Oslo, Norway, October 25, 2024

Dear Andrew Jones, Recommender at Peer Community In Registered Reports,

Thank you for the opportunity to improve the methods of our proposed study and to revise and resubmit our Stage 1 manuscript based on the helpful comments from expert reviewers and you, the recommender. Please find our point-by-point responses below. Recommender's and reviewers' comments are marked in bold and followed by our responses in regular font. Relevant excerpts from the revised Stage 1 manuscript and supplementary materials are italicized, with key changes highlighted in red.

Sincerely on behalf of the authors,

Martin Trøstheim (corresponding author)
E-mail: m.o.trostheim@psykologi.uio.no

Kongsberg Hospital, Vestre Viken Hospital Trust, Kongsberg, Norway
Department of Psychology, University of Oslo, Oslo, Norway

RECOMMENDER COMMENTS

Recommender, summary

You can see two expert reviewers have provided some comments on your stage one report. Overall, they are supportive of the research and approach. I agree with this. Both reviewers reflect on the point around access to technology, and considering and how well your sample might reflect the population of individuals with OUD. A related point is also the quality of data you might get from these online questionnaires. There may be instances of poor-effort / careless-responding responding and you may consider ways to tap into this (especially considering responses were mandatory)? https://www.annualreviews.org/content/journals/10.1146/annurev-psych-040422-045007

One final point I noted was how you plan to include living sitation as a covariate in your analysis. It's three catagories in your questionnaire, do you plan to dummy code or simplify (eg. place to live vs no place to live).

Thank you for highlighting key points made by the reviewers about representativeness and generalizability. We have updated the sampling and data collection protocol and now plan to provide necessary digital equipment and technical support to mitigate digital exclusion. Please see our responses to **Reviewer #1, comment #3**, **Reviewer #1, comment #5** and **Reviewer #2, comment #3** for details. We have also updated the analysis plan with a separate analysis to determine the representativeness of the sample by means of comparing demographic and clinical sample characteristics with near population level values collected by the Norwegian Centre for Addiction Research. Please see our response to **Reviewer #1, comment #3** and **Reviewer #2, comment #3** for details.

We are also grateful for you providing us with a good starting point for determining how to deal with careless responding. We have followed these recommendations and will be identifying and excluding careless responders based on low response times, incorrect responses to an instructed response item and a bogus item, and significantly deviating responses across valid items (Mahalanobis distance). The following description of this strategy has been added to the methods section (manuscript, p. 8):

We followed current recommendations for dealing with careless responding and used a combination of (moderate) a priori and (minimal) post hoc screening methods prior to data analysis. This involved excluding data from participants who either spent on average < 2 seconds responding to each item, responded incorrectly to the instructed response item or

the bogus item, or whose responses to the valid items had a statistically significant robust Mahalanobis-Minimum Covariance Determinant (MMCD) distance, calculated with a sub-sample of 0.75 and assessed with a χ^2-test and an α-level of 0.001. Use of invariance and consistency indicators to detect careless responders was considered unfeasible due to the use of questionnaires with few items, varying numbers of response options and little semantic overlap.

We have also added the following descriptions of the bogus and instructed response items to the methods section (manuscript, p. 11):

To detect careless responders, we included a binary bogus item with one highly improbable response option, and an instructed response item with a single correct response option on a 5-point scale (eAppendix III and IV). We also recorded the total time it took for participants to complete the survey and divided this by the number of items to obtain the average response time per item.

Exact items and response options are available in Norwegian and English in eAppendix III (eSupplement, p. e26) and eAppendix IV (eSupplement, p. e43), respectively.

Regarding living situation, approximately 96% of patients in the Norwegian opioid substitution treatment program have a place to live according to the Norwegian Centre for Addiction Research (see the status report for 2023). Simplification of this variable might therefore mask a potentially important nuance in living situation such as living alone versus living with someone. Living situation will therefore retain its three categories and be handled by means of dummy-coding in the secondary and planned exploratory analyses. This is now indicated in the following part of the methods section (manuscript, p. 11):

To achieve covariate balance and thereby adjust for potential confounders, we included age, sex, BMI, SES, living situation (dummy-coded) and diagnostic load in the models testing the associations between medication group and time use profile, and between time use profile and well-being, experienced stigma and life satisfaction. Medication group (dummy-coded; eAppendix II) was also included in the latter models of well-being, experienced stigma and life satisfaction. In the analyses using multimodel inference and model averaging to evaluate the importance of specific aspects of time use for predicting well-being, experienced stigma and life satisfaction, we adjusted for potential confounders by including age, sex, BMI, SES, living situation (dummy-coded), diagnostic load and medication group (dummy-coded) as fixed terms across all possible models.

