Introduction

Personal Sensing

Personal sensing is a longitudinal method for in situ (embedded in the context of day to day lives) data collection [1,2]. It involves the process of sensing raw data streams (e.g., sensor or log data) via a personal sensing platform and forming measures (e.g., geolocation, cellular communications, physiology) to be used as indicators for constructs which can be outcomes or predictors depending on the research or clinical goals.

A personal sensing platform consists of all hardware (e.g., smartphone, wearable sensor) and software (e.g., apps, cloud storage, API access) needed for the generation and retrieval of raw data streams. A platform can be limited to sensing a single raw data stream or it can be flexible enough to capture multiple sources of data.

Personal sensing can be coarsely categorized as active or passive in terms of effort. Active methods require frequent (i.e., at least daily) interactions with the personal sensing platform and as such place more effort on the individual using the platform. Passive methods require minimal interactions (i.e., installing an app and setting permissions once) with the personal sensing platform and operate passively in the background.

Importantly, a personal sensing method can be active in different places. One example is during data generation. An EMA method is active at this point in that it requires individuals to generate the data via self-report. On the other hand, a method for sensing geolocation is passive in that people do not need to input their location. Another place where a method can be active is at the point where the researcher or clinician needs to access or retrieve the data. In this case raw data streams that are automatically uploaded to cloud storage would be considered passive, but manual uploads would be considered active.

Additionally, two personal sensing methods may sense the same raw data stream but differ in activeness. In our study, our method for sensing sleep quality and duration was active (i.e., participants had to

interact with the platform to turn it on each night and off each morning). As technology advances, we can easily imagine a sleep monitoring method that is more passive (e.g., a one-time installment of a blue tooth enabled sensor under the mattress). Thus, while the activeness of the raw data stream is relatively stable, there may be room for improving the personal sensing platform and ultimately reducing burden on the individual using the sensing platform.

Personal sensing methods are gaining popularity in the field of clinical psychology where early detection of and monitoring for mental health symptoms enables treatment and intervention [1,3]. Researchers have begun to identify measures indicative of various mental health constructs including social anxiety [4], depression [5–10], bipolar disorder [11–13], schizophrenia [14–17], and substance use [18–22] from these signals.

The potential predictive power of these methods could have direct implications for clinical settings [3,23]. Reliable identification of mental health constructs could facilitate more effective screening. Screening from passively collected data can eliminate self-report bias. Removing the need for an office visit can improve the reach of mental health screening services. Additionally, personal sensing methods can provide a continuous stream of real-time data to inform a screening outcome instead of relying on information collected during a single office visit [24].

The ability to predict the presence or absence of these constructs prospectively can also open the door for more efficient monitoring. Current treatment approaches lack the necessary resources required for long-term monitoring of symptom onset and overall mental health stability. Yet detecting a symptom lapse early on could have benefits for both an patient's prognosis [25,26] and overall cost [27,28]. Additionally, mental health disorders are not static states [29]. Someone might be stable for months or years before re-experiencing symptoms [30,31]. Integrating personal sensing methods into clinical settings may be one way to improve the reach of treatment resources and may allow for intervention prior to a full symptom relapse.

Acceptability

Acceptability of Personal Sensing among People with Alcohol Use Disorder

Using one or more personal sensing methods in clinical settings requires that people accept their use and will sustain the behaviors they require. For example, using personal sensing methods to detect increasing severity of symptoms will not work if itinterferes with people's daily lives or if people simply dislike the signal.

Two key aspects of personal sensing methods that may affect their acceptability are the sensitivity of the data being collected [32,33], and the active effort required by the individual [34,35]. Importantly, these two dimensions explicitly differ from one personal sensing measure to the next. For example, text message content monitoring is completely passive, requiring no active effort from the participant, but the data are highly sensitive. On the other hand, an EMA method may ask generic questions that are not perceived to be sensitive, but the high level of effort required (e.g., lengthy questionnaire, high prompt frequency) may be too burdensome. As a result, acceptability of personal sensing methods may be best assessed at the individual level of the measure being sensed.

Sensitivity of information may be an especially salient concern when working with clinical populations. Public opinion often holds a stigmatizing view of mental health disorders [36–39]. As a result, people may fear that their employer or peers will become aware of their diagnosis. These fears can become further compounded when illicit factors like drug use are involved. A negative opinion by neighbors or consequences at work are often cited reasons for not seeking treatment for substance use [40,41]. Thus, the type and sensitivity of data that a personal sensing method collects may be an important contributor to the acceptability of that measure.

This component of sensitivity, however, can be difficult to quantify. It is possible to vary within group and within measure. For example, it is expected that some people will be more private than others. Some people may not feel comfortable sharing any information or they may only feel comfortable sharing information under certain contexts. Additionally, a single personal sensing measure could be viewed as having different degrees of sensitivity depending on the specific information it is collecting. An EMA that asks participants about their mood compared to drug-related behaviors is likely going to be seen as less sensitive. A study using focus-groups to assess reactions to a hypothetical personal-sensing smartphone application in a sample of men who use methamphetamine and are living with HIV found that participants expressed privacy concerns with both location and self-reported drug use information citing fears of legal consequences and unintentional

disclosure of their drug use or HIV status to family and friends. [32]. Thus, the threshold of what level of sensitivity is acceptable is not easy to operationalize. Instead, these individual and group differences may be reflected as a strong dislike for a particular measure or perhaps a willingness to use the measure for several weeks as a paid research participant but an unwillingness to continue using the measure in other settings (e.g., as a recovery plan with a clinician) or for longer durations (e.g., one year).

As mentioned, personal sensing methods also differ in the level of active effort needed from the participant for data collection. Daily or more frequent self-report surveys (e.g., EMA) delivered to the participant's smartphone are active in nature, requiring substantial involvement. In contrast, background monitoring of cellular communications, location, or other sensor or log data are passive signals that are collected automatically with minimal to no engagement needed from the individual.

Not surprisingly, much of what we know about how active effort plays into the acceptability of personal sensing measures revolves around applications of EMA methods. People have been shown to be able to comply with relatively intensive EMA protocols (e.g., 5-6 daily prompts over several weeks) when the survey is brief (1-3 minutes) [35,42]. Studies assessing user experience and associated burden with EMA in both clinical and non-clinical populations suggest participants perceive EMA to be an acceptable signal [43–47]. However, these studies all assessed user experience over short periods of participation (2-4 weeks).

EMA studies that stretch over longer durations often infer acceptability from EMA compliance [48]. While this is likely a contributing factor for acceptability, on its own it may be a poor indicator. Conflicting findings have been found in studies manipulating the active effort an EMA protocol requires. For example, increases in prompt frequencies and questionnaire length have both been shown to be related to increased subjective burden, but only questionnaire length resulted in decreased compliance [35,43]. Other studies, including a meta-analysis of EMA compliance in people with substance use disorders, found that assessed burden did not influence compliance [49,50]. However, missing data from non-compliance is not the only way bias can be introduced into the data. Data provided irrespective of burden could result in hurried or careless response patterns likely due to study fatigue [35,51]. For example, when reimbursement is structured around compliance rates, participants may answer questions from memory or strategically answer items to avoid follow-up questions and shorten the length of the survey [52].

Acceptability of Personal Sensing among People with Alcohol Use Disorder

Overall, compliance tends to be low in substance using populations and even lower for people with a clinical diagnosis of a substance use disorder [44,46,50,53]. The meta-analysis of substance use studies reported a pooled compliance rate of 75% [50]. This falls below the traditionally accepted compliance threshold of at least 80% [50,54]. Additionally, persons with substance use disorders may find highly active personal sensing signals, like EMA, less acceptable over longer periods. In a four-week study where participants completed four daily EMAs to document alcohol use, researchers reported a steep drop in compliance in week four despite consistently low assessment burden [46]. These findings highlight the potential challenges of implementing personal sensing protocols in substance use populations and the need to examine personal sensing acceptability over longer durations of time.

We know far less about compliance and overall acceptability with more passive personal sensing measures. Preliminary research suggests passive personal sensing signals may be acceptable [44,55–57]. However, these studies focused on a single personal sensing measure or used aggregate acceptability ratings across multiple personal sensing measures. This makes it difficult to parse out differences across measures that may vary in active effort and sensitivity.

