Acceptability of Personal Sensing Among People with Alcohol Use Disorder: Observational
Study

Kendra Wyant[1], Hannah Moshontz[1], Stephanie B. Ward[1], Gaylen E. Fronk[1], & John J.
Curtin[1]

[1] Department of Psychology, University of Wisconsin - Madison

Author Note

Correspondence concerning this article should be addressed to John J. Curtin, 1202
West Johnson St, Madison, WI 53706. E-mail: jjcurtin@wisc.edu

## Abstract

**Background:** Personal sensing may improve digital therapeutics for mental health care by facilitating early screening, symptom monitoring, risk prediction, and personalized/adaptive interventions. However, further development and use of personal sensing first requires better understanding of its acceptability to people targeted for these mental health applications. **Objective:** We assessed the acceptability of both active and passive personal sensing methods in a sample of people with moderate to severe alcohol use disorder using both behavioral and self-report measures. **Method:** Participants ($N = 154$; 50% female; mean age = 41; 87% White, 97% Non-Hispanic) in early recovery (1 – 8 weeks of abstinence) from alcohol use disorder were recruited from the Madison, WI area to participate in a 3-month longitudinal study. Participants engaged with active (ecological momentary assessment; EMA, audio check-in, and sleep quality) and passive (geolocation, cellular communication logs, and text message content) personal sensing methods. We assessed 3 behavioral indicators of acceptability: participants' choices about their participation in the study at various stages in the study procedure, their choice to opt-in to provide data for each personal sensing method, and their adherence to a subset of the active methods (EMA, audio check-in). We also assessed 3 self-report measures of acceptability (interference, dislike, and willingness to use for 1 year) for each method. **Results:** All but 1 of the individuals who were eligible to participate consented to the personal sensing procedures. Most of these individuals (88%) also returned 1 week later to formally enroll in the study and begin to provide these data. All participants (100%) opted-in to provide data for EMA, sleep quality, and all passive methods (geolocation, cellular communication logs, text message content). Three participants (2%) did not provide any audio check-ins while on study. The average completion rate for all requested EMAs (4X daily) was 80% for 4x daily and 94% for 1x daily. The completion rate for the daily audio check-in was 54%. Aggregate participant ratings indicated all personal sensing methods to be significantly more acceptable (all $P$'s < .05) compared to neutral across subjective measures of interference, dislike, and willingness

to use for 1 year. Participants did not significantly differ in their dislike of active compared to passive methods ($P = .23$). However, participants reported a higher willingness to use passive methods for 1 year compared to active methods ($P = .04$). **Conclusions:** The results of our study suggest that both active and passive personal sensing methods are generally acceptable to people with alcohol use disorder over a longer period than has previously been assessed. This was true even for data streams that contained potentially more sensitive information (e.g., geolocation, cellular communications). Important individual differences were observed both across people and methods which indicate opportunities for future improvements.

*Keywords:* personal sensing, digital therapeutics, mobile health, smartphone, alcohol use disorder

Acceptability of Personal Sensing Among People with Alcohol Use Disorder: Observational

Study

## Introduction

### Personal Sensing

The World Health Organization's Global Observatory for eHealth has concluded that "the use of mobile and wireless technologies to support the achievement of health objectives has the potential to transform the face of health service delivery across the globe" [1]. This conclusion applies to research and care for mental health as well as other traditional health services. These opportunities are now possible in part because of rapid advances in smartphone and related mobile technologies [2] and high levels of smartphone access across race, socioeconomic status, geographic region, and other demographic characteristics [3].

Personal sensing may become an important component of these digital health advances [4]. Personal sensing is a method for longitudinal measurement in situ; i.e., real-world measurement that is embedded in individuals' day to day lives [5–7]. Raw data streams are collected by smartphones, wearable sensors, or other smart devices. These raw data streams can consist of self-reports or more novel data streams such as geolocation, cellular communications, social media activity, or physiology. Subsequent processing can extract psychiatric or health relevant measures of thoughts, feelings, behavior, and even interpersonal interactions.

Ecological momentary assessment (EMA), a personal sensing method that collects brief self-reports about momentary states multiple times per day, has been used for many years in short-term longitudinal studies of psychiatric disorders. For example, EMA research on substance use disorders has identified proximal causes and risk factors for drug craving and relapse [8–10]. It has also characterized the time course and nature of drug withdrawal [11,12]. EMA research on major depressive and bipolar disorders has identified predictors of

daily mood fluctuations, documented treatment efficacy, and been paired with other human neuroscience methods to explore biological mechanisms [13]. Much of this research could not have been accomplished with other measurement methods.

More recently, mental health relevant research using personal sensing of raw data streams other than self-report is emerging. This includes methods to sense geolocation [14–16], cellular communications [16–20], sleep [20], and physiology [21,22], as examples. These alternative personal sensing methods provide benefits and opportunities not possible with EMA. For example, many of these data streams can be sensed passively such that they have very low assessment burden. This may allow their use for long-term longitudinal monitoring of participants that would not be feasible with EMA, which requires more active effort for data collection.

Personal sensing is a powerful tool for mental health research [23]. These data are inherently longitudinal, which allows observation of the temporal ordering for putative etiologic mechanisms and their effects. Longitudinal measurement is also critical for many mental health constructs that display meaningful, and often frequent, temporal variation within person (e.g., psychiatric symptoms). Measures based on personal sensing data generally have high ecological validity because they are collected in situ. Personal sensing measures also have low retrospective bias because they are often collected in real-time. Furthermore, personal sensing can derive measures from raw data streams (e.g., in situ behavior, physiology, interpersonal interactions) that are difficult or even impossible to obtain through other traditional research measurement methods.

Personal sensing may have even higher value in the future for mental health applications that target patient mental health care than it does for research [7,24,25]. Data collected by personal sensing methods may be used for preliminary screening for psychiatric disorders [19,26]. These methods can also be used to monitor psychiatric symptoms or even predict future risk for symptom recurrence, relapse, or other harmful behaviors (e.g., suicide

attempts) [14,27–30]. Personal sensing measures or risk indicators may be shared, with patient consent, to health care providers to allow for cost-effective, targeted allocation of limited mental health resources to patients with the greatest or most urgent need [31]. Personal sensing has the potential to support precision mental health care by adapting and timing interventions based on characteristics of the patient and the moment in time [32–34]. To be clear, these applications of personal sensing are currently aspirational rather than available for clinical implementation today. However, clinical research is advancing us rapidly toward these goals [16,20,28,35].

Mental health research and applications with emerging, often more passively sensed, novel data streams like geolocation and cellular communications is still nascent. This research has predominately involved "proof-of-concept" studies that typically include only healthy controls or other convenience samples rather than people with psychiatric disorders [17–19]. It has also often used very small sample sizes and/or short monitoring periods [15,17,21,22]. Recent reviews of this emerging literature have highlighted gaps in reporting on participant exclusions, attrition, and adherence that are necessary to assess selection biases and feasibility of these more novel personal sensing methods [36–38].

**Acceptability of Personal Sensing**

Further development and use of personal sensing necessitates better understanding of its acceptability to research participants and patients targeted for mental health applications. Will individuals consent to the use of personal sensing methods? Will they opt-in to allow for passive measurement methods? Can they sustain the behaviors necessary for active measurement methods for longer periods of time? Do they perceive specific personal sensing methods as burdensome or dislike them? Answers to these questions about the acceptability of personal sensing methods are central to its feasibility for both mental health research and applications.

The acceptability of a personal sensing method may be influenced by the degree of active effort required from the participant or patient to collect the raw data (i.e., the method's assessment burden) and other factors (e.g., the sensitivity of the data collected). As such, acceptability may vary across different personal sensing methods and comparisons across methods within the same individuals are thus warranted. Furthermore, comprehensive assessment of both behavioral measures (e.g., adherence) and subjective perceptions of acceptability may better anticipate potential issues for recruitment, consent, adherence, and attrition when they are used for either research or clinical applications.

Much of what is known about the acceptability of personal sensing is limited to EMA. Studies that have accessed participants' perceptions of EMA methods have generally concluded that it is acceptable to participants from both non-clinical and clinical samples [39–43]. Similarly, participants display moderate or better adherence with respect to response rates even with relatively high sampling density (e.g., 6 – 9 daily assessments) [39,44,45]. However, these studies generally assessed participants' perceptions and adherence over short monitoring periods (i.e., 2 – 6 weeks). Less is known about the use of EMA over longer duration monitoring periods (e.g., months) as would be necessary for clinical applications.

Assessment burden may impact the relative acceptability of EMA. EMA assessment burden can increase as more active effort is required from participants when EMA survey density (i.e., frequency of surveys) or depth (number of items/length of surveys) increases. However, the impact of active effort due to density and depth may not be comparable. Increased density does not appear to robustly impact perceptions or adherence [45]. However, greater survey depth may decrease both perceptions of acceptability and adherence [44].

The duration of the EMA monitoring period may also increase assessment burden and affect acceptability but findings have been mixed. Some studies suggest that as duration increases beyond only a few weeks, adherence decreases [42,47,48]. Yet other studies find no changes in adherence over time [40,45]. Importantly, very few studies have looked at

adherence over longer duration (e.g., > 6 weeks). One recent study asked people about their perceptions of EMA studies (without deploying the EMA protocol). They found shorter study duration to be associated with greater willingness to participate and a perception that the study would be more enjoyable. However, even in this hypothetical scenario the long duration condition was only 3 weeks [49].

The context of an EMA protocol likely moderates acceptability; existing research raises some concern about perceptions and adherence to EMA protocols in patients with substance use disorders relative to other groups. Specifically, a recent meta-analysis confirmed decreased adherence to EMA protocols in patients with substance use disorder diagnoses vs. recreational substance users [46]. However, a separate meta-analysis by [47] showed that adherence rates did not differ between healthy, physically ill, mentally ill, and mixed samples, which suggests that adherence concerns may be limited to applications with patients with substance use disorders rather than all psychiatric disorders more generally.

Far less is known about participants' perceptions and adherence to more passive personal sensing methods. Some research has presented hypothetical scenarios to participants to assess their perceptions about personal sensing methods [50–52]. Participants' willingness to share sensed data appears to vary by the data type (e.g., sleep, geolocation, social media activity). However, it is difficult to determine how well participants' perceptions in these hypothetical scenarios would generalize to real world collection of these data. And, of course, it is impossible to measure attrition and adherence outside of explicit implementation of these sensing methods.

Preliminary research has begun to examine perceptions and adherence during real world use of passive personal sensing methods. However, this research has generally been limited by small sample sizes [53,54], use of convenience samples (e.g., students) [40,53,55], short monitoring duration [21,53,54,56], and coarse, incomplete, or aggregate reporting of perceptions, adherence, and related participant behaviors [40,53,54]. These are important

first efforts but more research into the feasibility of personal sensing methods is clearly warranted.

## Study Goals

This study reports on the acceptability of both active and passive personal sensing methods in a sample of participants with moderate to severe alcohol use disorder. These participants were enrolled early in their recovery (i.e., $1 - 8$ weeks after becoming abstinent) and followed for 3 months. We used active personal sensing methods to collect EMA, daily audio check-ins, sleep quality, and selected physiology. We used primarily passive methods to collect geolocation, cellular communications logs, and text message content. We assessed participants' choices about their participation in the study at various stages in the study procedure (e.g., consent, enrollment, data collection), their choice to opt-in to provide data associated with each personal sensing method, and their reasons for discontinuation when available. For active measures, we also assessed their adherence for providing those raw data streams for up to 3 months of their study participation. Finally, we assessed participants' subjective perceptions of the acceptability of each of these personal sensing methods, separately, by self-report. We believe these data provide an unparalleled and transparent window into the feasibility of using numerous personal sensing methods with individuals with alcohol use disorder, a highly stigmatized psychiatric disorder.

## Methods

### Research Transparency

We value the principles of research transparency that are essential to the robustness and reproducibility of science [57]. Consequently, we maximized transparency through several complementary methods. First, we report how we determined our sample size, all data exclusions, all manipulations, and all available measures in the study [58]. Second, we completed a transparency checklist, which can be found in the supplement of this paper

(Multimedia Appendix 1) [59]. Third, we made the data, analysis scripts and annotated results, self-report surveys, and other study materials associated with this report publicly available [60].

**Participants**

We collected the study data between 2017 – 2019 as part of a larger grant-funded parent project (R01 AA024391). This study represents secondary data analyses of those data. The sample size was determined based on power analyses for the aims of that project. We used all available participants for this study. All procedures were approved by the University of Wisconsin-Madison Institutional Review Board (Study #2015-0780).

We recruited participants in early recovery (1 – 8 weeks of abstinence) from alcohol use disorder in Madison, Wisconsin, USA, to participate in a 3-month longitudinal study. Participants were recruited through print and targeted digital advertisements and partnerships with treatment centers. We required that participants:

1. were 18 years of age or older,

2. were able to write and read in English,

3. had at least moderate alcohol use disorder (>= 4 DSM-5 symptoms[1]),

4. were abstinent from alcohol for at least 1 week but no longer than 2 months,

5. were willing to use a single smartphone (their personal phone or one provided by us) while enrolled in the study.

We also excluded participants if they exhibited severe symptoms of psychosis or paranoia[2].

We assessed inclusion and exclusion criteria using a brief phone screen followed by a more detailed in person screening visit. One hundred ninety-two participants were eligible for enrollment. Of these participants, 191 consented to participate in the study at the screening

session and 169 subsequently enrolled in the study at the enrollment visit which occurred approximately 1 week later. Fifteen participants discontinued prior to the first monthly follow-up visit. The remaining 154 participants provided study measures for 1 (N = 14), 2 (N = 7) or 3 (N = 133) months. We provide a study participation flow chart in Figure 1.

