

11th August 2022

PCI Registered Report: Stage 1 Resubmission

Dear Chris Chambers and PCI Review Panel,

Please find submitted our revised Stage 1 manuscript ‘*Does childhood adversity alter opioid drug reward? A conceptual replication in outpatients before surgery*’. We have made extensive changes to the manuscript following the helpful comments from the three Reviewers. We hope that the Stage 1 manuscript will now be considered for in-principle acceptance.

The overarching change to the manuscript is that we now specify two explicit confirmatory hypotheses that specifically relate to the conceptual replication of the previous study. Our first hypothesis (1) is that childhood adversity is associated with a greater opioid-induced mood boost (feeling good), and our second hypothesis (2) is that childhood adversity is associated with greater liking of the drug effects. We now specify one single predictor (childhood trauma score), and two main outcomes (*feeling good* and *drug liking*).

Other outcomes will be considered exploratory. *Anxiety relief* was not measured in the previous placebo-controlled study and will thus be considered exploratory, and as discussed in a preprint of the main effect of drugs on the operating table, [10.31234/osf.io/pq7dh](https://doi.org/10.31234/osf.io/pq7dh), the Norwegian translation of *feeling high* may be better considered a measure of intoxication and does not covary with measures of positive drug effects in our Norwegian population. We also removed socioeconomic status as a pre-specified predictor and will also consider this exploratory, since it was not measured in the prior study. Any exploratory analyses will be considered hypothesis-generating, and any significant effects from these analyses will be interpreted as preliminary.

Other overarching changes include greater specificity of the analysis plan, interpretations, and methods to increase robustness of the findings, in addition to the inclusion of simulated data and our analysis script. We have also added Martin Trøstheim as a contributing author, adjusted the author sequence, and made other minor edits to wording throughout (all in tracked-changes).

We have included the reviewed manuscript and a point-by-point response to the reviewer comments below. We also thank the Reviewers for taking the time to provide insightful suggestions for our manuscript.

Kind regards,

Dr Molly Carlyle

Postdoctoral Research Fellow
LABlab, University of Oslo
molly.carlyle@psykologi.uio.no

Your decision for round #1 : *Revision needed*
Major Revision – Response to Reviewers by Authors

by [Chris Chambers](#), 17 May 2022 09:03

Manuscript: https://osf.io/3vd4f/?view_only=4238d2ee3d654c4f908a94efea82a027

I now have three very helpful and constructive reviews of your submission. As you will see, the reviewers are broadly positive about the prospects of your manuscript, although some significant work will be needed to meet the [Stage 1 criteria](#) and achieve in-principle acceptance (IPA).

Among the main concerns are:

1. The logical coherence of the introduction and rationale, including making clear how reduced mu-opioid receptor density is relate to increased reward sensitivity (a point raised in slightly different ways by two of the reviewers).

We agree that the link between density and reward is not clear and we understand why this was raised by the reviewers. We have now removed the two sentences on opioid receptor density, since the link is not necessary for the aims of this study. We have also adjusted the end of the paragraph on page 3 to ensure flow, as follows:

~~“Heightened reward responses among animals with early adversity were also associated to reductions in mu-opioid receptor density, and a reduced analgesic response to the drug (8, 10). Reduced mu-opioid receptor expression after early life stress has also been reported for mice (11), and preliminary positron emission tomography (PET) evidence also linked reductions in resting mu-opioid receptor availability to insecure childhood attachment styles. There are several potential mechanisms to help explain this heightened reward response after early adversity. Panksepp (12) proposed that opioid drugs may mimic the pleasure experienced from caring social bonds by binding to the mu-opioid attachment circuitry, and that exposure to adverse social factors (such as isolation) may increase the desirability of opioids. Accordingly, this may be one explanation for an and that this enhanced pleasure response to opioids will may be greater among for those with limited early experiences of stable-caring social bonds in childhood. However, support for this theory has scarcely been translated from preclinical findings to humans.”~~

2. Considering the potentially confounding effects of expectancy.

We agree with the Reviewers that expectancies are critically important to consider in the context of drug effects. As the previous placebo-controlled study (Carlyle et al. 2021) included a placebo condition, and responses to the placebo injection were comparable in the two groups, differing expectancies cannot be the sole driver of altered subjective effects to opioids after trauma.

We do appreciate that the current study is within a different context, and expectancies cannot be ruled out as a contributing mechanism if we do find a significant effect

childhood trauma. We were particularly interested by Reviewer 1's comment regarding greater placebo analgesia in people with childhood trauma, and this is something we will include within the discussion of the Stage 2 manuscript. However, we believe the results would be informative in this medical context, even if expectancies may be part of the underlying mechanisms.