However, to facilitate comparisons of characteristics between the sample and the overall population to identify constraints on generality (see our responses to **Reviewer #1, comment #3** and **Reviewer #2, comment #3**), homelessness would have to be derived as a binary simplification of the living situation variable. We now indicate this in eAppendix II (eSupplement, p. e#). This part reads:

The characteristics to be compared between the study sample and the overall population with χ^2-tests included 1) the proportion of women, 2) the proportions of individuals aged < 31 years, 31-40 years, 41-50 years and > 50 years, 3) the proportion of individuals who are working or studying (derived from time spent on occupational and educational activities for the study sample; eAppendix III and IV), 4) the proportion of homeless individuals (derived from living situation for the study sample; eAppendix III and IV), 5) the proportions of individuals experiencing depression, anxiety and psychosis, and 6) the proportions of individuals who are satisfied, dissatisfied and neither satisfied nor dissatisfied with their treatment situation.

REVIEWER COMMENTS

Reviewer #1 (anonymous)

Reviewer #1, comment #1

1A. The scientific validity of the research question(s):
The research questions are scientifically valid and are exploratory in nature. I woudl have also liked to see research questions handling the specific anticipated findings for each medication for opioid use disorder. I found the concept of "time use" to be a little woolly and thought that the authors could better articulate this and group in to substance related versus recovery related. I also thought that rather than just looking at the quantity of the activity (how many days followed up by a rating of the quality of that aspect).

Thank you for suggesting additional avenues to explore in our proposed study, and helpful ways to clarify the concept of time use.

We currently have no strong hypotheses about exact outcomes for specific medications for opioid use disorder. Due to differences in pharmacological properties, administration methods and administration frequency, it is possible that the various medications affect patients differently. For instance, the various medications may have side effect profiles that

differ in quality and intensity and that make patients prioritize their time differently (e.g., seeking relief from side effects and/or avoiding certain activities) or experience different levels of well-being and life satisfaction. Data will nevertheless be collected in sufficient detail to describe outcomes for specific medications for opioid use disorder.

Regarding the concept of time use, we appreciate the usefulness of grouping the different activities into substance-related and recovery-related activities. While some activities are either clearly substance related (e.g., seeking/using opioids or other illicit substances) or more likely recovery related (e.g., education, work, housekeeping, personal care, or physical activity), others do not necessarily fit into either of these categories (e.g., digital entertainment), or may fit into both depending on the context (e.g., social activities). To add some necessary contextual nuance, we have divided social activities into social activities with people who use illicit substances, and with people who do not use illicit substances. For convenience and ease of understanding, we have rearranged the time use variables in Table 2 (manuscript, p. 25) and added subheadings so that each activity is now sorted under one of the four categories 1) primarily substance use-related, 2) primarily treatment-related, 3) primarily recovery-related, and 4) other. We have also updated the description of the time use measures in the methods section (manuscript, p. 6) to use these terms. It now reads:

Time use was assessed with a custom 17-item questionnaire asking participants to indicate on how many days during the past week (0-7) they had engaged in various activities (e.g., substance use-related, treatment-related, and recovery-related activities; Table 2).

Likewise, we have updated the following part about the statistical analyses in the methods section (manuscript, p. 9) to also use the same terms:

We expected time use profiles in the current sample to broadly represent combinations of high and low amounts of time spent on 1) substance use-related activities, 2) treatment-related activities, and 3) recovery-related activities (possible combinations: 2×2×2 = 8).