## Procedure

Participants completed 5 study visits over the course of approximately 3 months. Participants first attended a screening visit where we determined eligibility, obtained informed consent, and collected self-report measures of individual differences (e.g., demographics, alcohol use history). We scheduled eligible and consented participants to enroll in the study approximately 1 week later. During this enrollment visit, we collected additional self-report and interview measures. Participants completed 3 additional follow-up visits that occurred about every 30 days. We collected self-report and interview measures and downloaded cellular communications logs (text message and phone call) at these visits. Finally, we collected various raw data streams (e.g., geolocation, cellular communication logs, EMA) using personal sensing to monitor participants throughout the 3-month study period. A full description of the procedure and data collected at each visit can be found at the study's OSF page [60].

## Personal Sensing

Personal sensing methods can be coarsely classified as active or passive. Active personal sensing requires active effort from the participant to provide the raw data streams whereas passive personal sensing data are collected automatically (either asynchronously or continuously) with little to no effort required by the participant. Our study obtained several active signals that varied somewhat in the amount of effort required by the participant. Specifically, we used active methods to collect EMA, daily audio check-ins, sleep quality, and selected physiology. We used primarily passive methods to collect geolocation, cellular

communications logs, and text message content. More detail about each raw data stream collected by personal sensing is provided below.

**Audio Check-in.** Participants recorded a diary-style audio response on their smartphone to an open-ended prompt each day following a reminder from us that was sent via text message. They responded to the prompt ("How are you feeling about your recovery today?"), which stayed the same throughout the entire study. We instructed them that their responses should be approximately 15 – 30 seconds in duration. These recordings were sent to us by text message.

**EMA.** Participants completed a brief (7 – 10 questions) EMA 4 times each day following reminders from us that were sent by text message. These text messages included a link to a Qualtrics survey that was optimized for completion on their smartphone. All 4 EMAs included items that asked about any alcohol use that had not yet been reported, current affective state (pleasantness and arousal), greatest urge to drink alcohol since the last EMA, any pleasant or positive events and any hassles or stressful events that occurred since the last EMA, any exposure to risky situations (i.e., people, places, or things) since the last EMA. The first EMA each day asked an additional 3 questions about how likely participants were to encounter a risky situation, encounter a stressful event, and drink alcohol in the upcoming week. The first and last EMAs of the day were scheduled within 1 hour of participants' typical wake and sleep times. The other 2 EMAs were each scheduled randomly within the first and second halves of the participants' typical day. All EMAs were separated from each other by at least 1 hour.

**Sleep Quality.** We collected information about participants' sleep duration, timing, and overall quality with a Beddit sleep monitor (Beddit Oy Inc., Espoo, Finland) that was placed in their beds and connected to their smartphones. We used an early version of the sleep monitor that required participants to actively start and stop the monitor when they entered and exited their bed each night and morning, respectively. These data are available for only 87 participants because Beddit Oy was acquired by Apple Inc. during data

collection for this study. Apple discontinued cloud support for data collection with the sleep monitor in November 2018, which prevented its further use for our remaining participants.

**Physiology.** We continuously monitored participants' physiology (heart rate, electrodermal activity, skin temperature) using an early version of the Empatica E4 wristband monitor (by Empatica Inc., Boston, MA). However, this early version did not adequately support Bluetooth streaming of data to the cloud. Instead, participants had to manually connect the wristband each night to a tablet we provided to upload their data. This and other software bugs made use of the wristband too complicated for many participants. Therefore, we discontinued use of the wristband after we collected data from 9 participants. Given this small sample size, we did not include the wristband in our primary analyses. We do provide self-reported acceptability ratings for this signal from this small sample in Multimedia Appendix 3 (Figure S1).

**Geolocation.** We continuously collected participants' moment-by-moment geolocation using location services on their smartphones in combination with commercial software that accessed these geolocation data and saved them in the cloud. At the start of the study, we used the Moves app (developed by ProtoGeo Oy, Helsinki, Finland). However, Facebook acquired ProtoGeo Oy and shut down use of the Moves app in July 2018. At this point, we switched to using the FollowMee GPS tracking mobile app (FollowMee LLC, Murphy, TX). Measurement of geolocation required only initial installation of the app by the participants. Subsequent measurement and transfer of the data to the cloud was completed automatically with no input or effort by the participant. Both apps allowed participants to temporarily disable location sharing if they deemed it necessary for short periods of time.

**Cellular Communication Logs.** We collected cellular communication logs that include meta-data about smartphone communications involving both text messages and phone calls. For each communication entry, these logs include the phone number of the other party, the type of call or message (i.e., incoming, outgoing, missed, rejected), the name of the party if listed in the phone contacts, the date and time the message or call occurred, whether

the log entry was read (text messages only), and the duration of the call (voice calls only).
These data are saved passively on the phone with no additional input or effort on the part of
the participant. We downloaded these logs from participants' phones at each monthly
follow-up visit. Participants were informed that they could delete any text message or voice
call log entries prior to the download if they desired.

**Text Message Content.**   We also collected the message content from participants'
text messages on their smartphone. As with the logs, content from individual text messages
is saved passively on the phone with no additional input or effort on the part of the
participant. We downloaded text message content at each monthly follow-up visit and
participants could delete text messages prior to the download. Note that we did not have a
parallel method to gain access to phone call content. Thus, we had meta-data from
communication logs for both text messages and phone calls but had the content of the
communication only for text messages.

## Measures

**Individual Differences.**   We collected demographic information and information
relevant to participants' alcohol use and DSM-5 alcohol use disorder symptoms at the
screening visit[3].

**Behavioral Measures of Acceptability.**   Coarse assessment of the acceptability of
the personal sensing methods can be made based on participants' behaviors. Specifically, we
assessed 3 categories of behavior. First, we assessed participants' choices about their
participation in the study at various stages in the study procedure (e.g., consent, enrollment,
data collection) and their reasons for discontinuation when available. Second, we assessed
their choice to opt-in to provide data associated with each personal sensing method.
Participants were allowed to participate in the study without opting-in to any specific
personal sensing method other than EMA. Instead, they were paid a monthly bonus for each
sensing method they chose to opt-in to that ranged from $10-$25 ($25/month for EMA,

$25/month for audio check-in, $15/month for cellular communication logs with text message content, $15/month for sleep quality data, and $10/month for geolocation). Finally, for a subset of the active measures (EMA, audio check-in), we assessed their behavioral adherence for up to 3 months of study participation.

**Self-reported Measures of Acceptability.**   To assess participants' subjective experience of the acceptability of the personal sensing methods in this study, each month they rated each method on 3 acceptability relevant dimensions (see Multimedia Appendix 2). Specifically, participants were asked to indicate how much they agreed with each of the following 3 statements on a 5-point bipolar scale (strongly disagree, disagree, undecided, agree, strongly agree) for the personal sensing signals[4]:

1. [Personal sensing method name] interfered with my daily activities.
2. I disliked [Personal sensing method name].
3. I would be willing to use [Personal sensing method name] for 1 year to help with my recovery.

The interference item (item 1) was collected only for the active methods because the passive methods require no effort and therefore cannot interfere with daily activities. Dislike and willingness to use for 1 year (items 2 & 3, respectively) were collected for all methods.

**Data Analytic Strategy**

We conducted all analyses in R version 4.1.1 [62] using RStudio [63] and the tidyverse ecosystem of packages [64].

**Behavioral Measures of Acceptability.**   We provide descriptive data on participants' choices about their participation in the study at various stages in the study procedure (e.g., consent, enrollment, data collection). We provide both coarse and more granular tabulation of their reasons for discontinuation when available. We report the percentages of participants who opted-in to provide us with the raw data streams we

collected via personal sensing. We also report adherence measures for 2 of the active personal sensing methods (EMA and audio check-in). Formal measures of adherence could not be calculated for geolocation, cellular communication logs, text message content, and sleep quality because it was not possible to distinguish between low volumes of data due to adherence (e.g., deleting phone calls or messages, turning off location services on the phone, failing to start sleep monitoring at bedtime) and valid reasons (no calls made during the day, no movement, erratic sleep patterns).

**Self-reported Measures of Acceptability.** Participants responded to the 3 self-report items related to acceptability (interference, dislike, and willingness to use for 1 year) on a 5-point bipolar scale (strongly disagree, disagree, undecided, agree, strongly agree). We retained these ordinal labels for visual display of these data in figures but ordered the labels such that higher scores represent greater acceptability (i.e., strongly agree for willingness to use for 1 year and strongly disagree for interference and dislike). For analyses, we re-coded these items to a numeric scale ranging from -2 to 2 with 0 representing the neutral (undecided) midpoint and higher scores representing greater acceptability.

Participants responded to these items at each monthly follow-up visit. Therefore, participants had up to 3 responses for each item depending on when they ended their participation. We analyzed their last available response in our primary analyses to allow us to include all participants and to represent their final perception of each personal sensing signal. However, mean responses across each time point remained relatively constant for all signals (see Figure S2 in Multimedia Appendix 3).

To detect mean perceptions of the personal sensing signals that diverge from neutral (i.e., mean responses to any items that are different from 0/undecided), we conducted one sample t-tests for the 3 self-report items for each personal sensing signal. To examine relative perceptions of the signals, we compared perceptions of the active vs. passive categories of signals using within-sample t-tests for dislike and willingness to use for 1 year[5].

We also report pairwise comparisons among all personal sensing signals using within-sample t-tests for each of the 3 self-report items in Table S1 in Multimedia Appendix 3.

Finally, we conducted 2 analyses to examine the consistency of perceptions across personal sensing signals (e.g., do participants who dislike 1 signal also dislike the other signals?). First, we calculated bivariate correlations among the personal sensing signals for each item. Second, we calculated intraclass correlations (single, case 3 [65]) separately for each item to quantify agreement in participants' perceptions across the signals.

## Results

### Participant Characteristics

A total of 154 participants completed at least 1 monthly follow-up visit and provided self-report acceptability ratings for interference, dislike, and willingness to use for 1 year. These participants serve as our primary sample for our analyses. Participants were mostly White (134/154; 87.0%) and non-Hispanic (150/154; 97.4%). Half of our participants were female (77/154) and the mean age was 41 years. Table 1 presents a full break-down of participants' demographic information. Table 2 characterizes information relevant to alcohol use for these participants. We compared demographics and alcohol use information for participants who were included in the analyses vs. eligible participants who did not provide study measures (i.e., did not enroll or discontinued prior to the first month follow-up; $N = 36^6$) and found no significant differences (see Table S2 in Multimedia Appendix 3 for more detail on these analyses).

Table 1

*Demographics*

|  | n | % | M | SD |
|---|---|---|---|---|
| Age |  |  | 41 | 11.9 |
| Sex |  |  |  |  |
| Female | 77 | 50.0 |  |  |
| Male | 77 | 50.0 |  |  |
| Race |  |  |  |  |
| American Indian/Alaska Native | 3 | 1.9 |  |  |
| Asian | 2 | 1.3 |  |  |
| Black/African American | 8 | 5.2 |  |  |
| White/Caucasian | 134 | 87.0 |  |  |
| Other/Multiracial | 7 | 4.5 |  |  |
| Hispanic, Latino, or Spanish Origin |  |  |  |  |
| Yes | 4 | 2.6 |  |  |
| No | 150 | 97.4 |  |  |
| Education |  |  |  |  |
| Less than high school or GED degree | 1 | 0.6 |  |  |
| High school or GED | 15 | 9.7 |  |  |
| Some college | 43 | 27.9 |  |  |
| 2-Year degree | 14 | 9.1 |  |  |
| College degree | 58 | 37.7 |  |  |
| Advanced degree | 23 | 14.9 |  |  |
| Employment |  |  |  |  |
| Employed full-time | 72 | 46.8 |  |  |
| Employed part-time | 27 | 17.5 |  |  |

| | | | | |
|---|---|---|---|---|
| Full-time student | | 7 | 4.5 | |
| Homemaker | | 1 | 0.6 | |
| Disabled | | 7 | 4.5 | |
| Retired | | 8 | 5.2 | |
| Unemployed | | 19 | 12.3 | |
| Temporarily laid off, sick leave, or maternity leave | | 3 | 1.9 | |
| Other, not otherwise specified | | 10 | 6.5 | |
| Personal Income | | | $34,233 | $31,543 |
| Marital Status | | | | |
| Never married | | 69 | 44.8 | |
| Married | | 33 | 21.4 | |
| Divorced | | 45 | 29.2 | |
| Separated | | 5 | 3.2 | |
| Widowed | | 2 | 1.3 | |

*Note:*      $N = 154$

Table 2

*Alcohol Related Characteristics for the Sample*

|  | n | % | M | SD |
|---|---|---|---|---|
| Alcohol Use Disorder Milestones | | | | |
| Age of first drink | | | 14.6 | 2.9 |
| Age of regular drinking | | | 19.5 | 6.5 |
| Age at which drinking became problematic | | | 27.9 | 9.6 |
| Age of first quit attempt | | | 31.6 | 10.4 |
| Number of Quit Attempts | | | 9.1 | 31.1 |
| Types of Treatment (Can choose more than 1) | | | | |
| Long-term residential (6+ mos.) | 8 | 5.2 | | |
| Short-term residential (< 6 mos.) | 51 | 33.1 | | |
| Outpatient | 77 | 50.0 | | |
| Individual counseling | 100 | 64.9 | | |
| Group counseling | 65 | 42.2 | | |
| Alcoholics Anonymous/Narcotics Anonymous | 96 | 62.3 | | |
| Other | 41 | 26.6 | | |
| Received Medication for Alcohol Use Disorder | | | | |
| Yes | 62 | 40.3 | | |
| No | 92 | 59.7 | | |
| DSM-5 Alcohol Use Disorder Symptom Count | | | 8.9 | 1.9 |
| Lifetime Drug Use | | | | |
| Tobacco products (cigarettes, chewing tobacco, cigars, etc.) | 122 | 79.2 | | |
| Cannabis (marijuana, pot, grass, hash, etc.) | 131 | 85.1 | | |
| Cocaine (coke, crack, etc.) | 86 | 55.8 | | |
| Amphetamine type stimulants (speed, diet pills, ecstasy, etc.) | 81 | 52.6 | | |

| | | |
|---|---|---|
| Inhalants (nitrous, glue, petrol, paint thinner, etc.) | 36 | 23.4 |
| Sedatives or sleeping pills (Valium, Serepax, Rohypnol, etc.) | 72 | 46.8 |
| Hallucinogens (LSD, acid, mushrooms, PCP, Special K, etc.) | 88 | 57.1 |
| Opioids (heroin, morphine, methadone, codeine, etc.) | 65 | 42.2 |
| Current Drug Use[a] | | |
| Tobacco products (cigarettes, chewing tobacco, cigars, etc.) | 84 | 54.5 |
| Cannabis (marijuana, pot, grass, hash, etc.) | 52 | 33.8 |
| Cocaine (coke, crack, etc.) | 4 | 2.6 |
| Amphetamine type stimulants (speed, diet pills, ecstasy, etc.) | 11 | 7.1 |
| Sedatives or sleeping pills (Valium, Serepax, Rohypnol, etc.) | 24 | 15.6 |
| Hallucinogens (LSD, acid, mushrooms, PCP, Special K, etc.) | 9 | 5.8 |
| Opioids (heroin, morphine, methadone, codeine, etc.) | 9 | 5.8 |

*Note:*  $N = 154$

[a] Current refers to past month drug use reported at follow-up visits 1 or 2

**Behavioral Measures of Acceptability**

**Participation.** Figure 1 shows participant attrition/discontinuation at each phase of the study. Of the 192 eligible participants at screening, only 1 did not consent after hearing the details of the study. Enrollment occurred during a second visit 1 week later. A total of 169 participants completed enrollment.