Lastly, on the operating table patients are told they will be given medication for pain, and medication for sleep. They are not told that it is an opioid or what type, which we hope will also reduce the impact of some expectancies. We have now added a sentence on this routine care instruction to patients into the manuscript methods (page 6), as follows:

“Patients were informed by the medical personnel that they would be given medication for pain and for sleep while on the operating table.”

3. Clarifying the precise details of the analysis plans and contingencies. For a revised manuscript, I would recommend generating and including analysis code on simulated data to verify suitability of the plans.

Following the several helpful comments from the reviewers, the analysis plan is now more detailed and precise with (a) two clear, testable primary hypotheses, (b) including the exact criteria we will use for analytical contingencies e.g., determining normality, and (c) included the simulated data and predefined analysis plan:

(a) A refined number of pre-specified analyses to focus on CTQ score on (1) feeling good, (2) drug liking. All other analyses will be considered exploratory.

While this overarching change is made throughout the manuscript document, the main changes are as follows.

Introduction page 4: *“Our primary hypothesis ~~is~~ ~~was~~ ~~ere~~ that patients with greater childhood adversity (higher trauma) ~~and lower socio-economic status scores~~ would 1) exhibit a larger mood boost (feeling good), and 2) paired with ~~express~~ greater liking of the drug effects ~~and feeling high~~ after the opioid administration, ~~conceptually replicating the previous findings.~~ ~~The~~ ~~We did not expect any effect on opioid disliking or feeling high in the minutes after infusion.~~ ~~For feeling high, translation the translation used was~~ ~~is~~ ~~was~~ ~~not~~ ~~deemed~~ ~~as~~ ~~a~~ ~~positive~~ ~~drug~~ ~~effect~~ ~~in~~ ~~a~~ ~~Norwegian~~ ~~population~~ ~~(17)~~-, and we did not expect any effect of childhood adversity on disliking or feeling high. Anxiety was not measured in the prior study, however ~~We~~ ~~A~~ ~~secondly~~ ~~secondary~~ ~~expected~~ ~~hypothesised~~ ~~was~~ ~~that~~ ~~childhood~~ ~~adversity~~ ~~would~~ ~~be~~ ~~associated~~ ~~with~~ ~~greater~~ ~~anxiety~~ ~~relief~~ ~~after~~ ~~opioid~~ ~~administration.~~ ~~Although~~ ~~Since~~ ~~anxiety~~ ~~is~~ ~~typically~~ ~~higher~~ ~~in~~ ~~people~~ ~~with~~ ~~childhood~~ ~~trauma~~ ~~and~~ ~~opioid~~ ~~use~~ ~~disorder~~ ~~(18)~~, and relief has been cited as a motivator for continued opioid use (19), we also explored the links between childhood trauma and anxiety pre- and post-drug.”*

Methods page 8-9: *“2.4.1 Primary analyses*

Multiple linear regressions were conducted ~~for the conceptual replication component of the analysis, to assess whether the primary predictor~~ The predictor variables for childhood adversity were childhood trauma (CTQ score), childhood SES (MSSS score), and a combined childhood adversity score that was calculated by standardising and computing the product of both CTQ and MSSS scores (where higher scores indicated higher trauma and lower SES). ~~was significantly positively associated with feeling good (H1), and drug liking (H2), disliking, and feeling high. Separate analyses were conducted for the predictor variables to assess both the independent (CTQ and SES) and combined effect on the outcomes, adjusted for demographic variables (age, gendersex), weight, opioid type, and surgery type were included in all analyses. The analyses forof feeling good and anxious were adjusted for the baseline pre-drug ratings by entering pre-drug responses as predictors in the regression, as thiseseey were was also measured before as well as after opioid administration. The regression equation for these analyses were:~~

$$\hat{Y}(\text{post-drug score}) = \beta^0 + \beta^1(\text{CTQ}) + \beta^2(\text{age}) + \beta^3(\text{sex}) + \beta^4(\text{opioid}) + \beta^5(\text{weight}) + \beta^6(\text{surgery}) + \beta^7(\text{pre-drug scores}) + \epsilon$$

Surgery type was categorical and dummy coded, where a regression coefficient was obtained for each level of the variable. Pre-drug scores in the regression equation were only relevant for feeling good. ”

- (b) **Provided greater precision on how we will determine normality, the criteria that will determine the need for bootstrapping, and how we will address missing data and outliers.**

