg, change of employment, promotion, and busy schedule)	107 (19.2)	N/A	555
Survey issues (eg, surveys too often, too long, or too many)	75 (13.5)	N/A	555
Privacy issues (eg, type or quantity of data collected and worries about data safety)	20 (3.6)	N/A	555
Difficulty with setup	48 (8.6)	N/A	555
Other—please describe	63 (11.4)	N/A	555
Was your compensation adequate for your participation in the study?			
1–Extremely inadequate	11 (1.9)	4.06 (1.02)	577
2–Somewhat inadequate	46 (8.0)	4.06 (1.02)	577
3–Neither adequate nor inadequate	81 (14.0)	4.06 (1.02)	577
4–Somewhat adequate	194 (33.6)	4.06 (1.02)	577
5–Extremely adequate	244 (42.3)	4.06 (1.02)	577
How useful was the compliance portal?			
I do not remember using the portal	51 (8.9)	N/A	576
Strongly disagree	17 (3.0)	N/A	576
Somewhat disagree	33 (5.7)	N/A	576
Neither agree nor disagree	100 (17.4)	N/A	576
Somewhat agree	227 (39.4)	N/A	576
Strongly agree	148 (25.7)	N/A	576
If you had issues, how satisfied were you with the timely resolution of any issues?			
1–Extremely dissatisfied	4 (0.7)	4.37 (0.88)	573
2–Somewhat dissatisfied	22 (3.8)	4.37 (0.88)	573
3–Neither satisfied nor dissatisfied	62 (10.8)	4.37 (0.88)	573
4–Somewhat satisfied	155 (27.1)	4.37 (0.88)	573
5–Extremely satisfied	329 (57.4)	4.37 (0.88)	573
Can we contact you if we run another study like this?			
Yes	548 (95.1)	N/A	576
No	28 (4.9)	N/A	576

[a]All percentages are calculated over the number of people who completed the item.

[b]Percentages do not sum up to 100% because more than 1 option is allowed.

[c]N/A: not applicable.

Results

Overview

The details and performance of the models addressing RQ1 and RQ2 are presented in Tables 7-9. To ensure that the models were valid, we followed the diagnostic tests outlined in the previous section. We calculated (generalized variance inflation factors) values for all models and found that all values were ≤1.27. Therefore, it is reasonable to assume that multicollinearity is not an issue. Diagnostic plots also did not raise any issues.

Table 7. Model descriptions for beta regression models trained on the nonblinded set predicting survey compliance.[a]

Category variables	Survey compliance					
	Model 1s		Model 2s		Model 3s	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Intercept[c]	1.40 (0.05 to 2.76)	.11	−1.50 (−2.73 to −0.26)	.07	−1.74 (−2.95 to −0.54)	*.02*
Demographics						
Age (years)	1.02 (1.00 to 1.03)	*.03*	1.01 (1.00 to 1.02)	.17	1.01 (1.00 to 1.02)	.39
Sex (male)	1.04 (0.70 to 1.27)	.91	1.12 (0.94 to 1.33)	.40	1.09 (0.92 to 1.30)	.49
Income (US $)						
50,000-75,000	1.04 (0.70 to 1.53)	.97	0.95 (0.68 to 1.34)	.82	0.81 (0.58 to 1.14)	.42
75,000-100,000	0.97 (0.65 to 1.43)	.97	0.89 (0.63 to 1.26)	.69	0.79 (0.56 to 1.11)	.39
100,000-150,000	0.9 (0.61 to 1.33)	.81	0.93 (0.66 to 1.32)	.78	0.82 (0.59 to 1.16)	.44
≥150,000	0.67 (0.44 to 1.01)	.13	0.73 (0.50 to 1.05)	.23	0.65 (0.45 to 0.93)	*.06*
Supervisor (yes)	0.65 (0.54 to 0.79)	*<.001*	0.81 (0.68 to 0.96)	*.07*	0.81 (0.69 to 0.96)	*.05*
English (as first language)	1.38 (1.05 to 1.80)	*.06*	1.15 (0.91 to 1.45)	.43	1.06 (0.84 to 1.33)	.74
Education level						
College degree	0.83 (0.57 to 1.20)	.49	0.91 (0.66 to 1.27)	.69	0.92 (0.67 to 1.26)	.74
Graduate degree	1.01 (0.69 to 1.48)	.97	1.13 (0.80 to 1.58)	.69	1.11 (0.80 to 1.55)	.76
Had a wearable						
Unknown	1.45 (0.94 to 2.25)	.17	1.12 (0.76 to 1.64)	.69	1.04 (0.71 to 1.52)	.88
Yes	1.25 (1.04 to 1.51)	*.06*	1.13 (0.96 to 1.34)	.29	1.05 (0.89 to 1.24)	.73
Personality						
Extraversion	0.74 (0.64 to 0.85)	*<.001*	0.82 (0.72 to 0.93)	*.02*	0.84 (0.74 to 0.96)	*.04*
Agreeableness	0.85 (0.70 to 1.02)	.16	0.93 (0.79 to 1.10)	.64	0.99 (0.84 to 1.16)	.89
Conscientiousness	1.25 (1.07 to 1.46)	*.02*	1.25 (1.09 to 1.42)	*.01*	1.26 (1.11 to 1.44)	*.003*
Neuroticism	0.9 (0.79 to 1.03)	.20	0.90 (0.80 to 1.02)	.23	0.91 (0.81 to 1.02)	.31
Openness	0.99 (0.85 to 1.16)	.97	1.00 (0.87 to 1.14)	.95	0.97 (0.85 to 1.11)	.75
Behavior						
Survey compliance in week 2 (%)	—[d]	—	1.03 (1.03 to 1.04)	*<.001*	1.02 (1.02 to 1.03)	*<.001*
Wearable compliance in week 2 (%)	—	—	—	—	1.01 (1.01 to 1.01)	*<.001*

[a]*P* values of <.10 are denoted in italics.

[b]OR: odds ratio.

[c]Intercept is not an odds ratio but an estimate.

[d]Variable was not included in the model.

Table 8. Model descriptions for beta regression models trained on the nonblinded set predicting wearable compliance (from least to most effort in data collection).[a]

Category variables	Wearable compliance					
	Model 1w		Model 2w		Model 3w	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Intercept[c]	1.40 (−0.04 to 2.84)	.10	−0.64 (−2.04 to −0.77)	.53	−1.19 (−2.51 to 0.12)	.14
Demographics						
Age (years)	1.02 (1.01 to 1.04)	*<.001*	1.02 (1.01 to 1.03)	*.009*	1.01 (1.00 to 1.02)	*.05*
Sex (male)	0.97 (0.78 to 1.20)	.76	1.01 (0.82 to 1.23)	.95	0.97 (0.80 to 1.17)	.80
Income (US $)						
50,000-75,000	1.19 (0.79 to 1.80)	.60	1.24 (0.84 to 1.83)	.45	0.97 (0.67 to 1.40)	.92
75,000-100,000	1.14 (0.75 to 1.74)	.60	1.19 (0.80 to 1.77)	.53	1.01 (0.70 to 1.46)	.96
100,000-150,000	1.19 (0.79 to 1.81)	.60	1.38 (0.93 to 2.05)	.23	1.16 (0.80 to 1.68)	.62
≥150,000	0.87 (0.56 to 1.35)	.60	1.09 (0.71 to 1.65)	.78	0.92 (0.62 to 1.36)	.79
Supervisor (yes)	0.66 (0.54 to 0.81)	*<.001*	0.77 (0.64 to 0.94)	*.03*	0.79 (0.66 to 0.95)	*.04*
English (as first language)	1.39 (1.05 to 1.85)	*.05*	1.24 (0.95 to 1.62)	.23	1.07 (0.83 to 1.37)	.78
Education level						
College degree	0.86 (0.58 to 1.27)	.60	0.88 (0.61 to 1.27)	.63	0.82 (0.58 to 1.16)	.42
Graduate degree	0.88 (0.58 to 1.32)	.60	0.88 (0.60 to 1.30)	.63	0.80 (0.56 to 1.15)	.41
Had a wearable						
Unknown	1.60 (1.01 to 2.55)	*.09*	1.33 (0.86 to 2.07)	.35	1.26 (0.83 to 1.91)	.42
Yes	1.50 (1.23 to 1.83)	*<.001*	1.33 (1.10 to 1.61)	*.011*	1.20 (1.01 to 1.44)	*.097*
Personality						
Extraversion	0.67 (0.57 to 0.78)	*<.001*	0.74 (0.64 to 0.86)	*.001*	0.79 (0.69 to 0.91)	*.005*
Agreeableness	0.68 (0.56 to 0.83)	*.001*	0.72 (0.60 to 0.87)	*.003*	0.79 (0.66 to 0.94)	*.03*
Conscientiousness	1.34 (1.14 to 1.58)	*.001*	1.31 (1.13 to 1.53)	*.003*	1.35 (1.17 to 1.56)	*<.001*
Neuroticism	0.85 (0.73 to 0.98)	*.05*	0.85 (0.74 to 0.97)	*.045*	0.85 (0.75 to 0.97)	*.04*
Openness	1.03 (0.87 to 1.22)	.75	1.02 (0.87 to 1.20)	.83	0.96 (0.83 to 1.12)	.78
Behavior						
Survey compliance in week 2 (%)	—[d]	—	1.02 (1.02 to 1.03)	*<.001*	1.01 (1.01 to 1.01)	*<.001*
Wearable compliance in week 2 (%)	—	—	—	—	1.02 (1.02 to 1.03)	*<.001*

[a]*P* values of <.10 are denoted in italics.

[b]OR: odds ratio.

[c]Intercept is not an odds ratio but an estimate.

[d]Variable was not included in the model.

Table 9. Model performance.

Test and metrics	Survey compliance			Wearable compliance		
	Model 1s	Model 2s	Model 3s	Model 1w	Model 2w	Model 3w
Likelihood ratio test						
df[a]	19	20	21	19	20	21
LogLik	247	405	435	214	283	370
Chi-square (*df*)	N/A[b]	316 (1)	59 (1)	N/A	140 (1)	174 (1)
P value	N/A	<.001	<.001	N/A	<.001	<.001
Nonblinded						
AIC[c]	−455	−767	−824	−389	−527	−698
R^2	0.14	0.44	0.48	0.19	0.37	0.53
5-fold cross-validation						
R^2 training set	0.15	0.45	0.49	0.20	0.38	0.54
R^2 testing set	0.11	0.44	0.48	0.18	0.33	0.50
MAE[d] training set	0.20	0.14	0.13	0.26	0.22	0.19
MAE testing set	0.21	0.14	0.14	0.27	0.23	0.19
Blinded						
R^2	0.16	0.58	0.62	0.25	0.50	0.66
MAE	0.21	0.12	0.11	0.25	0.20	0.16

[a]Degrees of Freedom.

[b]N/A: not applicable.

[c]AIC: Akaike information criterion.

[d]MAE: mean absolute error.

Associations With Compliance

The results in model 1s/1w generally replicated previous findings of associations with compliance: conscientiousness (model 1s: odds ratio [OR] 1.25, 95% CI 1.07-1.46, *P*=.02; model 1w: OR 1.34, 95% CI 1.14-1.58, *P*=.001) and age (model 1s: OR 1.02, 95% CI 1.00-1.03, *P*=.03; model 1w: OR 1.02, 95% CI 1.00-1.02, *P*<.001) were positively associated with compliance [13,31,34], sex did not have an effect on compliance [31], and extraversion had a negative effect (model 1s: OR 0.74, 95% CI 0.64-0.85, *P*<.001; model 1w: OR 0.67, 95% CI 0.57-0.78, *P*<.001) in compliance [31]. However, agreeableness (model 1w: OR 0.68, 95% CI 0.56-0.83, *P*=.001) and neuroticism (model 1w: OR 0.85, 95% CI 0.73-0.98, *P*=.05) were negatively associated with compliance, and income, education, and openness were not found to be statistically significant, which contradicts previous studies [13,31,33,34], possibly because of the following methodological differences: we treated compliance as a ratio and not a binary variable, our models included more controls than previous studies, and our sample was one of information workers, not of students, that also under sampled lower-income workers.

English as a native language was associated with higher compliance in our sample (model 1s: OR 1.38, 95% CI 1.05-1.80, *P*=.06; model 1w: OR 1.39, 95% CI 1.05-1.85, *P*=.05). Command of the English language could have facilitated participation in our study, given how all manuals, surveys, and communications with participants were written in English. Having a supervisory role, which we speculated could be an indicator of busyness, was negatively associated with compliance as expected. Participating in the study is one more competing need in a busy schedule that could preclude participants from dedicating as much time as they would have done otherwise. Finally, having had a wearable was positively associated with compliance (model 1s: OR 1.25, 95% CI 1.04-1.51, *P*=.06; model 1w: OR 1.50, 95% CI 1.23-1.83, *P*<.001). People who have had wearables before were familiar with the technology and demonstrated interest in them, which could serve as motivation to use it more, thus staying more compliant.

The significance of age, supervising role, and having had a wearable changed in the models as early compliance variables were added, possibly because of early compliance being highly correlated with long-term compliance (Table 2), and these other variables not helping above and beyond early compliance as controls.

RQ1: Participant Characteristics and Compliance Before the Study Starts

When it comes to goodness of fit, the results from model 1s and model 1w show that with a demographics and personality survey, we can explain 19% of the wearable compliance

variance (model 1w) and 14% of the daily survey compliance variance (model 1s). Cross-validated and blinded set R^2 shows that the models are not substantially overfitting with R^2 values close (15% versus 11% and 20% versus 18%) or slightly higher (14% versus 16% and 19% versus 25%) than the values for the data used in fitting the model (training versus testing and nonblinded versus blinded). Cross-validated results show a testing set MAE of 0.21 (model 1s) and 0.27 (model 1w), indicating that when using 80% of the data for training and predicting on new data, the predictions are on average within 0.21 and 0.26 of the actual compliance. The correlation between the predicted and actual values slightly decreases when comparing training and testing sets, which is expected. However, the MAE also indicates that the model is not substantially overfitting (even when training on 80% of the data) as the error increase from training to testing is ≤0.01.

In the blinded set, the error of model 1s was similar to that of the cross-validated version when comparing MAE (0.21). In the case of model 1w, the MAE was lower in the blinded set (0.20 versus 0.23).

