

Understanding Participant Needs for Engagement and Attitudes towards Passive Sensing in Remote Digital Health Studies

Samantha Kolovson
Human Centered
Design & Engineering,
University of Washington
kolovson@uw.edu

Ryan Allred
Psychiatry & Behavioral Sciences,
University of Washington
rallred@uw.edu

Abhishek Pratap
Biomedical Informatics &
Medical Education,
University of Washington
Sage Bionetworks
apratap@sagebase.org

Sean A. Munson
Human Centered
Design & Engineering,
University of Washington
smunson@uw.edu

Jaden Duffy
Psychiatry & Behavioral Sciences,
University of Washington
jduffy25@uw.edu

Patricia A. Areán
Psychiatry & Behavioral Sciences,
University of Washington
parean@uw.edu

ABSTRACT

Digital psychiatry is a rapidly growing area of research. Mobile assessment, including passive sensing, could improve research into human behavior and may afford opportunities for rapid treatment delivery. However, retention is poor in remote studies of depressed populations in which frequent assessment and passive monitoring are required. To improve engagement and understanding participant needs overall, we conducted semi-structured interviews with 20 people representative of a depressed population in a major metropolitan area. These interviews elicited feedback on strategies for long-term remote research engagement and attitudes towards passive data collection. Our results found participants were uncomfortable sharing vocal samples, need researchers to take a more active role in supporting their understanding of passive data collection, and wanted more transparency on how data were to be used in research. Despite these findings, participants trusted researchers with the collection of passive data. They further indicated that long term study retention could be improved with feedback and return of information

based on the collected data. We suggest that researchers consider a more educational consent process, giving participants a choice about the types of data they share in the design of digital health apps, and consider supporting feedback in the design to improve engagement.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods; Human computer interaction (HCI); KEYWORDS**

mental health; depression; mental health interventions; remote study; interviews; qualitative research; digital health

ACM Reference Format:

Samantha Kolovson, Abhishek Pratap, Jaden Duffy, Ryan Allred, Sean A. Munson, and Patricia A. Areán. 2020. Understanding Participant Needs for Engagement and Attitudes towards Passive Sensing in Remote Digital Health Studies. In *14th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '20)*, May 18–20, 2020, Atlanta, GA, USA. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3421937.3422025>

1 INTRODUCTION & RELATED WORK

There is a growing interest in the use of mobile technology to improve mental health assessment, treatment, and services [4, 7–10, 17, 21, 29, 31, 32, 43, 44, 50]. According to the World Health Organization [52], digital or mobile health tools are defined as the use of wireless and mobile technologies to support and improve health outcomes, services, quality of care and health research. In particular, smartphones and wearable sensors offer an opportunity to assist clinical research, by increasing the number of people scientists can reach and by collecting moment-to-moment information on social and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

PervasiveHealth '20, May 18–20, 2020, Atlanta, GA, USA

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7532-0/20/05...\$15.00

<https://doi.org/10.1145/3421937.3422025>

physical activity, sleep, heart rate, environmental exposures as well as send surveys on a regular basis. Though the recruitment of large samples is feasible, the ability to retain these large samples over time has been highly challenging, with the majority of studies retaining less than 10% of the sample [39]. This presents a problem related to bias in data collection and the ability to draw conclusions from the study [12, 31, 39].

Reasons for poor adherence include technical problems downloading, installing and getting started with the app [8], perceived user-concerns about data privacy (what kind of data are being collected and at what frequency) [37, 53], whether users fully understand and feel informed on what passive data are being collected from their phones [32, 37], ambiguity of study tasks, and overall likeability and usability of the app.

In adapting interventions to a digital setting, researchers and developers often attempt to create tools that look and function exactly like original pen and paper tools, without making use of the features and ease of use that technology platforms can offer [31]. By using models and theories designed around infrequent points of contact, researchers might also miss opportunities for better interventions [41].

Human Centered Design (HCD) can uncover opportunities, challenges, and concerns associated with long-term participation in a remote digital health research, through inclusion of participants in the design or co-design of systems. Prior long-term remote digital health research shows how effective this approach can be [4, 9, 10, 43].

Using qualitative HCD methods, we build upon prior work by including a diversity of perspectives from participants with depression and through our contribution of participant perspectives on issues raised in prior work: (1) Participant needs related to burden [9, 43, 47] and engagement [2, 29, 32]; and (2) Privacy concerns and how to preserve privacy in systems that rely on collecting personal data [29, 43].

Our contributions also include suggestions for researchers developing a remote digital health study:

- Design the remote study and digital health app to balance participant burdens with research goals
- Support engagement by providing feedback on study data
- Develop an educational consent process for passive data collection

2 RESEARCH QUESTIONS

The aim of this study was to understand participant burden, engagement, and reactions to passive sensing related to the delivery of a remote digital health study. Through this understanding we sought to improve the design of a digital health research app and to improve the design of the study that

the app supports before conducting a large-scale long-term remote digital health study.

Therefore, we sought to address the following research questions, where *participants* refers to potential participants for a remote digital health study:

- RQ1: *Design*: How can design improve upon the delivery of typical remote study tasks to keep participant burden low?
- RQ2: *Engagement*: What barriers do participants perceive to engagement in long-term remote digital health research studies?
- RQ3: *Passive Data*: Do participants understand what passive data are? What are participant perspectives on sharing their passive data with researchers?

3 METHODS

To understand participant burden, engagement, and reactions to passive sensing, we conducted a qualitative interview study that drew from protocol analysis [14, 18] and a phenomenological approach [33]. Thus, the interviews involved questions that elicited participant thoughts while performing tasks on mobile app prototypes and questions that elicited participant experiences and expectations for using a research study app. For this study, we interviewed 20 participants to gain perspectives from a diverse population representing people who would participate in a study that aims to access depression remotely. This study was approved by the University of Washington Human Subjects Division.

Recruitment

Participants were recruited from the Seattle area through flyers on the University of Washington campus and medical center as well as nearby Seattle neighborhoods. To be eligible to participate in the study, participants had to be older than 18 years, use a smartphone (Android or iPhone), have clinically significant symptoms of depression determined by Patient Health Questionnaire-8 (PHQ-8) score ≥ 10 [28], and be able to visit an off-campus site leased by the research team. Additionally, as the Seattle area is not particularly racially diverse [3], we sought to recruit participants in a stratified manner to increase participation by Asian, Hispanic, and African American minorities as well as keeping a gender balance. Participants who completed the interview were paid \$60 for up to 90 minutes of their time.

Study Procedure

At the start of the interview session, participants were told that the purpose of the study was to inform the design of digital health application to support mental health and facilitate a 12-week-long research study (see Appendix 1). Participants were provided with a smartphone, on the platform they used

regularly (either Android or iOS), for the duration of the interview session; the phone had the designs loaded so they could interact with the prototypes. All sessions were audio recorded, with participant consent.

We designed the semi-structured interview protocol [42] to address each of our research questions. RQ1 focuses on app and study design. We asked participants to think aloud while performing tasks using interactive mobile design prototypes so we could elicit participant experiences and understand attitudes towards app and study design [51]. Tasks were followed by questions to assess participant-perceived burdens of the study design (e.g., how many notifications they felt were reasonable to receive in a day) and to get feedback about engagement with the app to address RQ2 (e.g., how they felt about the app overall and feedback they have for improving it). Lastly, we used designs of potential mobile app screens as probes to generate discussion around passive data to address RQ3.

These methods — think aloud, design probes, and semi-structured sen to elicit participant thoughts, reactions, emotions, and feedback. The prototype designs gave participants something specific to talk about and react to; this helped guide discussions with participants and make them feel like they were participating in something that would come to life.

Part 1: Design Tasks, Burden, Engagement. The first half of the interview session focused on assessing participant experiences with tasks that the app would deliver in the hypothetical remote study. This included the burden of completing the tasks, their engagement with the app, and their expected engagement with the study. Following Suh et al.'s dimensions of user burden, we considered burden as the emotional, mental, physical, and time demands that computing systems might place on the user [47].

Participants were asked to think aloud while completing tasks on three mobile app designs of a digital health app aimed at assessing depression remotely (see Appendix A). The designs were presented in a random order, to counter-balance any effect of seeing one before another.

The first two tasks were a short ecological momentary assessment (EMA) asking about the participant's mood. We asked participants to select from a list of options to describe how their mood and then either to record a voice memo or write a memo elaborating on how they are feeling. With two tasks, participants experienced both the voice recording and the written memo. The third task was a survey prompting the participant to answer the PHQ-8 depression assessment [28].