Methods page 8: ~~“Prior to analyses, data were checked for n~~ Normality of residuals using the Shapiro-Wilk and Kolmogorov-Smirnov tests. Bootstrapping using random sampling with replacement (5000 iterations) was used if any of the two tests were significant (p>.01 for the Shapiro-Wilk and p>.05 for the Kolmogorov-Smirnov). were deemed acceptable if p>.01 and p>.05, respectively. The threshold for the standard Shapiro-Wilk test was adjusted due to overestimates of non-normality in samples when n>50 (31), and Tests were also followed by visually inspection of residuals using histograms and Q-Q plots to determine the nature of non-normality., in addition to assessing for o ~~Outliers were for the CTQ scores were assessed using boxplots. and missing data and were~~ Some extreme values were expected as there is typically a reduction in variation in CTQ scores for the moderate-severe range, however these will were be retained and reported. Extreme values were not expected for drug effect outcomes as these were bounded between 0 and 10 (11-point integers). Only p ~~Patients who have with a both pre and post-drug ratings for a given outcome and a CTQ score will were be included in that analysis. Patients with more >50% missing data for the one of the primary outcomes were was excluded from that analyses, Missing values were treated as missing. The alpha criterion for significance was p<.05~~

and p-values were corrected for multiple testing using the Holm-Bonferroni correction.”

(c) **Included simulated data and the planned analysis.** We have simulated hypothetical data and the analyses for the main analysis using means, standard deviations, and the expected distributions for the data to ensure that our analytical procedure is appropriate (submitted as a R Markdown html alongside the revised manuscript).

4. Clarifying the precise conditions that will confirm or disconfirm the predictions (which may entail the removal of redundant analyses). At present, the design plan does not sufficiently prespecify the conditions under which different conclusions will be drawn. This will require revision to both the main text and the study design table (while keeping the design table as succinct as possible).

We have removed redundant analyses and added the precise conditions for confirming/disconfirming the main analyses – both in the Analysis section and design table:

Analyses page 9: “The findings were interpreted as a full conceptual replication if both H1 and H2 were confirmed by a significant positive association between CTQ score with -post-drug feeling good and drug liking, or a partial conceptual replication if one of the two were significant. Regression coefficients (betas) were interpreted for effect size. For non-significant findings or significant associations in the opposite direction than hypothesised, we concluded that the conceptual replication was unsuccessful, and this was interpreted in the context of the methodological differences including: CTQ as a continuous measure instead of pre-stratified groups with fewer people in the moderate-severe range, and differences in drug type, dose, administration route, and the context of use (open-label surgery setting compared to placebo-controlled research study).”

5. Clarification of the level of bias control in the manuscript. In the submission checklist you selected Level 2: At least some data/evidence that will be used to answer the research question has been accessed and partially observed by the authors, but the authors certify that they have not yet observed the key variables within the data that will be used to answer the research question AND they have taken additional steps to maximise bias control and rigour (e.g. conservative statistical threshold; recruitment of a blinded analyst; robustness testing, multiverse/specification analysis, or other approach). Please add a section to the manuscript that makes clear the level of prior data observation that has taken place (and confirms the corresponding level of bias control achieved under the PCI RR taxonomy). The second part of the Level 2 definition does not appear to be tackled in your plans: additional steps to maximise bias control and rigour (e.g. conservative statistical threshold; recruitment of a blinded analyst; robustness testing, multiverse/specification analysis, or other approach). This will need to be comprehensively addressed to achieve IPA.

We have now added clarification of the level of bias to the end of the analysis section, in addition to how we intend to maximise bias control and rigour:

Page 10:

“2.5 Level of bias and control

As a registered prospective analysis, we have designated a Level 2 bias control because the wider dataset (n = 269) has been acquired and partially observed as part of the broader research project (17). However, the main predictor, CTO scores, and the exploratory variables, have not been accessed or observed, nor do any of the authors know which individuals make up the subset of participants (n = 155, 71%) that provided data for the current analysis. Steps to reduce bias include: (i) The submission of the pre-specified analysis script to provide transparency on the analytical plan and contingencies before this data has been observed; (ii) calculating the posterior probabilities using a Bayesian framework to assess the robustness of the results; (iii) using the Holm-Bonferroni alpha correction on the confirmatory tests; (iv) ensuring the lead authors of the manuscript responsible for analysis have had limited exposure to the data that has already been accessed as part of the broader research project.”

