

Dear Dr. Buis,

We are pleased to submit a revised version of our manuscript "Acceptability of personal sensing among people with alcohol use disorder: Observational study" (Manuscript No. JMU 41833) for your review at JMIR mHealth and uHealth. We thank the reviewers and you for the thoughtful commentary. We believe that these suggestions have greatly improved our manuscript. We have included detailed responses to each of the reviewers' concerns beginning on the following page.

If we can provide any additional information that can be of assistance, please do not hesitate to ask. Thank you in advance for your consideration.

Sincerely,

Kendra Wyant, Hannah Moshontz, Stephanie Ward, Gaylen Fronk, & John Curtin

Below, we delineate the reviewers' and action editor's questions, concerns, and requests regarding our original submission, and we detail the ways in which we have addressed each item. All comments have been copied verbatim to avoid misinterpretation. For brevity, we have removed comments regarding the strengths of the manuscript, but we appreciated these complimentary remarks.

Action Editor's Comments

AE.1: Please address the formatting issues identified in the checklist above.

AE.1.1: Please add subheadings under Introduction/Methods/Results/Discussion (if you use WinWord, apply the style "Heading 2" to IMRD headings, and the styles "Heading 3" to subsequent subheadings). Do not use italics or bold keywords or sentences in paragraphs in lieu of subheadings/sub-subheadings. We have revised and/or added headings and sub-headings to each IMRD section as requested. We have also removed all use of bold or italics through the manuscript.

AE.1.2: Shorten the paper. Some of the material/tables/formulas can be moved to a Multimedia Appendix.

We have substantially shortened the introduction to the paper - cutting four long paragraphs and making other edits throughout. As noted later in this cover letter, we also edited the introduction to clarify when we were reviewing empirical findings from the broader psychiatric literature vs. the narrower literature on alcohol and other substance use disorders.

We shortened the result section by moving one less focal table (previously Table 3 for characteristics of discontinued participants) to Multimedia Appendix 3. We also combined three additional tables (previously Tables 4-6) into a single table 3, saving space related to redundant titles and table notes.

We shortened the discussion section, in part by moving all the participant free response qualitative data to a table, along with other edits. However, we also were required to add content to the discussion in response to limitations that reviewers wanted explicitly acknowledged.

We believe all of these changes combine to produce a shorter but also more focused and clear paper.

AE.1.3: For all results for which you provide a relative result (percentage), you should also provide the absolute number, e.g. "132 out of 264 participants (50%) said that. . .". If n is less than 100, do not use decimal points in your percentages. Otherwise, do not use more than one decimal place.

We now report absolute numbers with all relative percentages and round percentages to one decimal place as requested.

AE.2: Please include relevant IRB approval numbers.

We have added the IRB approval number (Study # 2015-0780) to the methods section (see paragraph 2).

AE.3: Consider reformatting Fig 1. into a more traditional CONSORT diagram:

<https://www.consort-statement.org/consort-statement/flow-diagram>

We have reformatted Figure 1 to fit a traditional CONSORT diagram.

AE.4: The qualitative data presented in the Discussion really belongs in the Results section.

We have moved the qualitative data from the Discussion to either Table 7 in the Results section or Tables S3-S7 in Multimedia Appendix 3.

AE.5: The paper is quite lengthy and I'm not convinced all material, especially the lengthy content in the introduction, is required.

We have shortened the paper with a focus on cuts to the Introduction and Discussion, and reformatting and moving tables to appendix in the Results as noted above in point AE.1.2.

Reviewer B Comments:

RB.1: A systematic methodology and rationale for selecting the specific personal sensing types explored was not presented.

We have now added a section (Parent Project for Study Data in the Method) where we better situate the current study goals within the aims of the broader grant-funded parent project aims. In that section, we briefly describe the rationale for the sensing methods we selected, the inclusion of both active and passive methods, and the sampling rates associated with these methods.

RB.2: Prior work applying personal sensing to intoxication in general and AUD specifically with clear motivation for each of the sensing methods selected could be improved.

We have added review and discussion in the Introduction of numerous additional sensing studies that have been used to predict AUD/SUD relevant outcomes including alcohol, cannabis, and opioid use, lapse/relapse, and craving. These studies focus both on active measures like EMA but also demonstrate the potential utility of passive measures with a focus on geolocation and cellular communications as reviewed in this study.