RQ2: Early Assessment of Compliance and Long-term Compliance

We constructed models 2s/2w and 3s/3w to address whether very early compliance would be indicative of future compliance, as compliance has been shown to be trait-like [12,31,74]. Early daily survey compliance in week 2 was a good predictor for both survey and wearable compliance (model 2s: OR 1.03, 95% CI 1.03-1.04, P<.001; model 2w: OR 1.02, 95% CI 1.02-1.03, P<.001), and thus a good proxy for trait-like compliance.

Models 2s/2w and 3s/3w show a significant improvement when compared with the corresponding model 1. Adding early wearable compliance in model 3s/3w improved the fit across all tests and provided an improvement in our blinded data set by reducing error. As more tasks are added, more trait-like compliance can be captured early on. A lower Akaike information criterion indicated that both models are of better relative quality than the corresponding model 1, and models 3s/3w are better than models 2s/2w. Likelihood ratio tests further confirmed this. Although a training increase in R^2 is expected in regression models as more variables are added, the improvement on the cross-validated and blinded set R^2 with respect to the respective model 1 values shows a far better fit to previously unseen data.

When comparing model 3s with model 2s in predicting survey compliance, the benefit of including wearable compliance in week 2 (model 3s: OR 1.01, 95% CI 1.01-1.01, P<.001) is minor, with only a small increase in fit and a decrease in error. In the case of predicting wearable compliance, model 3w shows an improvement over model 2w by including wearable compliance in week 2 (model 3w: OR 1.02, 95% CI 1.02-1.03, P<.001). Cross-validated MAE reduced by 0.04, and R^2 increased by 0.17, whereas the blinded set R^2 increased by 0.16, and MAE decreased by 0.04.

Overall, survey compliance can be predicted with less error than wearable compliance. The cross-validated and blinded set MAE of both predictions is within a reasonable value to be useful, and the relatively high R^2 indicates that the predictions correlate highly with actual compliance.

Finally, as we predicted aggregated compliance, we needed to make sure that compliance is relatively stable to make sure that the predictions of these models could be generalized to studies of varying lengths. We know from Table 2 that week 2 compliance and long-term compliance are correlated. Therefore, we calculated intraclass correlation coefficients (ICC) and found that there was good agreement [75] between weekly compliances throughout the study in the entire data set (blinded and nonblinded) for both variables: wearable compliance, ICC was 0.695 (95% CI 0.673-0.718), and survey compliance, ICC was 0.656 (95% CI 0.630-0.682). Furthermore, to test the agreement of aggregate measures when surveys were no longer required, we aggregated wearable compliance until day 56 and correlated it with study-long compliance showing good agreement as well r=.82 (95% CI 0.80-0.85). The correlations in Table 2—ICCs—along with previous findings [31], suggest that any long-term aggregate of compliance would significantly correlate with the aggregation of compliance used in the analyses (56 days for surveys and year-long for wearable).

RQ3: Self-assessment of Compliance

Among the participants who answered our questions related to compliance and study participation, 80.9% (466/576) thought that it was easy to participate in the study, 75.9% (438/577) thought that they were extremely well- or somewhat well-compensated, and 95.1% (548/576) said that we could contact them if we were to run another study like this one. Notably, 65.1% (375/576) of the participants agreed that the portal was useful, 17.4% (100/576) neither agreed nor disagreed that the portal was useful, and only 17.5% (101/576) found the portal not useful or did not recall using it. 84.5% (484/573) of the participants felt somewhat or extremely satisfied with the way issues were resolved throughout the study. Overall, participants found that the biggest obstacles toward compliance were technical issues (280/555, 50.5%), followed by personal reasons (203/555, 36.6%), work reasons (107/555, 19.2%), survey issues (75/555, 13.5%), difficulty with setup (48/555, 8.6%), privacy (20/555, 3.6%), and other reasons (63/555, 11.4%).

In addition, portal use and perceptions were correlated with compliance. Portal usefulness was positively correlated with wearable compliance (r=0.139; P=.002) and survey (r=0.217; P<.001) compliance. Perceived timely resolution was correlated with wearable compliance (r=0.250; P<.001) and survey compliance (r=0.128; P=.002). Ease of study participation was associated with wearable compliance (r=0.355; P<.001) and survey compliance (r=0.252; P<.001). Finally, compensation was positively correlated with wearable compliance (r=0.281; P<.001) and survey compliance (r=0.190; P<.001). Taken together, these correlations demonstrate that, in general, perceived usefulness, timely issue resolution, ease of study participation, and adequate compensation are associated with higher compliance. Thus, effective feedback and timely problem resolution are useful goals for researchers looking to maximize compliance.

Discussion

Recommendations for Future Study Design

Although researchers cannot change the characteristics of participants found to be associated with compliance (ie, demographics and personality), there are still different strategies based on the models presented throughout this work and the exploration of the RQs proposed.

Oversampling

One initial strategy could be oversampling groups likely to drop out or be less compliant (eg, those who had higher extraversion or lower conscientiousness). We do not recommend simply excluding participants based on these variables, as this would introduce bias. We showed that a short survey of demographics and personality traits could predict compliance early on, with an MAE of 0.26 for wearable compliance (model 1w) and 0.20 for daily survey compliance (model 1s). Furthermore, our findings related to RQ2 suggest that there is value in an initial 2-week pilot in which participants fill out a subset of actual tasks or rough equivalents. We do not think the items themselves matter but only that the surveys are short (<2 minutes to complete). We would expect this to generalize as follows: completion of *some* of the full set of study tasks during the pilot will allow researchers to observe a *trait-like* characteristic of compliance in the participants, which will be indicative of how much they comply with study tasks in general. Alternatively, the pilot could include the full set of tasks (eg, survey completion and wearable use). This proved to have the best fit in our models for both kinds of compliance studied. However, in the case of a study involving wearables, providing the devices to the pilot participants could entail higher costs with minimal benefit (Table 9; models 3s/3w versus model 2s/2w). Nevertheless, a 2-week pilot would be cost-effective for studies that pay more as more data get collected or studies that require a certain level of data regardless of the population being sampled. Using Tesserae payment as an example, participants were paid US $50 to complete enrollment at week 1, US $150 at the end of week 12, US $200 at the end of 24 weeks, and US $350 at the conclusion of the study. A total of 107 participants dropped out of the study. In many study designs, data from participants who do not complete the study might be excluded because of insufficient data. If all 107 participants were compliant through week 12 and subsequently became noncompliant or dropped out, participants would have been paid US $200 for data that may not be useful in achieving the study goals. If a survey-only pilot lasting 2 weeks was conducted that screened 107 noncompliant participants, they would only have cost the study US $50. With the US $150 savings per participant, plus the savings of US $550 not paid to participants who dropped out, Tesserae researchers could have recruited an additional 99 fully compliant participants without deviating from the original budget. Researchers who have a specific budget for participant payments can thus maximize data collection by estimating compliance through a short pilot study.

Targeted Participant Engagement

If a study is not long enough or the budget does not allow for oversampling, using the models early could suggest to researchers which participants are likely to require extra support to engage with them properly or perhaps provide flexibility or an adaptive schedule in the completion of tasks, such as EMAs that interrupt ongoing activities unlike passive sensing [37]. Approximately 14% of the participants who answered what the biggest difficulties in maintaining compliance were identified *surveys too often, too long, and too many* as one of their difficulties. However, researchers need to carefully consider the interventions in the study to prevent them from being counterproductive. Throughout our study, the interventions were kept to a minimum. Participants were reminded of syncing their wearable only once a week (Mondays) in the event of missing data (most likely because of delayed data syncing from the wearable) and only if initial data from the wearable had not been received. Similar interventions were sent in the case of missing smartphone agent data, lack of beacon sightings in a significant period, or consistent periods of not responding to daily surveys. Despite limited interventions, the compliance for the study was quite high, with median compliance rates of 85.7% (649/757) for the daily surveys, 84.7% (641/757) for the wearables, and 93.7% (709/757) for the smartphone agent among all participants. Obtaining adequate compliance with minimal interventions can reduce experimenter effort and save experimenter resources while reducing participant interaction. More interventions do not necessarily produce better compliance [19,31], and too many notifications can lower their importance to participants [76].

Providing a Study Portal

Finally, we recommend providing support in the form of troubleshooting and a compliance tracking portal that can help participants stay compliant [77]. The portal in our study comprised an issue tracker and dashboards for the researchers and participants, with study researchers being able to track compliance as a study aggregate as well as per participant. Participants were able to track their compliance throughout the study and easily contact the researchers in cases where they saw a discrepancy between the compliance levels shown on the portal and their own expectations. Participants confirmed the usefulness of the portal, with 65.1% (375/576) believing that it was useful and only 17.5% (101/576) thinking that it was not useful or not remembering having used it.

Limitations and Future Work

It is important to note the several limitations of this work. As the participants were largely drawn from a population of information workers from 4 organizations, these participants may not represent all information workers or the general working population. For example, lower-income individuals were underrepresented in the sample. Although ethnicity was previously not found to be associated with compliance [31], we did not collect information about race; thus, we do not have a way of knowing if the sample was diverse and representative of the broader US population.

In addition, the study design could not explore how maintaining compliance on one stream affected compliance for other streams. Thus, it is possible that a 2-week pilot with only surveys may not have the same effect on compliance as a 2-week pilot with all streams.

Finally, there is the aspect of rewards and compensation that we could not examine or control for in our analysis because of study requirements and the fact that all participants that received lottery payments belonged to a single cohort. Although Musthag et al [78] found no differences among the 3 payment schemes in a study comparing the effect of incentives on compliance, the authors believed that if the incentives had lower values, they would have observed differences in the compliance rate. Harari et al [33] found a markedly different compliance across the 3 incentive schemes that relied on course credit and feedback, compensation and feedback, and a prize reward—keeping the wearable—at the end of the study. Given that at least 9.9% (57/577) of our participants found compensation to be inadequate, with a slight majority (35/57, 61%) of that 9.9% having received lottery payments, it is possible that a future study would find stipend compensation to be more effective than lottery-based payments.

In addition to addressing the above limitations, future work could examine the rate of dropouts through a survival analysis using time-varying covariates, as well as whether periods or onset of noncompliance are marked by spikes in stress or changes in sleep patterns. In this work, as we focused on the early prediction of long-term compliance instead of ongoing prediction, we did not include time-varying covariates. Furthermore, developing purely predictive models based on the findings of this work and such time-series analyses could lead to the development of effective study design and management tools that support decisions before and during the study to maximize compliance in studies.

Conclusions

Our work is an extensive analysis of sensor compliance for a longitudinal study of a population of information workers from multiple organizations and across the United States. We presented predictions of compliance in the Tesserae study along with a detailed description of the methodology of the study. We considered 31 variables and presented 6 beta regression models, with 14 selected variables that evaluated the association between compliance and participants' demographics, personality, and trait-like compliance early in the study. We presented participants' challenges in maintaining compliance, their satisfaction with troubleshooting issues, and their assessment of the portal that provided feedback on compliance and help with troubleshooting. From this work, we draw recommendations for future longitudinal studies that aim to improve efficiency by maximizing the amount of data collected. Ultimately, our work provides insights to improve the experimental setup of a study to maximize the quantity of data collection.

Acknowledgments

This paper is based on work supported in part by the Office of the Director of National Intelligence, Intelligence Advanced Research Projects Activity (IARPA), via IARPA contract number 2017-17042800007. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Office of the Director of National Intelligence, IARPA, or the US government. The US government is authorized to reproduce and distribute reprints for governmental purposes, notwithstanding any copyright annotation therein.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Other variables collected for exploratory analysis.
[DOCX File , 128 KB - mhealth_v9i11e22218_app1.docx]

Multimedia Appendix 2
Exploratory analysis.
[DOCX File , 43 KB - mhealth_v9i11e22218_app2.docx]

References

1. Arigo D, Rohde P, Shaw H, Stice E. Psychosocial predictors of physical activity change among college students in an obesity prevention trial. J Phys Act Health 2017 Jul;14(7):513-519 [FREE Full text] [doi: 10.1123/jpah.2016-0515] [Medline: 28290744]
2. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. USA: ACM; 2014 Presented at: UbiComp '14: The 2014 ACM Conference on Ubiquitous Computing; September 13 - 17, 2014; Seattle, Washington p. 3-14. [doi: 10.1145/2632048.2632054]
3. Harari GM, Müller SR, Aung MS, Rentfrow PJ. Smartphone sensing methods for studying behavior in everyday life. Curr Opin Behav Sci 2017;18:83-90. [doi: 10.1016/j.cobeha.2017.07.018]
4. Tong C, Harari G, Chieh A, Bellahsen O, Vegreville M, Roitmann E, et al. Inference of big-five personality using large-scale networked mobile and appliance data. In: Proceedings of the 16th Annual International Conference on Mobile Systems,