In line with adding the Bayesian component for robustness, we have added the details of this in the analysis:

Page 9: “Bayesian posterior probabilities were calculated to assess the robustness of the findings, using the ‘rstan’ (32) and ‘rethinking’ packages (33). Quadratic approximation was used to calculate the posterior probabilities [outcome_i ~ Normal(μ_i , σ)] for the centered linear relationships with CTO score [$\mu_i = \alpha + \beta(\text{CTO}_i - \bar{x})$]. Priors were constructed using the means, standard deviations and slopes from the previous study (13) and were tested using prior predictive simulations, with drug liking: $\alpha \sim \text{Normal}(30,15)$, $\sigma \sim \text{Uniform}(0,20)$ and $\beta \sim \text{Normal}(0,1)$; and feeling good (measured as ‘euphoria’ in the previous study): $\alpha \sim \text{Normal}(20,10)$, $\sigma \sim \text{Uniform}(0,10)$ and $\beta \sim \text{Normal}(0,1)$. The posterior mean, 89% credible interval, and 89% highest posterior density intervals (HPDI; the narrowest interval containing 89% of the probability mass) for betas were reported alongside each regression. Posterior predictive checks were also conducted to assess the reliability of the Bayesian models. Such that the results from the Bayesian analysis did not concur with the frequentist analysis, potential reasons for the lack of robustness were discussed.”

Overall, I believe the manuscript is sufficiently promising to invite a Major Revision. Your proposal addresses a scientifically valid question, and (from my own reading) strikes me as a innovative and valuable use of pre-existing data. Should you wish to revise, please ensure that you respond comprehensively to all of the issues raised above and in the reviews, including a point-by-point response to every comment of the reviewers, and a fully tracked-changes version of the revised manuscript.

Thank you, we have found the Reviewer comments very helpful for improving this manuscript submission and we hope that we have sufficiently addressed all comments.

Review 1 by anonymous reviewer, 10 May 2022 20:40

In this manuscript, the authors describe a study in which they explore to what extent childhood adversity predicts acute subjective responses (“reward”) to mu-opioid agonists administered in a medical setting. This is a very interesting and important topic, a nice follow-up from the authors’ previous study, and well-written start to a manuscript. The study has good scientific validity, and the hypotheses seem rational. However, there are some small matters that require clarification as described below.

Thank you for your interest in our study, and for your insightful comments on our Stage 1 manuscript.

Introduction

1. In the introduction, the authors state that early adversity is associated with reduced mu-opioid receptor density. It is not clear, however, how reduced mu-opioid receptor density relate to increased reward after exogenously administered opioids?

Thank you for raising this point. We agree that the link between density and reward is not clear and we understand why this was raised by the reviewers. We have now removed the two sentences on opioid receptor density, since the link is not necessary for the aims of this study. We have also adjusted the end of the paragraph on page 3 to ensure flow, as follows:

~~“Heightened reward responses among animals with early adversity were also associated to reductions in mu-opioid receptor density, and a reduced analgesic response to the drug (8, 10). Reduced mu-opioid receptor expression after early life stress has also been reported for mice (11), and preliminary positron emission tomography (PET) evidence also linked reductions in resting mu-opioid receptor availability to insecure childhood attachment styles. There are several potential mechanisms to help explain this heightened reward response after early adversity.- Panksepp (12) proposed that opioid drugs may mimic the pleasure experienced from caring social bonds by binding to the mu-opioid attachment circuitry, and that exposure to adverse social factors (such as isolation) may increase the desirability of opioids. Accordingly, this may be one explanation for an and that this enhanced pleasure response to opioids will may be greater among for those with limited early experiences of stable caring social bonds in childhood. However, support for this theory has scarcely been translated from preclinical findings to humans.”~~

2. The authors write, “Here, we examined whether childhood adversity increases risk of opioid misuse via enhanced positive drug effects.” Are the authors actually planning to measure opioid misuse? Otherwise, this statement should be modified, as it does not accurately describe the experimental question.

Thank you for this point, we agree that although there are implications for opioid misuse, this study does not examine this. We have now adjusted the wording, as follows:

“Here, we examined whether childhood adversity increases ~~risk of opioid misuse via positive drug effects of opioids given in a medical context.~~ Positive drug effects are considered a sign of higher abuse liability (5). ~~and these results could may help ensure at risk patients receive the best practice treatment, and informing more individualised approaches to the prescribing of opioid analgesics.~~”

3. In the intro, describe mechanism of action of two drugs (ie do they act as pure mu agonists? How do the doses used compare to one another?).

Thank you for this point. We have now added more specific details on the pharmacological action and medical use of both drugs in the study in the introduction and methods:

Introduction Page 4: “Both drugs are opioid agonists that are fast-acting and primarily stimulate the μ -opioid receptor subtype, and are frequently used both pre- and post-operatively to provide quick and effective pain relief.”

Opioid administration section, methods Page 6: “Both opioids led to comparable subjective intoxication, as reported in the broader research trial (18).”

Effects of the two drugs has been explored as part of the broader research trial and is reported in this pre-print: [10.31234/osf.io/pq7dh](https://doi.org/10.31234/osf.io/pq7dh) . While there are some differences in subjective responses to the two opioids, the current analysis is a conceptual replication of a previous study that uses a different opioid (morphine). For this study we are not planning to compare the two opioids, but will rather adjust for opioid type in the analysis, and include interpretation of this in context of the results in the discussion.