After participants completed the tasks completed, we asked participants about points of confusion while capturing emo-

tion, voice memos, notifications, and then solicited suggestions for the app and study design.

Part 2: Passive Data. The second half of the interview session focused on collection of passive data. We used three screen mockups to initiate and guide discussion about participant understanding and attitudes towards passive data collection. Participants were told that these were screens that the research team wanted to add to the research study app design they had previously reviewed.

The first screen (Figure 1a) provided information about passive data collection to participants: what passive data are, and why the research team wanted to collect passive data. After participants read this screen, we asked them to describe passive data in their own words. The purpose of this exercise was to assess participant understanding of passive data based on a short description.

The other two screens (Figures 1b and 1c) were variations of a settings page that listed a few types of passive data being collected with switches to allow participants to opt-in or opt-out of them. The first of these screens listed screen time, application usage, and GPS. The second screen allowed further adjustment of GPS subcategories (steps walked, weather in your location, unique locations visited, and distance travelled from home). After seeing each screen, participants were asked about their willingness to share data with the research team and if they had any privacy concerns. The purpose of these screens was to use a familiar construct to elicit participants perspectives on which types of data they would or would not share.

Analysis

All interviews were transcribed using the transcription service Temi [48]. The first and second authors began with a thematic analysis of the transcripts [5], developing a high-level coding scheme based on our research questions. A random selection of 12 interviews were coded using this scheme while also allowing for the development of inductive codes as sub-themes to further develop the coding scheme to give depth to the original codes. After these 12 transcripts were coded, the research team felt that the codes were sufficient to characterize themes related to our research goals. The coding scheme was set. The first author revisited these 12 interviews to consistently apply the coding scheme and also used it to code the remaining 8 interview transcripts. While coding these last 8 interview transcripts, the first author remained open to any new or surprising data. No changes to the coding scheme were made, but these interviews did add perspective on existing codes and confirm existing perspectives in the data.

4 RESULTS

The first author conducted all 20 interviews in August 2018. The interviews took place at an off-campus site leased by the research team in a small and quiet room set up for the purpose of conducting one-on-one interviews. The interviews were designed to take 90 minutes but ranged from 66 to 94 minutes (average: 74 minutes). Given the length, participants were encouraged to take breaks as needed.

Participant Characteristics

A total of 20 participants who matched the inclusion criteria were recruited for the study. Participants ranged in age from 19 to 76 years old (mean: 40), 11 were female (55%) and 9 were male (45%). They identified as Caucasian, Asian, Hispanic/Latino, Black/African-American, and more than one race. Participant PHQ-8 scores ranged from 10-21 (mean: 14.5). Table 1 presents participant demographics.

Number of Participants	20	
Gender (Female)	11	55%
Age (mean)	40	sd=17
Marital status		
Married or Partnered	8	40%
Never Married	6	30%
Divorced	3	15%
Separated	1	5%
Widowed	2	10%
Race/Ethnicity		
Caucasian	11	55%
Asian	4	20%
Black/African-American	1	5%
Hispanic/Latino	3	15%
Choose Not To Answer	1	5%
Years of Education (mean)	16	sd=3
Employed	11	55%
Medicaid	3	16%
Income		
<\$20,000	7	37%
20,000–40,000	6	32%
40,000–60,000	2	11%
60,000–80,000	1	5%
>\$80,000	3	16%
PHQ-8 Score (mean)	15	sd=4

Table 1: Participant Characteristics. In this table, Medicaid refers to the United States government program providing health coverage to eligible low-income adults, children, pregnant women, elderly adults and people with disabilities [19]. And, income refers to the annual household income for the participants.

RQ1: Design: How can design improve upon the delivery of these typical remote study tasks to keep participant burden low?

In this section, we will describe potential opportunities for researchers to increase participation in their studies by considering the needs and values of the participants while designing their studies.

Daily acceptable burden for mood assessments. Participants anticipated that three study tasks a day would be too burdensome (P2, P5, P6, P13, P14), even where all tasks were EMAs that took participants an average of 1 minute and 27 seconds. Participants instead suggested that one or two tasks per day would be more reasonable (P2, P5, P6, P14). In addition to feeling that the time spent on three tasks per day would be burdensome, participants felt that three task would be especially burdensome if the tasks were the same, for instance if they were asked to rate their mood three times per day. Participants did not feel their mood changed enough during the day to warrant more than one or two reports (P5, P13). The type of task required was also key to keeping the perceived burden low. For example, participants felt that having to write or record a memo as part of an EMA was very burdensome (P1, P2, P11) because of the time and effort required and the burden of voice recording (see the next section). However, this burden may be different for participants who find a task meaningful to them or have more flexible work hours (P14), or if participants felt the financial incentive was great enough (P5, P17).

As for notifications from the study app, several participants indicated that they would be willing to receive 1-2 notifications per day, a few were willing to receive 3-4 per day, and one participant said they wanted to receive only 1 notification a week. Additionally, participants acknowledged that too many notifications would be annoying which would cause them to delete the app (P1, P4, P5, P16).

In addition to preferences about the number of notifications, participants described preferences for notification schedules and notification customization. For example, only sending notifications in the morning and/or at night when participants had downtime to do study tasks (P1, P2), sending notifications at random intervals within a pre-specified time frame (P15), or allowing notifications to be snoozed for an hour or longer (P16). Mainly participants wanted any notifications to fit into their daily routine (P8, P12, P15).

These participant preferences suggest a remote study app should limit the number of notifications sent and perhaps support further notification customization.

Burden of voice-based assessments. 17 participants said they would not use the voice recording feature in the research study app. Their immediate reasoning was that they do not

like to hear the sound of their own voice, but participants also described practical reasons for not using this feature, such as effort and cognitive burden. Specifically, P7 and P12 said that they would need to sit and think for a few minutes before recording to compose their thoughts and avoid mistakes in voice memos:

I like getting the right words and so I don't know, sometimes it takes me a little while to think of how I properly feel rather than just some people, 'I feel down.' I really like words that really nail the exact feeling (P7).

I feel just a little bit more effort to say it out loud, even though it sounds like it would be more effort to type it. You have to really know what you're saying, and you have to be concise and say it like all at once. But when you're typing it you can take time and pause and delete things and you don't have to like fully start over if you mess anything up or whatever (P12).

For these participants, voice recording seemed like an unnecessary stress or burden, and potentially less useful for their own reflection. As an alternative, participants emphasized the practicality of text memos (typed or speech to text) compared to speaking. Participants also found voice recording problematic because of the inability to edit as you go, a feature seen as useful in text memos (P6, P7, P12). Similarly, participants thought text memos would be easier and quicker to review than voice recordings:

I could see a little bit more use in typing out something than recording it because you could see it quickly and easily if you typed it out verses recording it would be more effort to get that information again (P11).

Despite the barriers, two participants said that they would do voice recordings if it was required as part of a study, even though they do not like recording and hearing their own voice (P2, P15). These participants preferred speech-to-text, so that they would not have to hear their own voice to review entries. Speech-to-text is a feature already available on smartphones, and so standard input controls already support it.

Participants further noted that they would be unlikely to use speech input in public, while they can use text input anywhere. Researchers and designers should take into account the need for privacy by making voice recording optional. They should also consider limiting the number of voice-based tasks or be clear in the study consent process about the voice-based tasks that are required and why voice samples are needed.

RQ2: Engagement: What barriers do participants perceive to engagement in long-term remote digital health research studies?

Participants expected both return of information (that researchers would share back the data they collected about them) and for the app or researchers to provide them feedback to make sense of and act on the data. Participants also considered return of information and feedback as more than incentives for participation; they saw these features as necessary for sustaining their engagement.

I constantly need something to remind me why I'm doing [the study]... I started this, why am I going to continue doing this?...it's like, I tried having emotions toward [the study], but I don't have any emotions. I don't care if I keep doing it. I don't care if I quit I'll feel the same. So finding ways to make it convenient and wanting to be able to do it. That's the biggest hurdle. Why should I keep doing this when I don't feel anything? (P16)

P16's quote summarizes the challenge of designing a digital health app that keeps participants engaged in a remote study assessing depression. Depression, the very symptoms that the study hopes to collect, are barriers to engagement. To overcome this, participants said that they need something that makes them care about the app, so they will want to return to it and use it.