Study phases where we lost eligible participants are highlighted in red in Figure 1. In addition, we coarsely tabulated participants stated reasons for discontinuation as due to acceptability, other reasons, or unknown in this figure. Eleven of the 192 eligible participants (5.7%) were lost due to acceptability-relevant causes (e.g., no longer interested, non-adherence to sensing methods, or citing study demands as too burdensome). Other reasons for discontinuation not related to the acceptability of the signals include circumstances such as moving or no longer wishing to abstain from alcohol. It should be noted that 31 of the 192 participants (16.1%) were lost to follow-up such that we had no information about their reasons for discontinuation. We provide more granular tabulation of these reasons for discontinuation in Table 3.

**Opt-In and Adherence.** All participants who completed through follow-up 1 (154/154; 100%) opted-in to provide data for EMA, sleep quality, and all passive personal sensing methods (geolocation, cellular communication logs, text message content). Three of the 154 participants (1.9%) did not provide any audio check-ins while on study.

Daily adherence rates were relatively high for EMA such that on 94.1% of study days participants completed at least 1 of the 4 EMAs. On average, participants completed 3.2 EMAs every day. The overall adherence rate for all requested EMAs was 79.8%. Participants' completion rate for the audio check-in was 54.3% (Figure S4 in Multimedia Appendix 3 contains more information on this distribution). That is, of their total days on study, participants completed an audio check-in on approximately half of them. Figure 2 shows mean weekly adherence to each of these methods for each week on study. In

Multimedia Appendix 3 we also report adherence for participants who completed the 3-month study compared to those who dropped out prior to completion (Figure S3 in Multimedia Appendix 3).

Table 3

*Characterization of Discontinued Participants*

|  | n | % |
|---|---|---|
| Eligible and consented participants discontinued prior to completing enrollment | | |
| Health concerns | 1 | 4.5 |
| No longer has transportation | 1 | 4.5 |
| **No longer interested** | **3** | **13.6** |
| Rescheduled multiple times before cancelling/no showing | 7 | 31.8 |
| Unknown | 10 | 45.5 |
| Total | 22 | 100.0 |
| Enrolled participants discontinued prior to first month follow-up | | |
| Family crisis | 1 | 6.7 |
| **No longer interested** | **1** | **6.7** |
| No longer sober or no longer wishes to abstain from alcohol | 5 | 33.3 |
| **Noncompliance with providing data** | **1** | **6.7** |
| Rescheduled multiple times before cancelling/no showing | 3 | 20.0 |
| Unknown | 4 | 26.7 |
| Total | 15 | 100.0 |
| Enrolled participants discontinued after the first month follow-up | | |
| Cell service shut off | 3 | 14.3 |
| **Cited study demands as too burdensome** | **1** | **4.8** |
| Moved out of state | 2 | 9.5 |
| **No longer interested** | **3** | **14.3** |
| No longer sober or no longer wishes to abstain from alcohol | 3 | 14.3 |
| **Noncompliance with providing data** | **2** | **9.5** |
| Rescheduled multiple times before cancelling/no showing | 4 | 19.0 |

| | | |
|---|---:|---:|
| Unknown | 3 | 14.3 |
| Total | 21 | 100.0 |

*Note:*     Bolded rows depict explicit acceptability-related discontinuation.

**Self-reported Acceptability**

**Interference.** Figure 3 shows the distribution of participant responses to the self-reported acceptability item about interference. Responses are grouped by personal sensing data stream and the amount of active effort required to collect it. One sample t-tests revealed that each mean interference score (depicted as the solid red line) was significantly more acceptable than 0 (gray dashed line indicating undecided). Table 4 reports the summary statistics for each one sample t-test and pairwise correlations between personal sensing data streams. An ICC (type 3) showed that, on average, interference ratings were moderately consistent across the data streams, ICC = .42, 95% CI = [.31 - .53].

**Dislike.** Figure 4 shows the distribution of participant responses to the self-reported acceptability item about dislike by personal sensing data stream and amount of active effort required to collect it. One sample t-tests revealed that each mean dislike score was significantly more acceptable than 0. Table 5 reports the summary statistics for each one sample t-test and pairwise correlations between personal sensing data streams. An ICC (type 3) showed that, on average, the dislike ratings were moderately consistent across the data streams, ICC = .42, 95% CI = [.35 - .48].

We also assessed the effect of active effort on dislike ratings (Figure 5). We conducted a paired samples t-test to compare the average dislike for active (audio check-in, EMA) vs. passive (geolocation, cellular communication logs, text message content) methods. Participants did not significantly differ in their dislike of active vs. passive methods, $t(153) = 1.21$, $P = 0.23$, $d = 0.10$.

**Willingness to Use for 1 Year.** Figure 6 shows the distribution of participant responses to the self-reported acceptability item about willingness to use for 1 year for each personal sensing data stream (Figure S5 in Multimedia Appendix 3 contains additional information about willingness to use a 1X daily EMA method for 1 year). One sample t-tests revealed that each mean willingness score was significantly more acceptable than 0. Table 6

reports the summary statistics for each one sample t-test and pairwise correlations between personal sensing data streams. An ICC (type 3) showed that, on average, the willingness ratings were moderately consistent across the data streams, ICC = .52, 95% CI = [.46 - .58].

We also assessed the effect of active effort on willingness ratings (Figure 7). We conducted a paired samples t-test of the average willingness to use for 1 year for active (audio check-in, EMA) vs. passive (geolocation, cellular communication logs, text message content) signals. Participants reported higher acceptability with respect to willingness for passive data streams ($M = 0.80$, $SD = 1$) relative to active data streams ($M = 0.70$, $SD = 1.10$), $t(153) = 2.12$, $P = 0.04$, $d = 0.17$.

Table 4

*Bivariate and Univariate Statistics for Interference by Personal Sensing Data Stream*

|  | 1 | 2 | N | M | SD | t | d |
|---|---|---|---|---|---|---|---|
| **Active** | | | | | | | |
| 1. Audio Check-in | – | | 154 | 0.46 | 0.53 | 9.05* | 0.73 |
| 2. EMA | .19 | – | 154 | 0.55 | 0.40 | 11.37* | 0.92 |
| 3. Sleep Quality | -.96 | -.47 | 87 | 0.39 | 0.55 | 14.86* | 1.59 |

*Note:*    Initial columns indicate bivariate correlations among data streams. Final columns represent the number of participants (*N*), mean and standard deviation (*M*, *SD*), t-statistic (*t*) and Cohen's d Effect size (*d*) for the one sample t-tests against 0 (undecided). Higher values represent higher acceptability. Active methods are displayed in red.    *$P < .05$

Table 5

*Bivariate and Univariate Statistics for Dislike by Personal Sensing Data Stream*

| | 1 | 2 | 3 | 4 | 5 | N | M | SD | t | d |
|---|---|---|---|---|---|---|---|---|---|---|
| **Active** | | | | | | | | | | |
| 1. Audio Check-in | – | | | | | 154 | 0.45 | 0.30 | 4.91* | 0.40 |
| 2. EMA | .56 | – | | | | 154 | 0.45 | 0.29 | 12.92* | 1.04 |
| 3. Sleep Quality | -.41 | -.37 | – | | | 87 | 0.41 | 0.30 | 9.45* | 1.01 |
| **Passive** | | | | | | | | | | |
| 4. Geolocation | -.58 | -.49 | -.10 | – | | 154 | 0.56 | 0.26 | 13.51* | 1.09 |
| 5. Cellular Communication Logs | -.67 | -.63 | -.21 | .72 | – | 154 | 0.53 | 0.31 | 11.45* | 0.92 |
| 6. Text Message Content | -.37 | -.48 | -.53 | .60 | .76 | 154 | 0.52 | 0.31 | 6.07* | 0.49 |

*Note:*     Initial columns indicate bivariate correlations among data streams. Final columns represent the number of participants ($N$), mean and standard deviation ($M$, $SD$), t-statistic ($t$) and Cohen's d Effect size ($d$) for the one sample t-tests against 0 (undecided). Higher values represent higher acceptability. Active methods are displayed in red and passive methods are displayed in blue.
* $P < .05$

Table 6

*Bivariate and Univariate Statistics for Willingness to Use for 1 Year by Personal Sensing Data Stream*

| | 1 | 2 | 3 | 4 | 5 | N | M | SD | t | d |
|---|---|---|---|---|---|---|---|---|---|---|
| **Active** | | | | | | | | | | |
| 1. Audio Check-in | – | | | | | 154 | 0.55 | 0.22 | 7.09* | 0.57 |
| 2. EMA | -.24 | – | | | | 154 | 0.53 | 0.23 | 6.47* | 0.52 |
| 3. Sleep Quality | -.25 | -.30 | – | | | 87 | 0.54 | 0.23 | 6.19* | 0.66 |
| **Passive** | | | | | | | | | | |
| 4. Geolocation | -.18 | -.28 | -.06 | – | | 154 | 0.63 | 0.19 | 9.83* | 0.79 |
| 5. Cellular Communication Logs | -.34 | -.36 | -.37 | .30 | – | 154 | 0.65 | 0.23 | 9.76* | 0.79 |
| 6. Text Message Content | -.20 | -.43 | -.50 | .24 | .90* | 154 | 0.61 | 0.26 | 8.21* | 0.66 |

*Note:* Initial columns indicate bivariate correlations among data streams. Final columns represent the number of participants ($N$), mean and standard deviation ($M$, $SD$), t-statistic ($t$) and Cohen's d Effect size ($d$) for the one sample t-tests against 0 (undecided). Higher values represent higher acceptability. Active methods are displayed in red and passive methods are displayed in blue.

* $P < .05$

## Discussion

This study evaluated the acceptability of active and passive personal sensing methods for a variety of raw data streams and associated methods. To this end, we assessed participants' choices/behaviors about both participating in the study and providing raw data streams for each method and their subjective perceptions of each sensing method. We focused on participants with moderate to severe alcohol use disorder because they might have been expected to be less willing to share sensitive, private information due to the stigma associated with their disorder [66]. However, if these sensing methods were acceptable to them, highly promising opportunities are now emerging to address their largely unmet treatment needs [67] with technological solutions that include digital therapeutics combined with personal sensing [68]. We organize our discussion around 7 key conclusions from our analyses.

## 1. Individuals with alcohol use disorder will generally accept the use of personal sensing methods

Based on our sample, it appears individuals with alcohol use disorder are indeed willing to provide these sensitive, personally sensed raw data streams based on their behavioral choices regarding consent, enrollment, and opt-in for data collection in this study. All but 1 of the individuals (191/192; 99.5%) who were eligible to participate consented to the personal sensing procedures. Most of these individuals also returned 1 week later to formally enroll in the study and begin to provide these data (169/191; 88%). Furthermore, all (169/169; 100%) of the participants who enrolled in the study explicitly opted-in to provide the 3 arguably most sensitive passive data streams - geolocation, cellular communication logs, and text message content.

These consent, enrollment, and opt-in numbers could be considered upper- and lower-bound estimates of the percentage of individuals who are willing to provide these raw data streams in a research setting. The very high percentage for consent may overestimate

willingness because some of these individuals may have reconsidered their initial decision on further reflection such that they did not return for the next study visit to enroll formally. However, the still quite high enrollment percentage may underestimate willingness to provide these data because some attrition was expected between consent and enrollment visits due to the instability associated with the early stages of recovery from alcohol use disorder. In fact, table 3 indicates that almost half of the participants who consented but did not enroll may have done so for reasons other than their willingness to provide these raw data streams (e.g., health concerns, no transportation to lab, made repeated attempts to reschedule before discontinuing).

Participants' explicit self-report of their perceptions about the acceptability of these personal sensing methods were also generally consistent with their behavior. Specifically, on average, participants rated all of the sensing methods more favorable than the neutral mid-point ("undecided") of the rating scales for all 3 dimensions we evaluated - interference, dislike, and willingness to use for 1 year. These self-report data combined with our behavioral measures to suggest that all of these sensing methods can be considered for use with the majority of individuals with alcohol use disorder.