Methods:

4. One potential difficulty with this design is that it is not clear what role expectancy effects play in this study. What were the patients told about the medication they would be receiving? Did they know they would be receiving an opioid? There is some evidence that childhood adversity predicts placebo response in the context of analgesia, so one concern is that differences in expectancy effects between subjects with low and high adversity could confound the analysis.

We agree with the Reviewers that expectancies are critically important to consider in the context of drug effects. As the previous placebo-controlled study (Carlyle et al. 2021) included a placebo condition, and responses to the placebo injection were comparable in the two groups, differing expectancies cannot be the sole driver of altered subjective effects to opioids after trauma.

We do appreciate that the current study is within a different context, and expectancies cannot be ruled out as a contributing mechanism if we do find a significant effect childhood trauma. We were particularly interested by your comment regarding greater placebo analgesia in people with childhood trauma, and this is something we will include within the discussion of the Stage 2 manuscript. However, even if

expectancies drive subjective differences in pre-operative context with no placebo condition, this is still critical to determine from a clinical and translational perspective.

Lastly, on the operating table patients are told they will be given medication for pain, and medication for sleep. They are not told that it is an opioid or what type, which we hope will also reduce the impact of some expectancies. We have now added a sentence on this routine care instruction to patients into the manuscript methods (page 6), as follows:

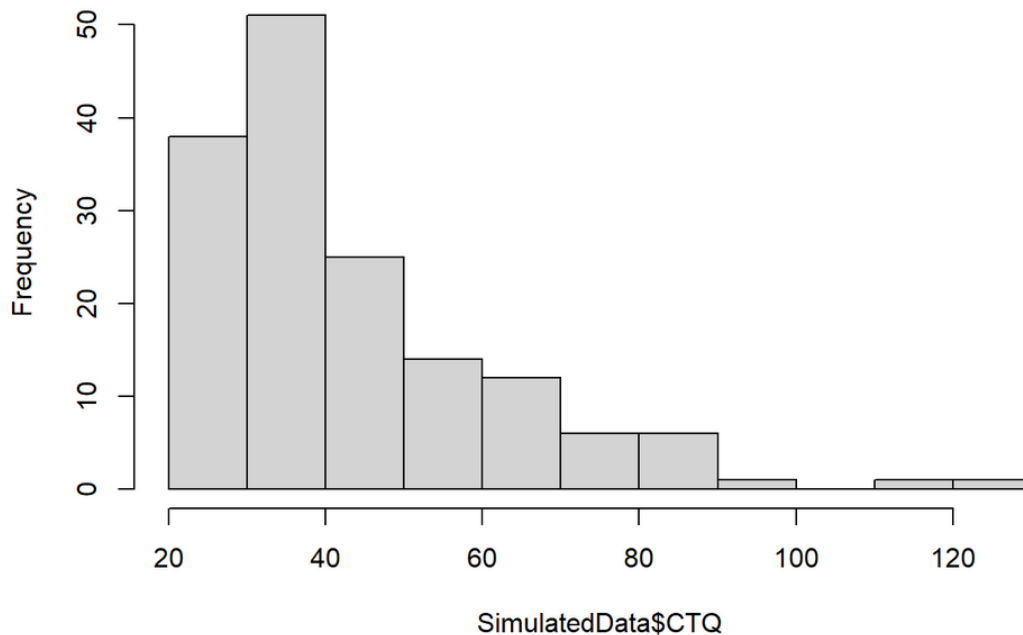
“Patients were informed by the medical personnel that they would be given medication for pain and for sleep while on the operating table.”

5. Another potential pitfall is that the authors’ sample of patients will not have sufficient variability on the CTQ to be able to conduct the planned analyses. Presumably most participants will not have any history of childhood trauma. How will the authors ensure a substantial enough range on the CTQ to obtain meaningful results?

This is an extremely important point regarding the feasibility of the analysis, and something we have considered to a great extent. We know that the childhood trauma scores will be highly positively skewed as severe childhood trauma is less prevalent. Nonetheless, we do expect enough variation to explore the research question for the following reasons:

(I) We have access to data collected in 1362 people who signed up to take part in psychopharmacology studies targeted at healthy people (collected in the Chicago area in the USA). This data has not yet been published, however we used the mean, standard deviation, and log distribution of this dataset to simulate a dataset for 155 people. We then used this simulated data to run our planned analyses (submitted with this revision). A histogram of the expected variation of childhood trauma is provided below:

Histogram of SimulatedData\$CTQ



(2) This distribution fits with the reports in the literature for the prevalence of childhood trauma:

- A UK-based study reported a 5-16% prevalence of severe childhood trauma, depending on the subdomain (<https://doi.org/10.1016/j.chiabu.2004.05.009>).
- A Norwegian study reported 9% of people scored above moderate on least one subdomain of childhood trauma among mentally-healthy people, and 51% in people with a mental health problem (<https://doi.org/10.3389/fpsyg.2017.01276>).
- An Australian study led by the lead author (MC) in a non-clinical sample of young people who use cannabis also reported 9% scored in the severe range, 17% moderate, 28% low, and 46% with none (<https://doi.org/10.1016/j.schres.2021.10.011>).