RB.3: In Table 3, the number of subjects who discontinued prior to completing enrollment whose reasons are unknown seems quite large. Feels like a lot of information is lost there.

We now acknowledge the large number of missing information as to why participants did not enroll as a limitation of our study in the Limitations and Future Directions section of the Discussion.

RB.4: It would have been great to rank user acceptance to various personal sensing streams.

Our new version of Table 3 in the Results provides the mean scores for each of the three self-report measures of acceptability (interference, dislike, willingness to use for a year), separately for each sensing method. These means can be used to rank the methods on their mean scores for each measure. In addition, Table S1 in Multimedia Appendix 3 provides statistical tests of pairwise comparisons among the methods for each of these three self-report measures.

RB.5: It would have been great to compare or propose comparing in future, the response of subjects with AUD with those without AUD. User willingness for subjects under treatment may differ from those not under treatment.

We have added a proposal to compare the acceptability of sensing methods across populations with vs. without AUD as a potential future topic of study (see Limitations and Future Directions section).

RB.6: Paper has several typos that should be checked and fixed including: - One a day would be great -> Once a day would be great.

We originally left grammatical errors and typos in the participant quotations to preserve authenticity. We have now lightly edited these quotations to make them more readable. These quotations have been moved from the discussion to Table 7 in the Results or Tables S3-S7 in the Supplemental materials (see point AE.4 above for more detail).

RB.7: The paper uncovered some interesting findings. However, the reasons behind some user responses and preferences were not clear. A strong qualitative study or follow-on focus group might have provided more answers. The paper does contain some qualitative research in the form of free form user responses but that falls short.

We now present the free form user responses more systematically. We present user responses that can be mapped onto key themes related to acceptability in Table 7 in the Results. We present all free form user

responses in Tables S3-S7 in the Supplemental materials for completeness and full transparency (also see point AE.4 above).

RB.8: The lack of racial or ethnic information of participants substantially influences the significance of study's findings as racial and ethnic differences are likely to be great and provide additional insights.

We previously reported racial and ethnic information of participants in Table 1. We now make our sample demographics more salient by providing a summary in the first paragraph of the results section in addition to referencing Table 1.

Reviewer E Comments:

RE.1: The biggest flaw regarding acceptability is the inability for formal measures of compliance on the more sensitive data collection measures (location, logs, text message content). Considering all participants opted-in to provide data... this is how one would know if they were quietly opting out (by turning the sensor collection off).

We now acknowledge our inability for formal measures of compliance on our passive sensing methods as a limitation of our study (see Limitations and Future Directions section). We also clarify how we differentiate between opting-in and compliance in the Methods section. Specifically, opting-in indicates participants were willing to provide us any data. This measure is available for both our active and passive sensing methods. It should be noted that although it is possible for participants to delete individual messages and to turn off tracking, the data do not suggest this was widespread. Participants provided geolocation data only almost all days across the sample and text and phone message logs contained many, many responses.

RE.2: The 50/50 split on sex is concerning. A 50/50 fit would imply the recruitment mechanism made this split. Thus how did you exactly assess choice at consent and enrollment when you probably had screeners built in to ensure the sex split was 50/50?

We used a stratified sampling approach that was designed to recruit approximately equal numbers of male and female participants because we believed that it was important to include both men and women in reasonable numbers to generalize findings across sexes. This did not affect the enrollment or consent process because those processes occurred after eligibility screening. We had a separate pipeline for recruiting eligible men and women. We simply stopped recruiting men once we had adequate numbers of male participants. At that point, additional men were no longer considered eligible to participate and simply informed that recruitment had been closed for men when they responded to our recruiting advertisements. All of our analyses start with the N=192 eligible participants. In that sample, we then report the number who consented (N = 191/192), enrolled (N = 169), and completed through at least one month (N=154). The fact that we had exactly matched numbers of men and women at one month (77 each) occurred by chance though we expected approximately equal numbers due to the stratified sampling approach to identify eligible potential participants.

RE.3: The inclusion of data in the discussion is a problem to me and the discussion section reads more like a qualitative review than a discussion. Also, there is no limitation section. I think the discussion should greatly be reduced to be comment solely on the data presented in the paper.

We have now moved the qualitative data from the discussion as noted earlier (see AE.4). Previously, we had commented on limitations throughout the discussion section. We have now re-organized the discussion to include a primary "Limitations and Future Directions" section to make the acknowledged limitations more salient. In addition, we also now include additional limitations based on reviewer critiques that were not previously acknowledged.