Applications, and Services. USA: ACM; 2018 Presented at: MobiSys '18: The 16th Annual International Conference on Mobile Systems, Applications, and Services; June 10 - 15, 2018; Munich Germany p. 1-530. [doi: 10.1145/3210240.3210823]
5. Mohamed R, Youssef M. HeartSense: ubiquitous accurate multi-modal fusion-based heart rate estimation using smartphones. New York, NY, United States: Association for Computing Machinery; 2017 Presented at: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; September 2017 p. 1-18. [doi: 10.1145/3132028]
6. Rabbi M, Aung MS, Gay G, Reid MC, Choudhury T. Feasibility and acceptability of mobile phone-based auto-personalized physical activity recommendations for chronic pain self-management: pilot study on adults. J Med Internet Res 2018;20(10):e10147 [FREE Full text] [doi: 10.2196/10147] [Medline: 30368433]
7. Mehrotra A, Musolesi M. Designing effective movement digital biomarkers for unobtrusive emotional state mobile monitoring. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services. USA: ACM; 2017 Presented at: MobiSys'17: The 15th Annual International Conference on Mobile Systems, Applications, and Services; June 23, 2017; Niagara Falls, New York USA p. 3-8. [doi: 10.1145/3089341.3089342]
8. Tseng V, Merrill M, Wittleder F, Abdullah S, Aung M, Choudhury T. Assessing mental health issues on college campuses: preliminary findings from a pilot study. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. USA: ACM; 2016 Presented at: UbiComp '16: The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 12 - 16, 2016; Heidelberg, Germany. [doi: 10.1145/2968219.2968308]
9. Rawassizadeh R, Momeni E, Dobbins C, Mirza-Babaei P, Rahnamoun R. Lesson learned from collecting quantified self information via mobile and wearable devices. J Sens Actuator Netw 2015 Nov 05;4(4):315-335. [doi: 10.3390/jsan4040315]
10. Harrison D, Marshall P, Bianchi-Berthouze N, Bird J. Tracking physical activity: problems related to running longitudinal studies with commercial devices. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. USA: ACM; 2014 Presented at: UbiComp '14: The 2014 ACM Conference on Ubiquitous Computing; September 13 - 17, 2014; Seattle, Washington p. 699-702. [doi: 10.1145/2638728.2641320]
11. DasMahapatra P, Chiauzzi E, Bhalerao R, Rhodes J. Free-living physical activity monitoring in adult us patients with multiple sclerosis using a consumer wearable device. Digit Biomark 2018;2(1):47-63 [FREE Full text] [doi: 10.1159/000488040] [Medline: 32095756]
12. Rowlands AV, Harrington DM, Bodicoat DH, Davies MJ, Sherar LB, Gorely T, et al. Compliance of adolescent girls to repeated deployments of wrist-worn accelerometers. Med Sci Sports Exerc 2018 Jul;50(7):1508-1517. [doi: 10.1249/MSS.0000000000001588] [Medline: 29474208]
13. Lee PH, Macfarlane DJ, Lam T. Factors associated with participant compliance in studies using accelerometers. Gait Posture 2013 Sep;38(4):912-917. [doi: 10.1016/j.gaitpost.2013.04.018] [Medline: 23688408]
14. Mukhopadhyay SC. Wearable sensors for human activity monitoring: a review. IEEE Sensors J 2015;15(3):1321-1330. [doi: 10.1109/JSEN.2014.2370945]
15. de Lima AL, Hahn T, Evers LJ, de Vries NM, Cohen E, Afek M, et al. Feasibility of large-scale deployment of multiple wearable sensors in Parkinson's disease. PLoS One 2017;12(12):e0189161 [FREE Full text] [doi: 10.1371/journal.pone.0189161] [Medline: 29261709]
16. Merilahti J, Pärkkä J, Antila K, Paavilainen P, Mattila E, Malm E, et al. Compliance and technical feasibility of long-term health monitoring with wearable and ambient technologies. J Telemed Telecare 2009;15(6):302-309. [doi: 10.1258/jtt.2009.081106] [Medline: 19720768]
17. Chen J, Lauderdale D. Cognitive function, consent for participation, and compliance with wearable device protocols in older adults. J Gerontol A Biol Sci Med Sci 2019;74(2):269-273 [FREE Full text] [doi: 10.1093/gerona/gly032] [Medline: 29579176]
18. Jeong H, Kim H, Kim R, Lee U, Jeong Y. Smartwatch wearing behavior analysis. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2017 Presented at: ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; September 2017 p. 1-31. [doi: 10.1145/3131892]
19. Booth BM, Mundnich K, Feng T, Nadarajan A, Falk TH, Villatte JL, et al. Multimodal human and environmental sensing for longitudinal behavioral studies in naturalistic settings: framework for sensor selection, deployment, and management. J Med Internet Res 2019 Aug 20;21(8):e12832 [FREE Full text] [doi: 10.2196/12832] [Medline: 31432781]
20. Intille S, Haynes C, Maniar D, Ponnada A, Manjourides J. µEMA: Microinteraction-based Ecological Momentary Assessment (EMA) using a smartwatch. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous. 2016 Sep Presented at: ACM International Joint Conference on Pervasive and Ubiquitous; September 12 - 16, 2016; Heidelberg, Germany p. 1124-1128 URL: http://europepmc.org/abstract/MED/30238088 [doi: 10.1145/2971648.2971717]
21. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
22. Schüz N, Walters J, Frandsen M, Bower J, Ferguson SG. Compliance with an EMA monitoring protocol and its relationship with participant and smoking characteristics. Nicotine Tob Res 2014 May 19;16 Suppl 2(Suppl 2):88-92. [doi: 10.1093/ntr/ntt142] [Medline: 24052500]
23. Chan L, Swain VD, Kelley C, de Barbaro K, Abowd GD, Wilcox L. Students' experiences with ecological momentary assessment tools to report on emotional well-being. In: Proceedings of the ACM on Interactive, Mobile, Wearable and

Ubiquitous Technologies. 2018 Mar 26 Presented at: ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; March 2018 p. 1-20. [doi: 10.1145/3191735]
24. Saha K, Chan L, De Barbaro K, Abowd GD, De Choudhury M. Inferring mood instability on social media by leveraging ecological momentary assessments. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2017 Presented at: ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; September 11, 2017; - p. 1-27. [doi: 10.1145/3130960]
25. De Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. In: Proceedings of the 5th Annual ACM Web Science Conference. 2013 Presented at: 5th Annual ACM Web Science Conference; May 2 - 4, 2013; Paris, France p. 47-56. [doi: 10.1145/2464464.2464480]
26. Das Swain V, Saha K, Rajvanshy H, Sirigiri A, Gregg JM, Lin S, et al. A multisensor person-centered approach to understand the role of daily activities in job performance with organizational personas. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2019 Presented at: ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; December 11, 2019; - p. 1-27. [doi: 10.1145/3369828]
27. Tseng V, Abdullah S, Costa J, Choudhury T. AlertnessScanner: what do your pupils tell about your alertness. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services. USA: ACM; 2018 Presented at: MobileHCI '18: 20th International Conference on Human-Computer Interaction with Mobile Devices and Services; September 3 - 6, 2018; Barcelona, Spain p. 1-11. [doi: 10.1145/3229434.3229456]
28. Monkaresi H, Bosch N, Calvo RA, D'Mello SK. Automated detection of engagement using video-based estimation of facial expressions and heart rate. IEEE Trans Affective Comput 2017 Jan 1;8(1):15-28. [doi: 10.1109/taffc.2016.2515084]
29. Liu S, Jiang Y, Striegel A. Face-to-face proximity estimation using bluetooth on smartphones. IEEE Trans on Mobile Comput 2014 Apr;13(4):811-823. [doi: 10.1109/tmc.2013.44]
30. Mundnich K, Booth BM, L'Hommedieu M, Feng T, Girault B, L'Hommedieu J, et al. TILES-2018, a longitudinal physiologic and behavioral data set of hospital workers. Sci Data 2020 Oct 16;7(1):354 [FREE Full text] [doi: 10.1038/s41597-020-00655-3] [Medline: 33067468]
31. Faust L, Jiménez-Pazmino P, Holland J, Lizardo O, Hachen D, Chawla N. What 30 days tells us about 3 years: identifying early signs of user abandonment and non-adherence. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. USA: ACM; 2019 Presented at: 13th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 20 - 23, 2019; Trento, Italy p. 216-224. [doi: 10.1145/3329189.3329196]
32. Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]
33. Harari GM, Müller SR, Mishra V, Wang R, Campbell AT, Rentfrow PJ, et al. An evaluation of students' interest in and compliance with self-tracking methods. Soc Psychol Personal Sci 2017 Jul 18;8(5):479-492. [doi: 10.1177/1948550617712033]
34. Evenson KR, Sotres-Alvarez D, Deng YU, Marshall SJ, Isasi CR, Esliger DW, et al. Accelerometer adherence and performance in a cohort study of US Hispanic adults. Med Sci Sports Exerc 2015 Apr;47(4):725-734 [FREE Full text] [doi: 10.1249/MSS.0000000000000478] [Medline: 25137369]
35. Wang W, Harari GM, Wang R, Müller SR, Mirjafari S, Masaba K, et al. Sensing behavioral change over time: using within-person variability features from mobile sensing to predict personality traits. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2018 Presented at: ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; September 18, 2018; - p. 1-21. [doi: 10.1145/3264951]
36. Wen C, Schneider S, Stone A, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017 Apr 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
37. van Berkel N, Goncalves J, Hosio S, Sarsenbayeva Z, Velloso E, Kostakos V. Overcoming compliance bias in self-report studies: a cross-study analysis. Int J Hum Comput 2020 Feb;134:1-12. [doi: 10.1016/j.ijhcs.2019.10.003]
38. Courvoisier DS, Eid M, Lischetzke T. Compliance to a cell phone-based ecological momentary assessment study: the effect of time and personality characteristics. Psychol Assess 2012 Sep;24(3):713-720. [doi: 10.1037/a0026733] [Medline: 22250597]
39. Dzubur E, Huh J, Maher J, Intille S, Dunton G. Response patterns and intra-dyadic factors related to compliance with ecological momentary assessment among mothers and children. Transl Behav Med 2018 Mar 01;8(2):233-242 [FREE Full text] [doi: 10.1093/tbm/ibx002] [Medline: 29381785]
40. Alshurafa N, Jain J, Alharbi R, Iakovlev G, Spring B, Pfammatter A. Is more always better?: discovering incentivized mhealth intervention engagement related to health behavior trends. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2018 Dec 27 Presented at: ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; December 27, 2018 p. 1-26 URL: http://europepmc.org/abstract/MED/32318650 [doi: 10.1145/3287031]
41. Martinez G, Mattingly S, Mirjafari S, Subigya N, Campbell A, Dey A, et al. WristSense 2020: 6th Workshop on Sensing Systems and Applications using Wrist Worn Smart Devices – Program. In: Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). 2020 Presented at: IEEE International

Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); March 23-27, 2020; Austin, TX, USA. [doi: 10.1109/percomworkshops48775.2020.9156169]
42. Mattingly S, Gregg J, Audia P, Bayraktaroglu A, Campbell A, Chawla N, et al. The tesserae project: large-scale, longitudinal, in situ, multimodal sensing of information workers. In: Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. USA: ACM; 2019 Presented at: CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK. [doi: 10.1145/3290607.3299041]
43. MOSAIC: multimodal objective sensing to assess individuals with context. Intelligence Advanced Research Projects Activity. URL: https://www.iarpa.gov/index.php/research-programs/mosaic [accessed 2020-11-11]
44. Saha K, Bayraktaraglu A, Campbell A, Chawla N, De CM. Social media as a passive sensor in longitudinal studies of human behavior and wellbeing. In: Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland, UK. [doi: 10.1145/3290607.3299065]
45. Soto CJ, John OP. The next Big Five Inventory (BFI-2): developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. J Pers Soc Psychol 2017 Jul;113(1):117-143. [doi: 10.1037/pspp0000096] [Medline: 27055049]
46. Spielberger C. State-trait anxiety inventory. In: The Corsini Encyclopedia of Psychology. Hoboken, New Jersey, United States: John Wiley & Sons; 2010.
47. Watson D, Clark LA. The PANAS-X: manual for the positive and negative affect schedule-expanded form. Iowa Research Online 1994:A. [doi: 10.17077/48vt-m4t2]
48. World Health Organization, Centers for Disease Control (U.S.) . Tobacco Questions for Surveys: A Subset of Key Questions From the Global Adult Tobacco Survey (GATS) : Global Tobacco Surveillance System. Geneva: World Health Organization; 2011.
49. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the alcohol use disorders identification test (audit): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
50. Hagströmer M, Oja P, Sjöström M. The international physical activity questionnaire (IPAQ): a study of concurrent and construct validity. Public Health Nutr 2006 Sep;9(6):755-762. [doi: 10.1079/phn2005898] [Medline: 16925881]
51. Patterson E. International physical activity questionnaire (October 2002): long last 7 days self-administered format. IPAQ. URL: https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnx0aGVpcGFxfGd4OjNhMzBjMDAyN2NkOGU0ZQ [accessed 2019-07-25]
52. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiat Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4]
53. Shipley W, Gruber C. Shipley-2 Manual. Los Angeles, CA: Western Psychological Services; 2009.
54. Williams LJ, Anderson SE. Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. J Manag 2016 Jun 30;17(3):601-617. [doi: 10.1177/014920639101700305]
55. Griffin MA, Neal A, Parker SK. A new model of work role performance: positive behavior in uncertain and interdependent contexts. Academy Manag J 2007 Apr;50(2):327-347. [doi: 10.5465/amj.2007.24634438]
56. Bennett RJ, Robinson SL. Development of a measure of workplace deviance. J Appl Psychol 2000 Jun;85(3):349-360. [doi: 10.1037/0021-9010.85.3.349] [Medline: 10900810]
57. Fox S, Spector P, Goh A, Bruursema K, Kessler S. The deviant citizen: measuring potential positive relations between counterproductive work behaviour and organizational citizenship behaviour. J Occup Organ Psychol 2012;85(1):220. [doi: 10.1111/j.2044-8325.2011.02032.x]
58. Diefendorff JM, Mehta K. The relations of motivational traits with workplace deviance. J Appl Psychol 2007 Jul;92(4):967-977. [doi: 10.1037/0021-9010.92.4.967] [Medline: 17638458]
59. Alley S, Schoeppe S, Guertler D, Jennings C, Duncan MJ, Vandelanotte C. Interest and preferences for using advanced physical activity tracking devices: results of a national cross-sectional survey. BMJ Open 2016 Jul 07;6(7):e011243 [FREE Full text] [doi: 10.1136/bmjopen-2016-011243] [Medline: 27388359]
60. Ashby FG, Isen AM, Turken AU. A neuropsychological theory of positive affect and its influence on cognition. Psychol Rev 1999 Jul;106(3):529-550. [doi: 10.1037/0033-295x.106.3.529] [Medline: 10467897]
61. Paterson TS, Yeung SE, Thornton WL. Positive affect predicts everyday problem-solving ability in older adults. Aging Ment Health 2016;20(8):871-879. [doi: 10.1080/13607863.2015.1043619] [Medline: 26033072]
62. Watson D, Clark LA. Negative affectivity: the disposition to experience aversive emotional states. Psychol Bull 1984;96(3):465-490. [doi: 10.1037/0033-2909.96.3.465]
63. Erdle S, Rushton JP. The general factor of personality, BIS–BAS, expectancies of reward and punishment, self-esteem, and positive and negative affect. Pers Individ Dif 2010;48(6):762-766. [doi: 10.1016/j.paid.2010.01.025]
64. Ball SA, Zuckerman M. Sensation seeking, Eysenck's personality dimensions and reinforcement sensitivity in concept formation. Pers Individ Dif 1990;11(4):343-353. [doi: 10.1016/0191-8869(90)90216-e]