Regarding the quality of the participants' time use, we do agree that this could be an important dimension to explore. We want to keep the survey relatively short to maintain a low threshold for participation and completion, and adding a quality rating after each quantity rating (i.e., 17 new items in total) would unfortunately substantially increase the survey length. This will likely deter more individuals from providing careful responses, completing the survey, or participating in the study at all. However, we do include two questions about which activities they would like to spend more/less time on (eSupplement, pp. e19 and e36),

giving us some indirect information about how satisfied they are with their engagement in the various activities. Likewise, the planned analyses will give some more indirect information about the potential quality of these activities through their relationships with well-being and life satisfaction.

Reviewer #1, comment #2

1B. The logic, rationale, and plausibility of the proposed hypotheses (where a submission proposes hypotheses):

While the hypotheses are logical to someone with knowledge in the area, I do not think that at present they are derived from what precedes them. You might imagine that type of medication would be related to time use pattern because time spent travelling to daily supervised consumption vs. a single long acting injection would dictate this. If the authors are implying that other factors affect this, they need to clearly provide a rationale in the introduction. Similarly with stigma, mental health and quality of life, a better case needs to be made for this relationship.

Thank you for pointing out that our rationales for the hypotheses need some clarification. We can confirm that the hypothesized link between medication type and time use is indeed based on the lower travel and supervision requirements associated with long-acting medication treatment compared to daily medication treatment. We have added an explicit statement about this to the following part of the introduction section (manuscript, p. 3):

Because of the drastically lower travel and supervision requirements, some patients are hopeful that treatment with long-acting medications will give them a more satisfactory life by reducing the stigma they experienced from frequently collecting medications and enabling them to spend more time on social, physical, educational, occupational and recreational activities. These activities are known to protect against the many mental and somatic health problems that often accompany opioid use disorder and that contribute to poor well-being (e.g., depression, anxiety, and chronic pain).

The hypothesized relationships between time use and well-being/experienced stigma/life satisfaction are primarily based on hopes for and concerns about treatment with long-acting medications that have been voiced by patients. For instance, patients report experiencing daily medication collecting as stigmatizing, implying that less time spent on this may be associated with lower experience of stigma. More time spent on various social, physical, educational, occupational and recreational activities may be associated with greater well-being among patients as such activities have been shown to be protective of well-being. As

indicated in the excerpt above, we have expanded this part of the introduction with additional details to help clarify the rationale for the hypotheses (manuscript, p. 3). To justify the non-directionality of the hypotheses, we have also added the following details about the potential negative outcomes to the introduction section (manuscript, p. 3):

However, patients have also voiced concerns about potentially reduced life satisfaction due to reduced contact with the healthcare system and difficulties adapting to the increase in spare time. The monitoring and social support offered by healthcare workers may encourage patients to stay abstinent and make positive lifestyle choices that benefit their well-being and life satisfaction. However, since the use of long-acting formulations obviates the need for frequent patient contact to administer medications, patients could potentially find themselves spending more time in social isolation and/or on previously discouraged activities (e.g., illicit substance use). In turn, this might help to maintain the experience of stigma, contribute to poor well-being, and ultimately reduce life satisfaction.

There may indeed be several factors mediating the relationships between time use and well-being/experienced stigma/life satisfaction (e.g., the degree of social support provided by healthcare workers in the opioid substitution treatment program). Such potential mediators are now being alluded to in the introduction section (manuscript, p. 3), as indicated in the above excerpt. It is however first necessary to establish whether there are relationships between time use and well-being/experienced stigma/life satisfaction before identifying and testing specific mediators to help further understand these relationships. We therefore consider the latter beyond the scope of this study.

Reviewer #1, comment #3

1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable): The methodology is sound and feasible, but I think it excludes a large proportion of people with OUD because it a) required people to be able to read and consent and b) required digital completion of questionnaires via a smartphone or device. Many people in recovery from OUD may not have access to digital tech, and thus the sample recruited and retained will be skewed heavily towards more stable people who might naturally make better use of their time and have higher wellbeing, and experience lower stigma.