RE.4: How much were participants paid for opting in—exactly. How much for each opt-in and where they only asked once to opt-in to tracking or could they add a feature on at any time?

We have now clarified how participants opted-in to provide each personal sensing method (see the specific sections for each method in the Personal Sensing section of the Method). In short, participants opted-in at the start of the study for all methods and could opt-out at any point for most methods or monthly at the study visit for cellular communications logs and text message content (which was collected at those visits). We also now clarify details about the participant compensation in the Compensation section of the Method. Participants were not compensated explicitly for opting-in to each sensing method. Instead, they were provided monthly sensing

method-specific bonuses for each method that ranged from \$10 to \$25 per month depending on the effort necessary to provide the raw data.

Reviewer K Comments:

RK.1: Would authors explain more about why certain sensing methods were compensated while others were not? In particular, the passive sensing methods require little effort were the ones more likely to be compensated. If anything, this provides greater support for their conclusions, still the question is why reimbursement there and not for other data?

We now clarify in the methods that we compensated participants for all sensing methods. We paid participants monthly bonuses for each method if they had 10% or less missing data for that method. The magnitude of those bonuses varied from \$10-\$25 per month. These details are described in the Compensation section of the Method. We varied these payment amounts to some degree because of the effort required to provide them to us (i.e., geolocation was completely passive whereas EMA and audio check-ins required effort each day).

RK.2: The questions about the willingness to wear devices ends with the phrase, “check-in for one year if it helped with my recovery.” Would the authors comment on the possible effect of that on participant responses? Alcohol use disorder is a relapsing disorder with few guarantees in treatment. How would people respond if the question were phrased, “I would be willing to check in for one year to TEST it helped with my recovery.”? My point is, the way the question is phrased may be affecting the answer, particularly in this early stage of testing.

We now discuss the impact the wording of our acceptability questions might have on participant responses in the Limitations and Future Directions section. Specifically, we acknowledge that asking participants if they would be willing to use a personal sensing method for 1 year to help with their recovery could imply there would be clinical benefit to using the method for 1 year. Such a perceived benefit may factor into their judgment of acceptability. However, it we have also now made clear in the manuscript (and supporting materials on OSF) that participants were informed at multiple points in the onboarding process that there were no benefits to them provided from sensing other than financial compensation. They did not receive any feedback from the sensed data. They were not provided tailored interventions based on the sensed data.

RK.3: There is a relevant paper that might be worth including on the longevity of tracker use:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6952057/>

#:~:text=Based%20on%20these%20results%20from,have%20accounted%20for%20the%20NE. How would the authors respond to this letter?

We now cite this paper in our discussion (see Limitations and Future Directions section). We acknowledge that while three months duration is an improvement to current studies on acceptability of personal sensing, it still may not be long enough to account for novelty effects. In fact, we are now conducting another study that uses sensing methods for up to one year with participants in recovery from SUD. We cite the published study protocol for that project as an important future direction.

RK.4: Can the authors comment on their use of “undecided” versus neutral as an anchor on the scale? They define it as ‘neutral’ but it is not necessarily clear that is so. For some questions, such terms could have different response sets.

We agree that other labels (e.g., neutral) could have been assigned to the mid-point of the scale. However, we did explicitly assess whether the individual measures retained at least ordinal properties given our choice of labels for the scale as part of the correlations and ICCs that were calculated among these self-report measures. Scatterplots were generally consistent with patterns expected given ordinal relationships among these measures when was undecided coded as the midpoint of each bipolar scale.

RK.5: How were the alcohol use disorder questions and diagnosis collected? was there a standard survey? There appear to be footnotes perhaps explaining this? but I did not find them.

We collected information about Alcohol Use Disorder symptoms using a self-report survey that was administered during the screening session. This survey included individual items for each of the DSM-5 symptoms for Alcohol Use Disorder. We describe this method in a footnote associated the Inclusion criteria listed in the Exclusion and Inclusion Criteria section of the Method. We provide access to this survey (and all

other surveys) on the study OSF page (<https://osf.io/du2an>). We describe how to access the study OSF page in the Research Transparency section of the Method.

RK.6: Is there a count of how many disabled locations or deleted texts/phone calls occurred? This is important bc as someone might near a bar, a warning text could be sent. however, if people can disable their location, then the method is nugatory.