Other studies have created hypothetical scenarios to gauge whether participants are willing to share passively collected data and found decent acceptability rates [58–60]. These rates are limited in their informativeness, however, because participants knew the question was hypothetical. A more accurate measure would be to look at metrics of enrollment and retention in a study that uses passive personal sensing measures in addition to self-report.

Our Study

Previous work assessing acceptability have been conducted in different settings and are difficult to generalize and compare across studies. Our study examines the acceptability of various personal sensing measures individually and in the same context. That is, we present reactions to using various personal sensing measures from participants with alcohol use disorder (AUD). Participants engaged daily with various personal sensing measures that differed in both the activeness of the measure (e.g., four-time daily EMA vs. location

tracking) and the sensitivity of the information being collected (e.g., sleep log data vs. text message content) for up to three months.

Our study focuses on individuals with moderate to severe AUD. As suggested earlier, personal sensing methods offer an opportunity for long-term monitoring in real-time. People with AUD may especially benefit from such monitoring. Unfortunately, in EMA studies, compliance rates have been shown to be lower for participants who report more alcohol use overall [61]. However, little is known about other measures of acceptability or compliance rates regarding more passive measures.

Additionally, we characterize acceptability across a range of subjective and objective measures instead of relying on a single proxy. Specifically, we evaluate self-reported ratings on how much participants dislike a personal sensing measure, how much it interfered with their daily activities, and whether they would be willing to continue using the measure for a year in addition to observed behaviors related to participation and compliance.

This multi-faceted characterization of acceptability combined with the wide array of personal sensing measures assessed will expand the current literature which predominantly focuses on EMA. Our study is the first that we know of to evaluate the perceived acceptability for a series of personal sensing measures covering a broad range of the active-passive continuum from participants actually using all of the measures. These analyses may help researchers and clinicians evaluate whether certain personal sensing measures will be accepted in substance use populations.

Acceptability of Personal Sensing among People with Alcohol Use Disorder

Method

Research Transparency

We value the principles of research transparency that are essential to the robustness and reproducibility of science [62]. Consequently, we maximized transparency through several complementary methods. First, we reported how we determined our sample size, all data exclusions, all manipulations, and all available measures in the study [63]. Second, we completed a transparency checklist, which can be found in the supplement of this paper (Multimedia Appendix 1) [64]. Third, we made the data, analysis scripts and annotated results, self-report surveys, and other study materials associated with this report publicly available [https://osf.io/cjsvk].

Participants

We recruited participants in early recovery (1 – 8 weeks of abstinence) from AUD within the Madison area to participate in a three-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 AUD ($\geq$ 4 DSM-5 AUD symptoms[1]),

4) were abstinent from alcohol for at least one week but no longer than two 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]. All procedures were approved by the University of Wisconsin-Madison Institutional Review Board.

We assessed eligibility and exclusion criteria using a brief phone screen followed by a more detailed in person screening visit. One hundred and ninety-two participants met criteria 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 one week later. Fifteen participants

discontinued prior to the first follow-up visit at one month. The remaining 154 participants provided study

measures for one (N = 14), two (N = 7) or three (N = 133) months. We provide a study participation flow chart
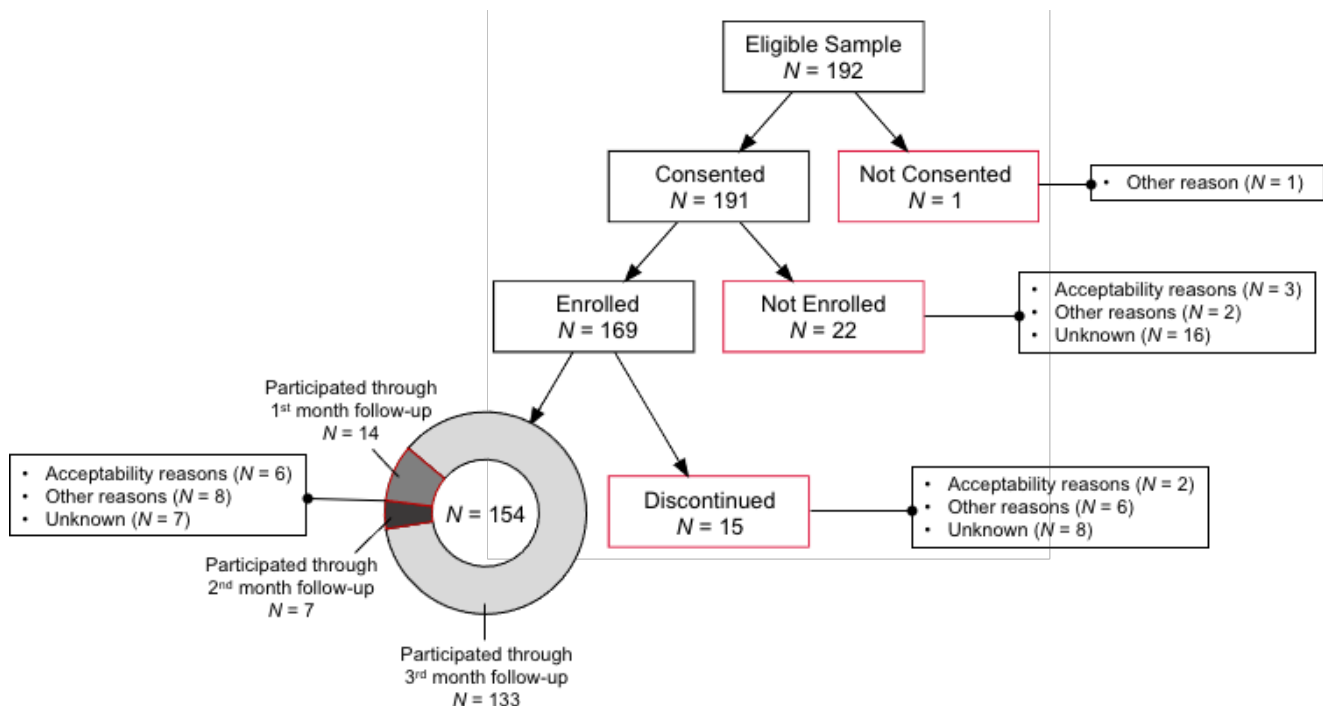
in Figure 1.

Figure 1. Flowchart of participant retention over the course of the three-month study.

We report retention and attrition of all eligible participants at various stages from consent though study completion. We

categorized reasons for attrition as due to acceptability, other reasons, or unknown. We present additional detail on

reasons for attrition in Figure 3.

Procedure

We collected the study data between 2017 – 2019 as part of a larger grant-funded parent project (RO1

AA024391). The sample size was determined based on power analyses for the aims of that project. We used

all available participants for this study. Participants completed five study visits over the course of

approximately three 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 one week later. During this enrollment visit, we collected additional self-report and interview measures. Participants completed an additional three 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, ecological momentary assessments) using personal sensing to monitor participants throughout the three-month study period. A full description of the procedure and data collected at each visit can be found at the study's OSF page [https://osf.io/cjsvk].

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 ecological momentary assessments, daily audio check-ins, sleep quality, and selected physiology. We used primarily passive methods to collect moment-by-moment 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.

Ecological Momentary Assessments (EMA). Participants completed a brief EMA four 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 four 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 three 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 one hour of participants' typical wake and sleep times. The other two 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 one 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 on only approximately 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 collected data from nine 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 2 (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 one-month 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 one-month 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 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 AUD symptoms at the screening visit (see surveys in Multimedia Appendix 1)[3].

Acceptability of Personal Sensing among People with Alcohol Use Disorder

Behavioral Acceptability. Coarse assessment of the acceptability of the personal sensing methods can be made based on participants' behaviors. Specifically, we assessed three 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. Finally, for a subset of the active measures (EMAs, daily audio check-ins), we assess their compliance with providing those raw data streams for up to three months of study participation.

Self-reported Acceptability. To assess participants' subjective experience of the acceptability of the personal sensing methods in this study, they rated each method on three acceptability relevant dimensions (see Acceptability Survey in Multimedia Appendix 1). Specifically, participants were asked to indicate how much they agreed with each of the following three statements on a five-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 one year to help with my recovery

The interference item was collected only for the active methods because they passive methods require no effort and therefore cannot interfere with daily activities. Dislike and willingness to use for one year were collected for all methods.