Despite the aggregate positive perceptions of the full sample, non-trivial percentages of participants did report individual ratings that were more negative than the neutral mid-point across sensing methods and specific self-report items. For example, 17.5% of participants (27/154) agreed or strongly agreed that the audio check-ins interfered with their daily activities. Approximately 25% of participants agreed or strongly agreed that they disliked both the audio check-ins (42/154; 27.3%) and providing access to the content of their text messages (33/154; 21.4%). And approximately 20% of participants disagreed or strongly disagreed that they would be willing to use our sensing methods for audio check-ins (25/154; 16.2%), EMA (35/154; 22.7%), and text message content (23/154; 14.9%) for 1 year to help their recovery. This suggests that there is still need to improve each of these sensing

methods to make them more acceptable to a larger percentage of individuals. The free response evaluations of each method (Tables S3 – S7 in Multimedia Appendix 3) provide a starting point to address participant concerns. That said, our participants did generally opt-in and adhere to our sensing methods despite reporting these concerns. Therefore, it is not clear yet at what threshold these concerns will translate to barriers for use or adherence to these methods.

## 2. Individuals can sustain the use of personal sensing for relatively long periods

Most enrolled participants were also able to sustain their commitment to provide these sensed data streams over time. More than 91% (154/169) provided at least 1 month of sensed data and a large majority (133/169; 78.7%) provided data for all 3 months. As with enrollment statistics, these numbers also likely underestimate participants' ability to sustain personal sensing because many of the participants who discontinued or did not complete the study reported reasons to stop their participation that were unrelated to personal sensing (e.g., family crisis, relapse, moved out of state). However, some participants ($N = 4$) did explicitly report reasons that appeared related to personal sensing (e.g., study demands too burdensome). Additionally, others who stopped participating may have been influenced by their experiences with personal sensing without formally reporting those concerns.

Participants who enrolled but then discontinued because of the personal sensing methods may have been influenced more by issues related to the burden associated with active sensing rather than more general issues related to data sensitivity/privacy. Participants concerned about sharing passively sensed private information such as their moment-by-moment location or cellular communications would likely have had these concerns from the beginning such that they would not have consented, enrolled, and then opted-in to provide these sensitive data. However, the burden associated with active sensing (e.g., 4x daily EMA, daily audio check-ins) may not have been clear to them until they tried to sustain those methods over time.

In our sample of participants, we saw evidence that many of our participants hardly thought about the passively sensed data streams. For example:

"I forgot I was being tracked so it was not a big deal to me."

"I went about my days as i normally would and never thought about it."

"I forgot this was even happening most of the time. It did not interfere with my life."

On the other hand, some participants reported more discontent with the actively sensed data streams as time progressed.

"I didnt mind the surveys but after the second month i felt like they became repetitive and I didn't feel like I got much utility from them."

"Became overwhelming over time."

"I did well with my recovery on my own and the surveys were more of a burden at the end."

Existing research assessing acceptability of sensing methods has been limited by short durations of monitoring, with very few studies extending beyond 6 weeks [21,54,56]. Additionally, adherence has been shown in some studies to decrease after only a few weeks [42,47,48]. The present study demonstrates that individuals can sustain their commitment to providing personally sensed data over time with limited drop-off. These findings suggest personal sensing methods may be viable in clinical settings, where consistent, sustained monitoring would be necessary. Given this promise, future research should expand to longer durations to assess self-reported and behavioral acceptability beyond 3 months. Our group is exploring this directly by employing personal sensing monitoring with individuals with opioid use disorder for a full year [16]. Methods that permit long-term monitoring are

particularly important for clinical applications for individuals with substance use disorders, who require lifelong care that can adapt as their risk for relapse and corresponding recovery needs fluctuate over time.

## 3. Some types of active personal sensing methods are generally acceptable and sustainable

Assessment burden may be expected to play a role in both the acceptability of active sensing methods and participant adherence to the associated procedures. Nonetheless, participants displayed relatively high adherence to the 4x daily EMA (on average 79.8% of EMAs completed). This is notable because our study duration of 3 months was substantially longer than typical studies using EMA, which often last only 2 – 4 weeks [46,47]. This increases confidence in the feasibility of this active sensing method for research and clinical applications that require longer monitoring periods. Of course, this level of adherence may be contingent on the measurement parameters used in our study (4x daily survey of 7 – 10 items). In fact, even higher adherence may have been observed if measurement was limited to 1 EMA per day given that on average participants completed at least 1 of the 4 EMAs on 94.1% of study days. Participants were also significantly more likely to report a willingness to use a 1X daily EMA compared to 4X daily EMA for 1 year. However, we must interpret these findings cautiously. Participant self-reports to a 1X daily EMA method are not based on experience since they were expected to adhere to the 4X daily EMA.

From free-response comments, we see evidence that many of our participants had no issues with the 4X daily EMA and some even enjoyed the frequent prompts. For example:

"They were quick and easy to complete."

"I had no issues for the daily surveys. I felt that it kept me in check and reminder to not drink. I would not change it."

"I liked the structure that it added to my day. For other activities,a being

deliberate about setting aside time for things has helped me maintain changes to

my life for the better. The surveys were no exception."

Still, other participants suggested less frequent prompts would be more practical. For

example:

"I felt it was an annoyance, another thing i had to do on top of everything else

when it was so frequent. I found one per day is tolerable by days end not a set

time. I am just too scattered with how my brain works."

"One daily would be great."

"Great tool for self check in. Not necessary to do four per day."

There was some evidence that participants found passive sensing methods to be more

acceptable than active sensing methods overall. Specifically, mean ratings for willingness to

use for 1 year were significantly higher for passive vs. active sensing methods. However, the

magnitude of this effect was small, and mean willingness was significantly greater than the

neutral mid-point for both active and passive methods. In addition, there was no difference

in mean dislike ratings for active vs. passive methods. Thus, differences between

acceptability of active and passive methods were small, inconsistent, and unlikely to be

clinically meaningful. These comparisons between active and passive methods increase our

confidence somewhat that the selective use of active measures, when necessary, may be

acceptable to participants for relatively long periods. Of course, from this study we cannot

speculate strongly beyond 3 months.

Some sensing methods (e.g., EMA, audio check-ins) will always require active input

from users but other methods may become more passive with further technological advances.

For example, our sensing of sleep quality in this study made use of an early version of the

Beddit sleep monitor that required participants to actively log when they entered and exited their bed during each period of sleep. However, later versions of the Beddit detect periods of sleep automatically. Similarly, we discontinued sensing of physiology with the Empatica E4 in an early phase of our study because participants had to manually connect the wristband each night to a tablet to upload their data. This proved too burdensome and complex for most participants. However, the current version of the Empatica E4 claims to have improved automatic bluetooth streaming of the data to the cloud, which if robust, would greatly reduce the burden associated with physiology sensing.

The acceptability of active sensing methods holds great clinical utility. Active personal sensing methods such as EMA offer unique insight into patient experiences, thoughts, and feelings that cannot always be captured accurately or comprehensively by passive methods. Self-report EMA in particular seems likely to maintain a role in risk monitoring and other, similar clinical applications. Thus, we were encouraged to find that even with relatively high active burden of 4X daily surveys, EMA was acceptable to participants as assessed via self-report and behavioral adherence.

## 4. Important individual differences in subjective perceptions exist both within and across personal sensing methods

We included a second and more novel daily active sensing method in this study, audio check-ins. These audio check-ins have high potential as a rich source of information about participants' daily experiences. Natural language processing of transcripts of their check-ins can provide a novel window into their thoughts [69–72]. These audio check-ins provided participants the opportunity to share more openly and candidly (i.e., without close-ended questions) their thoughts, feelings, and progress towards recovery without being limited to researcher-selected prompts. Analyses of the acoustic characteristics of their check-ins may yield independent measures of their affective state [73,74], including the potential for measuring affect outside of a participant's conscious awareness.

Unfortunately, overall participant adherence to the daily audio check-ins was relatively low (on average 54.3% of audio check-ins completed) and 1.9% of the sample (3/154) did not complete any check-ins throughout their entire study period. Participants' free-response evaluations of this method highlighted some concerns that could be addressed in the future to increase adherence (e.g., timing of the check-ins, technical issues with recording and sending check-ins, use of the same prompt for all check-ins). However, privacy issues related to recording the audio check-in were also reported by many participants. As examples of this concern, participants reported:

> "It takes time out of your day where you have to completely switch locations just so you can do it in private. I don't like that people could hear me and the topic wherever and whenever so I stopped using it."

> "I also have to keep privacy in mind since it's not something I'd like to complete within earshot of others."

> "The only negative part about the daily check in was my mom, who I live and work with, listening in while I was talking. I would have to find a time when she was preoccupied or away from me to talk freely."

These privacy concerns represent an inherent challenge to using this method as implemented, but accommodations could be made to gather some if not all of the same information. For example, using less frequent prompting with wider time completion windows (i.e., a weekly audio check-in) may increase individuals' ability to find a private moment. Additionally, allowing individuals to type their response as an alternative completion method could assuage concerns. This alternative would prevent acoustic analysis, but it would still permit natural language processing of open-ended responses. These accommodations could encourage greater adherence among those who completed few or no audio check-ins as well as individuals who missed check-ins sporadically due to privacy

concerns.

Finding ways to assuage privacy concerns and accommodate individual preferences may be useful as many other participants valued and believed they benefited from recording these daily audio check-ins. For example:

"I loved this part! It was like journaling kind of where i would discover things that were hidden in my subconscious. I have a recording app on my phone and i will continue doing the voice check in as a form of checking in with myself."

"This was my favorite part of the study. It helped me to set a good intention towards my recovery."

"I liked the daily check in. It gave me a way to vent without being judged."

Consistent with this somewhat polarized evaluation of the audio check-ins, a more nuanced consideration of distribution for adherence across participants suggested it was somewhat bi-modal. Participants tended to either adhere well or very poorly with this method.

More broadly, participants' self-reported perceptions were only moderately consistent across the different sensing methods. This can be seen in the moderate ICCs (and bivariate correlations) across methods for each self-report item. In other words, high dislike ratings for 1 sensing method by a specific participant did not strongly indicate that this same participant would also dislike the other sensing methods. This is also true for ratings of interference and willingness to use for 1 year items. Participants could dislike (or be unwilling to use, etc.) 1 method but not others. To the degree to which concerns are method-specific, opportunities may exist to tailor sensing systems to user preferences. In other words, participants could opt-out of methods they deemed unacceptable but provide those other sensing methods that were acceptable to them. For example, our behavioral

adherence data suggest that some participants would not have completed the study if daily audio check-ins were required, yet they were willing to provide data via other personal sensing methods. Algorithms that use sensed data for clinical applications could then be developed for different combinations of available raw data streams. Participants could be educated that personalized algorithms will likely perform better if given access to more raw data streams. This education will allow them to make an informed choice as to the threshold they set for themselves to opt-out and the potential consequences of not providing that data source. However, allowing them to opt-out of some methods may increase the number of participants who will agree to provide sensed data.

## 5. Benefits likely matter

The overall acceptability of personal sensing to research participants and patients is likely a function of both the perceived costs and benefits for those individuals [75–77]. However, we focused on measuring only perceived costs (e.g., privacy, burden) associated with personal sensing because the benefits to participants from the sensed data collected in this research study were minimal. Participants were provided with modest financial incentives to complete the EMAs ($25/month) and to provide access to the 2 passively sensed raw data streams ($10/month for geolocation and $15/month for cellular communication logs with text message content). These sensed data streams were not used to provide any clinical benefit to participants' recovery in our study.

Monetary incentives are commonly used in research to provide a more favorable cost/benefit ratio surrounding specific methods or overall participation. Such monetary incentives are commonplace and recommended when using active personal sensing methods like EMA [78]. However, the incentives to provide access to passively sensed geolocation and cellular communications in our study may have contributed to the acceptance of these methods and the success we had collecting those sensitive data from participants. This may be particularly true given the relatively low socioeconomic status of many of our participants.

For example, the median personal income for our participants was $34,233, with 12.3% (19/154) of individuals reporting current unemployment and 25.3% (39/154) reporting an annual income below the 2022 federal poverty level.

Monetary incentives to increase the acceptability of personal sensing do not need to be limited to research settings. Incentives can also be used as part of treatment or continuing care in clinical settings. For example, the use of monetary incentives or equivalents (e.g., prizes) as part of a contingency management program is well established to support abstinence from alcohol or other drugs [79,80]. If personal sensing proved useful for the treatment or ongoing support of patients' recovery, similar incentives could be established to encourage patients to provide these sensed data.

Incentives may be less necessary in clinical settings when more direct clinical benefits from personal sensing are available. For example, research has suggested that privacy concerns associated with personal sensing may be reduced if participants perceive that they will benefit from the sensed data [6,52,77]. There was some evidence for this perspective in the free response comments from our participants as well. For example:

"I felt all right having my location tracked. If it were used in a way to keep me from relapsing my feeling about it would be even more positive."

"I'm a somewhat typical introverted, slightly paranoid, grump, so the idea of being tracked is automatically negative. That being said, if it's something that helps people that can be a really good thing."

"It seems this was more to aid with the study research than to help with my recovery. If i could see a way it would be prioritized to help me, i would be more willing to have them tracked longer."

Given this, the acceptability of personal sensing may be higher than observed in our

study if the sensing system was implemented as part of their treatment or continuing care during their recovery. Digital therapeutics are particularly well-positioned to use sensed data to select, personalize, or time the delivery of interventions and other supports to improve clinical outcomes. Future research should evaluate the acceptability of personal sensing in contexts where its use directly benefits those providing the sensed data. In these contexts, benefits can also be explicitly measured. It may even be possible to manipulate the benefits from personal sensing across participants to evaluate their contribution to acceptability more rigorously.

## 6. Trust likely matters

Trust is also likely to affect the overall acceptability of personal sensing data, which are inherently private and sensitive in nature. Acceptability may depend on who employs personal sensing and who has access to the raw and processed data [51,77,81–83]. The available evidence suggests that people are more comfortable sharing private, sensitive information with researchers and their doctor and less comfortable sharing information with family members, electronic health record databases, and third party apps and websites [81–83].