65. Erez A, Isen AM. The influence of positive affect on the components of expectancy motivation. J Appl Psychol 2002 Dec;87(6):1055-1067. [doi: 10.1037/0021-9010.87.6.1055] [Medline: 12558213]
66. Barber L, Grawitch MJ, Munz DC. Are better sleepers more engaged workers? A self-regulatory approach to sleep hygiene and work engagement. Stress Health 2013;29(4):307-316. [doi: 10.1002/smi.2468] [Medline: 23086901]
67. Barnes CM. Working in our sleep: sleep and self-regulation in organizations. Organ Psychol Rev 2012;2(3):234-257. [doi: 10.1177/2041386612450181]
68. Ferrari S, Cribari-Neto F. Beta regression for modelling rates and proportions. J Appl Stat 2004 Aug;31(7):799-815. [doi: 10.1080/0266476042000214501]
69. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika 1965 Dec 01;52(3-4):591-611. [doi: 10.1093/biomet/52.3-4.591]
70. Smithson M, Verkuilen J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol Methods 2006 Mar;11(1):54-71. [doi: 10.1037/1082-989x.11.1.54]
71. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
72. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2019. URL: https://www.R-project.org/ [accessed 2020-11-07]
73. Fox J, Monette G. Generalized collinearity diagnostics. J Am Stat Asso 1992;87(417):178. [doi: 10.2307/2290467]
74. Maurice A, Howell A, Evans DG, O'Neil AC, Scobie S. Predicting compliance in a breast cancer prevention trial. Breast J 2006;12(5):446-450. [doi: 10.1111/j.1075-122x.2006.00295.x]
75. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
76. Sahami SA, Henze N, Dingler T, Pielot M, Weber D, Schmidt A. Large-scale assessment of mobile notifications. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 26 - May 1, 2014; Toronto, Canada p. 3055-3064. [doi: 10.1145/2556288.2557189]
77. Hsieh G, Li I, Dey A, Forlizzi J, Hudson S. Using visualizations to increase compliance in experience sampling. In: Proceedings of the 10th international conference on Ubiquitous computing. USA: Association for Computing Machinery; 2008 Presented at: 10th international conference on Ubiquitous computing; September 21 - 24, 2008; Seoul, Korea. [doi: 10.1145/1409635.1409657]
78. Musthag M, Raij A, Ganesan D, Kumar S, Shiffman S. Exploring micro-incentive strategies for participant compensation in high-burden studies. In: Proceedings of the 13th international conference on Ubiquitous computing. USA: ACM; 2011 Presented at: 13th international conference on Ubiquitous computing; September, 17 - 21, 2011; Beijing, China. [doi: 10.1145/2030112.2030170]

Abbreviations

EMA: ecological momentary assessment
IARPA: Intelligence Advanced Research Projects Activity
ICC: intraclass correlation coefficient
MAE: mean absolute error
OR: odds ratio
RQ: research question

Edited by L Buis; submitted 30.11.20; peer-reviewed by L König, A Ponnada; comments to author 06.01.21; revised version received 23.04.21; accepted 24.09.21; published 12.11.21.

Please cite as:
Martinez GJ, Mattingly SM, Robles-Granda P, Saha K, Sirigiri A, Young J, Chawla N, De Choudhury M, D'Mello S, Mark G, Striegel A
Predicting Participant Compliance With Fitness Tracker Wearing and Ecological Momentary Assessment Protocols in Information Workers: Observational Study
JMIR Mhealth Uhealth 2021;9(11):e22218
URL: https://mhealth.jmir.org/2021/11/e22218
doi:10.2196/22218
PMID:34766911

Review

Smartphone-Delivered Ecological Momentary Interventions Based on Ecological Momentary Assessments to Promote Health Behaviors: Systematic Review and Adapted Checklist for Reporting Ecological Momentary Assessment and Intervention Studies

Kim Phuong Dao[1,2], MPH; Katrien De Cocker[3*], PhD; Huong Ly Tong[1*], MRes; A Baki Kocaballi[4], PhD; Clara Chow[1], MD, PhD; Liliana Laranjo[1], MD, MPH, PhD

[1]Westmead Applied Research Centre, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[2]Capital Health Network, Canberra, Australia

[3]Institute for Resilient Regions, Centre for Health Research, University of Southern Queensland, Springfield Central, Australia

[4]School of Computer Science, Faculty of Engineering & Information Technology, University of Technology Sydney, Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Liliana Laranjo, MD, MPH, PhD
Westmead Applied Research Centre
Faculty of Medicine and Health
The University of Sydney
Rm No 2041, Research & Education Network, Westmead Hospital
Cnr Darcy and Hawkesbury Roads
Sydney, 2145
Australia
Phone: 61 413461852
Email: liliana.laranjo@sydney.edu.au

Abstract

Background: Healthy behaviors are crucial for maintaining a person's health and well-being. The effects of health behavior interventions are mediated by individual and contextual factors that vary over time. Recently emerging smartphone-based ecological momentary interventions (EMIs) can use real-time user reports (ecological momentary assessments [EMAs]) to trigger appropriate support when needed in daily life.

Objective: This systematic review aims to assess the characteristics of smartphone-delivered EMIs using self-reported EMAs in relation to their effects on health behaviors, user engagement, and user perspectives.

Methods: We searched MEDLINE, Embase, PsycINFO, and CINAHL in June 2019 and updated the search in March 2020. We included experimental studies that incorporated EMIs based on EMAs delivered through smartphone apps to promote health behaviors in any health domain. Studies were independently screened. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed. We performed a narrative synthesis of intervention effects, user perspectives and engagement, and intervention design and characteristics. Quality appraisal was conducted for all included studies.

Results: We included 19 papers describing 17 unique studies and comprising 652 participants. Most studies were quasi-experimental (13/17, 76%), had small sample sizes, and great heterogeneity in intervention designs and measurements. EMIs were most popular in the mental health domain (8/17, 47%), followed by substance abuse (3/17, 18%), diet, weight loss, physical activity (4/17, 24%), and smoking (2/17, 12%). Of the 17 studies, the 4 (24%) included randomized controlled trials reported nonstatistically significant effects on health behaviors, and 4 (24%) quasi-experimental studies reported statistically significant pre-post improvements in self-reported primary outcomes, namely depressive (P<.001) and psychotic symptoms (P=.03), drinking frequency (P<.001), and eating patterns (P=.01). EMA was commonly used to capture subjective experiences as well as behaviors, whereas sensors were rarely used. Generally, users perceived EMIs to be helpful. Common suggestions for improvement included enhancing personalization, multimedia and interactive capabilities (eg, voice recording), and lowering the

EMA reporting burden. EMI and EMA components were rarely reported and were not described in a standardized manner across studies, hampering progress in this field. A reporting checklist was developed to facilitate the interpretation and comparison of findings and enhance the transparency and replicability of future studies using EMAs and EMIs.

Conclusions: The use of smartphone-delivered EMIs using self-reported EMAs to promote behavior change is an emerging area of research, with few studies evaluating efficacy. Such interventions could present an opportunity to enhance health but need further assessment in larger participant cohorts and well-designed evaluations following reporting checklists. Future research should explore combining self-reported EMAs of subjective experiences with objective data passively collected via sensors to promote personalization while minimizing user burden, as well as explore different EMA data collection methods (eg, chatbots).

Trial Registration: PROSPERO CRD42019138739; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=138739

(JMIR Mhealth Uhealth 2021;9(11):e22890) doi:10.2196/22890

KEYWORDS

ecological momentary assessment; ecological momentary intervention; behavior change; health behavior; mHealth; mobile health; smartphone apps; mobile phone

Introduction

Background

Mobile technologies have become popular approaches to promote behavior change and improve health outcomes, offering the ability to reach large populations in an easy, rapid, and low-cost manner [1,2]. Until recently, mobile behavior change interventions were limited to providing automated and predefined generic or minimally tailored messages, mainly based on estimates of *baseline* or *usual* behaviors and their determinants [3]. As people's behaviors are driven by individual and contextual factors that vary across time [4,5], there is a need to make behavior change interventions that are more adaptive to the users' evolving needs and context. Such an adaptive and dynamic intervention approach might help maintain participant engagement, sustain and support continued behavior change for longer durations, and thereby achieve greater health benefits [4-6].

Ecological momentary interventions (EMIs) are behavior change interventions that deliver support in real time, when most needed [7], for example, when the user is most likely to engage in unhealthy behaviors. To provide the information or treatment in real time and in real settings, EMIs are often based on repeated user reports collected via questionnaires, that is, ecological momentary assessments (EMAs) [8]. These EMA self-reports are usually real time or near real time and can focus on behaviors, contexts, emotional states, beliefs, attitudes, perceptions, exposures, events, or experiences in naturalistic settings (eg, "How are you feeling right now?", "What are you doing right now?", and "Are you near anyone smoking?") [9]. EMAs originated in psychology a few decades ago, when these self-reports were primarily paper-based [8,9].

It has been suggested that tailoring EMIs based on EMAs may lead to higher user engagement and intervention effectiveness [7,10,11]. Given the ubiquity of smartphones [12,13], researchers are starting to explore the use of these mobile technologies to collect EMAs and deliver EMIs [14-17]. Previous systematic reviews of EMAs have focused on sedentary behavior, physical activity, and diet, mixing different EMA media for data collection, such as smartphones, PDAs (precursors of smartphones, now discontinued), and paper-and-pencil diaries [18-22]. The few existing systematic reviews on EMIs have focused on mental health and have also included studies with mixed media for EMIs, such as telephone, SMS text messaging, in-person counseling, computers, PDAs, and smartphones (a minority of included studies) [23-25]. To date, no studies have synthesized the current evidence on the use of smartphone-delivered EMIs using EMAs and their impact on health behaviors, user perspectives, or engagement.

Objective

The overall objective of this study is to systematically review the evidence and characteristics of smartphone-delivered EMIs to promote behavior change, using self-reported EMAs, specifically (1) their effects on health behaviors in any health domain, (2) user engagement, and (3) user perspectives. Although not the original aim of this systematic review, another objective arose upon data extraction and analysis—developing a reporting checklist (adapted from an existing checklist [22]) to facilitate interpretation and comparison of findings and enhance transparency and replicability of future studies using EMAs and EMIs.

Methods

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were used when conducting and reporting this systematic review. The protocol was registered in PROSPERO (International Prospective Register of Systematic Reviews; CRD42019138739).

Search Strategy for Identification of Studies

A literature search was conducted in June 2019 (and updated in March 2020) using MEDLINE (via PubMed interface), Embase, PsycINFO, and CINAHL. Search strings included a combination of free terms and controlled vocabulary when supported (complete search strategy available in Multimedia Appendix 1). The reference lists of relevant articles were also screened to ensure that all eligible studies were included. The authors were contacted if there was a need for any additional information about the included studies.

Study Selection Criteria

The eligibility criteria were developed using the PICO (Participants, Intervention, Comparator, Outcomes; Multimedia Appendix 2). Participants included healthy individuals or patients with chronic conditions. We included all experimental studies that incorporated EMIs to improve health behaviors in any health domain. For the purposes of this review, an EMI must have been delivered in real time through smartphone apps and must have been based on data collected from users' repeated reports in their natural context (ie, EMAs) and also via smartphone apps. Outcomes included any measures that illustrated the effects on health behavior changes (eg, changes in step counts and diet changes). Secondary outcomes included perspectives on EMIs and user engagement behaviors with different types of EMIs, including retention rate. No limiting criteria were used regarding comparison groups. Peer-reviewed studies published in English were included, and no restrictions were set regarding publication dates.

We excluded protocols, reviews, opinion pieces, and design and development papers without user evaluation of EMIs. Studies that used EMAs only for the purpose of data collection or outcome measurement were also excluded. Other exclusion criteria included interventions that relied solely on the automated data collected (eg, only through sensors and no user-reported EMAs) and interventions that were not based on data submitted by the participants (ie, EMAs) via smartphone apps or wearable devices.

Screening, Data Extraction, and Synthesis

A pilot screening of the studies was completed before the actual screening process began. The title and abstract screening and full-text screening were conducted by 2 independent investigators. A third researcher resolved disagreements. Cohen κ was applied to measure the intercoder agreement in each screening phase.

An investigator extracted the information from the included studies into a standardized form, and another researcher reviewed the form for consistency. The data collected from each study included the first author, year of publication, location, health domain, intervention aim, study design and duration, participants' settings and characteristics, EMA data collection characteristics (eg, type of information collected from participants, prompting design and frequency—following the CREMAS [Checklist for Reporting EMA Studies] reporting checklist [22]), intervention components (eg, app, website, and therapy sessions), smartphone-based EMI characteristics (eg, frequency), health-related outcomes, user's perspectives regarding EMIs and EMAs, and user engagement. Behavior change techniques (BCTs) were coded by 2 researchers using the BCT taxonomy [26]. Included randomized controlled trials (RCTs) were appraised by 2 researchers using the Cochrane risk of bias tool [27]. Nonrandomized studies were appraised by 2 researchers using the *Risk Of Bias In Non-randomized Studies of Interventions* tool [28]. A narrative synthesis was conducted for all included studies.

Results

Description of Included Studies

The search returned 2824 results (Figure 1). Of the 2824 studies, after removing duplicates, 2162 (76.56%) studies underwent title and abstract screening. Of the 2162 studies, there were 81 (3.75%) studies for full-text screening; of the 81 studies, 66 (81%) were excluded for not meeting the inclusion criteria (reasons for exclusion are presented in Multimedia Appendix 3). Cohen κ scores were 0.3 and 0.5 for abstract and full-text screening, respectively. We included 15 papers from the original search and 4 additional papers from other sources (reference lists of included studies and database search updates), corresponding to 19 articles, describing 17 unique studies (Table 1).