Participants expect return of information. Participants made it clear that they expected to have access to the passive data that researchers collect about them. Thus, access to collected data should be seen as a standard component of studies involving passive data collection.

Participants introduced their expectation of return of information before we introduced it in interview. Questions about return of information were discussed after all three screens in Figure 1 were reviewed, but many participants became excited about seeing their data while reviewing screen 2 (Figure 1b). For example, when P5 saw screen 2, she started discussing how interesting it would be to see her data and how it could help her change her behavior or reinforce behavior that helps her:

Having something there that shows you, well, you normally do these things on a good day, that could be super helpful. [The app I use] also helped to show me that exercise was improving my mood because I thought it wasn't. I hate exercising, so it was like, this is doing nothing. It's just a waste of time. But then I saw, wait a sec, my mood actually is improving. So I guess I got to keep exercising. (P5)

In this example, P5 was using an her own app (a commercial app with no association to our study) that convinced her that higher levels of physical activity lead to her being in a better mood. Seeing her data helped P5 see how exercising was helping her, and so, despite “hating” exercising, she decided to exercise more because she came to see the benefits.

Other participants also had goals for how they would use their data if they were able to see it. P1 wanted to use feedback on her total screen time as an indication of her depression and when she needs to change her behavior. P14 wanted to use distance travelled from home to affirm her successes in getting out of the house more and motivate her to continue this behavior:

People spend a lot of time when they're depressed looking through Facebook or disappearing into their phones. So seeing that I used my phone for 18 out of 24 hours, would you know, tell me that I needed to stop (P1).

My weekly activity, if you guys track that, I would tell myself, 'oh good job,' I went from staying home all day to doing some other stuff. So it's like motivating me to do more. So tracking this is for my benefit for sure. So I am agreeing collecting this information is for my best interests and making it visible for me to go back to it and see it (P14).

These three previous examples illustrate situations where return of information is beneficial. However, participants also identified potential dangers of being able to review their passive data, noting that overwhelming negative data, even if it revealed something important, may worsen their depression:

I could imagine that if you are in a downward spiral and you're struggling to get out, knowing all the negative information, like, 'oh, you didn't go out today and the weather is bad and you're staying near home too much and you haven't gone anywhere new lately'. If I had a bunch of spiraling downward information, I'm not sure if I would want to feedback or if it'd be more benign to kind of leave me in ignorance or what. (P2)

In addition to clearly negative data, there is also a risk of participants assuming that correlation means causation. For example, a visualization that shows participants their mood in relation to the weather may appear to show that their mood worsens with bad weather. This may or may not be true, but a participant might assume that their mood is tied to the weather and develop a negative attitude towards bad weather that worsens their depression when bad weather strikes.

This caused some participants to say that they would not want to see negative data, which aligns with previous work by Meng et al. [31]. However, for some participants negative data could motivate a beneficial change in behavior:

I personally would because sometimes you need that kind of like concrete evidence to really motivate you to make a change. Um, and that I think could be motivating for me even though it might be hard to look at. (P5)

Feedback about depression symptoms was important to participants. Our designs did not present participants with feedback about their depression symptoms. To our knowledge, this is common in study designs: there is a separation between research apps designed to assess depression and therapeutic apps designed to help manage depression. Despite this gap and lack of feedback in our design, participants made it clear that feedback would be important for engagement with the app in a future study.

Though participants were excited about the idea of contributing to science and technology that could benefit them and others in the future, they felt that feedback about the information they share with researchers would help them feel that they are getting something out of the effort they put in:

I mean because I mean if all it is going to do is collect data, it has to do more than that and it has me give you back the data in a way that's meaningful to you. Something that, that makes it exciting (P8).

I assumed that they were going to take the data, analyze, and then kind of give me feedback on, 'oh, we noticed that when you do this, your mood drops' or 'here's some suggestions based on your activity.' Like I assumed that was gonna be part of this. That's my expectations if I was going to do this (P2).

Participants had many suggestions about what feedback would be meaningful to them. Ideas included feedback given after a study task has been completed, a visualization of the data being collected where they could see trends in their symptoms, support for identifying triggers for their depression symptoms, and contextually appropriate cues to action.

Participants were more excited about visualizations and trigger identification than simple feedback. P5 suggested that a graph or other visualization that would help her see what activities she does on a good day might help her focus on those things on a bad day:

Since [the app I use] has all the different activities I can include in there, it at least helps to kinda correlate: what do I do on a good day? What kinds

of things are making me happy? With depression, it's really easy to forget that (P5).

P14 took this idea further, looking for the app to suggest activities based on her reported mood, the weather, and possibly activities she enjoys:

Based on your entry, the app could sense if it's a beautiful day or not. I don't know if it's going to be interactive with the weather and stuff, but how about suggesting going for a walk. On a rainy day, ...everybody has a neighborhood coffee shop that can sit there and block your electronics and sit there for hours with the book next to a fireplace and you might see a neighbor, they can tell you a story that puts a smile on your face, but sitting in a home when nobody knocks on your door because you're not actually a social person since you're dealing with depression, you're giving yourself an opportunity to speak to other people and that's was the goal. Get out of the house (P14).

Providing feedback and supporting actions to take in light of that feedback may give participants a reason to keep completing study tasks. Without that feedback and support, participants may be left wondering, “So what?!” (P8).

RQ3: Passive Data: Do participants in remote studies understand what passive data are? What are participant perspectives on sharing their passive data with researchers?

Interviews revealed that a short description of passive data is insufficient for helping participants understand passive data, that participants were willing to share most data with researchers without a monetary incentive, and that participants valued transparency and choice in which data they wanted to share.

Participants relied on the app's description and examples to understand passive data. After reading a description of passive data (Figure 1a), most participants exhibited an understanding of passive data that was grounded in the one example we provided in the description: “what apps you use each day and for how long.” For example:

It seems like it's just your phone's log of your activities (P12).

It's going to monitor in the background what apps you're using, for how long throughout the day. So you can kind of track if you're playing games or what you're doing (P17).

Although participants understood that the study app would track their phone usage (e.g., Facebook was used for 30 minutes), these participants did not expect that we would be interested in GPS or other data until the next screen (Figure

1b). Screen time was also not universally understood. For example, P14 understood it as an desire for study participants who could spend a lot of time using the study app:

Screen time. A person who had a lot of time on their hands would be the ideal person to use your app a lot because he already has time devoted in the day for that.

After reading the description we provided (Figure 1a), three participants understood passive data and even researchers' intended purpose for it:

It's not asking for any particular user input in this case, rather it is tracking what you do passively, what you do on your phone in a passive sort of behind the scenes sort of way in order to maybe glean insights about particular phone habits that could be indicative of things like depression (P15).

From this participant's explanation, it is clear they understand the difference between active and passive data and why depression researchers would be interested in passive data collection.

The majority of participants relied on the examples to ground their understanding of passive data collection, this suggests that examples are important to helping participants understand and make decisions about whether they are willing to participate. However, repeating back one or two examples, does not mean participants fully understand the concept of passive data collection. To ensure participants are giving informed consent, we suggest discussing with participants all the types of data that will be collected, why, and how they will be used during the recruitment and onboarding processes. We expand on this further in the discussion.

Participants trust researchers with their data. Almost all participants were open to sharing their passive data with researchers. Only one participant (P19) had a strong reaction and was unwilling to share her data. P19 said, “I don't see how it's necessary, I don't think this would tell you anything about my depression,” and it made her “feel uncomfortable.”

On the other hand, P1 and P17 explained why they would opt-in to sharing all the information represented on Screen 3 (Figure 1c):

I'd opt into all of them. I don't see a problem with it. As long as all that information isn't like being sold straight to someone else (P1).

Yeah, I mean these are all things I would probably opt in. Almost every app I have has GPS enabled, screen time, none of that really bothers me as far as what you're getting information from me (P17).