We saw evidence of the role of trust in the free response comments from our participants. For example:

"I trusted the study group to not use my personal information for any other use."

"I don't mind turning this info over and I trust the people running the study."

The research setting may come with relatively greater trust because of the high level of transparency regarding risks and protections associated with obtaining informed consent. Some protections may only be feasible for research as well. For example, NIH funded research that collects identifiable, sensitive information is automatically issued a Certificate

of Confidentiality that prohibits disclosing this information to anyone not connected to the research except when the participant consents or in a few other limited situations. Certificates of Confidentiality can also be requested for similar research not funded by NIH.

Our participants appeared to recognize and appreciate these protective measures. For example:

"I don't mind this at all either, as long as I know this information is protected and anonymous."

"Didn't bother me since its protected."

"Just in general this feels intrusive but as part of the study it is fine!"

Implementations of personal sensing for treatment inside and outside of clinical care settings [33] will need to carefully consider how to establish similar, high levels of trust. Clinical applications of personal sensing may sit at an intersection of sharing data with doctors (with which individuals tend to be comfortable) and with electronic health record databases and apps (with which individuals tend to be less comfortable) [82,83]. For example, it may be necessary to protect against the subpoena of sensitive information in civil and criminal proceedings. Patients will also likely need to be assured that sensed data used for their clinical care will not also be shared with their health insurance provider with associated risks related to higher insurance premiums or dropped coverage. These issues of data access and unauthorized secondary use of otherwise private information are often cited concerns regarding personal sensing [77,84].

Regardless of the setting, trust may be lower in stigmatized groups that could otherwise benefit from personal sensing. For example, individuals with mental illness still experience substantial stigma that could impede willingness to share personal, sensitive information with researchers or clinical care providers [85–88]. In fact, we focused on

individuals with alcohol use disorder in this study to evaluate the acceptance of personal sensing methods in a population that we expected might have barriers associated with trust. Of course, trust may be lower still among individuals with other substance use disorders that involve drugs whose use is illegal. That said, many of our participants reported ongoing use of drugs other than alcohol throughout the study (75/154; 48.7% reported past month illicit drug use), as expected given high rates of poly-substance use among individuals with substance use disorders. Furthermore, we have had promising, preliminary success recruiting patients with opioid use disorder for an NIH funded study on personal sensing in this population [89]. This suggests that our results regarding the acceptance of personal sensing may generalize across substance use disorders.

Trust and related privacy concerns may also be more difficult to overcome in historically marginalized groups that have experienced systemic racism and other stigma or exclusions [90]. These individuals may find it more difficult to achieve privacy in their daily lives, and they may hold very different perspectives on the costs vs. benefits of surveillance in the context of personal sensing or more generally. Unfortunately, our sample was not diverse with respect to race and ethnicity. Future research on personal sensing must specifically recruit for such diversity to better understand its acceptance in communities of color. We have learned from the present study and adjusted our recruiting efforts accordingly to recruit a sample that is more diverse with respect to race, ethnicity, and geographical region for our ongoing personal sensing project with individuals with opioid use disorder.

## 7. Feasibility is a function of more than participant perceptions of acceptabilty

Of course, user acceptance of personal sensing methods is necessary but not sufficient to expand use of these methods in research and clinical implementations. A variety of other key issues may facilitate or present barriers to wider use of personal sensing. These include cost and accessibility, stability over time, and the utility of personal sensing relative to other more traditional methods.

The smartphone itself is arguably the best available sensing system today. Today's smartphone contains numerous sensors and other raw data streams that can be used for personal sensing. In our study, we took advantage of GPS and other location services to track geolocation, and the microphone for daily check-ins. We accessed smartphone call and text message logs for communications meta-data and message content. The smartphone also provided a convenient platform to collect self-report EMA.

Smartphones also provide a relatively accessible platform for personal sensing. Despite their high cost, 85% of adults in the US already own a smartphone. Equally important, this level of ownership is relatively consistent across race/ethnicity, geographic regions (e.g., urban, suburban, rural) and income level [3]. In fact, only 11 of the 169 eligible participants for our study (6.5%) did not already own a contemporary smartphone. This is notable given that individuals with alcohol use disorder may have been expected to present with more barriers to smartphone ownership than that of the general population. In a research setting, we were able to provide individuals with a smartphone if they did not already have one. Like monetary incentives, this practice need not be limited to research; smartphones could be provided to permit personal sensing-based clinical support.

Personal sensing can also be done with wearable or other sensors outside of the smartphone. We used Empatica and Beddit systems to sense physiology and sleep, respectively. The use of watches (e.g., Apple Watch) and wristbands (e.g., Fitbit) for sensing activity and some physiology is also increasing [91,92]. However, some of these systems can be expensive, and - unlike smartphones - none have been adopted widely enough to assume that most users will already own said devices. For research applications, this limitation can be overcome by providing the hardware to participants as needed. Though not impossible to do the same in clinical settings, the large number of patients who would require this technology may either limit or increase the cost to scale the sensing system.

Both research and clinical applications of sensing systems require some guarantee that

the hardware and software will remain available and supported for the duration of the intended use. Unfortunately, there are currently high levels of churn among the companies that support these systems given the rapid innovation occurring at this time. We collected data for approximately 2.5 years between 2017 – 2019. During this time, Apple bought the company that developed the Beddit Sleep Monitor and discontinued support for previous users. Apple re-introduced the sleep sensing system for iPhone users in late 2018, but discontinued it again in early 2022. For these reasons, we were able to collect sleep sensing data on fewer than half of our research participants.

During this same data collection period, there was also churn in the software that we used for sensing geolocation. We used the Moves app at the start of the study, but needed to switch to use FollowMee when Facebook acquired the company that developed Moves and discontinued its support. However, this software churn was less disruptive because both apps relied on smartphone sensors to acquire the raw geolocation data stream. This suggests yet another reason to prefer systems that make use of generic smartphone sensors rather than propriety hardware.

High rates of churn can also affect the perceived acceptability of the software. For example, it could be inconvenient to have to adapt to frequent changing of app platforms. Additionally, software may be left unmonitored for periods of time leaving new bugs unresolved. We saw in our own sample of participants how frustrating technological issues were. For example:

> "I experienced issues with this audio check-in from the very beginning and discontinued early on. I was unable to get the message beyond 8-10 seconds with close to 6 attempts each day. Was not worth the time or stress for me."

> "My bigger problem with the tracker app. Was more the fact that my battery power was getting depleted so rapidly, sometimes within 2 hours post-full charge.

I found myself charging my phone at least 5 times a day. And at times, this became very stressful."

"There were problems with the sleep monitor in that I was not able to activate the monitor when I was going to sleep."

**Limitations and Future Directions**

The present study demonstrates the acceptability of several personal sensing methods. These methods were acceptable 1) over a longer period of time than has previously been assessed, 2) across active and passive methods, 3) despite the sensitivity of the data, 4) among individuals with alcohol use disorder who may have greater privacy concerns, and 5) without explicit clinical benefits to the participants. These findings suggest personal sensing methods are poised as accessible, feasible avenues to collect data about individuals to be used for clinical applications. More work is needed to determine the predictive utility of the data that can be collected via personal sensing, but our study shows that this work will be worthwhile to pursue.

Personal sensing is acceptable, and the technology to collect it (namely, the smartphone) is widely accessible. Personal sensing can make digital therapeutics - smartphone and web-based apps that provide mental health care - smart. These methods can personalize care for individuals such that they receive the specific interventions and supports they need at the time they need them. Smart digital therapeutics can be scaled widely to provide treatment to the overwhelming majority of individuals who do not currently receive mental health care. They can reach those who have been historically excluded from or have otherwise faced barriers to care. With personal sensing powering digital therapeutics, we are positioned for a paradigm shift in mental health care. The present study brings us 1 step closer to this goal, ensuring that the methods we hope to use to revolutionize care are acceptable to the patients who will use them.

## Acknowledgments

**Conflicts of Interest**

None declared.

# References

1. WHO Global Observatory for eHealth. mHealth: New horizons for health through mobile technologies: Second global survey on eHealth. World Health Organization; 2011.

2. Majumder S, Deen MJ. Smartphone Sensors for Health Monitoring and Diagnosis. *Sensors* Multidisciplinary Digital Publishing Institute; 2019 Jan;19(9):2164. [doi: 10.3390/s19092164]

3. Center PR. Mobile Fact Sheet. 2021.

4. Health C for D and R. What is Digital Health? *FDA* FDA; Tue, 09/22/2020 - 11:26;

5. Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual Review of Clinical Psychology* 2017;13(1):23–47. PMID:28375728

6. Klasnja P, Consolvo S, Choudhury T, Beckwith R, Hightower J. Exploring Privacy Concerns about Personal Sensing. In: Tokuda H, Beigl M, Friday A, Brush AJB, Tobe Y, editors. *Pervasive Computing* Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. p. 176–183. [doi: 10.1007/978-3-642-01516-8_13]

7. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: A timely opportunity to consider purpose, quality, and safety. *npj Digital Medicine* 2019 Dec;2(1):88. [doi: 10.1038/s41746-019-0166-1]

8. Morgenstern J, Kuerbis A, Muench F. Ecological Momentary Assessment and Alcohol Use Disorder Treatment. *Alcohol Research : Current Reviews* 2014;36(1):101–110.

PMID:26259004

9. Fronk GE, Sant'Ana SJ, Kaye JT, Curtin JJ. Stress Allostasis in Substance Use Disorders: Promise, Progress, and Emerging Priorities in Clinical Research. *Annual Review of Clinical Psychology* 2020;16(1):401–430. PMID:32040338

10. Schultz ME, Fronk G, Jaume N, Magruder K, Curtin JJ. Stressor-elicited smoking and craving during a smoking cessation attempt. *Journal of Abnormal Psychology* In press; [doi: 10.31234/osf.io/h2cbe]

11. Piasecki TM, Jorenby DE, Smith SS, Fiore MC, Baker TB. Smoking withdrawal dynamics: I. Abstinence distress in lapsers and abstainers. *Journal of Abnormal Psychology* 2003;112(1):3–13.

12. McCarthy DE, Piasecki TM, Fiore MC, Baker TB. Life before and after quitting smoking: An electronic diary study. *Journal of Abnormal Psychology* 2006 Aug;115(3):454–466. [doi: d8k3rz]

13. aan het Rot M, Hogenelst K, Schoevers RA. Mood disorders in everyday life: A systematic review of experience sampling and ecological momentary assessment studies. *Clinical Psychology Review* 2012 Aug;32(6):510–523. PMID:22721999

14. Epstein DH, Tyburski M, Kowalczyk WJ, Burgess-Hull AJ, Phillips KA, Curtis BL, Preston KL. Prediction of stress and drug craving ninety minutes in the future with passively collected GPS data. *npj Digital Medicine* 2020 Dec;3(1):26. [doi: ghqvcw]

15. Palmius N, Tsanas A, Saunders K, Bilderbeck A, Geddes J, Goodwin G, de Vos M.

Detecting Bipolar Depression From Geographic Location Data. *IEEE Transactions on Biomedical Engineering* 2016 Oct;PP:1–1. [doi: 10.1109/TBME.2016.2611862]

16. Moshontz H, Colmenares AJ, Fronk GE, Sant'Ana SJ, Wyant K, Wanta SE, Maus A, Jr DHG, Shah D, Curtin JJ. Prospective Prediction of Lapses in Opioid Use Disorder: Protocol for a Personal Sensing Study. *JMIR Research Protocols* JMIR Publications Inc., Toronto, Canada; 2021 Dec;10(12):e29563. [doi: 10.2196/29563]

17. Jacobson NC, Summers B, Wilhelm S. Digital Biomarkers of Social Anxiety Severity: Digital Phenotyping Using Passive Smartphone Sensors. *Journal of Medical Internet Research* JMIR Publications Inc., Toronto, Canada; 2020 May;22(5):e16875. [doi: 10.2196/16875]

18. Wang Y, Ren X, Liu X, Zhu T. Examining the Correlation Between Depression and Social Behavior on Smartphones Through Usage Metadata: Empirical Study. *JMIR mHealth and uHealth* JMIR Publications Inc., Toronto, Canada; 2021 Jan;9(1):e19046. [doi: 10.2196/19046]

19. Razavi R, Gharipour A, Gharipour M. Depression screening using mobile phone usage metadata: A machine learning approach. *Journal of the American Medical Informatics Association* 2020 Apr;27(4):522–530. [doi: 10.1093/jamia/ocz221]

20. Bai R, Xiao L, Guo Y, Zhu X, Li N, Wang Y, Chen Q, Feng L, Wang Y, Yu X, Xie H. Tracking and Monitoring Mood Stability of Patients With Major Depressive Disorder by Machine Learning Models Using Passive Digital Data: Prospective Naturalistic Multicenter Study. *JMIR MHEALTH AND UHEALTH* 2021;15.