The 17 included studies (19 papers) involved a total of 652 participants [29-47] (Table 1). Most studies (13/17, 76%; 15/19, 79% papers) were conducted in the United States [29-32,34,37-45,47]. Publication years ranged from 2011 to 2020 (13/17, 76% studies were published from 2016 onward). Study duration ranged from 2-15 weeks, and the average duration was 4 weeks. Sample size varied from 7-121 participants (mean 35.2, SD 33.3; 67% women). The health domains covered were: mental health [29-36], smoking cessation [37-39], and substance abuse control [40-42], as well as diet, weight loss, and physical activity [43,44,46,47]. Studies in the mental health domain mostly recruited patients from outpatient clinics diagnosed with a mental health problem (major depressive disorder [29], schizophrenia [34,36], bipolar disorder [31], and other conditions [30,32,35]), and only 1 study focusing on mood and anxiety management recruited participants without a diagnosis [33]. Studies focusing on smoking recruited participants from smoking cessation clinics [37-39]. Studies on substance abuse control recruited individuals currently in treatment for an alcohol disorder from the community [40], college students with problematic drinking [41], and marijuana users from primary care clinics [42]. Finally, studies on diet, weight loss, and physical activity recruited obese individuals undergoing assessment for bariatric surgery [43], overweight or obese participants from the community [44,45], university students interested in well-being [46], and African American women after breast cancer treatment, recruited from the community [47].

Of the 17 studies, there were 4 (24%; 5/19, 26% papers) RCTs [39,42,44,45,47] and 12 (71%; 13/19, 68% papers) quasi-experimental studies (all with a single-arm design [29-35,37,38,40,41,43,46] except for one with 2 arms [36]). Of the 17 studies, participant retention was reported in 14 (82%; 16/19, 84% papers) studies, ranging from 62.1%-100% in the intervention arm [29-31,34,36-47], with 11 (65%; 13/19, 68% papers) studies having retention rates >75% [29-31,34,36-38,40,41,44-47]. The risk of bias of the 4 RCTs was assessed as unclear for most of the *risk of bias* tool categories (Multimedia Appendix 4). Overall risk of bias in nonrandomized studies was assessed as *serious* for most studies (Multimedia Appendix 5).

Figure 1. Flowchart of included studies.

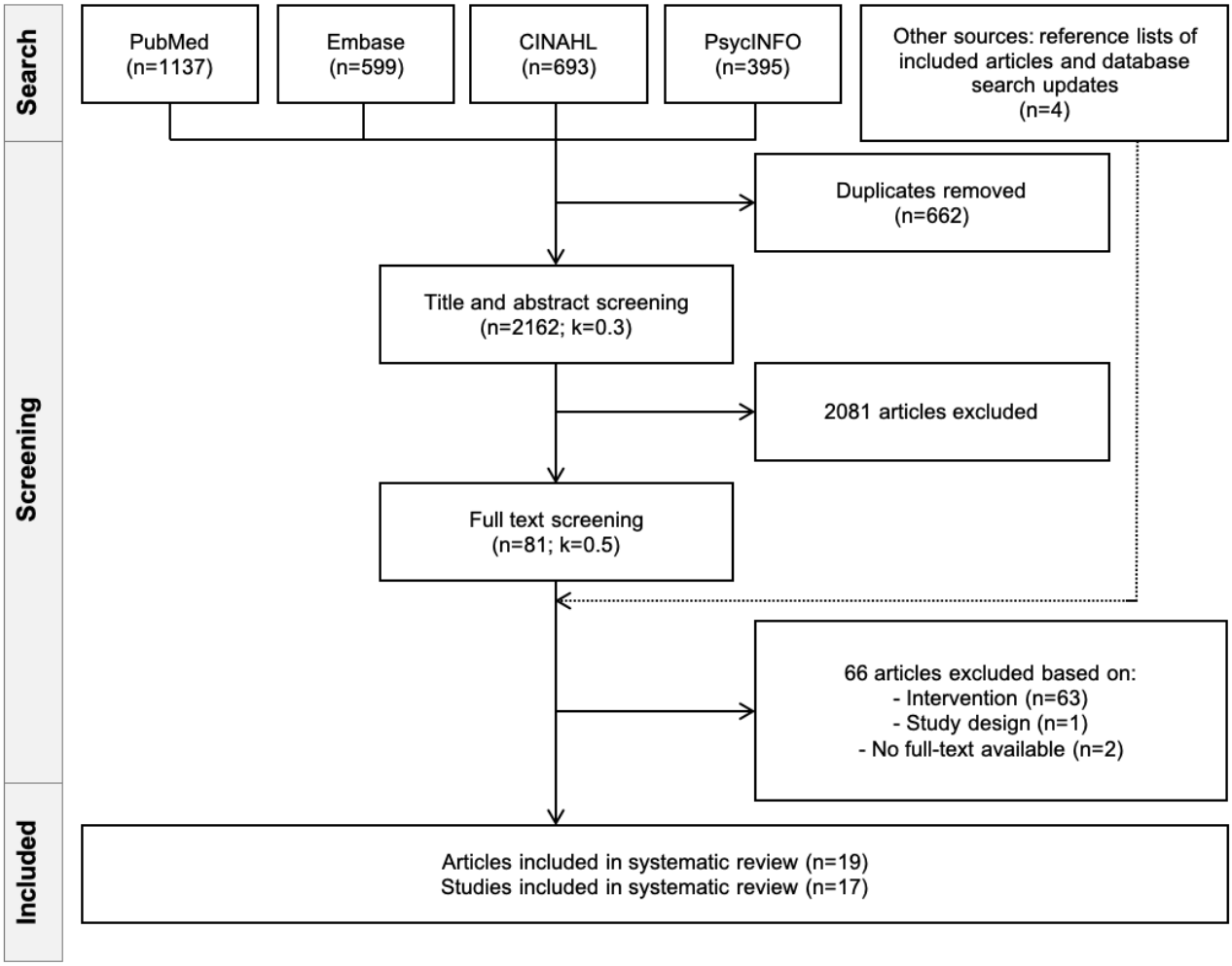


Table 1. Characteristics of included studies.

Author, year, location	Intervention aim	Study design	Duration (weeks); Sample size (I[a];C[b]); Age (years), mean (SD); Women (%);Characteristics	Study retention rate (%)	Intervention components	Health-related outcomes	*P* value
Mental health							
Burns, 2011, United States [29]	Mood disorders management	Quasi-experimental, 1 arm	7; 8 (N/A[c]); 37.4 (12.2); 87%; Adults with major depressive disorder recruited on the web	87	App, website, phone coaching, emails, and sensors	Depressive symptoms decreased postintervention	<.001
Bush, 2014, United States [30]	Mood and anxiety disorders management	Quasi-experimental, 1 arm	2; 8 (N/A); NR[d]; 37%; Military personnel under treatment for behavioral health issues	100	App	NR	—[e]
Wenze, 2016, United States [31]	Bipolar disorder management	Quasi-experimental, 1 arm	12; 8 (N/A); 44 (11.6); 65%; Patients with bipolar disorder from a psychiatric hospital (inpatient and outpatient)	100	App and therapy sessions (4 weekly during 1 month)	NS[f] Change in symptoms or adherence	—
Shrier, 2017, United States [32]	Impulse control disorder management	Quasi-experimental, 1 arm	4; 16 (N/A); 19.6 (NR); 100%; Primary care patients with depressive symptoms and at increased HIV risk	NR	App and therapy sessions	NR	—
Bakker, 2018, Australia [33]	Mood and anxiety disorders management	Quasi-experimental, 1 arm	4; 44 (N/A); 36 (13); 82%; Participants recruited on the web (no diagnosis needed)	NR	App	NR	—
Kreyenbuhl, 2019, United States [34]	Promote antipsychotic medication adherence	Quasi-experimental, 1 arm	2; 7 (N/A); 47.6 (10.4); 0%; African American men with schizophrenia from an outpatient mental health clinic	100	App and clinician appointment	Participants reported taking their antipsychotic medication in 100% of the adherence EMAs[g] to which they responded	—
Vaessen, 2019, The Netherlands [35]	Psychotic disorders management	Quasi-experimental, 1 arm	Results for intervention arm of randomized controlled trial; 16 (N/A); NR; NR; First episode psychosis in the past 3 years, recruited from mental health clinics	NR	App and acceptance and commitment therapy sessions (weekly)	NR	—
Hanssen, 2020, The Netherlands [36]	Schizophrenia spectrum disorders management	Quasi-experimental, 2 arms	3; 64 (NR; NR); 37.9 (8.6); 33%; Patients with schizophrenia spectrum disorder, recruited from hospitals and clinics	78	App	Psychotic symptoms significantly decreased postintervention in the intervention group compared with control (b=−0.005; 95% CI −0.01 to −0.0006)	.03
Smoking cessation							
Businelle, 2016, and Hebert, 2018, United States [37,38]	Smoking cessation and relapse prevention	Quasi-experimental 1 arm	13; 59 (N/A); 52 (7); 54%; Individuals attending a first visit at a smoking cessation clinic	78	App, group counseling, and cessation pharmacotherapy	Abstinence rate decreased over time (41% in week 1 and 20% in week 12)	—

Author, year, location	Intervention aim	Study design	Duration (weeks); Sample size (I[a];C[b]); Age (years), mean (SD); Women (%);Characteristics	Study retention rate (%)	Intervention components	Health-related outcomes	*P* value
Hebert, 2020, United States [39]	Smoking cessation and relapse prevention	Randomized controlled trial, 3 arms	13; 81 (28; 28; 28); 49.6 (11.9); 50%; Individuals referred to a smoking cessation clinic	66	App, group counseling, and cessation pharmacotherapy	Abstinence rate NS between groups	—
Substance abuse control							
Dulin, 2014, United States [40]	Alcohol abuse treatment	Quasi-experimental 1 arm	5; 28 (N/A); 33.6 (6.5); 46%; Individuals currently in treatment for an alcohol disorder, recruited from the community	100	App and sensor	Decrease in percentage of heavy drinking days postintervention (56% vs 25%; Cohen *d*=1.0)	<.001
Leonard, 2017, United States [41]	Alcohol abuse prevention and management	Quasi-experimental, 1 arm	3; 10 (N/A); 20.7 (NR); 100%; College students with problematic drinking not under treatment	100	App, two counseling sessions, and sensor	NR	—
Shrier, 2018, United States [42]	Marijuana use cessation	Randomized controlled trial, 3 arms	12; 70 (NR; NR; NR); 20.7 (NR); 60%; Marijuana users from primary care clinics	66	App and counseling sessions	Percentage of days abstinent, NS between arms	—
Diet and physical activity							
Mundi, 2015, United States [43]	Promote healthy lifestyles to prepare for bariatric surgery	Quasi-experimental, 1 arm	15; 30 (N/A); 41.3 (11.4); 90%; Patients with obesity undergoing assessment for bariatric surgery	67	App	Nutrition knowledge and engagement with healthy lifestyles: NS improvements	—
Goldstein, 2018 and 2020, United States [44,45]	Diet adherence	Randomized controlled trial, 2 arms	10; 121 (62; 59); 47.2 (13.4); 100%; BMI ≥25 kg/m^2 recruited from the community	84.3	App and Weight Watchers program	Weight loss: NS improvements; Lapse frequency: NS improvements	—
Pentikäinen, 2019, Finland [46]	Diet adherence	Quasi-experimental, 1 arm	4; 74 (N/A); 36.2 (12.5); 61%; Individuals interested in well-being, recruited from universities	79	App	The average interval between meals increased; the number of daily eating occasions decreased	.003; .01
Allicock, 2020, United States [47]	Promote physical activity and diet adherence	Randomized controlled trial, 2 arms	8; 22 (13;9); 52 (9); 100%; African American women post breast cancer treatment, recruited from the community	100	App	Reduced sedentary time by 4.37 (SD 7.14) hours/day versus controls; waist circumference, BMI change, physical activity, diet: NS improvements	<.05

[a]I: intervention.

[b]C: control

[c]N/A: not applicable.

[d]NR: not reported.

[e]Not available.

[f]NS: not supported.

[g]EMA: ecological momentary assessment.

Intervention Characteristics

The commonly collected types of EMA data were affect-related (eg, emotions, feelings, and mood; 12/17, 71% studies; 14/19, 74% papers) [29-33,35-39,41,42,44,45], related to behaviors (eg, self-reported diet, physical activity, alcohol use, and medication adherence; 12/17, 71% studies; 14/19, 74% papers) [31-39,42-47], related to cognitions (eg, reasons for nonadherence and reasons to quit alcohol; 10/17, 59% studies; 12/19, 63% papers) [29,31,33-39,42-45], and related to social and environmental context (eg, distractions while eating and interaction with someone smoking; 9/17, 53% studies; 11/19, 58% papers; Table 2) [29,36-45]. In addition to user-reported EMAs, 18% (3/17) of studies also had sensor-collected data (eg, GPS and accelerometer) [40,48,49]. User-reported data collection was initiated either by the app (user would be prompted by the app to provide certain data) [29,31,32,34-39,41-45] or by the user (eg, as users saw fit; after a certain event, such as a meal) [29,30,33,37-41,44-46], sometimes with more than one modality in the same study [37-39,41,44,45,47]. The daily frequency of EMA prompts in app-initiated data collection was reported in 11 (65%; 13/19, 68% papers) studies [29,31,32,35-39,42-45,47], ranging from 2-8 times, with the most common being 4 to 5 times daily [29,32,37-39,42,43]. In 12% (2/17) of studies, the daily frequency was variable, depending on the number of times the participant needed to take medication daily [34] and depending on a trigger from a sensor [41]. The time window allowed for responding to EMA prompts was reported in 29% (5/17) of studies and varied between 1 and 130 minutes (Multimedia Appendix 6) [31,36-38,43,44].

EMIs consisted mostly of suggesting coping strategies (eg, use of cognitive-behavioral skills) [29,31-33,35-45,47], followed by motivational feedback (eg, positive reinforcement and supportive messages) [29,32,34,37-43,47] and informational feedback (eg, user-tailored graphs; Table 2) [30,31,44-47]. EMI characteristics were poorly reported and were not described in a standardized manner across studies, rarely detailing the decision mechanism (eg, algorithm). The EMI mechanism was not reported in 7 (41%, 8/19, 42% papers) studies [31-33,35-39], predetermined in 8 (47%) studies [30,34,40-43,46,47], and adaptive in 2 (12%; 3/19, 16% papers) studies [29,44,45]. The delivery format was in the form of text in most studies [29,31-34,36-45,47]; approximately 12% (2/17) of studies used tailored graphs [30,46], and 6% (1/17) of studies used texts and images [35]. Most interventions used other components in addition to the app, the most common one being counseling sessions with a therapist, either face-to-face or by telephone [29,31,34,35,37-39,41,42].