Unfortunately, we are unable to see if and how often people disabled their locations or deleted phone calls and text messages. We agree this is a limitation of our study and we now acknowledge it in our Limitations and Future Directions section of the Discussion. It should be acknowledged that many of the sensed features are not limited to obvious risky locations like bars. Instead, we can sense time spent in treatment settings, time spent out of the home, time spent in homes of supportive others, time spent in settings that are self-described as rewarding. Preliminary research with these features suggest they have high predictive value in lapse prediction algorithms and it is arguably less likely that participants/patients would censor these locations.

RK.7: The compensation for participants should be in the methods section esp as it has bearing here.

We now describe participant compensation in the Compensation section of the Method.

RK.8: Reference for this: 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.

We have removed this statement from the manuscript. Although we had believed this might be true, there is also published research suggesting barriers are not higher in patients with AUD. We now cite that study instead.

“Furthermore, people with SUD also have generally high rates of mobile technology use (Collins et al., 2016).”

RK.9: Providing cellphones for those who did not have one should also go in the methods.

We now indicate that we provided smartphones for the duration of the study to participants who needed them. This information has been added to the Compensation section of the Method.

RK.10: Suggest that authors switch to word ‘adhere’ instead of ‘comply’ as that is the more accepted term for many reasons.

We now use the word adhere rather than comply, where appropriate, as requested.

Reviewer O Comments:

RO.1: The abstract should include more information about method relevant to the current aim.

We have edited the abstract to include more method detail within the limits of word counts.

RO.2: The title of the larger project suggests that sensors would be used to prevent relapse. Is there a risk that participants have been motivated by clinical benefit even if that was not the intention?

Participants were explicitly instructed at numerous points in the study that the only benefit for participation was financial compensation. They were informed that the personal sensing would not be used to monitor their current risk or to select, deliver, time, or adaptive any interventions to their needs. The Procedure section of the Methods has now been updated and supporting materials (i.e., consent form, recruiting flyer) added to our OSF page to more clearly reflect this.

RO.3: The different methods are presented as opt-in, but it is important to know in detail how this was presented. Were the participants free to choose what they wanted and interested in or were they more or less expected to agree and more had to opt-out of certain methods?

We have now clarified how participants opted-in to provide each personal sensing method (see the specific sections for each method in the Personal Sensing section of the Method). In short, participants opted-in at the start of the study for all methods and could opt-out at any point for most methods or monthly at the study visit for cellular communications logs and text message content (which was collected at those visits). We also now clarify details about the participant compensation in the Compensation section of the Method. Participants were not compensated explicitly for opting-in to each sensing method. Instead, they were provided monthly sensing method specific bonuses for each method that ranged from \$10 to \$25 per month depending on the effort necessary to provide the raw data.

RO.4: Background provides a good overview of the use of mobile sensors in psychiatry. But it is sometimes unclear when the statements refer to alcohol or substance use and what is more general in mental health.

We have edited the introduction to clearly identify when we are reviewing/citing studies that refer to samples with broader psychiatric disorders vs. narrower alcohol and other substance use disorders.

RO.5: The authors write regarding EMA that “compliance concerns may be limited to applications with patients with substance use disorders rather than all psychiatric disorders more generally”. This is confusing when the study is about substance use.

Personal sensing is being used with psychiatric samples generally and samples with alcohol and other substance use disorders. Meta-analyses have suggested that sensing is less acceptable to patients with substance use disorders but not other mental illness. We cite these meta-analyses as part of the rationale to look at acceptability in more detail in our sample with alcohol use disorder. We have clarified this rationale further in the introduction.

RO.6: What previous research have been made on the different sensing methods for alcohol use disorder? (Bae 2018, Stevenson 2020, Scott 2020...)

We have expanded our review of sensing methods for alcohol and other substance use disorder. This expanded review includes the papers cited here along with other relevant papers. See point RB.2 above.

RO.7: The data as it is presented should not be presented as “transparent window into the feasibility”.

We have removed this language from the manuscript.

RO.8: Thank you for making documentation available via OSF. Could be improved by adding participant information about the project and recruitment messages.

We have uploaded a copy of our recruitment flyer that was posted at various in-person and digital community locations, such as AA meetings, Metro buses, UW hospitals and clinics, Facebook, Craigslist, and UW Health AppSpace screens, to our OSF page (<https://osf.io/2pcjb>). We have also uploaded a copy of the consent form to OSF as well(<https://osf.io/54nvq>).