Data Analytic Strategy

We conducted all analyses in R version 4.0.3 [66] using RStudio [67] and the tidyverse ecosystem of packages [68].

Behavioral 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).

Acceptability of Personal Sensing among People with Alcohol Use Disorder

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 compliance measures for two of the active personal sensing methods (EMAs and daily audio check-ins). Formal measures of compliance could not be calculated for geolocation, cellular communication log, and sleep quality because it was not possible to distinguish between missing data due to compliance (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 Acceptability. Participants responded to the three self-report items related to acceptability (interference, dislike, and willingness to use for one year) on a five-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 one year and strongly disagree for interference and dislike. For analyses, we recoded 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 three 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 2).

To detect polarized perceptions of the personal sensing signals (i.e., mean responses to any items that are different from 0/undecided), we conducted one-sample t-tests for the three 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 one year[5]. We also report pairwise comparisons among all personal sensing signals using within-sample t-tests for each of the three self-report items in Table S2 in Multimedia Appendix 2.

Finally, we conducted two analyses to examine the consistency of perceptions across personal sensing signals (e.g., do participants who dislike one signal also dislike the other signals?). First, we calculated bivariate correlations among the personal sensing signals for each item. Second, we calculated intra-class correlations (single, case 3 [69]) separately for each item to quantify aggregate agreement in participant perceptions across the signals.

Results

Participant Characteristics

One hundred and fifty-four participants completed at least one monthly follow-up visit and provided self-report acceptability ratings for interference, dislike, and willingness to use for one year. These participants serve as our primary sample for our analyses. Table 1 presents demographic information these participants. Table 2 characterizes alcohol use and AUD-relevant information for these participants. We compared demographics and AUD 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 Multimedia Appendix 2 for more detail on these analyses).

Acceptability of Personal Sensing among People with Alcohol Use Disorder

Table 1

Demographic Characteristics for the Sample

	%	*M* (*SD*)
Age		41.0 (11.9)
Sex		
Female	50.0%	
Male	50.0%	
Race		
American Indian/Alaska Native	1.9%	
Asian	1.3%	
Black/African American	5.2%	
White	87.0%	
Multiracial or not listed	4.5%	
Hispanic, Latino, or Spanish Origin		
Yes	2.6%	
No	97.4%	
Education		
Less than high school or GED degree	0.6%	
High school or GED	9.7%	
Some college	27.9%	
2-Year degree	9.1%	
College degree	37.7%	
Advanced degree	14.9%	
Employment		
Employed full-time	46.8%	
Employed part-time	17.5%	
Full-time student	4.5%	
Homemaker	0.6%	
Disabled	4.5%	
Retired	5.2%	
Unemployed	12.3%	
Temporarily laid off or on leave	1.9%	
Other, not otherwise specified	6.5%	
Income		34,233 (31,543)
Marital Status		
Never married	44.8%	
Married	21.4%	
Divorced	29.2%	
Separated	3.2%	
Widowed	1.3%	

Note: *N* = 154 participants

Table 2

Alcohol Related Characteristics for the Sample

	%	M (SD)
AUD 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 (more than 1 possible)		
Long-term residential (6+ mos.)	5.2%	
Short-term residential (< 6 mos.)	33.1%	
Outpatient	50.0%	
Individual counseling	64.9%	
Group counseling	42.2%	
Alcoholics Anonymous/Narcotics Anonymous	62.3%	
Other	26.6%	
Received Medication for AUD		
Yes	40.3%	
No	59.7%	
AUD DSM-5 Symptom Count		8.9 (1.9)
Lifetime Drug Use		
Tobacco products (cigarettes, chewing tobacco, cigars, etc.)	79.2%	
Cannabis (marijuana, pot, grass, hash, etc.)	85.1%	
Cocaine (coke, crack, etc.)	55.8%	
Amphetamine type stimulants (speed, diet pills, ecstasy, etc.)	52.6%	
Inhalants (nitrous, glue, petrol, paint thinner, etc.)	23.4%	
Sedatives or sleeping pills (Valium, Serepax, Rohypnol, etc.)	46.8%	
Hallucinogens (LSD, acid, mushrooms, PCP, Special K, etc.)	57.1%	
Opioids (heroin, morphine, methadone, codeine, etc.)	42.2%	

Note: N = 154 participants

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 one did not consent after hearing the details of the study. Enrollment occurred during a second visit one 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 participants (5.7%) were lost due to acceptability-relevant causes (e.g., no longer interested, non-compliance with data, 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 thirty-one 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 Compliance. All participants (100%) opted-in to provide data for all personal sensing methods. Daily compliance rates were relatively high for EMAs such that 94% of participants completed at least one of the four EMAs every day. On average, participants completed 3.1 EMAs every day. The overall compliance rate for all requested EMAs was 77%. Participants' completion rate for the daily audio check-in was 53%. That is, of their total days on study, participants completed an audio check-in on approximately half of them.

Table 3

Reasons for Discontinuation…

A. Prior to Completing Enrollment

	N	%
Health concerns	1	0.05
No longer has transportation	1	0.05
No longer interested	**3**	**0.14**
Rescheduled multiple times before cancelling/no showing	6	0.29
Unknown	10	0.48
Total	21	1.00

Note: These participants are depicted as "Not Enrolled" in Figure 1. Bolded rows represent acceptability-related discontinuation.

B. Prior to Completing First Month Follow-up

	N	%
Family crisis	1	0.06
No longer sober or no longer wishes to abstain from alcohol	5	0.31
No longer interested	**1**	**0.06**
Noncompliance with providing data	**1**	**0.06**
Rescheduled multiple times before cancelling/no showing	4	0.25
Unknown	4	0.25
Total	16	1.00

Note: These participants are depicted as "Discontinued" in Figure 1. Bolded rows represent acceptability-related discontinuation.

C. After Completing First Month Follow-up

	N	%
Cell service shut off	3	0.14
Cited study demands as too burdensome	**1**	**0.05**
No longer sober or no longer wishes to abstain from alcohol	3	0.14
Moved out of state	2	0.10
No longer interested	**3**	**0.14**
Noncompliance with providing data	**2**	**0.10**
Rescheduled multiple times before cancelling/no showing	4	0.19
Unknown	3	0.14
Total	21	1.00

Note: These participants are depicted as "Participated through 1st month follow-up" or "Participated through 2nd month follow-up" in Figure 1. Bolded rows represent acceptability-related discontinuation.

Self-reported Acceptability

      Interference. Figure 2 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 3 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 4). 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 different significantly in their dislike of active vs. passive methods, $t(153) = 1.21$, $p = .23$, $d = .10$.

      Willingness to Use for One Year. Figure 5 shows the distribution of participant responses to the self-reported acceptability item about willingness to use for one year for each personal sensing data stream. 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].

Acceptability of Personal Sensing among People with Alcohol Use Disorder

We also assessed the effect of active effort on willingness ratings (Figure 6). We conducted a paired samples t-test of average the average willingness to use for one 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 = 0.84, SD = 1.0) relative to active data streams (M = 0.69, SD = 1.1), $t(153)$ = 2.12, p = .04, d = .17.