Based on research on eHealth services for opioid use disorder, access to necessary digital equipment does indeed seem to be a likely barrier for participation in the current study. We

have therefore updated the sampling and data collection protocol to state that we will be recruiting participants at the local sites of the services and organizations they typically interact with, and that we will be providing digital equipment and technical support to participants when necessary. This will likely mitigate, although not completely eliminate, skewed sampling (see our responses to **Reviewer #1, comment #5** and **Reviewer #2, comment #3**). To evaluate the representativeness of our sample, we will compare characteristics of the study sample against the near population-level values obtainable from the Norwegian Centre for Addiction Research's yearly status report on the Norwegian opioid substitution treatment program. Conclusions will be calibrated according to identified constraints to generality. The following explicit plans to evaluate and discuss constraints on generality additional analyses have been added to the methods section (manuscript, p. 12):

We followed current recommendations and interpreted the results in light of constraints on generality. To help identify constraints beyond inclusion criteria and material selection, we used χ^2-tests to test for statistically significant divergence in key demographic and clinical characteristics of the current sample from near population-level values for individuals with OUD in Norway (eAppendix II).

Further details are available in the following, newly added part of eAppendix II (eSupplement, p. e6):

Near population-level values for demographic and clinical characteristics of individuals with opioid use disorder in Norway were obtained from the Norwegian Centre for Addiction Research's latest yearly survey (see e.g., Nesse et al.) of all patients enrolled in the Norwegian opioid substitution treatment program (i.e., ~89% of all individuals with opioid use disorder in Norway). The characteristics to be compared between the study sample and the overall population with χ^2-tests included 1) the proportion of women, 2) the proportions of individuals aged < 31 years, 31-40 years, 41-50 years and > 50 years, 3) the proportion of individuals who are working or studying (derived from time spent on occupational and educational activities for the study sample; eAppendix III and IV), 4) the proportion of homeless individuals (derived from living situation for the study sample; eAppendix III and IV), 5) the proportions of individuals experiencing depression, anxiety and psychosis, and 6) the proportions of individuals who are satisfied, dissatisfied and neither satisfied nor dissatisfied with their treatment situation. To account for potential oversampling of patients treated with certain medications, we adjusted the sample proportions used in these tests according to the frequencies of patients receiving each medication in the population according to data from the Norwegian Centre for Addiction Research's latest yearly survey.

We have also updated the content of the column “Interpretation given different outcomes” in the study design template (manuscript, pp. 20-24) to calibrate potential conclusions in advance according to constraints on generality. See also our response to **Reviewer #2, comment #3**.

Reviewer #1, comment #4

1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses:
Yes, this is detailed and would be easy to replicate.

Thank you for confirming that our proposed methods are reported in sufficient detail for close replication. As requested in **Reviewer #1, comment #5** and by the recommender (**Recommender, summary**), we have added more details on how both medication group (manuscript, pp. 10-11; eSupplement, p. e4) and living situation (manuscript p. 11) will be handled in the analyses to further increase the replicability and reproducibility of this study.

Reviewer #1, comment #5

1E. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).
See above regarding exclusion of groups without access to digital tech. The statistical methods are robust, but modifying the sampling and data collection protocol, and clearly stating how group (i.e. type of medication) will be handled in the analysis would improve this.

Thank you for highlighting specific aspects of our proposed methods that we can improve on. As suggested, we have revised the sampling and data collection protocol. We now plan to recruit participants at the local sites of the services and organizations they typically interact with and provide them with necessary digital equipment and technical support to mitigate digital exclusion. This is now explicitly stated in the following two parts of the methods section (manuscript, pp. 5 and 5-6; see also our responses to **Reviewer #1, comment # 3** and **Reviewer #2, comment #3**):

Potential participants were recruited at local sites of services and organizations that they were already interacting with (e.g., clinics, other treatment providers, interest groups, and low-threshold services) to maintain their privacy.

The digital survey format obviated the need to travel and enabled responses via a smartphone, tablet or computer while patients were situated in a natural and familiar setting (e.g., at home or at the local sites of services and organizations they interact with). Studies of eHealth services for OUD indicate high usability among patients, but that access to digital devices and network connection are perceived as the main barriers to use. To mitigate sampling bias due to digital exclusion, participants were able to request necessary digital equipment and technical support from staff or visiting study personnel to complete the digital survey.