There were 35 BCTs identified across the studies (Multimedia Appendices 7 and 8). The most popular BCTs were social support (unspecified; 13/17, 76% studies; 15/19, 79% papers) [29,31,32,34,35,37-45,47], followed by prompt or cue (10/17, 59% studies; 11/19, 58% papers) [29,30,32,34-36,40,43-45,47], problem solving (9/17, 53% studies; 11/19, 58% papers) [29,31,33,36-41,44,45], feedback on behavior (6/17, 35% studies) [31,34,36,41,43,46], self-monitoring of behavior (7/17, 41% studies) [29,31,34,40,42,46,47], and social support (emotional; 6/17, 35% studies) [31-33,40-42]. The most commonly mentioned theories, frameworks, or models were cognitive behavioral therapy [31,33,41,42] and motivational interviewing [40-42].

Table 2. Characteristics of EMA[a] data collection and EMI[b] in included studies[c].

Author, year, location	EMA data collection				EMI		
	Type of user-reported data	Mechanism[d]	Format (input mode)	Sensors[e]	Type of intervention content	Mechanism[d]	Format (delivery mode)
Mental health							
Burns, 2011, United States [29]	Affect-related (mood), cognitions, social and environmental context, and motivational states	App-initiated (predetermined, ≥5 times daily at random times between 7 AM and 10 PM, depending on participant preference) and user-initiated (frequency as users see fit)	Likert scales and multiple choice	38 sensors (eg, GPS and accelerometer)	Coping strategies (suggested activities) and motivational feedback (message to reinforce improvement)	App-initiated and adaptive (eg, suggested activities when a user's self-reported mood was outside their typical range, based on a machine-learning algorithm built from EMA and sensor data); frequency, interval, and time allowed: NR[f]	Text
Bush, 2014, United States [30]	Affect-related and mental health-related symptoms and events (stress, head injury, depression, anxiety, well-being)	User-initiated (frequency as users see fit)	Slide bar to rate emotions and states	—[g]	Informational feedback (access to customized reports of mood data and personalized graphs of EMA data)	App-initiated; predetermined; frequency, interval, and time allowed: N/A[h]	Graph
Wenze, 2016, United States [31]	Affect-related, behaviors (daily medications and appointments and adherence behaviors), cognitions (risk factors for nonadherence), and bipolar disorder symptoms (eg, sleep)	App-initiated (time-contingent; 2/day, 9 AM and 9 PM; time allowed: 12 min)	Likert scale and multiple choices	—	Coping strategies and informational feedback	App-initiated; NR	Text
Shrier, 2017, United States [32]	Affect-related, behaviors (sexual behavior), and self-efficacy for safer sex behavior	App-initiated (predetermined, at random times, 4 times daily; time allowed: NR)	NR	—	Coping strategies and motivational feedback (provided supportive messages and prompted use of cognitive-behavioral skills)	App-initiated; NR	Text
Bakker, 2018, Australia [33]	Affect-related (mood), cognitions, and physiological response	User-initiated (frequency as users see fit)	Multiple choice and sliding bars	—	Coping strategies; upon completion of activities, gamified rewards were issued	NR	Text
Kreyenbuhl, 2019, United States [34]	Behaviors (medication adherence at scheduled times throughout the day) and cognitions (reasons for nonadherence)	App-initiated (predetermined, event-contingent and dependent on the number of times the participant needs to take medication daily)	Multiple choice	—	Motivational feedback based on self-reported adherence	App-initiated and predetermined (If-Then, depending on individual responses); frequency and interval dependent on EMA; time allowed: NR	Text

Author, year, location	EMA data collection				EMI		
	Type of user-reported data	Mechanism[d]	Format (input mode)	Sensors[e]	Type of intervention content	Mechanism[d]	Format (delivery mode)
Vaessen, 2019, Netherlands [35]	Affect-related (current mood), behaviors (activity), and symptoms	App-initiated (predetermined, random times, 8 times daily); time allowed: NR	NR	—	Coping strategies (suggested exercise to train general acceptance and commitment therapy principles)	App-initiated; NR	Image and text
Hanssen, 2020, Netherlands [36]	Affect-related (feelings and moods), thoughts, behaviors, cognitions, social and environmental context, and symptoms	App-initiated (predetermined, random, 6 times daily between 10 AM and 10 PM, intervals >130 min; time allowed: NR)	Likert scale, multiple choices, and yes/no answers	—	Coping strategies (provided suggestions for a certain activity or behavior change based on previous EMA answers in the following categories: psychotic symptoms, social engagement, health behavior, and mood and emotion)	App-initiated; mechanism NR; frequency: 2 prompts/day; interval and time allowed: NR	Text
Smoking cessation							
Businelle, 2016 [37] and Hebert, 2018, United States [38]	Affect-related, behaviors (recent alcohol consumption), cognitions (motivation to quit), social and environmental context (eg, cigarette availability and interaction with someone smoking), and urge to smokes	Three types of EMA with three different frequencies: Daily diary (app-initiated; once daily, 30 min after waking; time allowed: 60 seconds); Random sampling (app-initiated; predetermined, random, 4 times daily; time allowed: NR); Event sampling (user-initiated; precessation smoking, urge, and postcessation lapse)	Click buttons to report smoking incidents	—	Coping strategies (provided risk-tailored messages to help participants cope with lapse triggers) and motivational feedback	App-initiated; mechanism NR; frequency and interval: NR; time allowed: NR	Text
Hebert, 2020, United States [39]	Affect-related, behaviors (recent alcohol consumption), cognitions (motivation to quit), social and environmental context (eg, cigarette availability and interaction with someone smoking), and urge to smoke	Three types of EMA with three different frequencies: Daily diary (app-initiated; 1/day, 30 min after waking); Random sampling (app-initiated; predetermined, random, 4 times daily; time allowed: NR); Event sampling (user-initiated; precessation smoking, urge, and postcessation lapse)	Click buttons to report smoking incidents	—	Coping strategies (provided risk-tailored messages to help participants cope with lapse triggers) and motivational feedback	App-initiated; mechanism NR; frequency and interval: NR; time allowed: NR	Text
Substance abuse control							

Author, year, location	EMA data collection				EMI		
	Type of user-reported data	Mechanism[d]	Format (input mode)	Sensors[e]	Type of intervention content	Mechanism[d]	Format (delivery mode)
Dulin, 2014, United States [40]	Social and environmental context (user-identified high-risk locations); cravings	User-initiated (frequency as users see fit)	NR	GPS	Coping strategies (provided audible alert and suggestions for maintaining control of drinking when a boundary was crossed around a GPS-triggered high-risk location) and motivational feedback	App-initiated; predetermined; frequency and interval: based on EMA and sensor data; time allowed: NR	Text
Leonard, 2017, United States [41]	Affect-related (current emotions and level of intensity) and social and environmental context	App-initiated (event-contingent; frequency and interval: based on trigger from sensor; time allowed: NR) and user-initiated	Multiple choice	Electrodermal activity and accelerometer	Coping strategies (based on cognitive behavioral therapy) and motivational feedback	App-initiated; predetermined; frequency and interval based on EMA and sensor data; time allowed: NR	Text
Shrier, 2018, United States [42]	Affect-related, behaviors (use of marijuana), cognitions (personal top three triggers for use and effort to avoid use), social and environmental context, and marijuana desire	App-initiated (random; 4-6 times daily; time allowed: NR)	NR	—	Motivational feedback (provided messages designed to support self-efficacy)	App-initiated; predetermined; frequency and interval based on EMA responses; time allowed: NR	Text
Diet and physical activity							
Mundi, 2015, United States [43]	Behaviors (frequency of eating or snacking and use of calorie-containing beverages, meal planning, frequency of foods not prepared at home, rate of eating, and quantity of physical activity), cognitions (barriers to physical activity), and social and environmental context (distractions while eating)	App-initiated (predetermined, time-contingent; five times daily; time allowed: 60 min)	NR	—	Coping strategies and motivational feedback; upon a study subject's response to the given EMA message, a tailored EMI message was electronically generated (if a patient endorsed a healthy lifestyle, they were sent a congratulatory and supportive message, and if a patient was struggling to make a positive lifestyle modification, they were sent a supportive message outlining some alternative behavioral strategies)	App-initiated; predetermined; frequency and interval based on EMA responses; time allowed: NR	Text

Author, year, location	EMA data collection				EMI		
	Type of user-reported data	Mechanism[d]	Format (input mode)	Sensors[e]	Type of intervention content	Mechanism[d]	Format (delivery mode)
Goldstein, 2018 and 2020, United States [44,45]	Affect-related, behaviors (dietary lapse), cognitions, and social and environmental context (variables known to predict lapses)	App-initiated (predetermined, six times daily; time allowed: 90 min) and user-initiated (after a lapse)	Likert scales and yes or no answers	—	Coping strategies and informational feedback (alert was issued when the algorithm classified a user to be at risk for lapsing, communicating (a) top three factors contributing to level of risk (context-awareness) and (b) strategies to cope with each specific risk factor)	App-initiated; adaptive; frequency and interval based on EMA responses; time allowed: NR	Text
Pentikäinen, 2019, Finland [46]	Behaviors (eating rhythm)	User-initiated (when participant had meal)	Two buttons to record types of eating occasion	—	Informational feedback (graphs of EMA data)	App-initiated; predetermined; frequency and interval based on EMA responses; time allowed: N/A	Tailored graph
Allicock, 2020, United States [47]	Behaviors (diet and physical activity)	Three types: Daily diary (app-initiated; 1/day, 30 min after waking; time allowed: NR); Random sampling (app-initiated; predetermined, random, 2 times daily; time allowed: NR); User-initiated (before and after meals or exercise)	NR	—	Informational, coping strategies, and motivational feedback (providing behavioral cues or prompting, increasing self-efficacy, building behavioral capability, and providing positive reinforcements to behaviors)	App-initiated; predetermined; frequency and interval based on EMA responses); time allowed: NR	Text

[a]EMA: ecological momentary assessment.

[b]EMI: ecological momentary intervention.

[c]EMA and EMI characteristics reported according to items specified in Table 3 based on information reported in the included studies.

[d]Initiative, mechanism, frequency and interval, and time allowed.

[e]Additional components for data collection.

[f]NR: not reported.

[g]Not available.

[h]N/A: not applicable.

Incentives, Adherence, Reported Outcome Measures, and User Perspectives

Participants in 64% (11/17) studies (13/19, 68% papers) received material (eg, movie tickets) or monetary compensations for participating in the study [31,32,36-42,44-47]. Of those 11 studies, 6 (55%) studies (8/13, 62% papers) had incentives associated with EMA completion [31,37-39,42,44,45,47]. Adherence to EMA prompts (ie, to self-reporting data) was mentioned in 59% (10/17) studies (12/19, 63% papers) [31,34,36-39,41-45,47], most often in the form of response rate (Multimedia Appendix 6). The response rate varied from 30.7%-87% (9/17, 53% studies; 11/19, 58% papers; average 64.7%) [31,34,36-39,42-45,47]. Studies with a time limit to respond to EMA (4/17, 24% studies; 5/19, 26% papers) had lower response rates (30.7%, 58%, 62.9%) [31,43-45], except for 6% (1/17) of studies (2/19, 11% papers), with a response rate of 87% and with a high financial incentive for participants (those who completed 50%-74% of assessments received a US $40 gift card; 75-89% completion, US $80 gift card; and >90%, US $120 gift card) [37,38]. Of the 17 studies, 2 (12%) studies (3/19, 16% papers) reported the average time spent on each EMA prompt (ie, time spent self-reporting data), which varied from 2-6 minutes [31,37,38]. Adherence to EMI was reported in 24% (4/17) of studies (6/19, 32% papers) [36-38,44-46], with different measurements in each study (Multimedia Appendix 6).

Of the 17 studies, health-related outcomes were reported in 12 (71%) studies (14/19, 74% papers) [29,31,34,36-40,42-47]. Of the 17 studies, the 4 (24%) included RCTs reported nonstatistically significant improvements in substance abstinence, diet, weight loss, and sedentary time compared with the control group [39,42,44,47], and only 4 (24%) quasi-experimental studies reported statistically significant pre-post improvements in self-reported primary outcomes, namely depressive (P<.001) [29] and psychotic symptoms (P=.03) [36], drinking frequency (P<.001) [40], and eating patterns (P=.01) [46].

Regarding user perspectives (Multimedia Appendix 9), all apps were perceived as useful in supporting behavior change, although to varying degrees. In half of the studies, apps' ease of use was assessed, with users rating the apps favorably [30-34,36,40-42]. The most helpful aspect of the apps, according to participants, was increasing awareness of their own behavior patterns [29,31-33,36-42,44-46]. The preferred and desirable features of the apps included personalization (eg, tailored prompts, tailored content, and feedback based on user responses) [29,31,32,35-40,44,45], communication with clinicians or coaches [29,30,47], multimedia and interactive capabilities, including voice recording [29,30,32,47], and an appealing design of the graphical user interface [30,32,33,36,40]. Common negative perspectives included EMA prompts being too frequent (more than five times daily), inopportune or tedious to complete [33,35,36,41-45], technical issues (eg, battery drainage and connectivity problems) [29-32,34,40-42], and repetitive content and feedback [31,32,42,44,45]. Of the 17 studies, 2 (12%) studies mentioned the potential negative impacts of momentary prompts on users' mental well-being, including increased anxiety and stress because of prompts being too frequent or too sudden [41], or prompts giving users an unpleasant degree of self-awareness [35].

Checklist for Reporting EMA- and EMI-Specific Aspects in Behavior Change Experiments

EMI and EMA components were rarely reported and were not described in a standardized manner across studies. We found that half of the studies failed to report EMA adherence rates, and this was even lower for EMIs. In addition, the mechanism details for EMAs and EMIs and incentives to complete EMAs and adhere to EMIs have been infrequently reported. On the basis of our findings and on an existing CREMAS [22], we developed a set of reporting items to include in the methods and results sections of EMA and EMI experiments (CREMAIs [Checklist for Reporting EMA- and EMI-specific aspects]; Table 3).

Table 3. Adapted checklist for reporting smartphone-delivered EMA[a]- and EMI[b]-specific aspects in behavior change experiments (CREMAIs[c])[d].