Acceptability of Personal Sensing among People with Alcohol Use Disorder
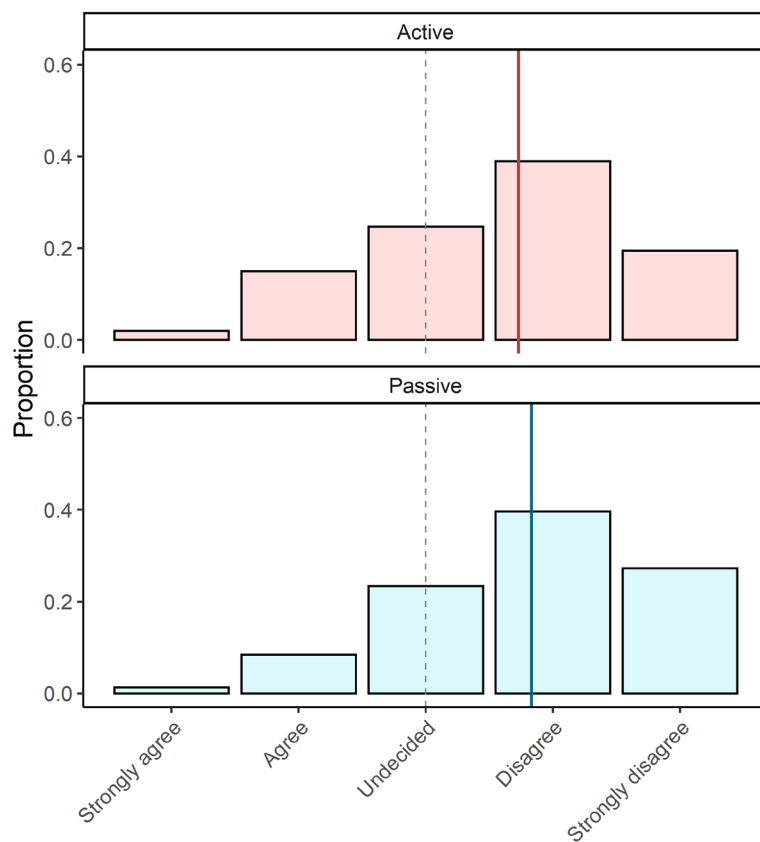
Figure 2

Interference Ratings by Personal Sensing Data Stream
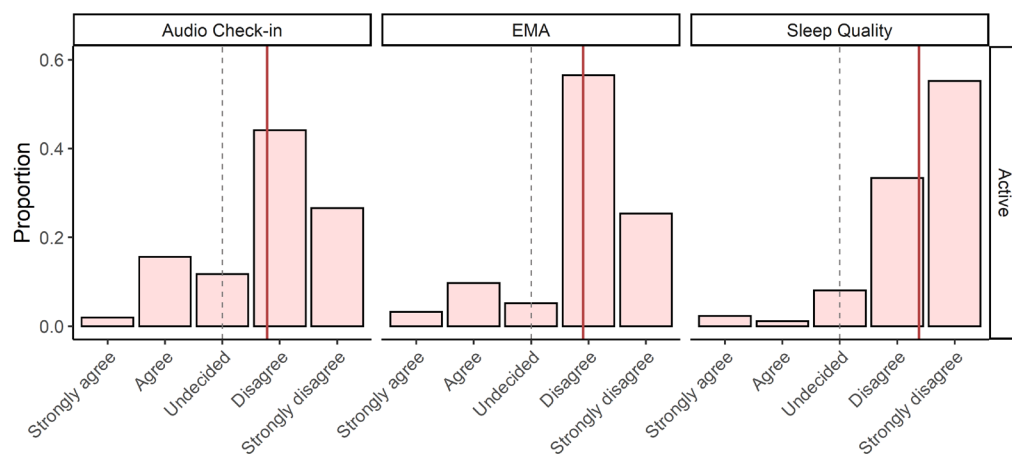
Notes: X-axes are ordered to display higher acceptability on the right side. N = 154 for all data streams except sleep quality (*N* = 87). Solid red line represents the mean and dashed black line represents the neutral midpoint (undecided). All raw data streams had a mean significantly higher than the neutral midpoint. Interference ratings were only collected for active methods.

Acceptability of Personal Sensing among People with Alcohol Use Disorder

Figure 3

Dislike **Ratings by Personal Sensing Data Stream**
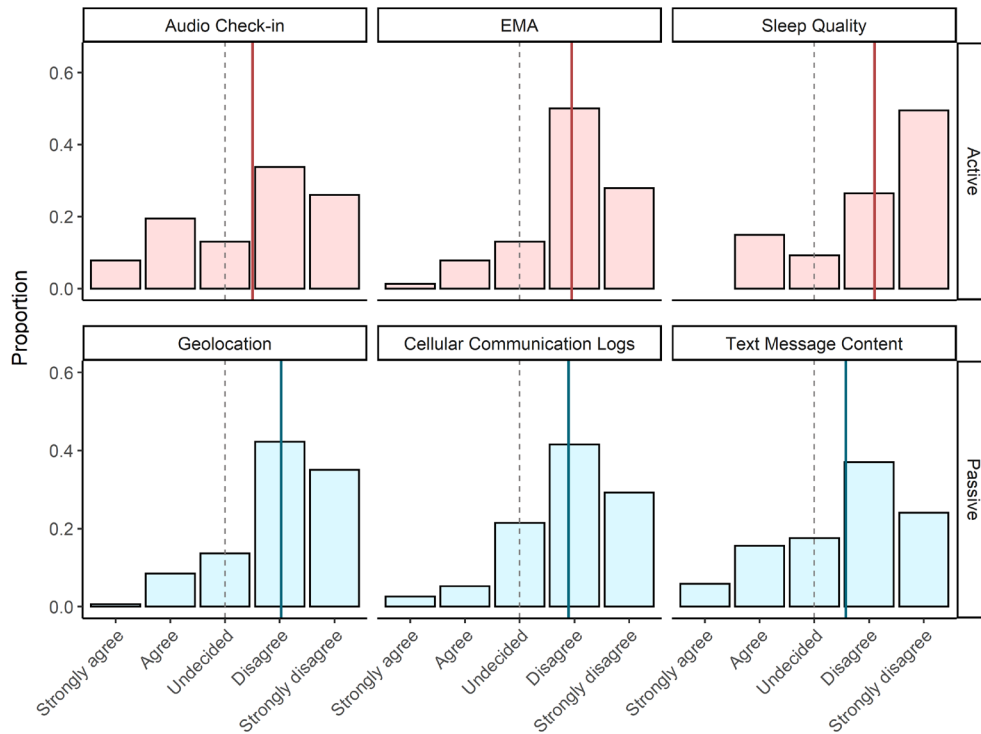
Notes: X-axes are ordered to display higher acceptability on the right side. N = 154 for all data streams except sleep quality (*N* = 87). Solid red or blue line represents the mean and dashed line represents the neutral midpoint (undecided). All raw data streams had a mean significantly higher than the neutral midpoint. Active methods are displayed in red and passive methods are displayed in blue.

Acceptability of Personal Sensing among People with Alcohol Use Disorder
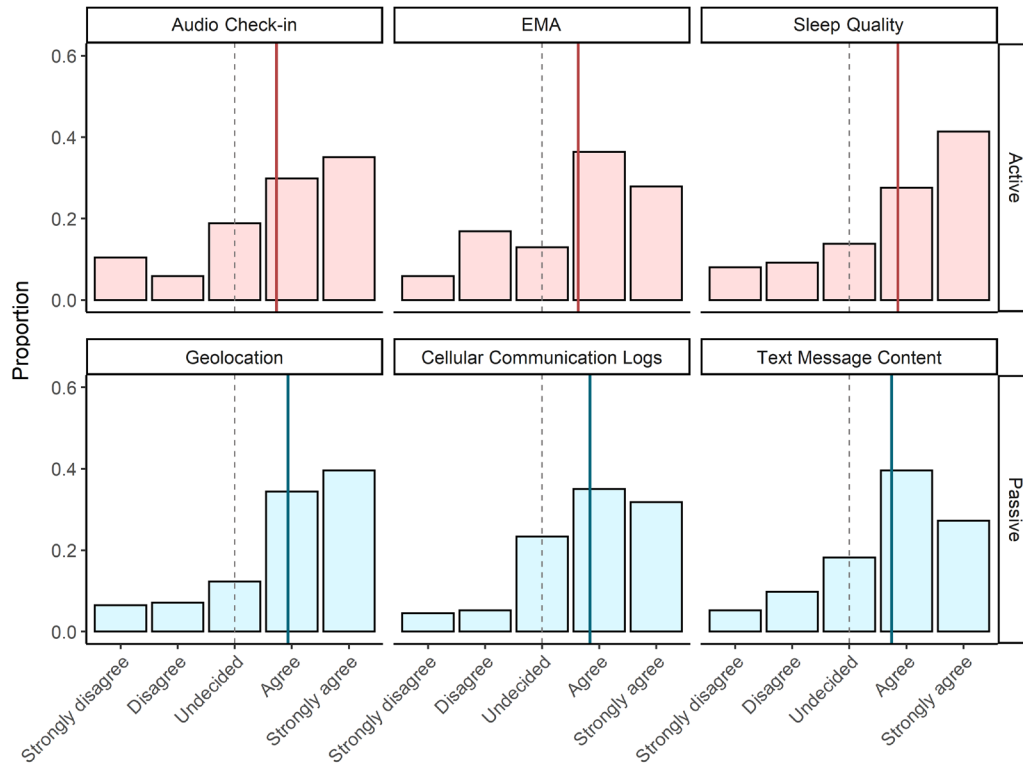
Figure 4

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 different significantly in their dislike of active vs. passive methods. N= 154. Active methods are displayed in red and passive methods are displayed in blue.