21. Kleiman E, Millner AJ, Joyce VW, Nash CC, Buonopane RJ, Nock MK. Using Wearable Physiological Monitors With Suicidal Adolescent Inpatients: Feasibility and Acceptability Study. *JMIR mHealth and uHealth* JMIR Publications Inc., Toronto, Canada; 2019 Sep;7(9):e13725. [doi: 10.2196/13725]

22. Kleiman EM, Bentley KH, Maimone JS, Lee H-IS, Kilbury EN, Fortgang RG, Zuromski KL, Huffman JC, Nock MK. Can passive measurement of physiological distress help better predict suicidal thinking? *Translational Psychiatry* Nature Publishing Group; 2021 Dec;11(1):1–6. [doi: 10.1038/s41398-021-01730-y]

23. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annual Review of Clinical Psychology* 2008;4(1):1–32. PMID:18509902

24. Onnela J-P, Rauch SL. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology* 2016 Jun;41(7):1691–1696. PMID:26818126

25. Torous J, Staples P, Onnela J-P. Realizing the Potential of Mobile Mental Health: New Methods for New Data in Psychiatry. *Current psychiatry reports* 2015 Aug;17(8):602. PMID:26073363

26. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, Asch DA, Schwartz HA. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences* 2018 Oct;115(44):201802331. PMID:30322910

27. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela J-P. Relapse prediction

in schizophrenia through digital phenotyping: A pilot study. *Neuropsychopharmacology* 2018 Jul;43(8):1660–1666. [doi: 10.1038/s41386-018-0030-z]

28. Chih M-Y, Patton T, McTavish FM, Isham AJ, Judkins-Fisher CL, Atwood AK, Gustafson DH. Predictive modeling of addiction lapses in a mobile health application. *Journal of Substance Abuse Treatment* 2014 Jan;46(1):29–35. PMID:24035143

29. Jashinsky J, Burton S, Hanson C, West J, Giraud-Carrier C, Barnes M, Argyle T. Tracking Suicide Risk Factors Through Twitter in the US. *Crisis* 2013 Oct;35:1–9. [doi: 10.1027/0227-5910/a000234]

30. Jacobson NC, Chung YJ. Passive Sensing of Prediction of Moment-To-Moment Depressed Mood among Undergraduates with Clinical Levels of Depression Sample Using Smartphones. *Sensors* 2020 Jun;20(12):3572. [doi: 10.3390/s20123572]

31. Quanbeck AR, Gustafson DH, Marsch LA, McTavish F, Brown RT, Mares M-L, Johnson R, Glass JE, Atwood AK, McDowell H. Integrating addiction treatment into primary care using mobile health technology: Protocol for an implementation research study. *Implementation science: IS* 2014 May;9:65. PMID:24884976

32. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, Murphy SA. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine* 2018 May;52(6):446–462. PMID:27663578

33. Aggarwal N, Ahmed M, Basu S, Curtin JJ, Evans BJ, Matheny ME, Nundy S, Sendak MP, Shachar C, Shah RU, Thadaney-Israni and S. Advancing Artificial Intelligence in Health

Settings Outside the Hospital and Clinic. *NAM Perspectives* 2020 Nov; [doi: 10.31478/202011f]

34. Kaiser J, 2015, Pm 4:15. Obama gives East Room rollout to Precision Medicine Initiative. Science https://www.sciencemag.org/news/2015/01/obama-gives-east-room-rollout-precision-medicine-initiative; 2015.

35. Sheikh M, Qassem M, Kyriacou PA. Wearable, Environmental, and Smartphone-Based Passive Sensing for Mental Health Monitoring. *Frontiers in Digital Health* 2021;3.

36. De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, Lavelle G, Matcham F, Pace A, Mohr DC, Dobson R, Hotopf M. Digital health tools for the passive monitoring of depression: A systematic review of methods. *npj Digital Medicine* Nature Publishing Group; 2022 Jan;5(1):1–14. [doi: 10.1038/s41746-021-00548-8]

37. Ortiz A, Maslej MM, Husain MI, Daskalakis ZJ, Mulsant BH. Apps and gaps in bipolar disorder: A systematic review on electronic monitoring for episode prediction. *Journal of Affective Disorders* 2021 Dec;295:1190–1200. [doi: 10.1016/j.jad.2021.08.140]

38. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: Status and considerations. *International Journal of Bipolar Disorders* 2018 Jan;6. PMID:29359252

39. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: Reactivity, compliance, and patient satisfaction. *Pain* 2003 Jul;104(1-2):343–351. PMID:12855344

40. Kirk GD, Linas BS, Westergaard RP, Piggott D, Bollinger RC, Chang LW, Genz A. The Exposure Assessment in Current Time Study: Implementation, Feasibility, and Acceptability of Real-Time Data Collection in a Community Cohort of Illicit Drug Users. *AIDS Research and Treatment* Hindawi; 2013 Nov;2013:e594671. [doi: 10.1155/2013/594671]

41. Ramsey AT, Wetherell JL, Depp C, Dixon D, Lenze E. Feasibility and Acceptability of Smartphone Assessment in Older Adults with Cognitive and Emotional Difficulties. *Journal of Technology in Human Services* Routledge; 2016 Apr;34(2):209–223. [doi: 10.1080/15228835.2016.1170649]

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

43. Moitra E, Gaudiano BA, Davis CH, Ben-Zeev D. Feasibility and acceptability of post-hospitalization ecological momentary assessment in patients with psychotic-spectrum disorders. *Comprehensive Psychiatry* 2017 Apr;74:204–213. [doi: 10.1016/j.comppsych.2017.01.018]

44. Eisele G, Vachon H, Lafit G, Kuppens P, Houben M, Myin-Germeys I, Viechtbauer W. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. 2020;

45. Wen CK, Schneider S, Stone A, Spruijt-Metz D. Compliance With Mobile Ecological Momentary Assessment Protocols in Children and Adolescents: A Systematic Review and

Meta-Analysis. *Journal of Medical Internet Research* 2017 Apr;19:e132. [doi: 10.2196/jmir.6641]

46. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IHA, Wen CKF, Field M. Compliance with ecological momentary assessment protocols in substance users: A meta-analysis. *Addiction (Abingdon, England)* 2019 Apr;114(4):609–619. PMID:30461120

47. Wrzus C, Neubauer AB. Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance Across Research Fields. *Assessment* SAGE Publications Inc; 2022 Jan;10731911211067538. [doi: 10.1177/10731911211067538]

48. Ono M, Schneider S, Junghaenel DU, Stone AA. What Affects the Completion of Ecological Momentary Assessments in Chronic Pain Research? An Individual Patient Data Meta-Analysis. *Journal of Medical Internet Research* 2019 Feb;21(2):e11398. [doi: 10.2196/11398]

49. Smyth JM, Jones DR, Wen CKF, Materia FT, Schneider S, Stone A. Influence of ecological momentary assessment study design features on reported willingness to participate and perceptions of potential research studies: An experimental study. *BMJ Open* British Medical Journal Publishing Group; 2021 Jul;11(7):e049154. PMID:34330860

50. Duncan DT, Park SH, Goedel WC, Sheehan DM, Regan SD, Chaix B. Acceptability of smartphone applications for global positioning system (GPS) and ecological momentary assessment (EMA) research among sexual minority men. Puebla I, editor. *PLOS ONE* 2019 Jan;14(1):e0210240. [doi: 10.1371/journal.pone.0210240]

51. Rieger A, Gaines A, Barnett I, Baldassano CF, Connolly Gibbons MB, Crits-Christoph

P. Psychiatry Outpatients' Willingness to Share Social Media Posts and Smartphone Data for Research and Clinical Purposes: Survey Study. *JMIR Formative Research* 2019 Aug;3(3):e14329. [doi: 10.2196/14329]

52. Bessenyei K, Suruliraj B, Bagnell A, McGrath P, Wozney L, Huguet A, Elger BS, Meier S, Orji R. Comfortability with the passive collection of smartphone data for monitoring of mental health: An online survey. *Computers in Human Behavior Reports* 2021 Aug;4:100134. [doi: 10.1016/j.chbr.2021.100134]

53. Lind MN, Byrne ML, Wicks G, Smidt AM, Allen NB. The Effortless Assessment of Risk States (EARS) Tool: An Interpersonal Approach to Mobile Sensing. *JMIR Mental Health* 2018 Aug;5(3):e10334. [doi: 10.2196/10334]

54. Ben-Zeev D, Wang R, Abdullah S, Brian R, Scherer EA, Mistler LA, Hauser M, Kane JM, Choudhury T, Campbell A. Mobile Behavioral Sensing in Outpatients and Inpatients with Schizophrenia. *Psychiatric services (Washington, DC)* 2016 May;67(5):558–561. PMID:26695497

55. Rooksby J, Morrison A, Murray-Rust D. Student Perspectives on Digital Phenotyping: The Acceptability of Using Smartphone Data to Assess Mental Health. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* Glasgow Scotland Uk: ACM; 2019. p. 1–14. [doi: 10.1145/3290605.3300655]

56. Raugh IM, James SH, Gonzalez CM, Chapman HC, Cohen AS, Kirkpatrick B, Strauss GP. Digital phenotyping adherence, feasibility, and tolerability in outpatients with schizophrenia. *Journal of Psychiatric Research* 2021 Jun;138:436–443. [doi: 10.1016/j.jpsychires.2021.04.022]

57. Schönbrodt FD, Maier M, Heene M, Zehetleitner M. Voluntary commitment to research transparency. http://www.researchtransparency.org; 2015.

58. Simmons JP, Nelson LD, Simonsohn U. A 21 Word Solution. http://dx.doi.org/10.2139/ssrn.2160588; 2012.

59. Aczel B, Szaszi B, Sarafoglou A, Kekecs Z, Kucharský Š, Benjamin D, Chambers CD, Fisher A, Gelman A, Gernsbacher MA, Ioannidis JP, Johnson E, Jonas K, Kousta S, Lilienfeld SO, Lindsay DS, Morey CC, Monafò M, Newell BR, Pashler H, Shanks DR, Simons DJ, Wicherts JM, Albarracin D, Anderson ND, Antonakis J, Arkes HR, Back MD, Banks GC, Beevers C, Bennett AA, Bleidorn W, Boyer TW, Cacciari C, Carter AS, Cesario J, Clifton C, Conroy RM, Cortese M, Cosci F, Cowan N, Crawford J, Crone EA, Curtin J, Engle R, Farrell S, Fearon P, Fichman M, Frankenhuis W, Freund AM, Gaskell MG, Giner-Sorolla R, Green DP, Greene RL, Harlow LL, de la Guardia FH, Isaacowitz D, Kolodner J, Lieberman D, Logan GD, Mendes WB, Moersdorf L, Nyhan B, Pollack J, Sullivan C, Vazire S, Wagenmakers E-J. A consensus-based transparency checklist. *Nature Human Behaviour* 2019 Dec;1–3. [doi: 10.1038/s41562-019-0772-6]

60. https://osf.io/cjsvk;

61. Derogatis LR, Lipman RS, Covi L. SCL-90: An outpatient psychiatric rating scale–preliminary report. *Psychopharmacology Bulletin* 1973 Jan;9(1):13–28. PMID:4682398

62. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.

63. RStudio Team. RStudio: Integrated Development for R. Boston, MA: RStudio, Inc;

2020.

64. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. Welcome to the Tidyverse. *Journal of Open Source Software* 2019 Nov;4(43):1686. [doi: 10.21105/joss.01686]

65. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 1979;86(2):420–428.

66. Kilian C, Manthey J, Carr S, Hanschmidt F, Rehm J, Speerforck S, Schomerus G. Stigmatization of people with alcohol use disorders: An updated systematic review of population studies. *Alcoholism: Clinical and Experimental Research* 2021;45(5):899–911. [doi: 10.1111/acer.14598]

67. Substance Abuse and Mental Health Services Administration. Key Substance Use and Mental Health Indicators in the United States: Results from the 2019 National Survey on Drug Use and Health. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2020. Report No.: HHS Publication No. PEP20-07-01-001.

68. TEDx Talks. Mental Healthcare at Our Fingertips. 2022.

69. Tausczik YR, Pennebaker JW. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 2010 Mar;29(1):24–54. [doi: 10.1177/0261927X09351676]

70. Tackman AM, Sbarra DA, Carey AL, Donnellan MB, Horn AB, Holtzman NS, Edwards TS, Pennebaker JW, Mehl MR. Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *Journal of Personality and Social Psychology* American Psychological Association; 2019 May;116(5):817–834. [doi: 10.1037/pspp0000187]

71. Jacobucci R, Ammerman BA, Wilcox KT. The use of text-based responses to improve our understanding and prediction of suicide risk. *Suicide and Life-Threatening Behavior* 2021;51(1):55–64. [doi: 10.1111/sltb.12668]

72. Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS. Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study. *Journal of Medical Internet Research* JMIR Publications Inc., Toronto, Canada; 2020 Oct;22(10):e22635. [doi: 10.2196/22635]

73. Belouali A, Gupta S, Sourirajan V, Yu J, Allen N, Alaoui A, Dutton MA, Reinhard MJ. Acoustic and language analysis of speech for suicidal ideation among US veterans. *BioData Mining* 2021 Dec;14(1):11. [doi: 10.1186/s13040-021-00245-y]

74. Faurholt-Jepsen M, Busk J, Frost M, Vinberg M, Christensen EM, Winther O, Bardram JE, Kessing LV. Voice analysis as an objective state marker in bipolar disorder. *Translational Psychiatry* 2016 Jul;6:e856. PMID:27434490

75. Pavlou PA. Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. *International Journal of Electronic Commerce* Taylor & Francis, Ltd.; 2003;7(3):101–134.

76. Schnall R, Higgins T, Brown W, Carballo-Dieguez A, Bakken S. Trust, Perceived Risk, Perceived Ease of Use and Perceived Usefulness as Factors Related to mHealth Technology Use. *Studies in Health Technology and Informatics* 2015;216:467–471. PMID:26262094

77. Atienza A, Zarcadoolas C, Vaughon W, Hughes P, Patel V, Chou W-Y, Pritts J. Consumer Attitudes and Perceptions on mHealth Privacy and Security: Findings From a Mixed-Methods Study. *Journal of health communication* 2015 Apr;20:1–7. [doi: 10.1080/10810730.2015.1018560]

78. Parkinson B, Meacock R, Sutton M, Fichera E, Mills N, Shorter GW, Treweek S, Harman NL, Brown RCH, Gillies K, Bower P. Designing and using incentives to support recruitment and retention in clinical trials: A scoping review and a checklist for design. *Trials* 2019 Nov;20(1):624. [doi: 10.1186/s13063-019-3710-z]

79. Prendergast M, Podus D, Finney J, Greenwell L, Roll J. Contingency management for treatment of substance use disorders: A meta-analysis. *Addiction (Abingdon, England)* 2006 Nov;101(11):1546–1560. PMID:17034434

80. Ginley MK, Pfund RA, Rash CJ, Zajac K. Long-term efficacy of contingency management treatment based on objective indicators of abstinence from illicit substance use up to 1 year following treatment: A meta-analysis. *Journal of Consulting and Clinical Psychology* US: American Psychological Association; 2021;89(1):58–71. [doi: 10.1037/ccp0000552]

81. Rendina HJ, Mustanski B. Privacy, Trust, and Data Sharing in Web-Based and Mobile Research: Participant Perspectives in a Large Nationwide Sample of Men Who Have Sex With Men in the United States. *Journal of Medical Internet Research* 2018 Jul;20(7).