In these examples, P1 and P17 indicated that they are not concerned with sharing their passive data. In this context, it

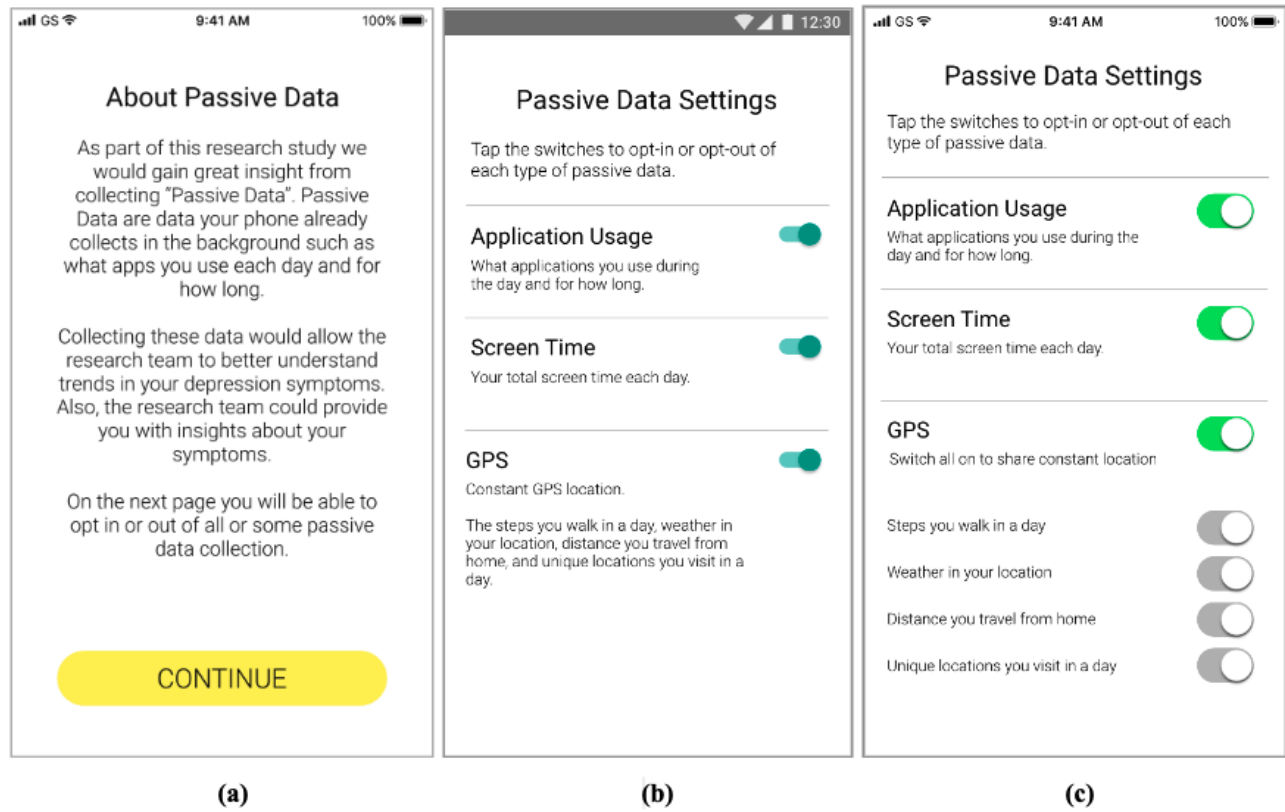


Figure 1: Mock ups used to generate discussions with participants about passive data.

is possible that participants are not concerned because, as other participants describe, they trust researchers with their data:

I trust the researcher teams to like keep it confidential and stuff like that...the privacy would be an issue in another context, but because this is research, I'm fine with it (P2).

It's just you want to trust that whoever is collecting all this stuff and why they're collecting it is for benevolent things, Not malevolent things (P4).

So participants were comfortable sharing their data with researchers, but would be less trusting of non-researchers. For example, further elaboration by P2 and other conversations (P6, P13, P16), uncovered concern about how a corporate institution might use their data. Participants specifically cited Facebook data sharing breaches [22] and Google continuing to track people's location even after they turn off constant location tracking [11] as examples of privacy violation and data misuse. These issues were current events at the time the interviews were conducted. Though this may have made participants in our study more skeptical of companies tracking data with apps, they still seemed to have trust in

researchers. Recent research has also shown that participant willingness to join remote research studies and share data is tied to their trust in the scientific team conducting the study including the institutional affiliations of researchers [38].

However, not all participants were willing to share all types of data with researchers, even if they had trust in researchers. Of the main types of data we discussed with participants (application usage, screen time, and constant GPS), participants were least willing to share constant GPS because they felt it would be “creepy” (P6), “too intrusive” (P8), or like someone would be watching where they are going (P1, P4).

In the discussion, we further broke down GPS to four categories: steps walked in a day, the weather in your area, unique locations you visit in a day¹, and distance traveled from home in a day. Of these four categories, participants were more comfortable sharing steps and weather than sharing unique locations and distance traveled from home (P4, P5, P6 P11). Participants were skeptical about sharing these

¹We intended this to mean the number of locations a person visited in a day, but participants interpreted this as the specific locations they visit (e.g., “work” and “the mall” rather than “2”). Participants felt more comfortable sharing this data point when this meaning was clarified.

types of data because these data types made them feel that someone would know where they are or where they go (like constant GPS).

Transparency and choice are important to reducing privacy concerns in passive data sharing. Though participants trusted researchers with their data, transparency and choice on what data could be shared were found to be important in further building trust and reducing privacy concerns overall.

All participants indicated that they want to know why researchers want to collect each type of data. For example, explaining that GPS data as a measure how often you leave your home could help researchers better understand depression:

The GPS makes sense. Like if I'm staying in my home all day or I'm, I'm walking in a park and unique location, like some people think it's stalking, like in the app will know if you went to the museum or you went to the mall today. Some people just don't want to be tracked... I have no problem. (P14)

Participants also indicated that they would be more likely to share data if they had a choice in what was shared. For example, participants preferred Figure 1c compared to Figure 1b as it allows more control over the types of GPS-derived data (“features”) they could choose to share.

5 DISCUSSION

This study provides insights on potential ways to enhance participant engagement in fully remote research studies, particularly studies that aim to enroll a depressed cohort and gather frequent self-report data and continuous passive data. Our findings can be grouped into three potential recommendations: (1) Ensure that study app tasks are designed with minimal burden to participants, but can still support researcher goals; (2) Support engagement by providing participants with feedback, visualizations, and action items based on active and passive data; and (3) Be clear about what data is being collected and why is it being collected.

Balancing participant burdens with research needs

In traditional ecological momentary assessment (EMA) studies, frequent assessment and reminders to complete surveys are common, and participants may be asked to complete as many as seven assessments a day [26]. Although these methods have been fruitful in smaller scale survey studies, in the larger, fully remote studies, these methods have yielded inconsistent completion rates [46]. We suggest that one way of facilitating completion of study tasks is for researchers to balance the burden of study tasks on participants with study's scientific needs.

According to our findings, the reasons for inconsistent adherence may be due to the mental, physical, or time burdens [47] of completing these surveys. Participants in our study indicated that frequency, timing of EMA, and meaningfulness would affect their willingness to complete a survey. Therefore, we recommend that researchers consider the number of assessments their studies require, as our participants anticipated that more than three per day would be too burdensome. Researchers should also consider the repetitiveness of daily assessments.

Researchers should also consider the mental and physical burden of voice-based assessments. Participants saw voice-based assessments as particularly burdensome due to privacy concerns and discomfort with recording themselves. If collecting vocal data is important for a research project, the researchers should consider: (1) Giving participants flexibility to complete voice-based tasks at times when participants have more privacy; and (2) Informing participants about the research need for these data and why these data should be shared (see similar suggestions for passive data in the subsection below, “Informed Consent in Passive Data Collection”).

However, future work should further examine participant attitudes towards voice-based assessments and access participant behavior when completing voice-based assessments in the wild. We collected participant reactions to recording a voice memo after responding to a EMA designed to collect their current mood. Participants may respond differently to another voice-based assessment or if they are presented with more information about how their voice data could be used. For example, research shows the promise of voice-based features as potential digital biomarkers for a variety of neurological conditions such as depression [34], bipolar disorder [15], cognitive impairment [27], and Parkinson's [23]. If presented with this information, participants may feel more positively or be willing to complete voice-based assessments.

Supporting engagement through feedback

Our findings suggest that the key to keeping participants engaged is giving them a reason to keep doing the study tasks. Particularly for people who are struggling with depression, there needs to be something to remind them why they are participating. Though this could take the form of a reward through the study app, our findings suggest including a feedback mechanism will help participants stay motivated and engaged with the study. Participants in our study expressed expecting feedback. They felt that given the data they were providing, they should get feedback about their condition or at least be able to review their own data. Additionally, participants in our study were looking for feedback that would help them understand the correlation between their behavior and their depression symptoms. This is the same need expressed

by student participants in the Meng et al. study for whom the authors stated that “integrated behavioral targets” would help them understand the same correlations [31]. Providing the feedback participants are looking for could increase self-awareness and motivate behavior change [6, 20].