Acceptability of Personal Sensing among People with Alcohol Use Disorder

Figure 5

Willingness to Use for One Year **Ratings by Personal Sensing Data Stream**

Notes: 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 6

Average *Willingness to Continue for One Year* by Active vs. Passive Methods
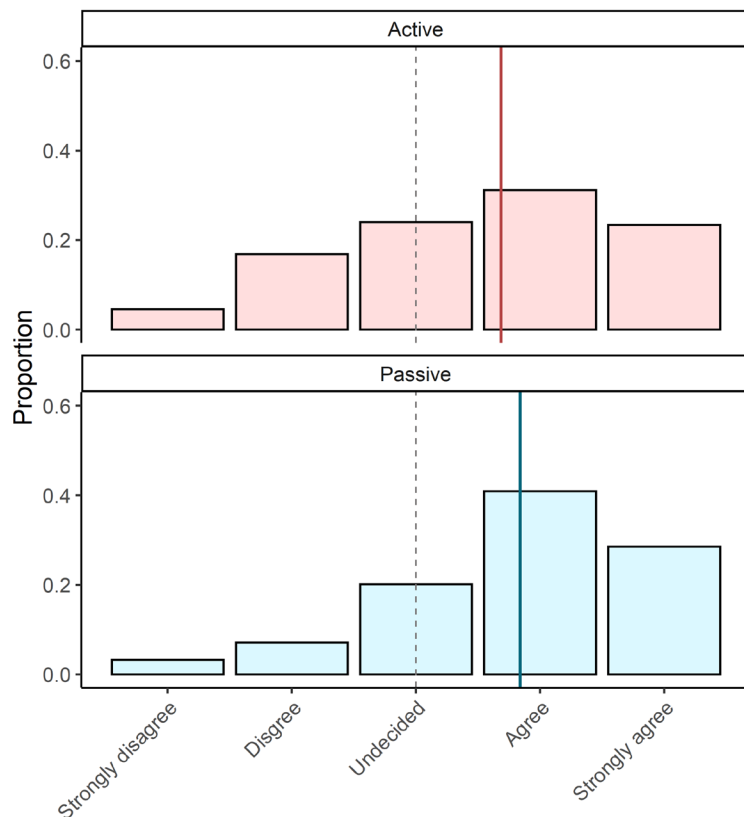
Notes: X-axes are ordered to display higher acceptability on the right side. Active methods (displayed in red) represent an average of audio check-in and EMA. Passive methods (displayed in blue) represent an average of geolocation, cellular communication logs and text message content. Solid red or blue 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 one year for passive compared to active methods. N= 154. Active methods are displayed in red and passive methods are displayed in blue.

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	0.78	1.07	9.05*	0.73
2. EMA	.42*	-	154	0.91	0.99	11.37*	0.92
3. Sleep Quality	-.06	.24*	87	1.38	0.87	14.86*	1.59

Notes: 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 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	0.51	1.28	4.91*	0.40
2. EMA	.57*	-				154	0.95	0.92	12.92*	1.04
3. Sleep Quality	.25*	.26*	-			87	1.10	1.09	9.45*	1.01
Passive										
4. Geolocation	.31*	.36*	.41*	-		154	1.03	0.94	13.51*	1.09
5. Cellular Communication Logs	.22*	.25*	.33*	.67*	-	154	0.90	0.97	11.45*	0.92
6. Text Message Content	.34*	.28*	.19	.62*	.69*	154	0.58	1.18	6.07*	0.49

Notes: 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 One Year* by Personal Sensing Data Stream

	1	2	3	4	5	*N*	*M*	*SD*	*t*	*d*
Active										
1. Audio Check-in	-					154	0.73	1.28	7.09*	0.57
2. EMA	.44*	-				154	0.64	1.22	6.47*	0.52
3. Sleep Quality	.44*	.41*	-			87	0.85	1.28	6.19*	0.66
Passive										
4. Geolocation	.51*	.48*	.53*	-		154	0.94	1.18	9.83*	0.79
5. Cellular Communication Logs	.47*	.47*	.47*	.65*	-	154	0.84	1.07	9.76*	0.79
6. Text Message Content	.47*	.39*	.37*	.60*	.84*	154	0.74	1.12	8.21*	0.66

Notes: Initial columns represent bivariate correlations among signals. Final columns represent the number of participants (*N*), mean acceptability score (*M*), standard deviation (*SD*), and t statistic (*t*) and Cohen's d Effect size (*d*) for a one sample t-test centered around a neutral/ambivalent score of 0 on a 5-point bipolar scale. Higher values represent higher acceptability. Active methods are displayed in red and passive methods are displayed in blue.

* *p* < .05

Discussion

Behavioral Acceptability

Overall, participants showed behaviors consistent with acceptance of our personal sensing methods. Only one participant did not consent to the study and all enrolled participants agreed to provide us their raw data streams for each of the actively (EMA, audio check-in, sleep) and passively (geolocation, cellular communication logs, text message content) sensed measures.

Participants were willing to opt-in to all passive methods despite the sensitive nature of the raw data streams being collected. Participants concerned about sharing passively sensed data would likely have had these concerns from the beginning (i.e., resulting in not consenting or not opting-in to sensitive personal sensing methods). As a result, any drop-out is likely more indicative of burden from personal sensing methods higher in active effort then from measures high in sensitivity.

Behavioral acceptability was a bit more nuanced for our active methods. Compliance differed based on active effort and type of personal sensing measure. Participants had low overall compliance for the audio check-in. If this type of personal sensing method is desirable, researchers and clinicians may find better compliance with less frequent prompts (i.e., less than daily). Participants did moderately well with overall compliance for the four-time daily EMA (77%). This falls below the 80% recommended compliance rate but is quite typical when working with substance use populations [50,54]. We also note that our study is longer than the average EMA study (spanning three months instead of 2-4 weeks) and this could account for a lower compliance rate. As a result, we can imagine compliance being even lower if participants were asked to provide EMA data for one year. Participants were able to comply with one-time daily EMA at high rates (94%). Thus, fewer prompts may be needed to maintain high compliance rates over longer durations.

Self-reported Acceptability

With respect to the self-reported measures of acceptability (interference, dislike, willingness to use for one year), all personal sensing methods were rated to be on the acceptable side of neutral. So, it appears

participants found these methods to be acceptable (or at least not unacceptable). More importantly, these findings suggest participants are willing to use these methods for up to a year if it is helping their recovery.

We found mild differences in self-reported acceptability for active and passive personal sensing methods. Active and passive methods were viewed equally acceptable in terms of disliking the methods. For willingness to use the methods for one year, we found a significant, but weak, effect. Participants reported they were more willing to use passive methods. This difference is not surprising. Methods requiring high active effort from the participant can be burdensome making it difficult to maintain long-term engagement with the personal sensing platform.

This self-reported difference in willingness to use active and passive methods was further reflected in compliance. Despite participants self-reporting the audio check-in to be acceptable, we found poor daily compliance. Likewise, we found EMA compliance to much higher if the threshold was lowered to a one-time daily method. Additionally, when asked about a one-time daily method, participants self-reported that they would much more likely to use a one-time daily EMA method compared to a four-time EMA method for one year.

Finally, across active and passive measures, we did see some agreement in terms of self-reported acceptability. ICC scores ranging from .42 - .52 suggest some shared variance across measures. However, there was also some individual variation from one measure to the next.

Active Effort and Sensitivity

We labeled our personal sensing methods as active or passive based on the amount of active effort our specific methods required. Small changes in a personal sensing method can shift where it falls on this active-passive continuum. For example, an EMA method will never be passive but by altering the length or prompt frequency it may become more or less active. On the other hand, our measure of sleep was on the low side of the active continuum, but with advances in technology passive methods for sensing sleep duration and quality now exist. Thus, our labels may not necessarily generalize to other methods collecting the same types of raw data streams.