Regarding the medication type, we will handle this variable in the analyses by means of dummy coding. The different medications will be grouped according to whether they are daily or long-acting formulations, and whether the active compound is methadone, buprenorphine, morphine, heroin or naltrexone. Individuals with opioid use disorder who are not currently in treatment will constitute its own group. This leaves us with 8 possible levels for the medication variable: 1) daily buprenorphine, 2) long-acting buprenorphine, 3) daily methadone, 4) daily morphine, 5) daily heroin, 6) daily naltrexone, 7) long-acting naltrexone, and 8) no medication. We now indicate that medications will be grouped according to formulation and active compound and handled by means of dummy coding in the methods section (manuscript, p. 10). This part now reads:

we used multinomial logistic regression implemented with the R package nnet to test the association between medication group (categorized and dummy-coded according to formulation and active compound; eAppendix II) and time use profile (dummy-coded)

The following details on how specific medications will be categorized have been added to eAppendix II (eSupplement, p. e4):

To facilitate statistical analysis without excluding patients who are receiving less common pharmacological treatments (e.g., daily oral levomethadone or long-acting implantable buprenorphine), we categorized participants into medication groups according to the formulation (i.e., daily or long-acting) and active compound (i.e., methadone, buprenorphine, morphine, heroin or naltrexone) of the medication they were receiving for opioid use disorder. These medication groups included 1) daily buprenorphine (i.e., oral buprenorphine

and buprenorphine/naloxone), 2) long-acting buprenorphine (i.e., injectable and implantable buprenorphine), 3) daily methadone (i.e., oral methadone and levomethadone), 4) daily morphine (i.e., 12- and 24-hour oral morphine), 5) daily heroin (i.e., injectable heroin), 6) daily naltrexone (i.e., oral naltrexone), 7) long-acting naltrexone (i.e., injectable naltrexone), and 8) no medication (i.e., currently not in opioid substitution treatment).

Reviewer #1, comment #6

Other points:

Abstract:

"Frees up considerable time" is a very vague statement when referring to benefits of long acting Buprenorphine. For who? In what way? Also amount of time does not necessarily = increased quality of time.

Thank you for pointing out that this statement was unclear. The statement referred to the amount of patients' time that would otherwise have been spent on seeking illicit opioids (if they are not in treatment) or collecting daily opioid substitution medications (if they are in treatment). We now specify this in the abstract (manuscript, p. 2). We agree that more time not necessarily equates to increased quality of time as time can be spent on activities that vary in how beneficial and detrimental they are to the individual's quality of life. We have therefore reworded the abstract so that statements about time use are neutral and thus in line with the non-directional hypotheses (see **Reviewer #2, comment #2**). The relevant part of the abstract (manuscript, p. 2) now reads:

Pharmacological treatment for opioid use disorder with new, long-acting medications (e.g., injectable and implantable buprenorphine) frees up a considerable amount of patients' time otherwise spent seeking illicit opioids or collecting daily opioid substitution medications. How much of this time patients treated with long-acting medications spend on activities related to recovery, substance use, and treatment is however currently unclear. Based on patients' hopes for and concerns about long-acting medications, we hypothesized that there is a relationship between medication type and time use, and between time use and well-being, experienced stigma and life satisfaction, in individuals with opioid use disorder.

Reviewer #2 (Chris Chambers)

Reviewer #2, summary

This is an interesting proposal to explore the potential effects of long-acting medications in opioid use disorder, focusing in particular on the time saving

experienced by patients and any consequent association with outcome measures including well-being, experienced stigma, and life satisfaction. I am not an expert in this topic and am providing a non-specialist review to evaluate the proposal in accordance with more general RR requirements. Overall, though, I thought it tackled a valid and interesting question using a useful approach (e.g. the use latent profile analysis is unusual for RRs).
I offer some comments below to help the authors refine and optimise their current plan.

Thank you for your comments and suggestions. These were indeed very helpful for optimizing our current plan for the study. Please see below for a detailed response to each comment.

Reviewer #2, comment #1
Perhaps my most substantive concern is that the design appears to lack a statistical sampling plan. The authors plan to recruit 500 patients (the minimum apparently required for the latent profile analysis), and their sampling plan is understandably limited by available resources; however, given the intended use of null hypothesis significance testing to evaluate hypotheses H2-H5, the design lacks the necessary sensitivity power analysis to confirm that this sample size is adequate. The authors really need to show what effect size this maximally obtainable sample size has sufficient power to detect for each of these predictions, and to then justify this minimally detectable effect accordingly (on a related note, the entries in the design table column "Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis" do not answer the question asked by this section of the table).