Paper section and item	Description	EMA	EMI
Methods			
Type	Details about the type of EMA and EMI	Type of data collected (eg, affect-related, behaviors, cognitions, and social and environmental context)	Intervention content (eg, coping strategies, motivational feedback, informational feedback, and other behavior change techniques)
Mechanism	Initiative	System (eg, app) and/or user-initiated EMA	System (eg, app) and/or user-initiated EMI
	Mechanism responsible for triggering the EMA/EMI	Predetermined (event-contingent, time-contingent and/or random) or adaptive (eg, using statistical/machine learning methods to adapt EMA prompting based on user data)	Predetermined (eg, IF *X* EMA response, THEN *Y* EMI) or adaptive (eg, using statistical/machine learning methods to adapt EMI based on previous EMA responses and other user data)
	Frequency and interval	Number of EMA prompts/day and time between each EMA	Number of EMI prompts/day and time between each EMI
	Time allowed	Total time allowed to answer/receive/perform EMAs before prompt expires	Total time allowed to answer/receive/perform EMIs before prompt expires
Format	Details about how EMAs/EMIs are delivered	Input mode (eg, Likert scales, yes/no answers, multiple choice, voice, free-text, and image)	Delivery mode (eg, voice, text, and image)
Additional components	Other components used in conjunction with the app (eg, sensors; face-to-face behaviors; and website)	Other components used in conjunction with the app (eg, sensors; face-to-face behaviors; and website)	Other components used in conjunction with the app (eg, sensors; face-to-face behaviors; and website)
Behavior change rationale	Theories/frameworks/models to inform the design of the intervention	Theories/frameworks/models to inform the design of the intervention	Theories/frameworks/models to inform the design of the intervention
Incentives	Incentives provided for EMA/EMI adherence	Incentives provided for EMA adherence	Incentives provided for EMI adherence
Results			
Response latency	Average time to respond to EMA/EMI prompt	Average time to respond to EMA prompt	Average time to respond to EMI prompt
Time spent per prompt	Average time spent per EMA/EMI prompt	Average time spent per EMA prompt	Average time spent per EMI prompt
Adherence rate	Response or adherence rate for EMA/EMI prompts, detailing the total number of prompts answered/EMI suggestions implemented, and the total number of prompts delivered	Response or adherence rate for EMA prompts, detailing the total number of prompts answered/EMI suggestions implemented, and the total number of prompts delivered	Response or adherence rate for EMI prompts, detailing the total number of prompts answered/EMI suggestions implemented, and the total number of prompts delivered
Missing data	Report whether EMA/EMI adherence is related to demographic or other variables (eg, prompt relevance)	Report whether EMA adherence is related to demographic or other variables (eg, prompt relevance)	Report whether EMI adherence is related to demographic or other variables (eg, prompt relevance)

[a]EMA: ecological momentary assessment.

[b]EMI: ecological momentary intervention.

[c]CREMAIs: checklist for reporting EMA and EMI-specific aspects.

[d]Adapted from Liao et al [22].

Discussion

Principal Findings

Although the potential for EMIs that build on EMA data for behavior change in the smartphone era seems promising, research on this approach is lacking. We identified 17 studies (only 4 RCTs), all with small sample sizes, short follow-up, and limited evaluation of efficacy. EMIs described were predominantly in mental health management, with a few addressing smoking cessation, substance abuse, diet, weight loss, and physical activity. The most common type of EMA data collected were related to subjective experiences, namely affective states and cognitions, indicating the usefulness of EMAs for this purpose. Behaviors were also often collected via EMAs, with sensors rarely being used. Adherence to collection of EMA data was a common barrier to implementation, with participants disliking the high frequency and tedious nature of EMA data collection. This suggests that EMAs could be gathered via other methods preferred by users (eg, voice). In addition, EMAs could be coupled with passively collected sensor

data whenever possible to decrease user burden while still enabling the collection of subjective experiences relevant to user-desired personalization.

Description of interventions and reporting of evaluation measures were heterogeneous in each health domain, and there were few studies per health domain, limiting any conclusion being made on their efficacy on health behaviors, engagement, and outcomes.

Comparison With Existing Literature

To our knowledge, this is the first systematic review of smartphone-delivered EMIs based on self-reported EMAs to support behavioral changes. Existing reviews of EMIs in the treatment of psychotic disorders [24], major depressive disorder [50], alcohol use [51], and eating disorders [14] found that most interventions were in the early stages of development, which aligns with the findings of this review. Notably, the present findings show that most uses of EMIs based on EMAs to date seem to be in the field of mental health, where emotional and cognitive states can vary considerably throughout the day and influence behaviors. Previous systematic reviews on EMIs have all focused on mental health, used mostly older technologies, and did not tailor EMIs based on EMAs, having found mixed results (2 meta-analyses [23,25] showing small but positive effect sizes and another systematic review demonstrating acceptability and feasibility [24]).

Our review found that EMI and EMA components were rarely reported and were not described in a standardized manner across studies, hampering progress in this field. EMA- and EMI-specific aspects, such as the triggering mechanism and incentives, are important determinants of intervention uptake, retention, and efficacy. Hence, this poor reporting makes it difficult to synthesize and replicate existing evidence. Thus, we developed a set of reporting items—a checklist for reporting EMA- and EMI-specific aspects in behavior change experiments (CREMAIs)—based on an existing reporting checklist for EMA studies (CREMAS) [22]. Given that our adapted checklist focuses exclusively on EMA and EMI aspects, it should be used in conjunction with other reporting guidelines, depending on the type of experimental study design [52-55]. Our findings extend on previous systematic reviews in the field and add to the CREMAS checklist [22] by providing a detailed description of both EMI and EMA components (not just EMA) and specifically with respect to interventions that use smartphones.

EMI users had negative feedback regarding technical issues, inopportune and repetitive alerts, and prompts not being tailored enough, which may decrease participant engagement. The most common recommendations for intervention design were to make the intervention more personalized and engaging (eg, personalized coping strategies) and to tailor data collection and reduce reporting burden and invasiveness. These perspectives expand on existing literature by showing that for sustained efficacy of behavior change interventions, user engagement is paramount [4,6,56]. Personalization has been commonly suggested as a way to make interventions more engaging, effective, and better received by users [57-60]. One example includes *just-in-time adaptive interventions*, which are system-triggered interventions that aim to provide the right type/amount of support, at the right time, by adapting to an individual's changing internal and contextual state (usually based on sensor-collected data) [61].

Strengths and Limitations

This review has several strengths. We developed and followed a protocol that was registered in the PROSPERO database at the start of the study. Intervention components were characterized in detail, including the coding of BCTs. However, the results of this review need to be interpreted in the context of certain limitations. Owing to the small number of RCTs, a meta-analysis was not conducted, and thus it was not possible to provide an estimation of preliminary efficacy. There was low to moderate agreement in screening, which reflects the difficulty in establishing whether a study met the inclusion criteria. Screening was complicated by incomplete intervention descriptions, particularly with regard to EMI and EMA reporting. Finally, the definitions of EMI and EMA are not consensual in the literature. Thus, the studies included in this review reflect the predefined definitions we adopted.

Implications for Future Studies

The use of smartphone-delivered EMIs based on EMAs in behavior change interventions is a novel area of research, where more RCTs are needed to determine efficacy. Given the ubiquity of smartphones, these interventions have the potential to support behavioral changes at scale. Nevertheless, it is still uncertain which populations may find the use of EMIs based on EMAs most acceptable and which populations and settings may benefit the most. So far, studies have focused on mental health, smoking, substance abuse, diet, weight loss, and physical activity, with mixed results. Appropriately powered clinical trials are needed to examine the use of EMIs tailored by EMAs in a range of populations and settings and to examine the impact on health outcomes and the longevity of these benefits.

Future studies should explore the combination of EMAs and sensor data to deliver more personalized and minimally burdensome EMIs. EMA involves manual data collection at several points in time, which can be burdensome for users, but remains important to gather individual data that sensors are currently unable to capture, such as subjectively perceived cognitive and affective states [62]. Capturing subjective experiences (eg, cravings, pain, and loneliness) enables a richer and deeper insight into a person's behavior and can foster the tailoring of an intervention to a person's needs, which in turn may increase the perceived relevance of EMIs. By combining self-reported EMAs of subjective experiences with additional objective data passively collected via sensors (eg, physical activity patterns and heart rate) [63,64], there is potential to promote a more engaging personalized intervention, as minimally burdensome as possible. Novel machine learning algorithms can further explore these different types of data to increase the precision of personalized interventions [65].

A more seamless EMA and EMI experience is crucial for engagement. User burden associated with data entry is the most reported reason why people stop using mobile health apps [66]. In addition to using sensors whenever possible, another possibility to reduce user burden is to optimize the design of

data collection modes. For instance, faster methods, such as speech-based data entry, may be used instead of requiring users to type in response [67]. Another option would be the use of a chatbot to enable data collection in a conversational and more engaging way. Other feasible options include data entry templates, such as dropdown menus, and the use of personalization to autopopulate some data fields [68] based on previous entries or other data sources [33]. Co-designing interventions with users may offer insights into the best options for data collection in each particular case, regarding the types and amount of data, and the mode, frequency, and timing of data collection [69].

Future research in this area should adhere to existing reporting standards, namely, what concerns the detailing of EMA- and EMI-specific characteristics. Reporting guidelines are essential in facilitating the evaluation of study validity and allowing for comparisons across interventions. Consistency and detail in reporting intervention characteristics enable replication efforts and allow for meta-analyses and meta-regression to explore the features associated with the highest user engagement and intervention efficacy. Advancements in the field of EMAs and EMIs and the higher scientific impact of published studies in this area are dependent on the consistent use of reporting guidelines.

Conclusions

This is the first systematic review of smartphone-delivered EMIs based on self-reported EMAs promoting health behaviors. The use of this approach in behavior change is an emerging area of research, with few studies evaluating efficacy and most interventions focusing on mental health management. EMAs were commonly used to capture subjective experiences, as well as behaviors, whereas sensors were rarely used. Future research should explore combining self-reported EMAs of subjective experiences with objective data passively collected via sensors to promote personalization. Studies should also explore the effects of different EMA data collection methods (eg, chatbots) on user burden, engagement, and efficacy. A reporting checklist was developed with the goal of facilitating interpretation and comparison of findings and enhancing transparency and replicability in future studies using EMAs and EMIs.

Acknowledgments

The authors thank Michelle Italiano, Ali Reza, Eddisson Munoz, and Jacqueline Maroon for their help with title, abstract, and full-text screening. The authors also thank Agustina Briatore for her input during the early discussion of the protocol.

Authors' Contributions

The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. Conception and design of the work were done by LL, KPD, HLT, and ABK; database search was conducted by KPD and HLT; title, abstract, and full-text screening was done by KPD, LL, and HLT; outcome data extraction was done by KPD, ABK, and HLT; the first draft was produced by KPD, LL, and KDC; data interpretation and critical revision of drafts for important intellectual content were made by KPD, LL, CC, KDC, ABK, and HLT; and final approval of the version to be published was given by KPD, LL, CC, KDC, ABK, and HLT.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strings.
[DOCX File , 17 KB - mhealth_v9i11e22890_app1.docx]

Multimedia Appendix 2
Inclusion and exclusion criteria.
[DOCX File , 16 KB - mhealth_v9i11e22890_app2.docx]

Multimedia Appendix 3
List of articles excluded after full-text review for not meeting inclusion criteria regarding the population, intervention, outcome or study design.
[DOCX File , 22 KB - mhealth_v9i11e22890_app3.docx]

Multimedia Appendix 4
Risk of bias of included randomized controlled trial.
[DOCX File , 16 KB - mhealth_v9i11e22890_app4.docx]

Multimedia Appendix 5

Risk of bias of nonrandomized studies.
[DOCX File , 18 KB - mhealth_v9i11e22890_app5.docx]

Multimedia Appendix 6
Ecological momentary assessment adherence, ecological momentary intervention adherence, and incentives used.
[DOCX File , 19 KB - mhealth_v9i11e22890_app6.docx]

Multimedia Appendix 7
Behavior change techniques present in the interventions of included studies.
[DOCX File , 18 KB - mhealth_v9i11e22890_app7.docx]

Multimedia Appendix 8
Behaviour change techniques used in each study and excerpts.
[DOCX File , 30 KB - mhealth_v9i11e22890_app8.docx]

Multimedia Appendix 9
User perspectives and suggestions.
[DOCX File , 26 KB - mhealth_v9i11e22890_app9.docx]

References

1. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
2. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, American Heart Association Publications Committee of the Council on Epidemiology and Prevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research, and Stroke Council. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213 [FREE Full text] [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
3. Neubeck L, Lowres N, Benjamin EJ, Freedman SB, Coorey G, Redfern J. The mobile revolution--using smartphone apps to prevent cardiovascular disease. Nat Rev Cardiol 2015 Jun;12(6):350-360. [doi: 10.1038/nrcardio.2015.34] [Medline: 25801714]
4. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016 Sep;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]
5. Conner M, Norman P. Health behaviour: current issues and challenges. Psychol Health 2017 Aug;32(8):895-906. [doi: 10.1080/08870446.2017.1336240] [Medline: 28612656]
6. Müller-Riemenschneider F, Reinhold T, Nocon M, Willich SN. Long-term effectiveness of interventions promoting physical activity: a systematic review. Prev Med 2008 Oct;47(4):354-368. [doi: 10.1016/j.ypmed.2008.07.006] [Medline: 18675845]
7. Heron K, Smyth J. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
8. Stone A, Shiffman S. Ecological momentary assessment (EMA) in behavorial medicine. Ann Behav Med 1994 Sep 01;16(3):199-204. [doi: 10.1093/abm/16.3.199]
9. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
10. Beckjord E, Shiffman S. Background for real-time monitoring and intervention related to alcohol use. Alcohol Res 2014;36(1):9-18. [Medline: 26258996]
11. Schueller SM, Aguilera A, Mohr DC. Ecological momentary interventions for depression and anxiety. Depress Anxiety 2017 Jun;34(6):540-545. [doi: 10.1002/da.22649] [Medline: 28494123]
12. Global mobile consumer survey US edition. Deloitte. 2018. URL: https://www2.deloitte.com/tr/en/pages/technology-media-and-telecommunications/articles/global-mobile-consumer-survey-us-edition.html [accessed 2021-09-22]
13. Global mobile consumer survey: UK cut - Plateauing at the peak: the state of the smartphone. Deloitte. 2019. URL: https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/technology-media-telecommunications/deloitte-uk-plateauing-at-the-peak-the-state-of-the-smartphone.pdf [accessed 2021-09-22]
14. Smith KE, Juarascio A. From ecological momentary assessment (EMA) to ecological momentary intervention (EMI): past and future directions for ambulatory assessment and interventions in eating disorders. Curr Psychiatry Rep 2019 Jun 04;21(7):53. [doi: 10.1007/s11920-019-1046-8] [Medline: 31161276]