We therefore do expect enough variation in CTQ scores to conduct the planned analyses, and will consider the distribution of trauma scores in the interpretation of the results. We have also commented on this within the analysis section of the manuscript:

Analyses page 9: ‘~~Outliers were for the CTQ scores were assessed using boxplots, and missing data and were~~ Some extreme values were expected as there is typically a reduction in variation in CTQ scores for the moderate-severe range, however these will be retained and reported.’

6. For the future submission: in the analysis, the group that got remifentanyl and the group that got oxycodone should be compared on subjective effects to make sure that the doses of the different drugs were matched on this metric.

This is an important point. Assessment of drug effects has already been conducted as part of the broader research trial in the larger sample of 269 people (preprint: [10.31234/osf.io/pq7dh](https://doi.org/10.31234/osf.io/pq7dh)). Mood and drug effects were largely comparable, although the ratio of drug liking to disliking was higher for oxycodone than for remifentanyl. To account for variance due to drug type, we will adjust for drug type in the analyses.

Review 2 by [Yuki Yamada](#), 06 May 2022 05:36

I read this study with interest even though I am not a complete expert on the topic, as it attempts to test as a natural experiment the hypothesis that human childhood detrimental situations, for which indirect evidence has been accumulated in very controlled situations, are associated with later opioid effects. Since this is an observational study in a natural setting, I am conservative as to whether the authors can draw conclusions about causality in their hypothesis here, but there is no doubt that the present study will still provide useful findings. Below is a list of points that I believe should be addressed in advance for a better protocol.

Thank you for your informative comments regarding our analysis and your interest in our research question. We have now attempted to address your comments below.

1. I understand that the present study is designed to analyze data that already exists. In such cases, I think the authors need to be clarified as to how much specific knowledge of the data they have. PCI RR has a set level of bias control, so please refer to that.

Thank you for raising this point. We have now added clarification to the end of the analysis section to clarify how much knowledge we have on the data, and how we intend to maximise bias control and rigour:

Page 10:

“2.5 Level of bias and control

As a registered prospective analysis, we have designated a Level 2 bias control because the wider dataset (n = 269) has been acquired and partially observed as part of the broader research project (17). However, the main predictor, CTO scores, and the exploratory variables, have not been accessed or observed, nor do any of the authors know which individuals make up the subset of participants (n = 155, 71%) that provided data for the current analysis. Steps to reduce bias include: (i) The submission of the pre-specified analysis script to provide transparency on the analytical plan and contingencies before this data has been observed; (ii) calculating the posterior probabilities using a Bayesian framework to assess the robustness of the results; (iii) using the Holm-Bonferroni alpha correction on the confirmatory tests; (iv) ensuring the lead authors of the manuscript responsible for analysis have had limited exposure to the data that has already been accessed as part of the broader research project.”

2. Are there any findings from previous studies that socioeconomic status affects opioid misuse, and so on? I think it needs to be explicitly explained why the authors are focusing on SES here. Furthermore, it is unclear at what point in the subjects' "childhood" the authors expect SES to have an effect, and it is also unclear about what

time of SES the subjects will respond about theirs. As SES can vary over time, shouldn't this point be specified specifically?

Thank you for raising this point. We have now moved the analyses using SES to exploratory to reduce the number of pre-specified analyses we conduct, and because there is currently no known existing evidence linking childhood SES with subjective effects. The initial hypothesis was guided by a large amount of evidence linking childhood SES with later drug misuse (e.g., <https://doi.org/10.1111/j.1465-3362.2008.00042.x>). However, given that we are assessing subjective effects (and not opioid misuse), we realise that it is a large leap to make.

Participants were asked to provide two estimates of SES, one from their childhood and one current estimate. The planned exploratory analyses focus on their estimate of childhood SES.

3. There are two predictor variables, CTQ and MSSS, which will not always show consistent results. In what cases does this mean that the hypothesis is supported?