Researchers should also consider unintended consequences of providing such feedback. It is difficult to provide human-in-the-loop personalized feedback based on data to hundreds or thousands of participants in a remote study. Further, many of these remote studies are designed to inform the creation of algorithmic screening or feedback, and preliminary use of such algorithms may lead to the wrong conclusions. Use of algorithmic feedback could also cause frustration if not designed for contestability [24]. We note that the community has much to learn about how to design such feedback. For example, negative feedback could be beneficial if it increases self-awareness and produces a behavior change, but it also comes with the alternative risk of being frustrating or even dangerous for depressed participants.

Future work could address how to design and incorporate feedback into remote digital health apps, including visualizations of participant data and suggestions for behavior changes. Though designing and incorporating feedback is challenging even beyond the potential unintended consequences, prior work has found that visualizations across more than one data type are valuable to users and can be used to identify opportunities for behavior change [5, 13]. Also, designing to support personalized feedback [40] and goal-directed self-tracking [45] can help avoid common self-tracking pitfalls (e.g., low motivation or self-efficacy) so that the desired outcome is achieved.

In future work, researchers should also consider that providing any kind of feedback will be an intervention which could conflict with their research goals. Additionally, research goals and individual participant goals may differ, and any study that provides feedback must take into consideration that it is also providing a sort of intervention. However, as our findings suggest that participants expect feedback and that it may help keep them engaged, deciding not to provide feedback may also be detrimental to study goals. Therefore, future work should also consider if there is a way to provide feedback without compromising research goals or how research goals could be modified so that feedback can be provided.

Informed Consent in Passive Data Collection

Our findings suggest that our app designs were not sufficient to inform participants about passive data collection. To feel confident that participants are able to give their informed consent, our design would need significant iteration and testing to be sure that participant understand what data are collected and how they are collected (from phone-based

sensors “continuously”). Furthermore, participants indicated needing further transparency about why researchers were collecting these data and having the choice to share select passive data that they feel most comfortable in sharing with researchers.

If for no other reason than ensuring a study is ethical, participants need to be informed about what they are consenting to share [1]. However, in the case of passive data collection, our results suggest that participants may need more than a short description to truly understand the risks to their personal privacy [25]. Repeating back the provided example of what type of data will be collected is not an indication that participants understand why it is being collected or how.

Therefore, in addition to reminding participants about passive data collection when they first download a research study app, we recommend including education about passive data collection in the recruitment process, the reason each type of data are being collected, and what sort of incentive (monetary or informational) they will receive. To do this effectively, researchers should design and test a script to deliver to potential participants over the phone. Nebecker et al. provide one example of working to ensure new research practices are consistent with ethical guidelines and to inform the informed consent process [36].

Based on our findings, we further recommend researchers allow participants to choose, to the extent it is feasible, which data types they are willing to share during the study onboarding. If a participant is unwilling to share data needed for the study, then researchers should revisit the consent process as it is likely better for that person to not consent to participate. Additionally, these data sharing choices should be built into the research app so that participants remain aware of the data they are sharing with the research team. We make these suggestions from the participant’s perspective, but flexibility and control is also supported by the designers perspective. Based on their work with mobile app developers and interaction designers, Vilaza and Bardram suggest giving users control over the level of detail reported in addition to which data will be shared [49]. They also suggest allowing users the option to delete data. However, such control could complicate data collection and so researchers should consider how to balance their needs for consistent data collection with participant needs for transparency, privacy, and control. Additionally, users often say they want more controls but, in practice, they do not use them [16, 30]. This is a paradox that should be considered here with privacy controls but also applies to controls such as customization of notifications.

Limitations

Limitations of methods. Our findings are limited by our method of a one-time in-person interview with each participant, based on prospective use of an application rather than actual enrollment in a long-term study. This method supports understanding attitudes of potential participants, predicting enrollment decisions, and understanding barriers to participation and participant knowledge of passive data. However, it does not describe behavior once enrolled in a remote study.

Limitations of probing about passive data. We had expected strong negative reactions or even fear about passive data collection because we thought participants would feel these data we proposed to collect are private. With only a small sample in one geographic location, it is possible this concern may be stronger in a different population or location. Also, participant attitudes about passive data collection may have been biased by the following factors: 1) The discussion about passive data was at the end of the interview, meaning that rapport was well established between participant and interviewer; and, 2) Three participants were researchers or scientists themselves. Two of these participants directly acknowledged that this made them more likely to want to help other researchers and scientists.

6 CONCLUSION

In our work, we have discussed considerations for designing studies, and related study software, across three previously identified challenges for digital health studies: (1) Participant needs and burdens involved in mood assessments; (2) Participant needs for continued engagement; and (3) Participant understanding of and attitude towards passive data collection.

Additionally, we suggest that remote studies using a smartphone app to assess mental health may be more successful in getting participants engaged if researchers consider participants needs and values in the design of the study and mobile application. Specifically, participant needs should have greater emphasis in the design of all remote data collection constructs, participant values of transparency and agency should be exhibited in passive data collection, and participants expect feedback from researchers based on these data shared by them and their progress to keep them engaged in long term remote studies.

Lastly, future work could seek to: (1) Collect and understand participants behavior (adherence) while enrolled in a remote digital health study; (2) Design and test a consent process for passive data collection to be sure it is effective in educating participants [35, 36] and know if it affects participants attitudes about sharing passive data; and (3) Design and test feedback, visualizations, and suggestions for a remote digital health study.

ACKNOWLEDGMENTS

We thank participants for their time and feedback. Research reported in this publication was supported by the National Institutes of Health (NIH) under grant P50MH115837 and the National Science Foundation under grant IIS-1553167. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NSF.

REFERENCES

- [1] Paul S Appelbaum, Charles W Lidz, and Alan Meisel. 1987. *Informed consent: Legal theory and clinical practice*.
- [2] Patricia A Arean, Kevin A Hallgren, Joshua T Jordan, Adam Gazzaley, David C Atkins, Patrick J Heagerty, and Joaquin A Anguera. 2016. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. *J. Med. Internet Res.* 18, 12 (Dec. 2016), e330.
- [3] Samuel Assefa. 2020. Population & Demographics: Census 2020. <http://www.seattle.gov/opcd/population-and-demographics>.
- [4] Jakob E Bardram, Mads Frost, Károly Szántó, Maria Faurholt-Jepsen, Maj Vinberg, and Lars Vedel Kessing. 2013. Designing mobile health technology for bipolar disorder: a field trial of the monarca system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2627–2636.
- [5] Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. 2013. Health Mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 5 (2013), 1–27.
- [6] Arpita Bhattacharya, Samantha Kolovson, Yi-Chen Sung, Mike Eacker, Michael Chen, Sean A Munson, and Julie A Kientz. 2018. Understanding pivotal experiences in behavior change for the design of technologies for personal wellbeing. *J. Biomed. Inform.* 79 (March 2018), 129–142.
- [7] Bruno Biagiante, Diego Hidalgo-Mazzei, and Nicholas Meyer. 2017. Developing digital interventions for people living with serious mental illness: perspectives from three mHealth studies. *Evid. Based. Ment. Health* 20, 4 (Nov. 2017), 98–101.
- [8] Ken Cheung, Wodan Ling, Chris J Karr, Kenneth Weingardt, Stephen M Schueller, and David C Mohr. 2018. Evaluation of a recommender app for apps for the treatment of depression and anxiety: an analysis of longitudinal user engagement. *J. Am. Med. Inform. Assoc.* 0, June (2018), 1–8.
- [9] Kevin Doherty, Marguerite Barry, José Marcano-Belisario, Bérenger Arnaud, Cecily Morrison, Josip Car, and Gavin Doherty. 2018. A Mobile App for the Self-Report of Psychological Well-Being During Pregnancy (BrightSelf): Qualitative Design Study. *JMIR Ment Health* 5, 4 (Nov. 2018), e10007.
- [10] Kevin Doherty, José Marcano-Belisario, Martin Cohn, Nikolaos Mastellos, Cecily Morrison, Josip Car, and Gavin Doherty. 2019. Engagement with Mental Health Screening on Mobile Devices: Results from an Antenatal Feasibility Study. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15.
- [11] Emily Dreyfuss. 2018. Google Tracks You Even If Location History's Off. Here's How to Stop It. *Wired* (Aug. 2018).
- [12] Katie L Druce, William G Dixon, and John McBeth. 2019. Maximizing Engagement in Mobile Health Studies: Lessons Learned and Future