Acceptability of Personal Sensing among People with Alcohol Use Disorder

We determined it would not be appropriate to group methods by where they fall on the high-low sensitivity continuum. Sensitivity cannot be defined by the personal sensing method alone. It is highly contextual depending on factors such as characteristics and preferences of the person providing their data, who is receiving the data (e.g., researcher, clinician, health insurance), and what it is being used for [70]. It is also dynamic so that a certain set of circumstances may be acceptable for sharing sensitive information at one point (e.g., when someone is stable in their recovery), but not at a later point (e.g., when someone is in the middle of a series of lapses or a relapse).

In our own sample we saw individual variation in how participants felt about the sensitivity of their data in free-response comments.

"I don't mind it at all. It's sort of nice to know I'm being monitored" [participant 047].

"I am concerned with data privacy and felt uncomfortable having my location actively tracked" [participant 196].

Individuals may experience privacy differently in their day to day lives. People belonging to historically marginalized groups may find it more difficult to achieve privacy in their daily life [71]. Relatedly, our participants are in recovery for AUD. For some, there may be high costs (e.g., legal consequences, or loss of job, relationships, or health insurance) if information related to their recovery (e.g., frequent lapses) gets in the wrong hands. As a result, personal experiences may affect how much one values privacy and as a result whether one would be willing to provide sensitive data streams.

Researchers and clinicians should consider these factors when designing their personal sensing method. For example, in theory the one-time daily audio check-in that could be completed at any point during the day and only had to be 30 seconds in length should require less effort then a four-time daily EMA that was to be completed when prompted. When examining our participants' free-text comments about their experience with the different methods, we found that many said it was difficult to find a private place to record their audio check-in.

"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" [participant 018].

Acceptability of Personal Sensing among People with Alcohol Use Disorder

"Sometimes I wouldn't be able to do it because I was with people too much which was

frustrating" [participant 037].

This lack of privacy compounded with the nature of their check-in (i.e., their recovery) likely contributed to the poor daily compliance. So, if obtaining a free response commentary (i.e., the content of the audio check-in) each day was crucial to our research goals it might have been in our better interest to provide participants with an alternate response format (e.g., typed text). On the other hand, if it was more important to get data related to vocal inflection, pace, and other information extracted from audio data, a less than daily prompt with reminders might be more effective.

Theoretical Framework

Whether someone finds a personal sensing method acceptable ultimately comes down costs, benefits, and trust [34,72,73]. Our study only looks at acceptability related to cost (time and privacy). However, it is important that we also consider benefits and trust.

Benefits. Participants in our study received minimum benefits. Benefits included financial compensation for their time and possible gratification for their contribution to a risk prediction algorithm with the aim to help other people in recovery for AUD and from the self-reflection required for some of the daily tasks (audio check-in, EMA).

Research suggests that privacy concerns about personal sensing methods may be mitigated by the perceived value of the sensing platform [2,34]. Free response comments from our participants support this.

"As long as my personal information was secure and only shared by study staff, and its

collection could help in my recovery, I'm comfortable with it" [participant 189].

"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" [participant 016].

As a result, individuals may find the personal sensing methods more acceptable if they were also receiving treatment through the application or receiving other benefits from using the application (e.g., resources, personalized notifications).

Trust. In our study, trust levels were high. We had several protections in place to keep participants' data confidential and we thoroughly explained them during the consent process. Additionally, participants came into our lab for their monthly visit where they often built report with a limited number of study staff. Participant comments suggest this trust influenced their perceived acceptability of our sensing methods.

> *"Invasive in theory, no problems actually because of confidentiality" [participant 193].*

> *"I trusted the study group to not use my personal information for any other use" [participant 166].*

It is not surprising that trust plays an integral role in a person's acceptance of these methods. We know form the literature who people are sharing their personal sensing data with matters [33,34,60,74,75]. Focus group studies suggest that people are most comfortable sharing data anonymously with researchers and with their doctor, and least comfortable sharing information with family members, electronic health record databases, and third party apps and websites [33,74,75]. This suggests trust is most closely related to trust that the information will be kept confidential.

Ethics

Acceptability of personal sensing methods should not be taken as minimizing the sensitive nature of these data. In fact, we use the term personal sensing to make explicit how personal and sensitive these data are [1]. Personal sensing has the potential to seep into almost every aspect of an individual's world. Research and applications using personal sensing methods must only be done with appropriate protections in place.

Our participants generally felt comfortable providing these data to us. We had the strictest possible measures in place to keep their data safe and as a result they trusted us. One of these safeguards, a certificate of confidentiality, protects research participants by preventing forced disclosure of any identifying information. However, this is unique to the research context. Implementing personal sensing platforms in applied settings could potentially put the end user at risk [76]. Companies and third-party applications could be subpoenaed and required to disclose sensitive information to be used as ammunition in a custody case or as evidence in a criminal trial. Likewise, a company might be providing a patient's healthcare data to both their provider and their health insurance company resulting in higher insurance premiums or dropped coverage.

This risk is not lost on the people being asked to provide their data and control over who can access their data and unauthorized secondary use of that information are often cited concerns [34,70].

The current study assesses the acceptability of personal sensing methods that start with the collection of raw data streams and end with indicators for constructs. The likely next step is to use these indicators as inputs into a statistical algorithm to create a predictive machine learning model. Predictive models built on personal sensing data can be valuable in clinical areas such as risk prediction, mental health screening and diagnosis, and treatment matching. However, to benefit the patient it is critical that these models are trained on large and diverse samples of participants.

Controlling who gets access to personal sensing data, preventing conflicts of interest, and creating equitable and fair machine learning models are only some of the ethical concerns that must be addressed before personal sensing methods can be implemented.

Feasibility

Acceptability of personal sensing methods is only the first step in using such methods in applied settings. It must also be feasible to implement a personal sensing platform or series of personal sensing methods. In other words, it must be possible to do in a convenient, realistic, and sustainable way. Feasibility of applying personal sensing methods can be broken down into feasibility for the participant, client, or patient and feasibility for the researcher, provider, or clinician.

Participant, client, or patient. The primary feasibility concern on the user side of personal sensing is the ability to access the personal sensing platform and implement it into their daily routine.

The minimum technology needed for access to a personal sensing platform (e.g., an app) is a smartphone. While this may have been a large financial burden in the past, today there are more affordable options. Smartphone ownership is estimated to be at about 81% of the population and is not much lower in substance use populations and low socio-economic groups [77,78].

Not only do people generally have the requisite technology needed to access personal sensing platforms, but they are also willing to sustain interaction with personal sensing methods for long durations of time. We found participants accepted the personal sensing methods and were able to incorporate them into their daily lives with little interference. Additionally, most participants were willing to use these methods for one

year to help their recovery. This is valuable in that they are reporting this after already having used these methods daily for about three months. So, they are using these methods and willing to continue using them for longer periods of time. This potential for sustained engagement with a personal sensing platform is promising if we are to ultimately use such methods in the real world.

      Researcher, provider, or clinician. Feasibility concerns on the data collection end revolve mostly around the technology. These include technical difficulties, expense, and churn with companies.

      We started out with three platforms (physiology wristband, sleep monitor, and study iPhone) and experienced varying levels of technological difficulties with each. Most notably we had to stop providing the Empatica E4 physiology wristband to participants almost immediately because it was too difficult to retrieve the data. The physiology wristband did not support Bluetooth streaming of data to the cloud and participants were required to manually upload the data each day on a study tablet. This process soon proved to be too burdensome, and it was too difficult to trouble shoot software problems that came up during the data upload process.

      Expense must also be considered when evaluating feasibility of personal sensing methods. Had we continued using the physiology wristband we would have been providing participants with two pieces of hardware (wristband and tablet) to collect a single data stream. Early participants were also offered a study iPhone if they did not have an eligible smartphone (Android or IOS). This can be expensive and discouraging. However, as technology improves, we see how more can be done with a smartphone as a single all-encompassing personal sensing platform. Additionally, many people now own a smartphone. We stopped offering study smartphones because so many people already had one. Only six participants were given a study iPhone and five participants were screened out for not having an eligible phone. It then seems that personal sensing technology is becoming more accessible and affordable as time goes on.