15. Schlicht W, Ebner-Priemer UW, Kanning M. Ecological momentary assessment and intervention in physical activity and well-being: affective reactions, social-cognitive factors, and behaviors as determinants of physical activity and exercise. Front Psychol 2013 Dec 06;4:916 [FREE Full text] [doi: 10.3389/fpsyg.2013.00916] [Medline: 24367347]
16. Miller CK. Adaptive intervention designs to promote behavioral change in adults: what is the evidence? Curr Diab Rep 2019 Jan 25;19(2):7. [doi: 10.1007/s11892-019-1127-4] [Medline: 30684109]
17. Parmar A, Sharma P. Ecological momentary interventions delivered by smartphone apps: applications in substance use treatment in indian scenario. Indian J Psychol Med 2017;39(1):102-103 [FREE Full text] [doi: 10.4103/0253-7176.198942] [Medline: 28250570]
18. Schembre SM, Liao Y, O'Connor SG, Hingle MD, Shen S, Hamoy KG, et al. Mobile ecological momentary diet assessment methods for behavioral research: systematic review. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e11170 [FREE Full text] [doi: 10.2196/11170] [Medline: 30459148]
19. Romanzini CL, Romanzini M, Batista MB, Barbosa CC, Shigaki GB, Dunton G, et al. Methodology used in ecological momentary assessment studies about sedentary behavior in children, adolescents, and adults: systematic review using the checklist for reporting ecological momentary assessment studies. J Med Internet Res 2019 May 15;21(5):e11967 [FREE Full text] [doi: 10.2196/11967] [Medline: 31094349]
20. Mason TB, Do B, Wang S, Dunton GF. Ecological momentary assessment of eating and dietary intake behaviors in children and adolescents: a systematic review of the literature. Appetite 2020 Jan 01;144:104465 [FREE Full text] [doi: 10.1016/j.appet.2019.104465] [Medline: 31541670]
21. Degroote L, DeSmet A, De Bourdeaudhuij I, Van Dyck D, Crombez G. Content validity and methodological considerations in ecological momentary assessment studies on physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2020 Mar 10;17(1):35 [FREE Full text] [doi: 10.1186/s12966-020-00932-9] [Medline: 32151251]
22. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREMAS). J Med Internet Res 2016 Jun 21;18(6):e151. [doi: 10.2196/jmir.4954] [Medline: 27328833]
23. Versluis A, Verkuil B, Spinhoven P, van der Ploeg MM, Brosschot JF. Changing mental health and positive psychological well-being using ecological momentary interventions: a systematic review and meta-analysis. J Med Internet Res 2016 Jun 27;18(6):e152 [FREE Full text] [doi: 10.2196/jmir.5642] [Medline: 27349305]
24. Bell IH, Lim MH, Rossell SL, Thomas N. Ecological momentary assessment and intervention in the treatment of psychotic disorders: a systematic review. Psychiatr Serv 2017 Nov 01;68(11):1172-1181. [doi: 10.1176/appi.ps.201600523] [Medline: 28669284]
25. Loo Gee B, Griffiths K, Gulliver A. Effectiveness of mobile technologies delivering ecological momentary interventions for stress and anxiety: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):221-229. [doi: 10.1093/jamia/ocv043] [Medline: 25997643]
26. 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]
27. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 6. New Jersey, United States: Wiley-Blackwell; 2019.
28. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Br Med J 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
29. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
30. Bush NE, Ouellette G, Kinn J. Utility of the T2 mood tracker mobile application among army warrior transition unit service members. Mil Med 2014 Dec;179(12):1453-1457. [doi: 10.7205/MILMED-D-14-00271] [Medline: 25469967]
31. Wenze SJ, Armey MF, Weinstock LM, Gaudiano BA, Miller IW. An open trial of a smartphone-assisted, adjunctive intervention to improve treatment adherence in bipolar disorder. J Psychiatr Pract 2016 Nov;22(6):492-504 [FREE Full text] [doi: 10.1097/PRA.0000000000000196] [Medline: 27824786]
32. Shrier LA, Spalding A. "Just take a moment and breathe and think": young women with depression talk about the development of an ecological momentary intervention to reduce their sexual risk. J Pediatr Adolesc Gynecol 2017 Feb;30(1):116-122. [doi: 10.1016/j.jpag.2016.08.009] [Medline: 27575408]
33. Bakker D, Kazantzis N, Rickwood D, Rickard N. Development and pilot evaluation of smartphone-delivered cognitive behavior therapy strategies for mood- and anxiety-related problems: moodmission. Cogn Behav Pract 2018 Nov 23;25(4):496-514 [FREE Full text] [doi: 10.1016/j.cbpra.2018.07.002]
34. Kreyenbuhl J, Record EJ, Himelhoch S, Charlotte M, Palmer-Bacon J, Dixon LB, et al. Development and feasibility testing of a smartphone intervention to improve adherence to antipsychotic medications. Clin Schizophr Relat Psychoses 2019;12(4):152-167 [FREE Full text] [doi: 10.3371/CSRP.KRRE.070816] [Medline: 27454213]

35. Vaessen T, Steinhart H, Batink T, Klippel A, Van Nierop M, Reininghaus U, et al. ACT in daily life in early psychosis: an ecological momentary intervention approach. Psychosis 2019 Mar 19;11(2):93-104. [doi: 10.1080/17522439.2019.1578401]
36. Hanssen E, Balvert S, Oorschot M, Borkelmans K, van Os J, Delespaul P, et al. An ecological momentary intervention incorporating personalised feedback to improve symptoms and social functioning in schizophrenia spectrum disorders. Psychiatry Res 2020 Feb;284:112695 [FREE Full text] [doi: 10.1016/j.psychres.2019.112695] [Medline: 31831201]
37. Businelle MS, Ma P, Kendzor DE, Frank SG, Vidrine DJ, Wetter DW. An ecological momentary intervention for smoking cessation: evaluation of feasibility and effectiveness. J Med Internet Res 2016 Dec 12;18(12):e321 [FREE Full text] [doi: 10.2196/jmir.6058] [Medline: 27956375]
38. Hébert ET, Stevens EM, Frank SG, Kendzor DE, Wetter DW, Zvolensky MJ, et al. An ecological momentary intervention for smoking cessation: the associations of just-in-time, tailored messages with lapse risk factors. Addict Behav 2018 Mar;78:30-35 [FREE Full text] [doi: 10.1016/j.addbeh.2017.10.026] [Medline: 29121530]
39. Hébert ET, Ra CK, Alexander AC, Helt A, Moisiuc R, Kendzor DE, et al. A mobile just-in-time adaptive intervention for smoking cessation: pilot randomized controlled trial. J Med Internet Res 2020 Mar 09;22(3):e16907 [FREE Full text] [doi: 10.2196/16907] [Medline: 32149716]
40. Dulin PL, Gonzalez VM, Campbell K. Results of a pilot test of a self-administered smartphone-based treatment system for alcohol use disorders: usability and early outcomes. Subst Abus 2014;35(2):168-175. [doi: 10.1080/08897077.2013.821437] [Medline: 24821354]
41. Leonard NR, Silverman M, Sherpa DP, Naegle MA, Kim H, Coffman DL, et al. Mobile health technology using a wearable sensorband for female college students with problem drinking: an acceptability and feasibility study. JMIR Mhealth Uhealth 2017 Jul 07;5(7):e90 [FREE Full text] [doi: 10.2196/mhealth.7399] [Medline: 28687533]
42. Shrier LA, Burke PJ, Kells M, Scherer EA, Sarda V, Jonestrask C, et al. Pilot randomized trial of MOMENT, a motivational counseling-plus-ecological momentary intervention to reduce marijuana use in youth. Mhealth 2018 Jul 30;4:29 [FREE Full text] [doi: 10.21037/mhealth.2018.07.04] [Medline: 30148142]
43. Mundi MS, Lorentz PA, Grothe K, Kellogg TA, Collazo-Clavell ML. Feasibility of smartphone-based education modules and ecological momentary assessment/intervention in pre-bariatric surgery patients. Obes Surg 2015 Oct;25(10):1875-1881. [doi: 10.1007/s11695-015-1617-7] [Medline: 25702141]
44. Goldstein S. Comparing effectiveness and user behaviors of two versions of a just-in-time adaptive weight loss smartphone app. Drexel University 2018:A. [doi: 10.17918/d8fh3r]
45. Goldstein SP, Thomas JG, Foster GD, Turner-McGrievy G, Butryn ML, Herbert JD, et al. Refining an algorithm-powered just-in-time adaptive weight control intervention: a randomized controlled trial evaluating model performance and behavioral outcomes. Health Informatics J 2020 Dec;26(4):2315-2331 [FREE Full text] [doi: 10.1177/1460458220902330] [Medline: 32026745]
46. Pentikäinen S, Tanner H, Karhunen L, Kolehmainen M, Poutanen K, Pennanen K. Mobile phone app for self-monitoring of eating rhythm: field experiment. JMIR Mhealth Uhealth 2019 Mar 27;7(3):e11490 [FREE Full text] [doi: 10.2196/11490] [Medline: 30916657]
47. Allicock M, Kendzor D, Sedory A, Gabriel KP, Swartz MD, Thomas P, et al. A pilot and feasibility mobile health intervention to support healthy behaviors in african american breast cancer survivors. J Racial Ethn Health Disparities 2021 Feb;8(1):157-165. [doi: 10.1007/s40615-020-00767-x] [Medline: 32385847]
48. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
49. Leonard NR, Silverman M, Sherpa DP, Naegle MA, Kim H, Coffman DL, et al. Mobile health technology using a wearable sensorband for female college students with problem drinking: an acceptability and feasibility study. JMIR Mhealth Uhealth 2017 Jul 07;5(7):e90 [FREE Full text] [doi: 10.2196/mhealth.7399] [Medline: 28687533]
50. Colombo D, Fernández-Álvarez J, Patané A, Semonella M, Kwiatkowska M, García-Palacios A, et al. Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: a systematic review. J Clin Med 2019 Apr 05;8(4):465 [FREE Full text] [doi: 10.3390/jcm8040465] [Medline: 30959828]
51. Beckjord E, Shiffman S. Background for real-time monitoring and intervention related to alcohol use. Alcohol Res 2014;36(1):9-18. [Medline: 26258996]
52. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Gesundheitswesen 2016 Mar;78(3):175-188. [doi: 10.1055/s-0041-111066] [Medline: 26824401]
53. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004 Mar;94(3):361-366. [doi: 10.2105/ajph.94.3.361] [Medline: 14998794]
54. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

55. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. Br Med J 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
56. Ory M, Smith ML, Mier N, Wernicke MM. The science of sustaining health behavior change: the health maintenance consortium. Am J Health Behav 2010;34(6):647-659 [FREE Full text] [doi: 10.5993/ajhb.34.6.2] [Medline: 20604691]
57. Tong HL. Can personalised mobile technology increase physical activity and healthy diet? A systematic review and meta-analysis. International Society of Behavioral Nutrition and Physical Activity, Auckland, NZ. 2020. URL: https://venuewest.eventsair.com/QuickEventWebsitePortal/isbnpa2020annualmeeting/eventinfo/Agenda/AgendaItemDetail?id=2be6426d-3c89-4c46-b959-9e8e0692aa88 [accessed 2021-10-01]
58. Lustria ML, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
59. Lustria ML, Cortese J, Gerend MA, Schmitt K, Kung YM, McLaughlin C. A model of tailoring effects: a randomized controlled trial examining the mechanisms of tailoring in a web-based STD screening intervention. Health Psychol 2016 Nov;35(11):1214-1224. [doi: 10.1037/hea0000399] [Medline: 27441869]
60. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
61. Nahum-Shani I, Smith S, Spring B, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
62. Boehner K, DePaula R, Dourish P, Sengers P. Affect: from information to interaction. In: Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility. 2005 Presented at: Aarhus05: 4th Aarhus Conference 2005 - Critical Computing; between Sense and Sensibility; August 20 - 24, 2005; Aarhus Denmark p. 59-68. [doi: 10.1145/1094562.1094570]
63. Insel TR. Digital phenotyping: technology for a new science of behavior. J Am Med Assoc 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
64. Bertz JW, Epstein DH, Preston KL. Combining ecological momentary assessment with objective, ambulatory measures of behavior and physiology in substance-use research. Addict Behav 2018 Aug;83:5-17 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.027] [Medline: 29174666]
65. Wongvibulsin S, Martin SS, Saria S, Zeger SL, Murphy SA. An individualized, data-driven digital approach for precision behavior change. Am J Lifestyle Med 2019 Apr 25;14(3):289-293 [FREE Full text] [doi: 10.1177/1559827619843489] [Medline: 32477031]
66. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
67. Ruan S, Wobbrock JO, Liou K, Ng A, Landay JA. Comparing speech and keyboard text entry for short messages in two languages on touchscreen phones. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2017 Jan 08 Presented at: ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; December 2017; - p. 1-23. [doi: 10.1145/3161187]
68. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec;31(12):1417-1426 [FREE Full text] [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
69. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019 May;45(2):159-172 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]

Abbreviations

BCT: behavior change technique
CREMAIs: checklist for reporting ecological momentary assessment and ecological momentary intervention-specific aspects
CREMAS: checklist for reporting ecological momentary assessment studies
EMA: ecological momentary assessment
EMI: ecological momentary intervention
PICO: Participants, Intervention, Comparator, Outcomes
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 26.07.20; peer-reviewed by S Dick, Y Huang, C Simons; comments to author 11.09.20; revised version received 06.11.20; accepted 26.07.21; published 19.11.21.

Please cite as:
Dao KP, De Cocker K, Tong HL, Kocaballi AB, Chow C, Laranjo L
Smartphone-Delivered Ecological Momentary Interventions Based on Ecological Momentary Assessments to Promote Health Behaviors: Systematic Review and Adapted Checklist for Reporting Ecological Momentary Assessment and Intervention Studies
JMIR Mhealth Uhealth 2021;9(11):e22890
URL: https://mhealth.jmir.org/2021/11/e22890
doi:10.2196/22890
PMID:34806995

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org