- Directions. *Rheum. Dis. Clin. North Am.* 45, 2 (May 2019), 159–172.
- [13] Daniel A Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean A Munson. 2014. Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In *Proceedings of the 2014 conference on Designing interactive systems - DIS '14*. ACM Press, New York, New York, USA, 667–676.
- [14] Karl Anders Ericsson and Herbert Alexander Simon. 1993. *Protocol Analysis: Verbal Reports as Data*. Bradford Books.
- [15] Maria Faurholt-Jepsen, Jonas Busk, Mads Frost, Maj Vinberg, Ellen M Christensen, Ole Winther, Jakob Eyvind Bardram, and Lars V Kessing. 2016. Voice analysis as an objective state marker in bipolar disorder. *Translational psychiatry* 6, 7 (2016), e856–e856.
- [16] Casey Fiesler and Blake Hallinan. 2018. We Are the Product: Public Reactions to Online Data Sharing and Privacy Controversies in the Media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. dl.acm.org, 53.
- [17] Joseph Firth and John Torous. 2015. Smartphone Apps for Schizophrenia: A Systematic Review. *JMIR mHealth and uHealth* 3, 4 (Nov. 2015), e102.
- [18] Marsha E Fonteyn, Benjamin Kuipers, and Susan J Grobe. 1993. A Description of Think Aloud Method and Protocol Analysis. *Qual. Health Res.* 3, 4 (Nov. 1993), 430–441.
- [19] Centers for Medicare & Medicaid Services in the United States Government. 2020. Medicaid. <https://www.medicare.gov/medicaid/index.html>. Accessed: 2020-3-30.
- [20] Elizabeth Pegg Frates, Margaret A Moore, Celeste Nicole Lopez, and Graham T McMahon. 2011. Coaching for behavior change in psychiatry. *Am. J. Phys. Med. Rehabil.* 90, 12 (Dec. 2011), 1074–1082.
- [21] Colleen Galambos, Marjorie Skubic, Shaung Wang, and Marilyn Rantz. 2013. Management of Dementia and Depression Utilizing In- Home Passive Sensor Data. *Gerontechnology* 11, 3 (2013), 457–468.
- [22] Kevin Granville. 2018. Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens. *The New York Times* (March 2018).
- [23] Brian Harel, Michael Cannizzaro, and Peter J Snyder. 2004. Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study. *Brain and cognition* 56, 1 (2004), 24–29.
- [24] Tad Hirsch, Kritzia Merced, Shrikanth Narayanan, Zac E Imel, and David C Atkins. 2017. Designing Contestability. In *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS '17*. 95–99.
- [25] Franz J Ingelfinger. 1979. Informed (But Uneducated) Consent. In *Biomedical Ethics and the Law*, James M Humber and Robert F Almeder (Eds.). Springer US, Boston, MA, 265–267.
- [26] Andrew Jones, Danielle Remmerswaal, Ilse Verveer, Eric Robinson, Ingmar H A Franken, Cheng K Fred Wen, and Matt Field. 2019. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. *Addiction* 114, 4 (April 2019), 609–619.
- [27] Lampros C Kourtis, Oliver B Regele, Justin M Wright, and Graham B Jones. 2019. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *NPJ digital medicine* 2, 1 (2019), 1–9.
- [28] Kurt Kroenke, Tara W Strine, Robert L Spitzer, Janet B W Williams, Joyce T Berry, and Ali H Mokdad. 2009. The PHQ-8 as a measure of current depression in the general population. *J. Affect. Disord.* 114, 1-3 (2009), 163–173.
- [29] Neal Lathia, Veljko Pejovic, Kiran K Rachuri, Cecilia Mascolo, Mirco Musolesi, and Peter J Rentfrow. 2013. Smartphones for Large-scale Behaviour Change Interventions. *IEEE Pervasive Comput.* 12, 3 (July 2013), 66–73.
- [30] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proc. ACM Hum. -Comput. Interact.* 2, CSCW (Nov. 2018), 102:1–102:31.
- [31] Jingbo Meng, Syed Ali Hussain, David C Mohr, Mary Czerwinski, and Mi Zhang. 2018. Exploring User Needs for a Mobile Behavioral-Sensing Technology for Depression Management: Qualitative Study. *J. Med. Internet Res.* 20, 7 (July 2018).
- [32] David C Mohr, Michelle Nicole Burns, Stephen M Schueller, Gregory Clarke, and Michael Klinkman. 2013. Behavioral Intervention Technologies: Evidence review and recommendations for future research in mental health. *Gen. Hosp. Psychiatry* 35, 4 (2013), 332–338.
- [33] Dermot Moran. 2002. *Introduction to phenomenology*. Routledge.
- [34] James C Mundt, Peter J Snyder, Michael S Cannizzaro, Kara Chappie, and Dayna S Geralts. 2007. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of neurolinguistics* 20, 1 (2007), 50–64.
- [35] Camille Nebeker, Tiffany Lagare, Michelle Takemoto, Brittany Lewars, Katie Crist, Cinnamon S Bloss, and Jacqueline Kerr. 2016. Engaging research participants to inform the ethical conduct of mobile imaging, pervasive sensing, and location tracking research. *Transl. Behav. Med.* 6, 4 (Dec. 2016), 577–586.
- [36] Camille Nebeker, John Torous, and Rebecca J Bartlett Ellis. 2019. Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC Med.* 17, 1 (July 2019), 137.
- [37] Jennifer Nicholas, Katie Shilton, Stephen M Schueller, Elizabeth L Gray, Mary J Kwasny, and David C Mohr. 2019. The Role of Data Type and Recipient in Individuals' Perspectives on Sharing Passively Collected Smartphone Data for Mental Health: Cross-Sectional Questionnaire Study. *JMIR mHealth and uHealth* 7, 4 (2019), 1–10.
- [38] Abhishek Pratap, Ryan Allred, Jaden Duffy, Donovan Rivera, Heather Sophia Lee, Brenna N Renn, and Patricia A Areán. 2019. Contemporary Views of Research Participant Willingness to Participate and Share Digital Data in Biomedical Research. *JAMA network open* 2, 11 (2019).
- [39] Abhishek Pratap, Elias Chaibub Neto, Phil Snyder, Carl Stepnowsky, Noémie Elhadad, Daniel Grant, Matthew H Mohebbi, Sean Mooney, Christine Suver, John Wilbanks, et al. 2020. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. *npj Digital Medicine* 3, 1 (2020), 1–10.
- [40] Mashfiqui Rabbi, Min Hane Aung, Mi Zhang, and Tanzeem Choudhury. 2015. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 707–718.
- [41] William T Riley, Daniel E Rivera, Audie A Atienza, Wendy Nilsen, Susannah M Allison, and Robin Mermelstein. 2011. Health behavior models in the age of mobile interventions: are our theories up to the task? *Translational Behavioral Medicine* 1, 1 (2011), 53–71.
- [42] Jane Ritchie, Jane Lewis, of Social Policy Jane Lewis, Carol McNaughton Nicholls, and Rachel Ormston. 2013. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. SAGE.
- [43] John Rooksby, Alistair Morrison, and Dave Murray-Rust. 2019. Student Perspectives on Digital Phenotyping: The Acceptability of Using Smartphone Data to Assess Mental Health. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14.
- [44] Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, and David C Mohr. 2015. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *J. Med. Internet Res.* 17, 7 (July 2015), e175.
- [45] Jessica Schroeder, Ravi Karkar, Natalia Murinova, James Fogarty, and Sean A. Munson. 2019. Examining Opportunities for Goal-Directed Self-Tracking to Support Chronic Condition Management. *Proc. ACM*

- Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 151 (Dec. 2019), 26 pages. <https://doi.org/10.1145/3369809>
- [46] Saul Shiffman. 2009. Ecological momentary assessment (EMA) in studies of substance use. *Psychol. Assess.* 21, 4 (Dec. 2009), 486–497.
- [47] Hyewon Suh, Nina Shahriaree, Eric B Hekler, and Julie A Kientz. 2016. Developing and Validating the User Burden Scale: A Tool for Assessing User Burden in Computing Systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, New York, New York, USA, 3988–3999.
- [48] Temi. 2020. Audio to Text Automatic Transcription Service & App. <https://www.temi.com/>. Accessed: 2020-3-30.
- [49] Giovanna N Vilaza and Jakob E Bardram. 2019. Sharing Access to Behavioural and Personal Health Data. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM Press.
- [50] Rui Wang, Min S H Aung, Saeed Abdullah, Rachel Brian, Andrew T Campbell, Tanzeem Choudhury, Marta Hauser, John Kane, Michael Merrill, Emily A Scherer, Vincent W S Tseng, and Dror Ben-Zeev. 2016. CrossCheck: Toward passive sensing and detection of mental health changes in people with schizophrenia. In *UbiComp 2016 - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 886–897.
- [51] Gordon B Willis. 2005. Cognitive interviewing in practice: think-aloud, verbal probing and other techniques. *Cognitive interviewing: a tool for improving questionnaire design*. London: Sage Publications (2005), 42–63.
- [52] World Health Organization. 2011. *mHealth: New horizons for health through mobile technologies*. Global Observatory for eHealth series, Vol. 3. WHO Press, 20 Avenue Appia, 1211 Geneva 27, Switzerland.
- [53] Dong Whi Yoo and De Choudhury. 2019. Designing Dashboard for Campus Stakeholders to Support College Student Mental Health. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM Press.

A DIGITAL HEALTH APP DESIGNS

Overview

As described in the Study Procedure, participants were told the purpose of the study was to inform the design of Brighten, a mobile application to support mental health and facilitate a 12-week-long research study, and that we were looking for their open and honest feedback. To elicit better reactions and feedback from participants, we wanted them to work with real or interactive designs. Design 1 is the baseline, while Designs 2 and 3 were re-designs of Design 1. In the study protocol, Design 1 was always shown to participants first, then Design 2 and 3 were shown. However, Design 2 and 3 were not always shown in order. Design 2 was shown first to half of the participants and Design 3 was shown first to the other half. This was to eliminate any bias of assuming the last design shown was the best. Additionally, Designs 2 and 3 were assigned names so that participants did not think that “Design 2” was meant to be better than “Design 3” or vice versa. Within this appendix, we provide screenshots of three different designs that participants used so that readers may have better clarity about the study procedures. Please

note that while all screenshots are taken on an iPhone, all three designs were also made for Android.

Design 1: “Brighten”

The first design participants saw was a fully functional mobile application used in a prior remote study for mental health [2]. The research team aimed to understand opportunities to improve this app for future studies.

The screenshots below (Figure 2) are representative of this design.

Design 2: “Brighten Messenger”

The second design we called Brighten Messenger because it was designed as if Brighten was a digital agent that messages you like a concerned friend and responds to the users input. The user choices were also designed like text message responses to maintain the messaging feel and show prior interactions like a typical message history. The other notable part of this design is the logo which was intended to look like a rising sun, but participants thought it was a taco or a beehive.

Note: This design is an interactive prototype created in AdobeXD. This means that the user can interact with the prototype to complete the given tasks, but outside of those tasks, the prototype may not function as the user would expect a fully functional app to do. Additionally, the keyboard is not functional in this prototype and so participants were asked to speak what they would type when a keyboard appeared.

The screenshots below (Figures 3 and 4) are representative of the design. To interact with the prototype used in the study, visit this link: <https://tinyurl.com/brighten-messenger>.

Design 3: “Re-Brighten”

The third design we called Re-Brighten as it was intended to be a simplified re-design of the original Brighten. The key ideas of this design were 1) To give users better context for what tasks they needed to do as part of the study through a dashboard, and 2) To improve the design of EMAs like the mood question and longer surveys like PHQ-8.

Note: This design is an interactive prototype created in AdobeXD. This means that the user can interact with the prototype to complete the given tasks, but outside of those tasks, the prototype may not function as the user would expect a fully functional app to do. Additionally, the keyboard is not functional in this prototype and so participants were asked to speak what they would type when a keyboard appeared.

The screenshots below (Figures 5 and 6) are representative of the design. To interact with the prototype used in the study, visit this link: <https://tinyurl.com/rebrighten>.

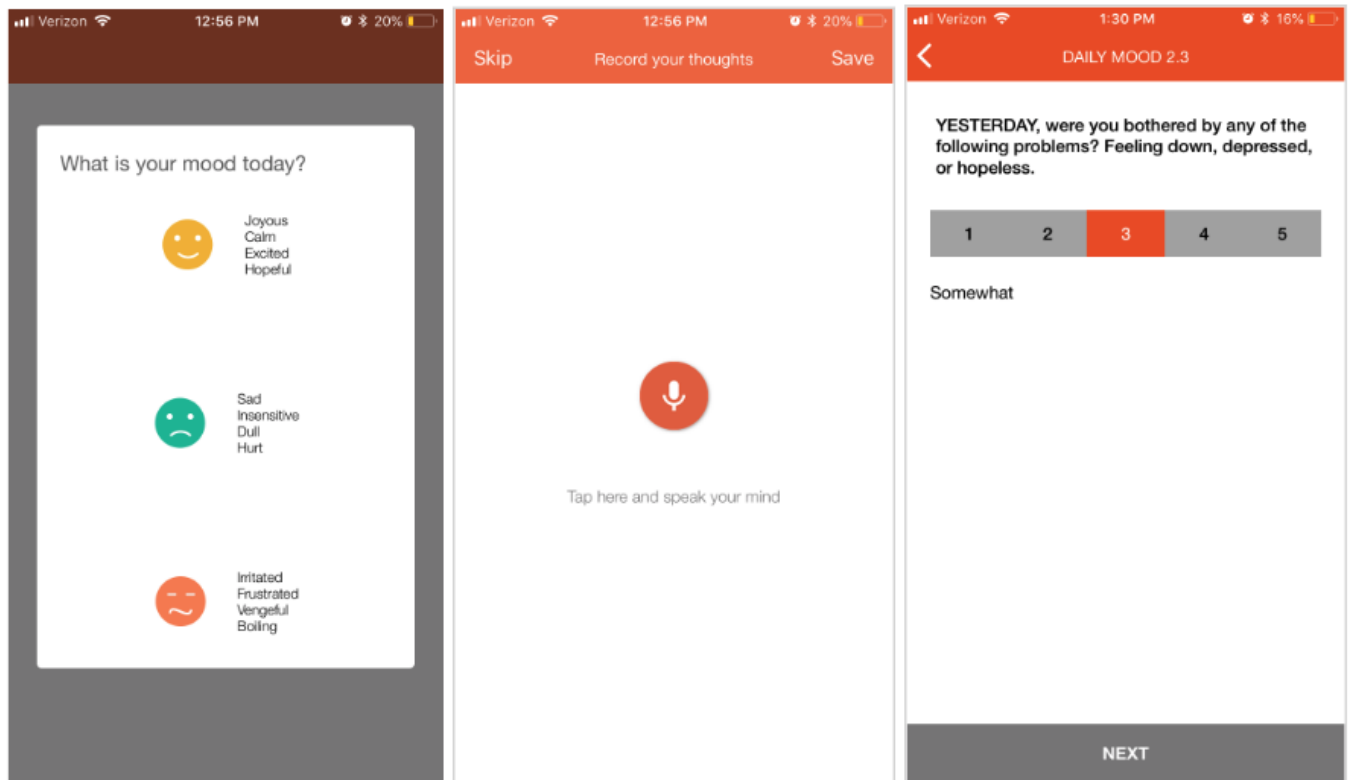


Figure 2: Three sample screens from Brighten Messenger (Design 1: “Brighten”).