PMID:29973332

82. Nicholas J, Shilton K, Schueller SM, Gray EL, Kwasny MJ, Mohr DC. 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* 2019 Apr;7(4):e12578. [doi: 10.2196/12578]

83. Prasad A, Sorber J, Stablein T, Anthony D, Kotz D. Exposing privacy concerns in mHealth. 2011.

84. Ackerman M, Mainwaring SD. Privacy Issues and Human-Computer Interaction. https://www.semanticscholar.org/paper/Privacy-Issues-and-Human-Computer-Interaction-Ackerman-Mainwaring/d36442dacea38d57ed2f0ef2b52eac763c2b3cfb; 2008.

85. Schomerus G, Lucht M, Holzinger A, Matschinger H, Carta MG, Angermeyer MC. The Stigma of Alcohol Dependence Compared with Other Mental Disorders: A Review of Population Studies. *Alcohol and Alcoholism* 2011 Mar;46(2):105–112. [doi: 10.1093/alcalc/agq089]

86. Barry CL, McGinty EE, Pescosolido BA, Goldman HH. Stigma, Discrimination, Treatment Effectiveness, and Policy: Public Views About Drug Addiction and Mental Illness. 2014;65(10):4.

87. Overton SL, Medina SL. The Stigma of Mental Illness. *Journal of Counseling & Development* 2008;86(2):143–151. [doi: 10.1002/j.1556-6678.2008.tb00491.x]

88. Parcesepe AM, Cabassa LJ. Public Stigma of Mental Illness in the United States: A

Systematic Literature Review. *Adm Policy Ment Health* 2013;16.

89. https://reporter.nih.gov/search/9c2I9yUXI0afivxf9QjnDA/project-details/10172881;

90. Marwick AE, Boyd D. Privacy at the Margins| Understanding Privacy at the Margins—Introduction. *International Journal of Communication* 2018 Mar;12(0):9.

91. Doryab A, Villalba DK, Chikersal P, Dutcher JM, Tumminia M, Liu X, Cohen S, Creswell K, Mankoff J, Creswell JD, Dey AK. Identifying Behavioral Phenotypes of Loneliness and Social Isolation with Passive Sensing: Statistical Analysis, Data Mining and Machine Learning of Smartphone and Fitbit Data. *JMIR mHealth and uHealth* JMIR Publications Inc., Toronto, Canada; 2019 Jul;7(7):e13209. [doi: 10.2196/13209]

92. Stevenson BL, Kunicki ZJ, Brick L, Blevins CE, Stein M, Abrantes AM. Using Ecological Momentary Assessments and Fitbit Data to Examine Daily Associations Between Physical Activity, Affect and Alcohol Cravings in Patients with Alcohol Use Disorder. *International Journal of Behavioral Medicine* 2021 Nov; [doi: 10.1007/s12529-021-10039-5]
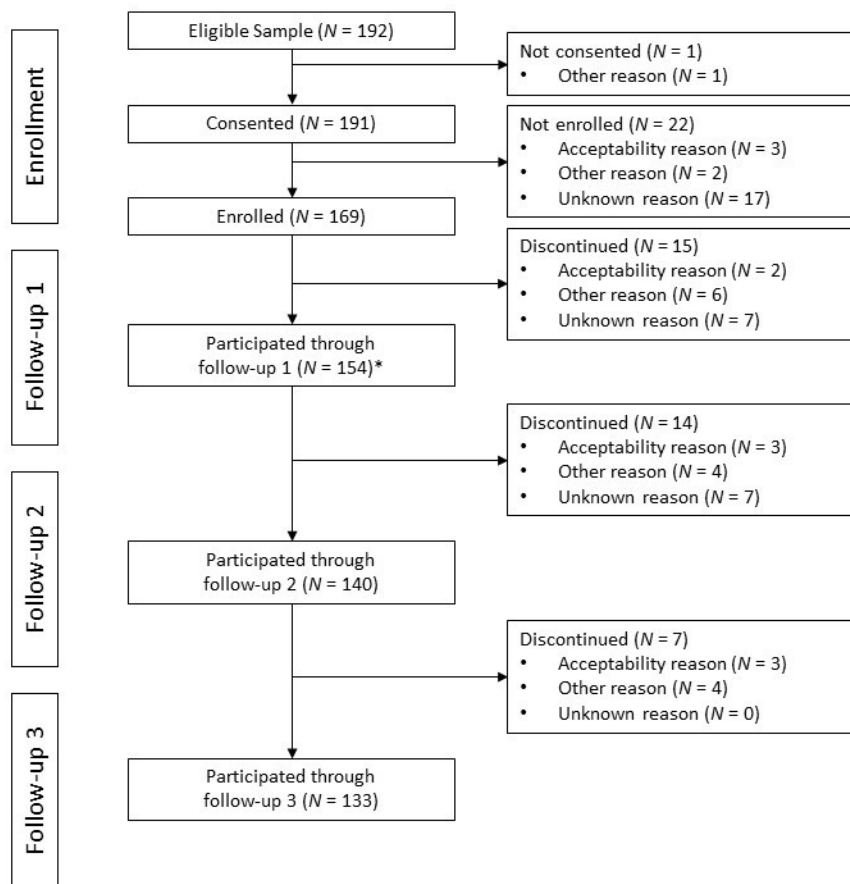
*Figure 1*. Flowchart of participant retention over the course of the 3-month study. This figure displays retention and attrition of all eligible participants at various stages from consent through study completion. It also displays the reasons for attrition categorized as due to acceptability, other reasons, or unknown. We present additional detail on reasons for attrition in Table 3.

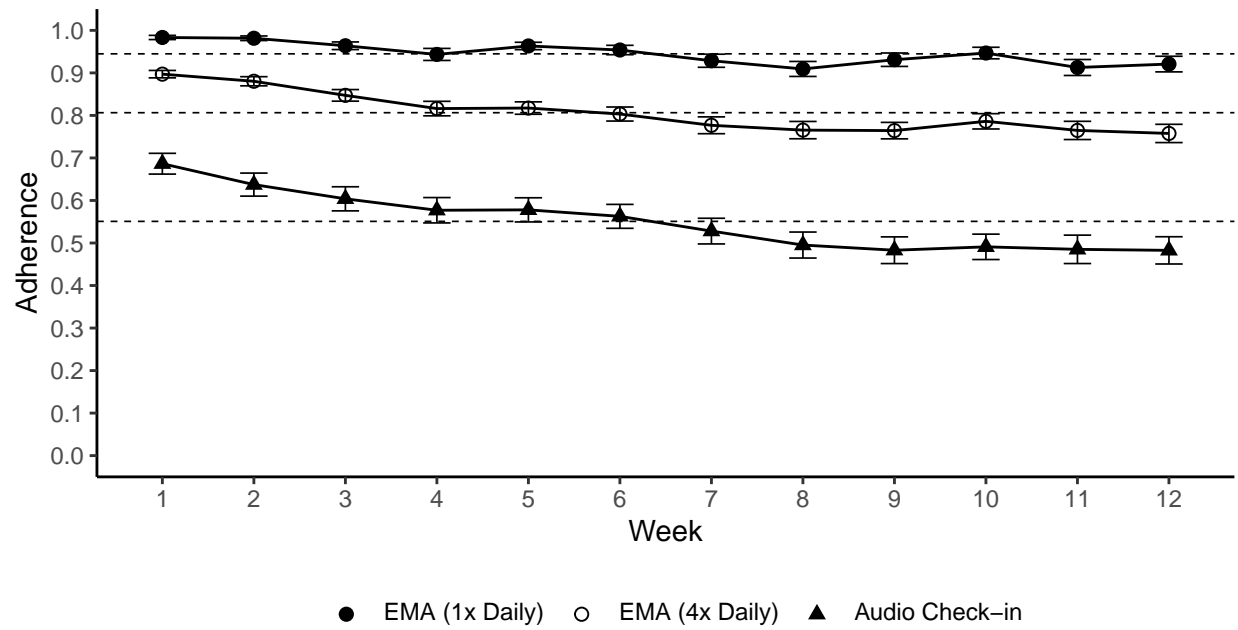*All participants who completed through follow-up 1 were used in the analyses.

*Figure 2*. Adherence over Time for EMA (1x daily), EMA (4x daily), and Audio Check-in. Notes: Mean adherence for each week on study. Mean standard error is depicted by the solid error bars. Overall mean adherence is depicted by the dashed line. $N = 154$.
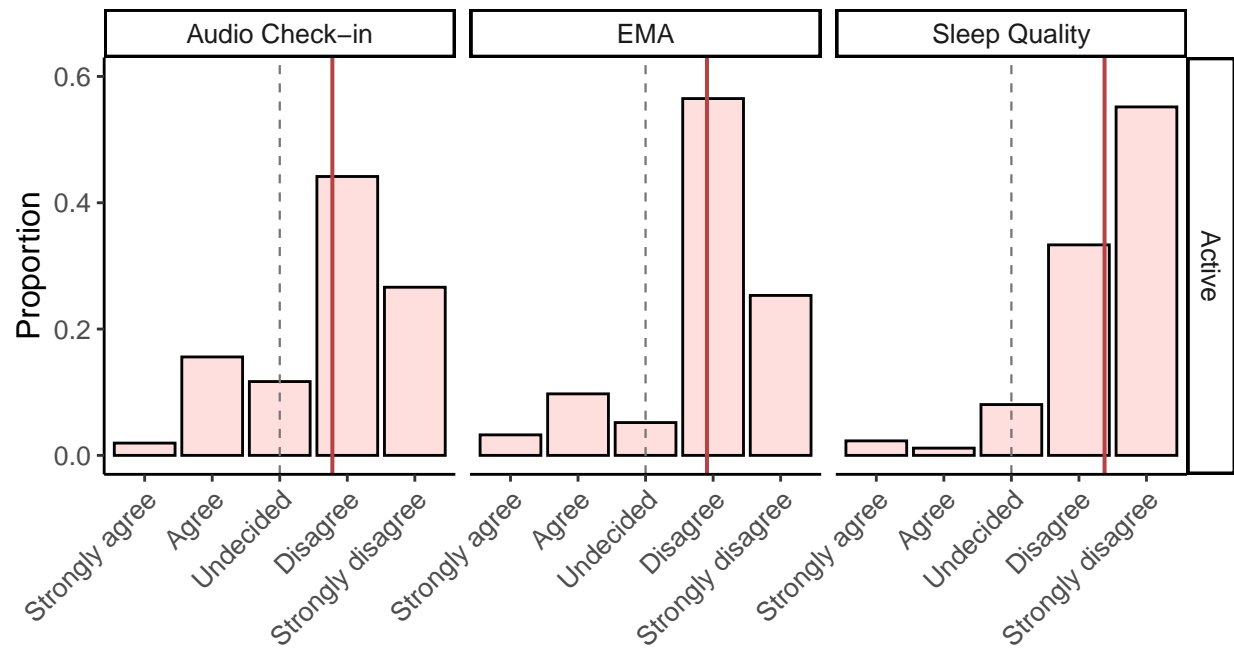
*Figure 3*. Interference Ratings by Personal Sensing Data Stream.

Notes: Mean responses to "[Personal sensing method name] interfered with my daily activities." X-axes are ordered to display higher acceptability on the right side. $N = 154$ for all data streams except sleep quality ($N = 87$). Solid red line represents the mean and dashed black line represents the neutral midpoint (undecided). All raw data streams had a mean significantly higher than the neutral midpoint. Interference ratings were only collected for active methods.
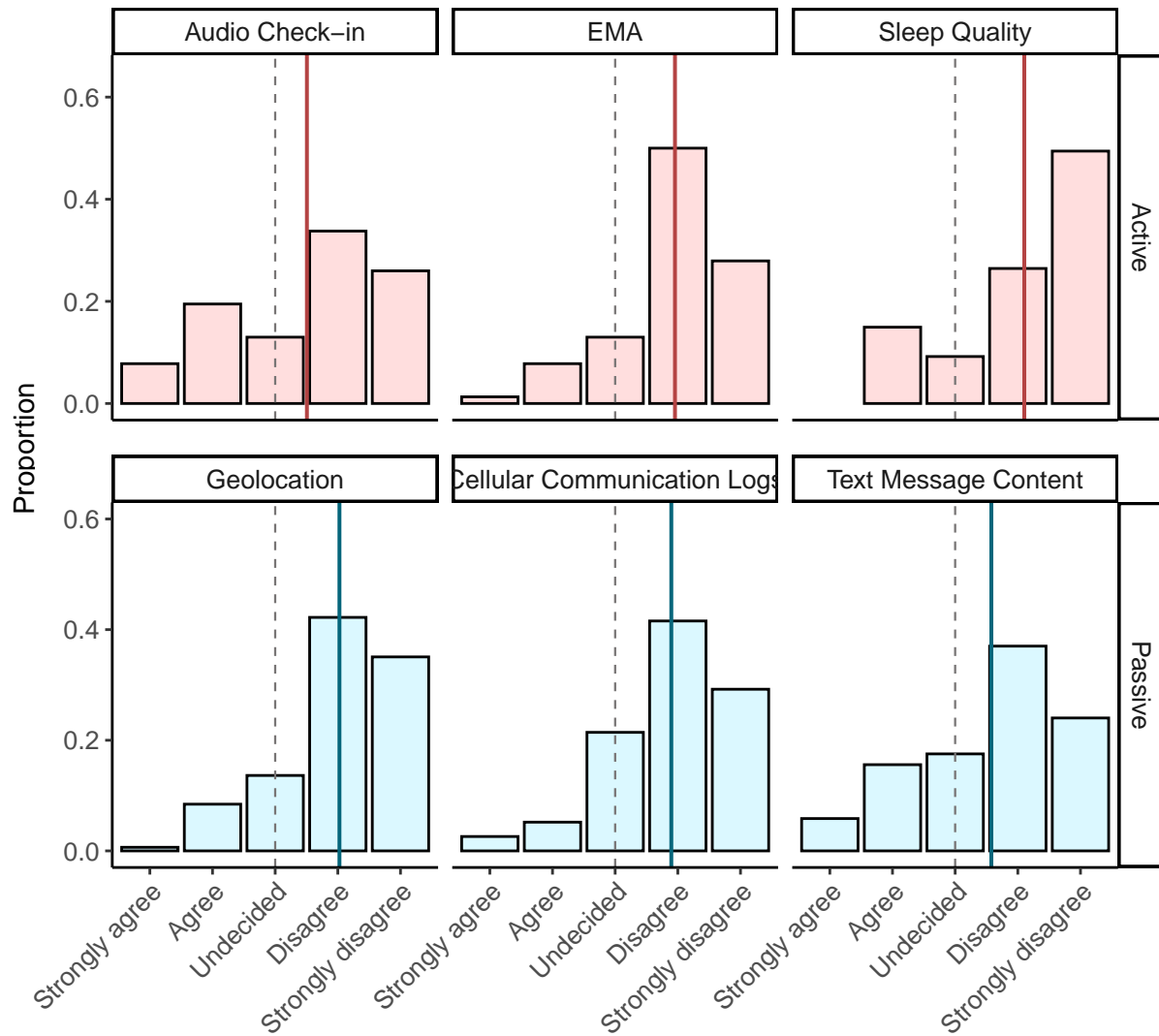
*Figure 4*. Dislike Ratings by Personal Sensing Data Stream.