Thank you for encouraging us to run sensitivity power analyses to evaluate our sampling plan. Descriptions of these analyses and their results have been added to the methods sections (manuscript p. 10) and eAppendix II (eSupplement, pp. e4-e5) along with a justification for the minimally detectable effect. The R script is available on the Open Science Framework project page (https://osf.io/56z8y). In brief, the LR χ^2-tests of H2-H5 are powered to detect small-to-medium effect sizes when n = 500 (minimum Cohen's w between 0.14-0.19 depending on degrees of freedom). We were also asked by the ethics committee if the planned analyses would be feasible with a smaller sample size, in the event that we are not able to reach the target sample size during the data collection period. Considering that a sample size as small as n = 300 can be sufficient for identification of latent profiles with

latent profile analysis, we also conducted sensitivity power analyses with the sample size set to n = 300. In this case, the planned LR χ^2-tests of H2-H5 would still be sufficiently powered to detect small-to-medium effect sizes (minimum Cohen's w between 0.19-0.25 depending on degrees of freedom). Due to the stark difference in travel and surveillance requirements between treatment with daily medications and treatment with long-acting medications (i.e., once every week, month, or 6 months), it seems reasonable to expect relatively large effects on patients' time use and experience of well-being, stigma and life satisfaction, and we therefore consider the study to be sufficiently powered. The relevant new description of the sensitivity power analyses in the methods section (manuscript, p. 10) reads:

Depending on the final number of medication groups (2-8) and identified latent time use profiles (2-8), these LR χ2-tests have 90% power to detect a statistically significant minimum effect size of Cohen's w = 0.14-0.19 (i.e., a small-to-medium effect size) at α = 0.05 with the target sample size of n = 500. Even with a smaller sample size of n = 300, these tests are still powered to detect small-to-medium effect sizes (minimum Cohen's w = 0.19-0.25; eAppendix II).

Further details about the sensitivity power analyses are available in following part of eAppendix II (eSupplement, pp. e4-e5):

An adapted version of Cohen's w has been proposed as a χ^2-based effect size index for comparisons of nested models as it accounts for both the sample size and the difference in degrees of freedom between the full model (i.e., alternative model) containing the parameter of interest (i.e., explanatory variable) and the reduced model (i.e., null model) not containing this parameter. We therefore conducted sensitivity power analyses in R with the function pwr.chisq.test from the pwr package to verify that the target sample size of n = 500 would sufficiently power the omnibus likelihood ratio (LR) χ^2-tests of hypotheses 2-5 (i.e., main effects of medication group and time use profile). For these analyses, we set power to 0.90, α to 0.05, and difference in degrees of freedom to one less than the possible numbers of medication groups (i.e., 2-8; Δdf = 1-7) or identified latent time use profiles (i.e., 2-8; Δdf = 1-7).

Although our target sample size was n ≥ 500, the final sample size depended on what was feasible to collect within the one-year data collection period. A sample size as small as n = 300 can still be sufficient for latent profile analysis. We therefore also conducted similar sensitivity power analyses with n = 300 to verify that the planned LR χ2-tests would still be sufficiently powered to detect small-to-medium effect sizes (i.e., Cohen's w between

0.1-0.3) in the event that the target sample size would not be reached within the limited data collection period.

The sensitivity power analyses indicated that, depending on the degrees of freedom, the LR χ^2-tests would have 90% power at $\alpha = 0.05$ to detect a statistically significant minimum effect size of Cohens' w = 0.14-0.19 with n = 500, and Cohens' w = 0.19-0.25 with n = 300 (eFigure 1; eTable 1). In both cases, these are small-to-medium effect sizes6. Considering the stark difference in travel and supervision requirements between daily opioid substitution treatment and treatment with long-acting medications (e.g., once every week, month, or 6 months), it seems reasonable to expect relatively large effects on patients' time use, well-being, experience of stigma, and general life satisfaction. We therefore deemed the study sufficiently powered.

As requested, we have also revised the contents of the column "Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis" in the study design template (manuscript, pp. 20-24) to refer to the sensitivity power analyses and correctly answer the question asked by this section of the table.