MSSS score has now been removed as a primary predictor and will only be examined in the exploratory analyses

4. Data are examined in a number of ways before multiple regression analysis, but I am not sure how and when each of these is determined. For example, the Shapiro-Wilk test and the Kolmogorov-Smirnov test do not return the same results, and the criteria for visual judgments are unclear and can be arbitrary. As for outliers, there is no indication of how to detect them.

Good point! We have now specified the decision steps in the analyses, as follows:

Page 8: "~~Prior to analyses, data were checked for n Normality of residuals using the Shapiro-Wilk and Kolmogorov-Smirnov tests. Bootstrapping using random sampling with replacement (5000 iterations) was used if any of the two tests were significant ($p > .01$ for the Shapiro-Wilk and $p > .05$ for the Kolmogorov-Smirnov). were deemed acceptable if $p > .01$ and $p > .05$, respectively. The threshold for the standard Shapiro-Wilk test was adjusted due to overestimates of non-normality in samples when $n > 50$ (31), and Tests were also followed by visually inspection of residuals using histograms and Q-Q plots to determine the nature of non-normality., in addition to assessing for θ Outliers were for the CTQ scores were assessed using boxplots. and missing data and were Some extreme values were expected as there is typically a reduction in variation in CTQ scores for the moderate-severe range, however these will were be retained and reported. Extreme values were not expected for drug effect outcomes as these were bounded between 0 and 10 (11-point integers). Only p Patients who have with a both pre and post-drug ratings for a given outcome and a CTQ score will were be included in that analysis. Patients with more $> 50\%$ missing data for the one of the primary outcomes were was excluded from that analyses; Missing values were treated as missing. The alpha criterion for significance was $p < .05$ and p -values were corrected for multiple testing using the Holm-Bonferroni correction."~~

5. In multiple regression analysis, it is stated that another regression analysis is performed using the product of two predictor variables, but I did not understand this clearly. This is stated to examine the "combined effect," but I thought this was usually done to examine interactions. What is this "combined effect"? By what background is the combined effect hypothesized? Also, since there are only two predictors, I thought that this could be discussed to some extent just by looking at the multiple correlation coefficient and the coefficient of determination, but is there any reason not to? Also, what would be the (single?) correlation analysis between this combined variable and the outcome variables? Is it a partial correlation analysis? What is the interpretation if there is no significant effect of individual variables and only the combined effect is significant, or vice versa?

For clarity, we have now removed the SES and 'combined effect' as primary predictors and will instead examine SES in the exploratory analyses (as outlined in response to your points 2-3). We no longer plan to assess the 'combined effect'.

6. The description of the baseline is ambiguous and it is unclear how it is to be set.

Thank you for this point. We have adjusted the wording from 'baseline' to 'pre-drug scores' throughout the manuscript, and attempted to increase clarity on when these pre-drug scores are collected, e.g., as follows:

Introduction page 4: "~~Patients~~ were asked to give verbal numerical ratings of mood immediately before and 1 minute after drug infusion."

Methods page 6: "Immediately prior to opioid administration, patients were asked by the medical personnel to verbally rate their mood for: (i) how good they felt; and (ii) how anxious they felt, on a scale from 0-10 (0 - not at all, 10 - very much) (pre-drug scores)."

Analysis page 8: "The analyses of mood items (feeling good and anxious) adjusted for the ~~baseline-pre-drug~~ ratings by entering pre-drug responses as predictors in the regression, as the ~~ses~~ were measured before as well as after opioid administration."

7. How do the results of a multinomial logistic regression analysis for changes in mood ratings support the hypothesis? Also, how do you reconcile results that are inconsistent with the multiple regression analysis that preceded it?

Thank you for this point, we have removed this analysis altogether following recommendations by Reviewer 3, point 6.

8. Two types of opioid analgesics are used, does this difference affect the testing of the hypothesis? If so, I think it needs to be clearly stated in the manuscript.

This is an important point. While there are reported differences in subjective responses to the two opioids, the current analysis is a conceptual replication of a previous study that uses a different opioid (morphine). For this study we are not planning to compare the two opioids, but will rather adjust for opioid type in the analysis, and include interpretation of this in context of the results in the discussion.

Furthermore, mood and drug effects have already been assessed as part of the broader research trial in the larger sample of 269 people (preprint: [10.31234/osf.io/pq7dh](https://doi.org/10.31234/osf.io/pq7dh)). These effects were largely comparable, although the ratio of drug liking to disliking was higher for oxycodone than for remifentanyl. To account for variance due to drug type, we include this information in all analyses.

Many of the questions I have raised here about analysis would not be particularly problematic if it were all exploratory analysis. However, if it is to be registered as a confirmatory analysis, please clarify each hypothesis and the criteria for evaluating and interpretation of the results.