RO.9: What was the aim of the parent project. Will/are the results from the parent study available?

We have now provided additional detail about the rationale, aims, and broader methods of the parent project to provide a better context within which to interpret the current study. These details have been added to a new section (Parent Project for Study Data) in the Method.

RO.10: What was the reason for the sobriety requirement? Many people with AUD prefer a moderation goal. Would not the sensors work just as well with moderate consumption?

Most future applications of sensing in the AUD/SUD domain target patients in recovery with plans to use sensing to help them monitor for lapses/relapses and/or select and adapt digital interventions to address their momentary needs. For this reason, we wanted our sample to consist of people who were in recovery rather than simply continuing harmful use. We used alcohol abstinence as a behavioral indicator of a commitment to recovery. Although, recovery may be possible without complete abstinence, clinicians typically recommend abstinence for patients who present with moderate or more severe AUD (see Eddie et al., 2022). We now provide this rationale for the abstinence inclusion criterion as a footnote.

RO.11: It is hard to understand the reason for using geolocation and cellular communication logs without reading the OSF material, and even after that it is somewhat unclear. It probably connects to the aim or the parent study, but should be made clear in methods here.

As noted earlier (see point RO.9), we now provide more detail about the parent project including the selection of sensing methods that were expected to be powerful predictors of alcohol lapse.

RO.12: Did the participants receive any additional information on the purpose of each method connected to their alcohol use?

No, we added the following sentence to the Procedure section in the Method to clarify this. “We informed participants that we were collecting these data to develop an algorithm that could be used in the future to

monitor for relapse risk. We did not provide them with any further information about how each sensed data stream might be used in this algorithm.”

RO.13: There are no references to the acceptability questions having been previously used or tested in research. The fact that the answers were significantly higher than neutral is interpreted as the method being acceptable. But without any other comparison it is impossible to know what explains this result. It is common in evaluations that participants give over average grade for many different reasons like social desirability or lack of willingness to give feedback.

We now acknowledge that these questions have not been used or tested in other research settings and the potential for biased responses in our Limitations and Future Directions section of the Discussion. Of course, it should be acknowledged that our analyses and conclusions regarding acceptability were not based only on the self-report measures but also on object participant behaviors as well. ## RO.14: All of the questions are framed in a lack of negative way rather than from a negative to a positive experience of the different methods. We have edited the description in the Method of the self-report questions posed to participants and directly shared the measure on the OSF study page. In fact, the participants were asked to “indicate how much you agree or disagree with each statement” on a 5-point bipolar scale (strongly disagree, disagree, undecided, agree, strongly agree). It is true that the framing of these statements was intended to detect burden, which is why they focused on dislike, interference, and willingness to use for a year. We did not expect that people would necessary like or otherwise experience enjoyment from sensing methods. They are not intended for that purpose. However, if they were burdensome, it may be that people would not use them even if they provided benefits in their use in later clinical applications. Nonetheless, as acknowledged elsewhere (see RO.16), we recognize that the use of novel measures that we created for this study is a limitation and report it as such in the Future Directions and Limitations section of the Discussion.

RO.15: How was the content of text messages analyzed?

Text message content was not analyzed as part of this study. It will be used in the parent project for feature engineering to predict alcohol lapses.

RO.16: Most of the result section show detailed information about the results on the acceptability questions. On most of the questions and sensing methods the users mean ratings are similar. Could this be an indication on problems with the questions and/or sampling rather than that the methods really are acceptable for people with AUD?

We do not believe that the data suggest problems with the self-report measures. Although the means for methods across the sample are similar, those means are, in fact, statistically different from each other in many instances (see pairwise comparisons among methods in Table S1 in Multimedia Appendix 3). In addition, the histograms of individual responding demonstrate that participants used the full range of all of the measures across the sample. Furthermore, significant correlations among measures (reported in the analyses) would be possible to detect if there was systematic variance across participants in their scores on these measures. Nonetheless, we do now acknowledge that the use of novel (previously unvalidated) measures of acceptability is a limitation of this study. That said, we are not aware of previously validated measures of acceptability that could have been used instead as this is an emerging area of research.

RO.17: In table 2, does Types of Treatment refer to recent treatment or lifetime?

We have clarified this heading in Table 2. It now reads “Lifetime History of Treatment (can choose more than 1)”

RO.18: In table 3 Not wanting to stay sober, cancelling no/show and unknown are clearly the main reasons for discontinuing. This is interpreted as not being about acceptability. Can you really know and say that?