Reviewer #2, comment #2

The hypotheses are non-directional, which suggests that there might be reasons to expect a negative association between long-acting medication use and the various outcome measures. Is this the case, and if so, could the authors explore the rationale for predicting positive or negative associations in the introduction? With my non-specialist hat on, it seemed intuitive for the hypotheses to be directional (predicting benefits), but perhaps there are reasons to expect effects in either direction.

Thank you for bringing to our attention that potential negative effects of long-acting medication use were underexplored in the introduction. While treatment with long-acting medications provides patients more spare time, it is not given that patients spend this time on activities that are beneficial for their well-being, life satisfaction and experience of stigma. It is possible that some patients will instead spend this time in social isolation or on other activities (e.g., illicit substance use) that could contribute to poor well-being and life satisfaction and help maintain their experience of stigma. We now explicitly refer to H2-H5 as non-directional and have updated the introduction section to highlight both potential positive and negative outcomes of long-acting medication use. The two relevant parts of the introduction section (manuscript, pp. 3 and 4) now reads:

Because of the drastically lower travel and supervision requirements, some patients are hopeful that treatment with long-acting medications will give them a more satisfactory life by reducing the stigma they experienced from frequently collecting medications and enabling them to spend more time on social, physical, educational, occupational and recreational activities. These activities are known to protect against the many mental and somatic health problems that often accompany opioid use disorder and that contribute to poor well-being (e.g., depression, anxiety, and chronic pain). However, patients have also voiced concerns about potentially reduced life satisfaction due to reduced contact with the healthcare system and difficulties adapting to the increase in spare time. The monitoring and social support offered by healthcare workers may encourage patients to stay abstinent and make positive lifestyle choices that benefit their well-being and life satisfaction. However, since the use of long-acting formulations obviates the need for frequent patient contact to administer medications, patients could potentially find themselves spending more time in social isolation and/or on previously discouraged activities (e.g., illicit substance use). In turn, this might help to maintain the experience of stigma, contribute to poor well-being, and ultimately reduce life satisfaction.

Next, we tested the non-directional hypotheses that there is a relationship between medication group and time use pattern (hypothesis 2), and between time use pattern and well-being (hypothesis 3), experienced stigma (hypothesis 4) and life satisfaction (hypothesis 5).

See also our response to **Reviewer #1, comment #6**.

Reviewer #2, comment #3

Data will be collected by digital format only. How severe a sampling bias does this introduce for this particular patient population? I think some additional discussion of this concern is warranted in the main text, ideally with reference to previous literature, e.g. considering the rate with which patients in this group have access to the necessary technical infrastructure to participate. I would also recommend including a Constraints on Generality section to ensure that all conclusions are appropriately calibrated in advance (see https://journals.sagepub.com/doi/10.1177/1745691617708630)

Thank you for asking this important question. Research on eHealth services for opioid use disorder indicates that access to necessary technical infrastructure is the main barrier to use, and it is therefore reasonable to assume that limited access to technical infrastructure will

also represent a barrier to participation in this digital survey-based study. As requested, we have added a brief discussion about digital exclusion (along with reference to previous literature) to the methods section (manuscript, pp. 4-5). We now also clarify that we will be recruiting participants at the local sites of the services and organizations they typically use (manuscript, p. 4) and that we will mitigate biased sampling due to digital exclusion by providing participants with the necessary digital devices and technical support to complete the survey (manuscript, pp. 4-5). The relevant parts of the methods section (manuscript, pp. 4 and 4-5) now read:

Potential participants were recruited at local sites of services and organizations that they were already interacting with (e.g., clinics, other treatment providers, interest groups, and low-threshold services) to maintain their privacy.

The digital survey format obviated the need to travel and enabled responses via a smartphone, tablet or computer while patients were situated in a natural and familiar setting (e.g., at home or at the local sites of services and organizations they interact with). Studies of eHealth services for OUD indicate high usability among patients, but that access to digital devices and network connection are perceived as the main barriers to use. To mitigate sampling bias due to digital exclusion, participants were able to request necessary digital equipment and technical support from staff or visiting study personnel to complete the digital survey.

See also our response to **Reviewer #1, comment #5**.