      One unexpected issue we faced was the high churn rate of companies. In just our two-year data collection period, we had to discontinue use of sleep monitor platform and our geolocation tracking app. The company that ran the sleep monitor platform, Beddit Oy, was bought out by Apple about halfway through the study. Apple discontinued cloud support and we were unable to continue accessing sleep monitor data. Additionally, the app we were using for geolocation tracking (Moves) was acquired and shut down by Facebook. As a result, we had to switch to a completely different app (FollowMee) mid-study.

While this churn is difficult in the middle of data collection, it does signal the fast rate of technology development and improvement. Some of our active measures, like sleep and physiology can now be done much more passively. Free response comments from participants about using the sleep monitor illustrate how a seemingly passive raw data stream (sleep) can still be active.

"Would be better if it would just connect automatically" [participant 021].

"There were problems with the sleep monitor in that I was not able to activate the monitor when I was going to sleep" [participant 001].

"Sometimes it recorded all night but when I pressed I'm up there was no data recorded" [participant 053].

"Would have to wake up to use" [participant 056].

Limitations and Future Directions

There are many underlying factors that influence people's view of personal sensing measures. Highly active measures may only be tolerable within short time frames. Personal data may differ in their perceived sensitivity from individual to individual and group to group. As such our study offers an initial insight into the acceptability of personal sensing methods. Future studies should use longer study durations (i.e., one year) and recruit nationally representative samples.

Our study also only generalizes to our specific personal sensing methods. Although we do not expect general self-reported acceptability about the types of raw data streams to change, different methods may increase or reduce user burden resulting in shifts of behavioral acceptability (i.e., compliance and retention). For example, participants were more willing to use passive personal sensing measures for one year compared to active measures. In clinical applications we may expect individuals to be engaging with these methods for years. Thus, it is important to test newer and more passive personal sensing methods as they become available.

Finally, acceptability is not limited to perceived costs but is an exercise of weighing costs, benefits, and trust. Future research should directly measure benefits (e.g., with self-report) and manipulate how the data is benefiting the participant (e.g., incorporating it into treatment). Trust should also be examined by applying this

research to other contexts (e.g., who is receiving the data, what level of relationship exists between participant and data collector). It is critical we understand the entire framework of personal sensing acceptability if we are to implement these methods in real treatment settings.

Conclusion

References

1. Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. Annu Rev Clin Psychol 2017 May 8;13(1):23–47. [doi: 10.1146/annurev-clinpsy-032816-044949]

2. 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 Comput [Internet] Berlin, Heidelberg: Springer Berlin Heidelberg; 2009 [cited 2021 Jan 3]. p. 176–183. [doi: 10.1007/978-3-642-01516-8_13]

3. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. Npj Digit Med 2019 Dec;2(1):88. [doi: 10.1038/s41746-019-0166-1]

4. Jacobson NC, Summers B, Wilhelm S. Digital Biomarkers of Social Anxiety Severity: Digital Phenotyping Using Passive Smartphone Sensors. J Med Internet Res 2020 May 29;22(5):e16875. [doi: 10.2196/16875]

5. 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 UHealth 2021 Jan 6;9(1):e19046. [doi: 10.2196/19046]

6. Razavi R, Gharipour A, Gharipour M. Depression screening using mobile phone usage metadata: a machine learning approach. J Am Med Inform Assoc 2020 Apr 1;27(4):522–530. [doi: 10.1093/jamia/ocz221]

7. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, Asch DA, Schwartz HA. Facebook language predicts depression in medical records. Proc Natl Acad Sci National Academy of Sciences; 2018 Oct 30;115(44):11203–11208. PMID:30322910

8. 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 UHEALTH 2021;15.

9. Reece AG, Danforth CM. Instagram photos reveal predictive markers of depression. EPJ Data Sci 2017 Aug 8;6(15):1–12. [doi: 10.1140/epjds/s13688-017-0110-z]

10. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. Proc SIGCHI Conf Hum Factors Comput Syst - CHI 13 [Internet] Paris, France: ACM Press; 2013 [cited 2018 Aug 30]. p. 3267. [doi: 10.1145/2470654.2466447]

11. Palmius N, Tsanas A, Saunders K, Bilderbeck A, Geddes J, Goodwin G, de Vos M. Detecting Bipolar Depression From Geographic Location Data. IEEE Trans Biomed Eng 2016 Oct 25;PP:1–1. [doi: 10.1109/TBME.2016.2611862]

12. Grünerbl A, Muaremi A, Osmani V, Bahle G, Ohler S, Troester G, Mayora O, Haring C, Lukowicz P. Smart-Phone Based Recognition of States and State Changes in Bipolar Disorder Patients. IEEE J Biomed Health Inform 2014 Jul 25;19. [doi: 10.1109/JBHI.2014.2343154]

13. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: status and considerations. Int J Bipolar Disord [Internet] 2018 Jan 23 [cited 2020 Dec 19];6. PMID:29359252

14. Mitchell M, Hollingshead K, Coppersmith G. Quantifying the Language of Schizophrenia in Social Media. 2015. p. 20. [doi: 10.3115/v1/W15-1202]

15. 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]

16. Adler DA, Ben-Zeev D, Tseng VW-S, Kane JM, Brian R, Campbell AT, Hauser M, Scherer EA, Choudhury T. Predicting Early Warning Signs of Psychotic Relapse From Passive Sensing Data: An Approach Using Encoder-Decoder Neural Networks. JMIR MHealth UHealth 2020 Aug 31;8(8):e19962. [doi: 10.2196/19962]

17. Mcmanus K, Mallory E, Goldfeder R, Haynes W, Tatum J. Mining Twitter Data to Improve Detection of Schizophrenia. AMIA Jt Summits Transl Sci Proc AMIA Summit Transl Sci 2015 Aug 26;2015:122–6.

18. Bae S, Ferreira D, Suffoletto B, Puyana JC, Kurtz R, Chung T, Dey AK. Detecting Drinking Episodes in Young Adults Using Smartphone-based Sensors. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Jun 30;1(2):1–36. [doi: 10.1145/3090051]

19. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. Addict Behav 2018 Aug;83:42–47. [doi: 10.1016/j.addbeh.2017.11.039]

20. 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 Digit Med 2020 Dec;3(1):26. [doi: 10.1038/s41746-020-0234-6]

21. Hébert ET, Suchting R, Ra CK, Alexander AC, Kendzor DE, Vidrine DJ, Businelle MS. Predicting the first smoking lapse during a quit attempt: A machine learning approach. Drug Alcohol Depend 2021 Jan;218:108340. [doi: 10.1016/j.drugalcdep.2020.108340]

22. Ding T, Bickel WK, Pan S. Social Media-based Substance Use Prediction. ArXiv170505633 Cs [Internet] 2017 May 16 [cited 2018 Jan 31]; Available from: http://arxiv.org/abs/1705.05633

23. 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

24. Torous J, Staples P, Onnela J-P. Realizing the Potential of Mobile Mental Health: New Methods for New Data in Psychiatry. Curr Psychiatry Rep 2015 Jun 16;17(8):61. [doi: 10.1007/s11920-015-0602-0]

25. McGlashan TH. Early Detection and Intervention in Schizophrenia: Research. Schizophr Bull 1996 Jan 1;22(2):327–345. [doi: 10.1093/schbul/22.2.327]

26. Scott J, Eccleston D, Boys R. Can we predict the persistence of depression? Br J Psychiatry J Ment Sci 1992 Nov;161:633–637. PMID:1422613

27. Kessler RC, Aguilar-Gaxiola S, Alonso J, Chatterji S, Lee S, Ormel J, Üstün TB, Wang PS. The global burden of mental disorders: An update from the WHO World Mental Health (WMH) Surveys. Epidemiol Psichiatr Soc 2009;18(1):23–33. PMID:19378696

28. Wang PS, Simon GE, Avorn J, Azocar F, Ludman EJ, McCulloch J, Petukhova MZ, Kessler RC. Telephone screening, outreach, and care management for depressed workers and impact on clinical and work productivity outcomes: a randomized controlled trial. JAMA J Am Med Assoc 2007 Sep 26;298(12):1401–1411. PMID:17895456