We have moved this table from the Results to Table S3 in Multimedia Appendix 3 to somewhat de-emphasize this information. To be clear, Table S3 did not suggest that unbolded categories were not about acceptability. Instead, they were simply not highlighted as clearly related to acceptability vs. other clearer categories (e.g., non-compliance with providing data, city study demands as burdensome, no longer interested). Patients with SUD often relapse such that it is not unexpected that many participants would discontinue because they no longer wanted to stay sober. Participants counted as multiple reschedules are participants that did repeatedly try

to come to study visits (they recontacted us after missed or re-scheduled appointments). If they found the sensing methods unacceptable it seems unlikely that they would not simply not reschedule with us after a missed visit. Nonetheless, we have updated the table note to more clearly indicate that bolded entries are clearly related to acceptability but no entries could be definitively designated as unrelated to acceptability.

RO.19: Have this high sampling density been used in substance use before? Would it be used clinically and if so for what reasons?

As noted earlier (see point RO.9), we now provide more detail about the parent project including the rationale for the high sampling density of many sensed raw data streams. The aims of that project were to develop a temporally precise (up to one hour precision) machine learning algorithm to predict future lapses back to alcohol use. Such high temporal resolution is innovative but also very important for the use of “just-in-time” interventions that are designed to be delivered at moments of greatest need.

RO.20: It is not clinically reasonable to remind people with 2-3 months abstinence about their alcohol use problems 4 times a day. This could have side effects on emotional state or drinking. Have you considered this?

The goals of most contemporary treatments (e.g., Cognitive-Behavioral Relapse Prevention, Mindfulness Based Relapse prevention) are to train and encourage patients to carefully monitor themselves and their environments for risks to their recovery. Alcohol and other substance use disorders are chronic relapsing disorders and patients must learn to sustain such monitoring long-term, if not for the rest of their lives. Careful, frequent monitoring is very important during early recovery (the period studied in our sample). Of course, frequent and/or long-term monitoring is very difficult to achieve. That is why we (and other research teams) are pursuing the use of sensing combined with machine learning prediction algorithms to support such monitoring automatically. We now highlight these goals of the parent project (see point RO.9). Our use of both active methods (that explicitly ask people about experiences that might affect their recovery via EMA) and passive methods that can potentially assess these same risks but without burden or explicit attention was intentional to allow us to compare the potential acceptability (this study), burden (this study), and predictive power (subsequent aims of parent project).

RO.21: There is a lack of data on cravings, alcohol use and on qualitative data from users in the result section.

The focus of this paper is on the acceptability of collecting sensed raw data streams over longer (up to three months) periods of time. The study does not provide participants with any feedback or interventions using these sensed data to target clinical outcomes like craving or alcohol use. As such, we do not report data on craving or alcohol use in this study. We limit analyses to behavioral and self-report measures of acceptability. In that context, we do report qualitative data on participants reactions to each of the sensing methods both in the body of the paper (see Table 4) and in Multimedia appendix 3.

RO.22: How was lapses/relapses among the participants handled?

We have added the following information to the Procedure section in the Method.

“There were no consequences for continued study participation if participants lapsed back to alcohol use during the study. However, for human subjects reasons, we did offer brief motivational interviewing interventions to participants if they reported any alcohol use to study staff. Participants were not required to participate in these interventions but we were offered it to them as support to maintain their recovery if desired.

RO.23: There are several indications that the individuals in this sample are different from other individuals with alcohol use disorder in regard to their perception of mobile sensing. The large proportion of people agreeing to participate in research probably depend a lot on where and with what message the participants were recruited and what information they got about the project. This should be described in much more detail for others to be able to get the same results and to be able to judge what, if anything, the high proportion of people agreeing say about personal sensing.

We have provided additional details about the recruiting, screening, and consenting process. This includes additional information in the Method in the Procedure section and the Compensation section. We also clearly enumerate all study exclusion and inclusion criteria (see Exclusion and Inclusion Criteria in the Method). In addition, we also now share the recruitment flyer and the consent form on the OSF study page to provide full

transparency about recruitment and consent (see point RO.8 for more detail).

RO.24: The authors write: “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.” This is probably true, not only for passively sensed private information, but for participation in the study as a whole. Those less likely to accept mobile sensing would not have consented, enrolled, and then opted-in.