Some sampling bias may however still be present. The planned analyses will adjust for covariates and thus strengthen generalizability within the range of our sample. However, generalization to the general population of individuals with OUD will likely be more limited. To determine the representativeness of the sample for the overall population of individuals with OUD in Norway and to help identify constraints on generality, we will compare characteristics of the study sample against the near population-level values obtainable from the Norwegian Centre for Addiction Research's yearly status report on the Norwegian opioid substitution treatment program. This analysis is now introduced in the methods section (manuscript, p. 12), where we also explicitly state that the results will be interpreted in light of constraints on generality. The relevant part reads:

We followed current recommendations and interpreted the results in light of constraints on generality. To help identify constraints beyond inclusion criteria and material selection, we

used χ^2-tests to test for statistically significant divergence in key demographic and clinical characteristics of the current sample from near population-level values for individuals with OUD in Norway (eAppendix II).

Further details about this analysis have been added to eAppendix II (eSupplement, p. e6). This part reads:

Near population-level values for demographic and clinical characteristics of individuals with opioid use disorder in Norway were obtained from the Norwegian Centre for Addiction Research's latest yearly survey (see e.g., Nesse et al.) of all patients enrolled in the Norwegian opioid substitution treatment program (i.e., ~89% of all individuals with opioid use disorder in Norway). The characteristics to be compared between the study sample and the overall population with χ^2-tests included 1) the proportion of women, 2) the proportions of individuals aged < 31 years, 31-40 years, 41-50 years and > 50 years, 3) the proportion of individuals who are working or studying (derived from time spent on occupational and educational activities for the study sample; eAppendix III and IV), 4) the proportion of homeless individuals (derived from living situation for the study sample; eAppendix III and IV), 5) the proportions of individuals experiencing depression, anxiety and psychosis, and 6) the proportions of individuals who are satisfied, dissatisfied and neither satisfied nor dissatisfied with their treatment situation. To account for potential oversampling of patients treated with certain medications, we adjusted the sample proportions used in these tests according to the frequencies of patients receiving each medication in the population according to data from the Norwegian Centre for Addiction Research's latest yearly survey.

We have also updated the column "Interpretation given different outcomes" in the study design template (manuscript, pp. 20-24) with conclusions that are calibrated in advance according to constraints on generality. As a consequence of this, constraints on generality will naturally be discussed in the discussion section of the Stage 2 manuscript. See also our response to **Reviewer #1, comment #3**.

Reviewer #2, comment #4

Please provide a link to the eSupplement in the main text (https://osf.io/5evuh)

Thank you for pointing out that a link to the eSupplement was missing from the manuscript. We have updated each initial mention of eAppendix I (manuscript, p. 4), II (manuscript, p. 5), and III and IV (manuscript, p. 6) with a link to the latest version of the eSupplement on the Open Science Framework project page (https://osf.io/za7en).

Reviewer #2, comment #5

The Stage 1 manuscript is written in past tense but is also entered as Level 6 in the submission checklist. I assume this is just a stylistic convenience to minimise text changes at Stage 2 (which is fine) but please confirm for sure in the response to reviewers whether any of the data has (yet) been collected. If some data has already been collected, this must be stated in the Stage 1 manuscript and the bias control level adjusted down accordingly. If no data has been collected (and none will be until after IPA, which I assume is the case) then a comment to this effect in the response to reviewers will suffice.

Thank you for bringing to our attention that the status of data collection was unclear as a result of the Stage 1 manuscript being written in past tense. We can confirm that the past tense is indeed a stylistic convenience to minimize text changes at Stage 2 and that no data has been collected. To avoid confusion, we now explicitly state that the study was preregistered as a Stage 1 registered report prior to data collection. The relevant part of the method section (manuscript, p. 4) now reads:

The study was approved by the Regional Committees for Medical Research Ethics South East Norway (REK South East case No. 777516) and Sikt – Norwegian Agency for Shared Services in Education and Research ([insert case number]), and preregistered as a Stage 1 registered report prior to data collection (see Table 1 for the study design template).

We have also updated eAppendix II (eSupplement, p. e4) with the following explicit statement that the Stage 1 was written in past tense, but that no data were collected prior to IPA:

The Stage 1 report was written in past tense as a convenience to minimize text changes at Stage 2. No data were collected prior to in-principle acceptance.