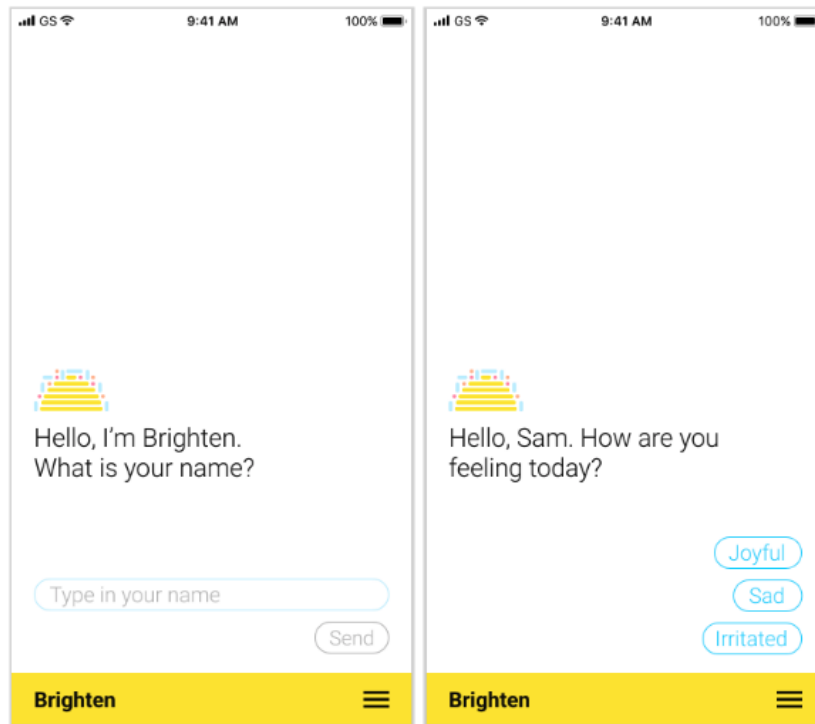


Figure 3: Two sample screens from Brighten Messenger (Design 2: “Brighten Messenger”).

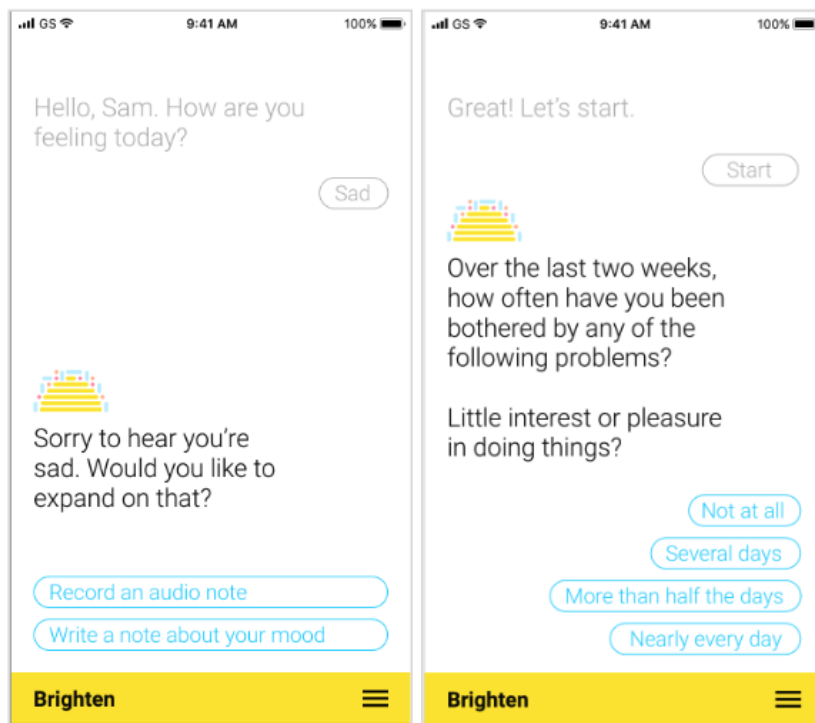


Figure 4: Two more sample screens from Brighten Messenger (Design 2: “Brighten Messenger”).

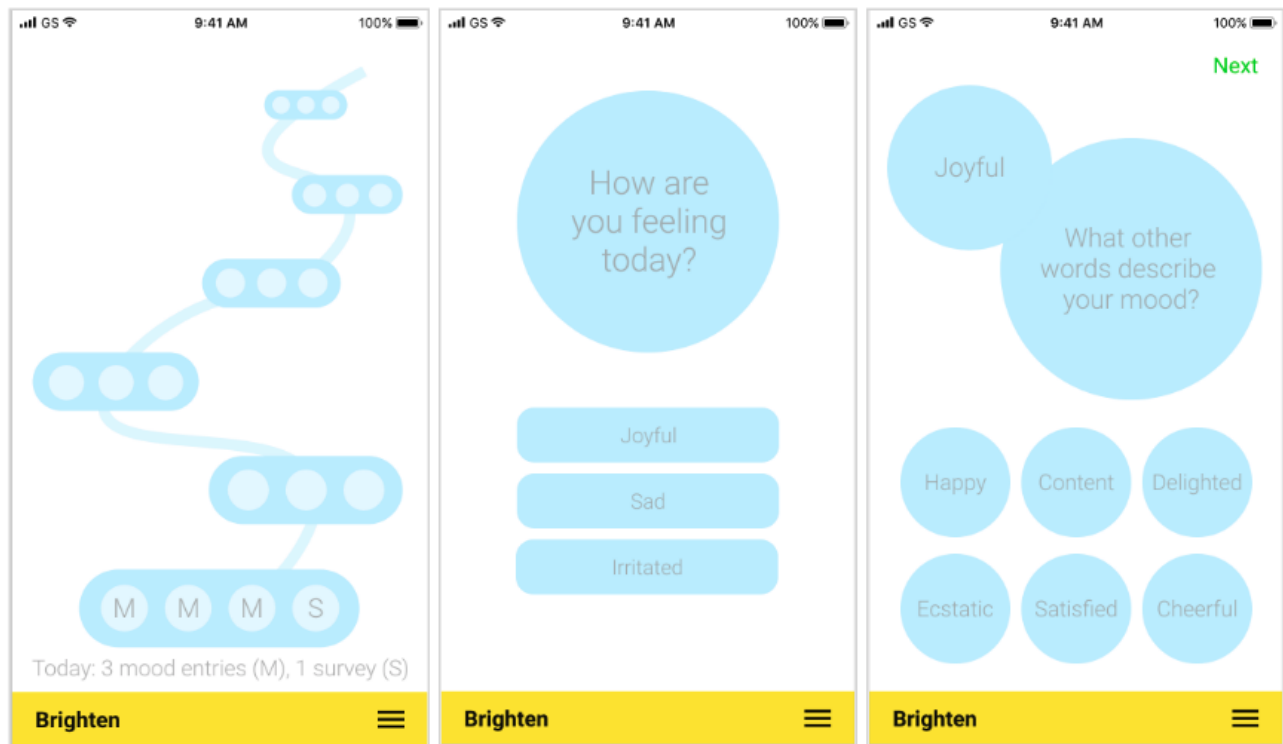


Figure 5: Three screens from Re-Brighten (Design 3: “Re-Brighten”).

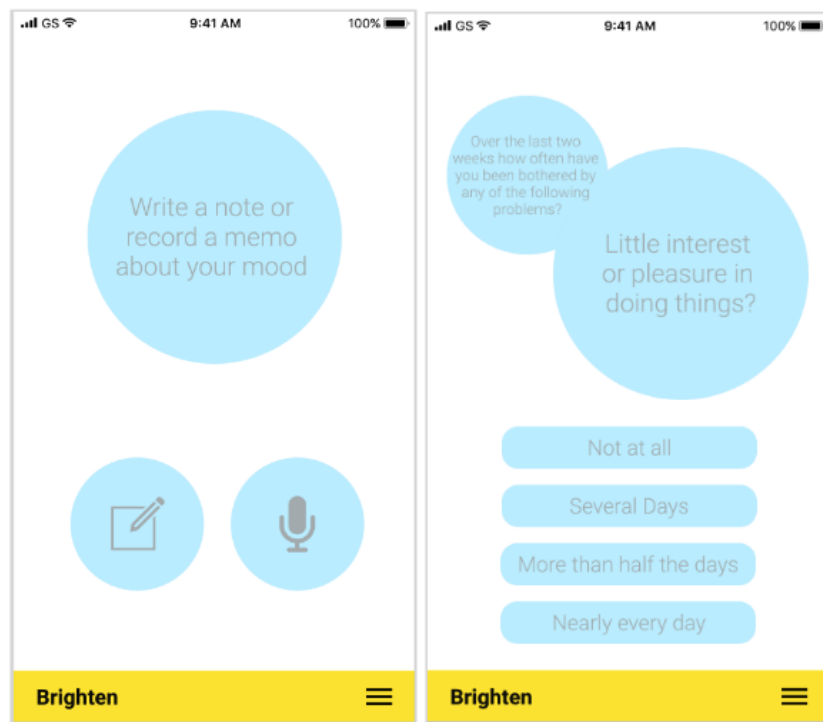


Figure 6: Two more sample screens from Brighten Messenger (Design 3: “Re-Brighten”).