Notes: Mean responses to "I disliked [Personal sensing method name]." X-axes are ordered to display higher acceptability on the right side. $N = 154$ for all data streams except sleep quality ($N = 87$). Solid red or blue line represents the mean and dashed line represents the neutral midpoint (undecided). All raw data streams had a mean significantly higher than the neutral midpoint. Active methods are displayed in red and passive methods are displayed in blue.
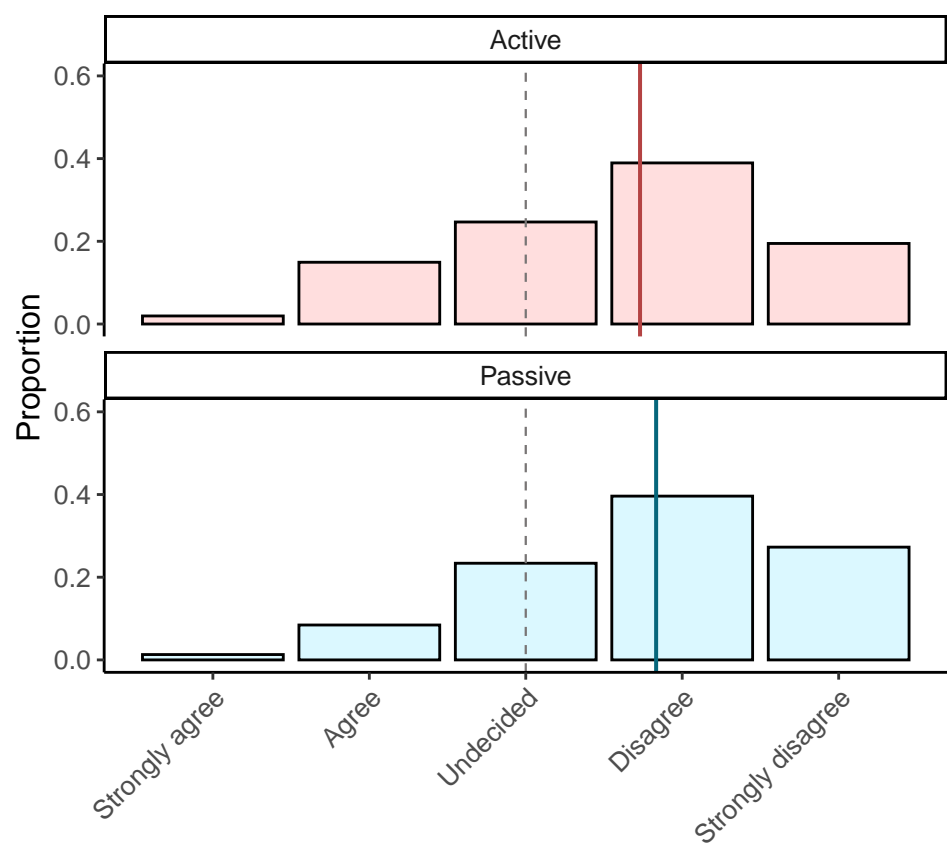
*Figure 5*. Average Dislike by Active vs. Passive Methods.

Notes: X-axes are ordered to display higher acceptability on the right side. Active methods (displayed in red) represent an average of audio check-in and EMA. Passive methods (displayed in blue) represent an average of geolocation, cellular communication logs, and text message content. Solid red or blue line represents the mean and dashed line represents the neutral midpoint (undecided). Participants did not differ significantly in their dislike of active vs. passive methods. $N = 154$.
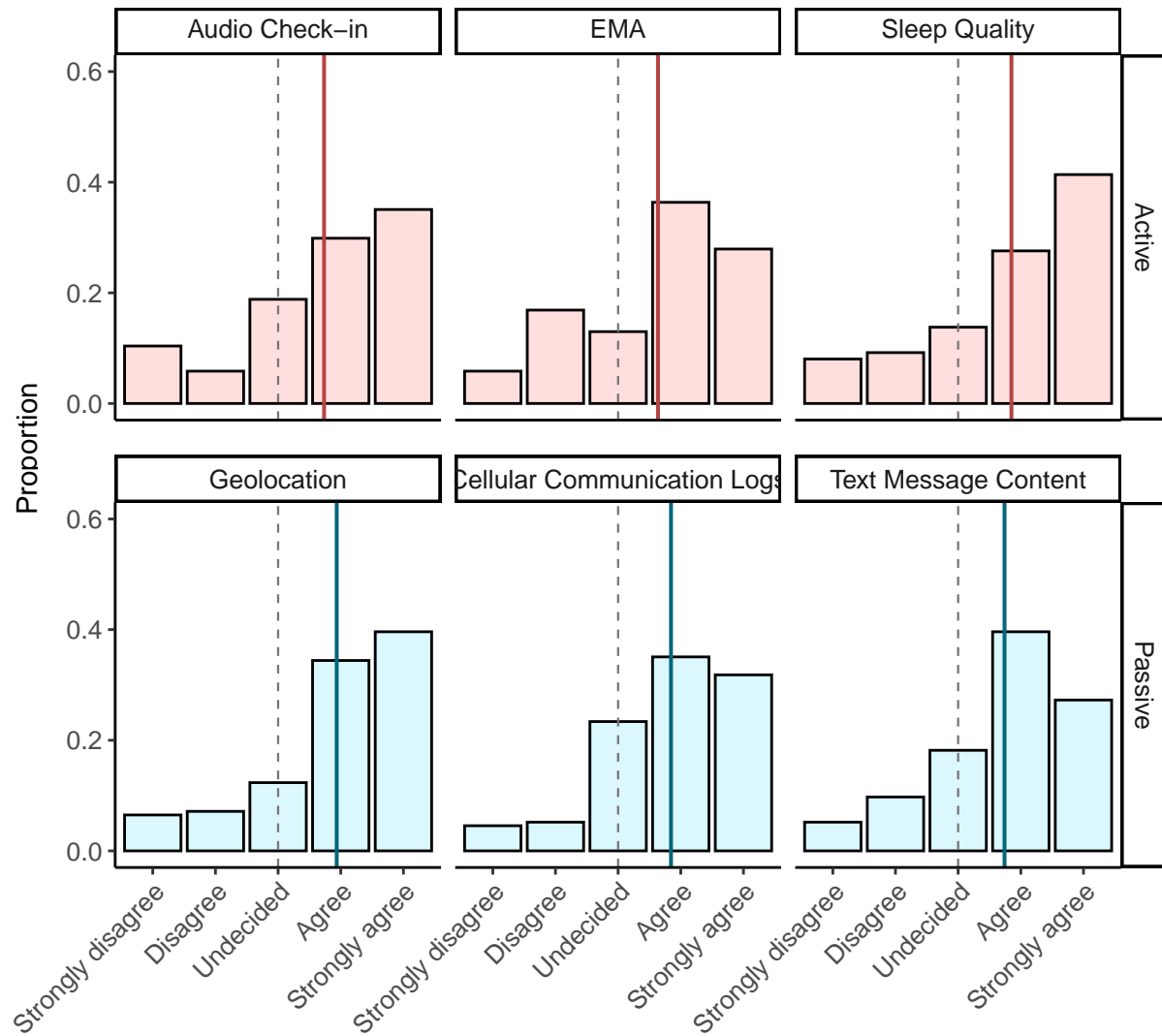
*Figure 6*. Willingness to Use for 1 Year Ratings by Personal Sensing Data Stream.

Notes: Mean responses to "I would be willing to use [Personal sensing method name] for 1 year to help with my recovery." X-axes are ordered to display higher acceptability on the right side. $N = 154$ for all data streams except sleep monitoring ($N = 87$). Solid blue or red line represents the mean and dashed line represents the neutral midpoint (undecided). All raw data streams had a mean significantly higher than the neutral midpoint. Active methods are displayed in red and passive methods are displayed in blue.
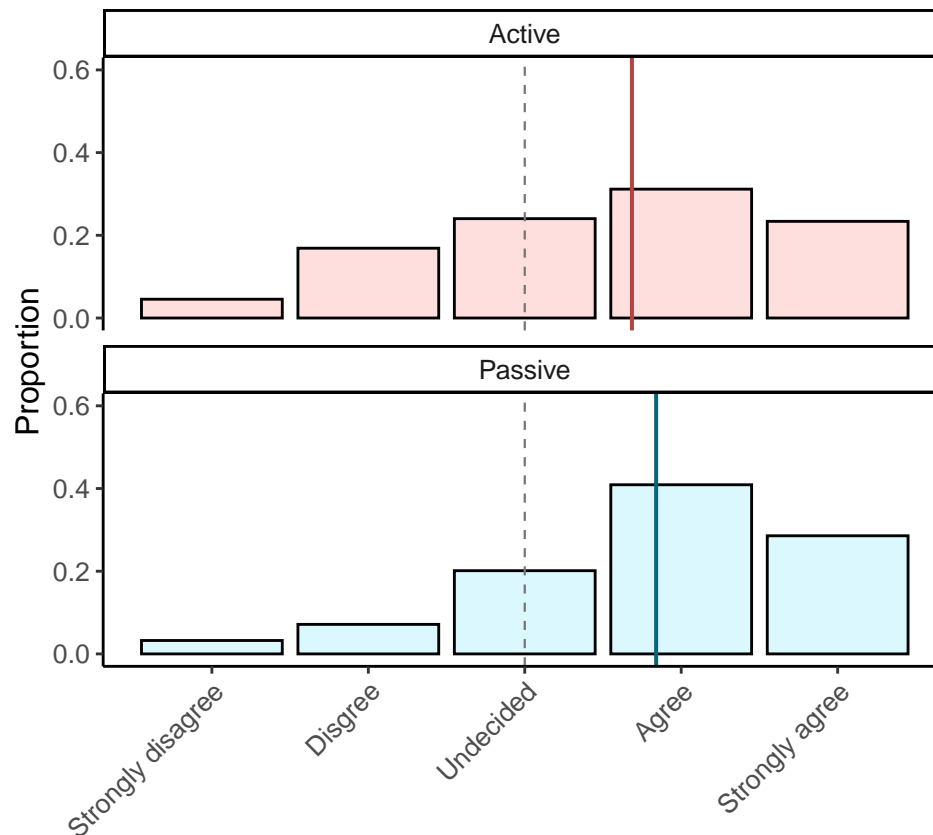
*Figure 7*. Average Willingness to Continue for 1 Year by Active vs. Passive Methods.

Notes: X-axes are ordered to display higher acceptability on the right side. Active methods (displayed in red) represent an average of audio check-in and EMA. Passive methods (displayed in blue) represent an average of geolocation, cellular communication logs, and text message content. Solid red or blue line represents the mean and dashed line represents the neutral midpoint (undecided). Participants reported on average significantly higher acceptability with respect to willingness to continue using for 1 year for passive compared to active methods. $N = 154$.

## Footnotes

[1]We measured DSM-5 symptoms with a self-report survey administered to participants during the screening visit.

[2]Psychosis and paranoia were defined as scores greater than 2.2 or 2.8, respectively, on the psychosis or paranoia scales of the on the Symptom Checklist – 90 (SCL-90) [61].

[3]Additional variables were measured as part of the parent project aims. We share the full surveys on OSF.

[4]Participants provided ratings of dislike and willingness separately for text message and phone call logs. However, participants' ratings for each item were highly correlated across the 2 logs (r = 0.83 for dislike; r = 0.79 for willingness). Furthermore, permissions for API access to text messages and phone call logs are linked for both Android and iOS operating systems such that researchers and app developers get access to both if permission is granted. Therefore, we decided to combine ratings of dislike and willingness for each of these 2 logs.

[5]Participants did not provide ratings of interference for passive signals so the comparisons of active vs. passive categories were limited to dislike and willingness to use for 1 year. Also, due to high proportion of missing data for the sleep monitor, we excluded this signal from these analyses and the intra-class correlations described next.

[6]We are missing demographic data for 1 participant who consented but did not subsequently enroll in the study.