29. Nelson B, McGorry PD, Wichers M, Wigman JTW, Hartmann JA. Moving From Static to Dynamic Models of the Onset of Mental Disorder: A Review. JAMA Psychiatry 2017 May 1;74(5):528. [doi: 10.1001/jamapsychiatry.2017.0001]

30. Hayes AM, Laurenceau J-P, Feldman G, Strauss JL, Cardaciotto L. Change is Not Always Linear: The Study of Nonlinear and Discontinuous Patterns of Change in Psychotherapy. Clin Psychol Rev 2007 Jul;27(6):715–723. PMID:17316941

31. Witkiewitz K, Marlatt GA. Modeling the complexity of post-treatment drinking: It's a rocky road to relapse. Clin Psychol Rev 2007 Jul 1;27(6):724–738. [doi: 10.1016/j.cpr.2007.01.002]

32. Pasipanodya EC, Kohli M, Fisher CB, Moore DJ, Curtis B. Perceived risks and amelioration of harm in research using mobile technology to support antiretroviral therapy adherence in the context of methamphetamine use: a focus group study among minorities living with HIV. Harm Reduct J 2020 Dec;17(1):41. [doi: 10.1186/s12954-020-00384-1]

33. 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. J Med Internet Res [Internet] 2018 Jul 4 [cited 2021 Mar 29];20(7). PMID:29973332

34. 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. J Health Commun 2015 Apr 14;20:1–7. [doi: 10.1080/10810730.2015.1018560]

35. 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;

36. 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 Alcohol 2011 Mar 1;46(2):105–112. [doi: 10.1093/alcalc/agq089]

37. 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.

38. Overton SL, Medina SL. The Stigma of Mental Illness. J Couns Dev 2008;86(2):143–151. [doi: https://doi.org/10.1002/j.1556-6678.2008.tb00491.x]

39. Parcesepe AM, Cabassa LJ. Public Stigma of Mental Illness in the United States: A Systematic Literature Review. Adm Policy Ment Health 2013;16.

40. 2019 National Survey on Drug Use and Health: Detailed Tables. Rockville, MD.: Substance Abuse and Mental Health Services Administration; p. 2472.

41. Stringer KL, Baker EH. Stigma as a Barrier to Substance Abuse Treatment Among Those With Unmet Need: An Analysis of Parenthood and Marital Status. J Fam Issues 2018 Jan;39(1):3–27. PMID:29307947

42. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med Publ Soc Behav Med 2002;24(3):236–243. PMID:12173681

43. 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

44. 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 Res Treat Hindawi; 2013 Nov 6;2013:e594671. [doi: 10.1155/2013/594671]

45. Ramsey AT, Wetherell JL, Depp C, Dixon D, Lenze E. Feasibility and Acceptability of Smartphone Assessment in Older Adults with Cognitive and Emotional Difficulties. J Technol Hum Serv Routledge; 2016 Apr 2;34(2):209–223. [doi: 10.1080/15228835.2016.1170649]

46. Yang C, Linas B, Kirk G, Bollinger R, Chang L, Chander G, Siconolfi D, Braxton S, Rudolph A, Latkin C. Feasibility and Acceptability of Smartphone-Based Ecological Momentary Assessment of Alcohol Use Among African American Men Who Have Sex With Men in Baltimore. JMIR MHealth UHealth 2015 Jun 17;3(2):e67. [doi: 10.2196/mhealth.4344]

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

48. 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. J Med Internet Res 2017 Apr 26;19:e132. [doi: 10.2196/jmir.6641]

49. van Genugten CR, Schuurmans J, Lamers F, Riese H, Penninx BWJH, Schoevers RA, Riper HM, Smit JH. Experienced Burden of and Adherence to Smartphone-Based Ecological Momentary Assessment in Persons with Affective Disorders. J Clin Med 2020 Jan 23;9(2):322. [doi: 10.3390/jcm9020322]

50. 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. Addict Abingdon Engl 2019 Apr;114(4):609–619. PMID:30461120

51. Meade A, Craig B. Identifying Careless Responses in Survey Data. Psychol Methods 2012 Apr 16;17:437–55. [doi: 10.1037/a0028085]

52. Freedman MJ, Lester KM, McNamara C, Milby JB, Schumacher JE. Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. J Subst Abuse Treat 2006 Mar;30(2):105–111. [doi: 10.1016/j.jsat.2005.10.005]

53. Epstein DH, Willner-Reid J, Vahabzadeh M, Mezghanni M, Lin J-L, Preston KL. Real-time electronic-diary reports of cue exposure and mood in the hours before cocaine and heroin craving and use. Arch Gen Psychiatry 2009 Jan;66(1):88–94. PMID:19124692

54. Stone AA, Shiffman S. Capturing momentary, self-report data: A proposal for reporting guidelines. Ann Behav Med 2002;24(3):8.

55. 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. Psychiatr Serv Wash DC 2016 May 1;67(5):558–561. PMID:26695497

56. 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 Ment Health 2018 Aug 28;5(3):e10334. [doi: 10.2196/10334]

57. 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 UHealth 2019 Sep 24;7(9):e13725. [doi: 10.2196/13725]

58. Duncan DT, Goedel WC, Williams JH, Elbel B. Acceptability of smartphone text- and voice-based ecological momentary assessment (EMA) methods among low income housing residents in New York City. BMC Res Notes 2017;10(1):517. [doi: 10.1186/s13104-017-2850-z]

59. 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 28;14(1):e0210240. [doi: 10.1371/journal.pone.0210240]

60. 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 Form Res 2019 Aug 29;3(3):e14329. [doi: 10.2196/14329]

61. Rintala A, Wampers M, Myin-Germeys I, Viechtbauer W. Momentary predictors of compliance in studies using the experience sampling method. Psychiatry Res 2020 Apr;286:112896. [doi: 10.1016/j.psychres.2020.112896]

62. Schönbrodt FD, Maier M, Heene M, Zehetleitner M. Voluntary commitment to research transparency. [Internet]. 2015. Available from: http://www.researchtransparency.org

63. Simmons JP, Nelson LD, Simonsohn U. A 21 Word Solution [Internet]. 2012 [cited 2016 Mar 21]. Available from: http://dx.doi.org/10.2139/ssrn.2160588

64. 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. Nat Hum Behav 2019 Dec 2;1–3. [doi: 10.1038/s41562-019-0772-6]

65. Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale--preliminary report. Psychopharmacol Bull 1973 Jan;9(1):13–28. PMID:4682398

66. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/

67. RStudio Team. RStudio: Integrated Development for R [Internet]. Boston, MA: RStudio, PBC; 2020. Available from: http://www.rstudio.com/

68. 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. J Open Source Softw 2019 Nov 21;4(43):1686. [doi: 10.21105/joss.01686]

69. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979;86(2):420–428.

70. Ackerman M, Mainwaring SD. Privacy Issues and Human-Computer Interaction [Internet]. 2008 [cited 2021 Jun 23]. Available from: https://www.semanticscholar.org/paper/Privacy-Issues-and-Human-Computer-Interaction-Ackerman-Mainwaring/d36442dacea38d57ed2f0ef2b52eac763c2b3cfb

71. Marwick AE, Boyd D. Privacy at the Margins| Understanding Privacy at the Margins—Introduction. Int J Commun 2018 Mar 1;12(0):9.

72. Pavlou PA. Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. Int J Electron Commer Taylor & Francis, Ltd.; 2003;7(3):101–134.

73. 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. Stud Health Technol Inform 2015;216:467–471. PMID:26262094

74. 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 UHealth 2019 Apr 5;7(4):e12578. [doi: 10.2196/12578]

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

76. O'Neil C. Weapons of math destruction: How big data increases inequality and threatens democracy. Harlow, England: Penguin Books; 2017. ISBN:978-0-14-198541-1

77. Masson CL, Chen IQ, Levine JA, Shopshire MS, Sorensen JL. Health-related internet use among opioid treatment patients. Addict Behav Rep 2019 Jun;9:100157. [doi: 10.1016/j.abrep.2018.100157]

78. Pew Research Center. Mobile Fact Sheet [Internet]. Available from: https://www.pewresearch.org/internet/fact-sheet/mobile/

---

[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) [65].

[3] We included additional self-report surveys at screening and later visits as part of the parent project aims. We do not report results from these additional surveys but we do share the 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 two 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 two 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 one 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 one participant who consented but did not subsequently enroll in the study.