We agree with this statement and had previously acknowledged it by providing participant retention at each point of the study (i.e., consent, enrollment, opting-in). For example, only one participant chose not to consent after hearing about our study. We now also explicitly acknowledge in our limitations and future directions section that participants who chose to not enroll in the study after initial consent may have done so for reasons related to acceptability.

RO.25: The qualitative data in the form of statements from participants on their experience is interesting. How were they collected and analyzed? Why are they not presented as results?

We now present these qualitative data based on participant responses in Table 7 in the Results and Tables S3-S7 in the Supplemental Material as requested by this and previous reviews (e.g., see AE.4). We also added detail to the Method to describe how these responses from participants were collected.

RO.26: The study demonstrates that active sensing methods are feasible for this type of research setting but not for clinical applications.

We now explicitly acknowledge that more research is needed to test the acceptability of these methods in applied clinical settings in our Limitations and Future Directions section of the Discussion.

RO.27: The overall aim of the use of these kind of sensors in treatment should be made more clear. Why should this information be gathered about people with alcohol use disorder? Is it for feedback, self-control and self-knowledge, for external control/pressure to stay abstinent and/or for other purposes?

As noted earlier (see point RO.9), we now provide more detail about the aims of the parent project. In both the Parent Project for Study Data section and in the Introduction, we highlight the potential future use of these sensed data to develop algorithms that can support self-monitoring by predicting future lapses/relapse and to select, time and adapt interventions to meet patients' momentary needs during their recovery.

RO.28: Who made repeated attempts to reschedule before discontinuing, researchers or participants?

We now clarify in Table 3 that participants made attempts to reschedule before ultimately discontinuing.

RO.29: The authors argue that reimbursing participant is common in many studies and even in clinical settings. Even if this is true it is a big limitation than should be clear from the abstract. The way the sample was recruited and informed about the project should be more clearly discussed under benefits.

We may not fully understand this issue. To be clear, it is almost universal in research studies in psychology to provide participants with some form of compensation (monetary or otherwise) for their participation in the study. Of course, the level of compensation has to be consistent with the tasks/effort expected from research participants and not so high as to be considered coercive. Such issues are evaluated explicitly when studies are evaluated by IRBs. We have clarified participant compensation for this study (see Compensation section in the Method) so that this information is clearly available to readers. We have also edited the abstract to make clear that participants were compensated to engaged in the personal sensing methods.

In the discussion, we acknowledge that it is possible for clinical interventions/treatments to include compensation/payments to patients. In fact, contingency management (e.g., <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3083448/>) is a treatment that explicitly includes compensation to patients to encourage them to engage in healthy behaviors (e.g., not using drugs) or adhere to other beneficial treatments. Given this, we believe that it may be possible (though it may not be necessary) within such a framework to compensate patients to provide some personal sensing data if those data can be used to improve their health as part of a monitoring or adaptive intervention system. We have attempted to clarify this discussion

further in the revised manuscript.

RO.30: The section on how trust is important is interesting. How might participants trust in researchers affect the results?

We now discuss how trust may affect acceptability and how it may differ in research vs clinical settings and according to how has access to the raw and processed data. This information is provided in section 6 (“Trust likely matters”) of the Discussion.

RO.31: How were the themes presented in the discussion chosen?

We organized our discussion around 7 key conclusions from that we believe were supported to varying degrees by the analyses reported in the results section. This was simply an organizational framework designed to make salient the important “take-home” messages that we believe are warranted given this study.

RO.32: Its contradictory to state that the sensing was acceptable “without explicit clinical benefits to the participants” and at the same time that they can be “used for clinical applications”. The study does not appear to be designed to test clinical application, but recruited a population that seem clinical, with ongoing substance use among many participants.

We now make clear in the method that participants did not receive any explicit clinical benefits from the personal sensing conducted as part of this study. They did not receive any feedback on the sensed data. The sensed data were also not used to provide or adapt any interventions for these participants. As such, personal sensing did not provide them with any clinical benefits in this study. Our hope is that we can use data collected via personal sensing to develop machine learning algorithms that can be used in FUTURE clinical applications of these methods to both help people monitor their recovery and/or deliver or adapt clinical interventions that are provided to them. This is a FUTURE possible application. These algorithms do not yet exist (and in fact, one of the aims of the parent project is to develop such algorithms for future use based on the data collected here).