We appreciate this point and the updated report only includes the key confirmatory analyses needed for replication of the prior results, with other analyses now explicitly considered exploratory.

Review 3 by [Zoltan Dienes](#), 02 May 2022 13:41

This is a very interesting study making good use of a naturalistic situation to look at whether childhood adversity affects how people respond subjectively to opioids. I didn't see any discussion of how bias is controlled, but I will presume the editor has this in hand.

Thank you for your interest and helpful comments for our study, and comments regarding the improvement of analytic flexibility. For the control of bias, we have now added clarification to the end of the analysis section, in addition to how we intend to maximise bias control and rigour:

Page 10:

“2.5 Level of bias and control

As a registered prospective analysis, we have designated a Level 2 bias control because the wider dataset (n = 269) has been acquired and partially observed as part of the broader research project (17). However, the main predictor, CTQ scores, and the exploratory variables, have not been accessed or observed, nor do any of the authors know which individuals make up the subset of participants (n = 155, 71%) that provided data for the current analysis. Steps to reduce bias include: (i) The submission of the pre-specified analysis script to provide transparency on the analytical plan and contingencies before this data has been observed; (ii) calculating the posterior probabilities using a Bayesian framework to assess the robustness of the results; (iii) using the Holm-Bonferroni alpha correction on the confirmatory tests; (iv) ensuring the lead authors of the manuscript responsible for analysis have had limited exposure to the data that has already been accessed as part of the broader research project.”

My main point is that there is still plenty of scope for analytic flexibility. Specifically:

1. Normality is to be checked in a range of ways. Under what conditions will normality be presumed good enough to proceed? If it is not good enough, what will be the exact bootstrapping procedure?

Good points! This was similarly raised by Reviewer 2 point 4, we have now specified the decision steps in the analyses and we hope we have now outlined the exact conditions as follows:

Page 8: “~~Prior to analyses, data were checked for n~~Normality of residuals using the Shapiro-Wilk and Kolmogorov-Smirnov tests. Bootstrapping using random sampling with replacement (5000 iterations) was used if any of the two tests were significant ($p > .01$ for the Shapiro-Wilk and $p > .05$ for the Kolmogorov-Smirnov). ~~were deemed acceptable if $p > .01$ and $p > .05$, respectively. The threshold for the standard Shapiro-Wilk test was adjusted due to overestimates of non-normality in samples when $n > 50$ (31).~~ and Tests were also followed by visually inspection of residuals using histograms and Q-Q plots to determine the nature of non-normality. ~~in addition to assessing for o~~Outliers ~~were for the CTQ scores were assessed using boxplots. and missing data and were~~Some extreme values were expected as there is typically a reduction in variation in CTQ scores for the moderate-severe range, however these ~~will were be~~retained and reported. Extreme values were not expected for drug effect outcomes as these were bounded between 0 and 10 (11-point integers). ~~Only p~~Patients who have with a both pre and post-drug ratings for a given outcome and a CTQ score ~~will were be~~included in that analysis. Patients with more $> 50\%$ missing data for the one of the primary outcomes were ~~was~~excluded from that analyses. Missing values were treated as missing. The alpha criterion for significance was $p < .05$ and p-values were corrected for multiple testing using the Holm-Bonferroni correction.”

We have also included our script in an RMarkdown html that includes the bootstrapping procedure for the analyses.

2. Childhood adversity is to be measured using three IVs. If any one is significant in predicting a DV, will there be presumed to be a relationship between adversity and that DV? This gives one three shots at that conclusion. Either pick one main predictor or adjust with Bonferroni (etc) - and adjust the power calculation accordingly.

Thank you for highlighting this, following these helpful reviews have we have now decided to limit our researcher degrees of freedom by reducing and refining both the number of IVs and DVs in the confirmatory section of the manuscript. We will now only use one predictor (childhood trauma scores) and two outcomes: (i) feeling good (ii) drug liking. All other analyses will be considered exploratory. These exact changes are summarised in response to Point 3 by Chris Chambers.

3. Specify exactly how demographic variables will be coded.

We have now added this to the measures section of the manuscript (page 7) as follows:

“Demographic data such as age (years), sex (male, female), and weight (kg), were collected, in addition to opioid type (oxycodone, remifentanyl), and surgery type (categorical and dummy coded).”

4. Specify exactly how ratings will be adjusted for baseline - e.g. will baseline ratings be entered as IVs?

Baseline ratings, which we now call pre-drug ratings for clarity, will be entered as IVs into the regression to control for them. We have now added this to the manuscript page 8:

“The analyses ~~for~~ feeling good ~~and anxious were~~ adjusted for the baseline-pre-drug ratings by entering pre-drug responses as predictors in the regression, as ~~thisesey were was also~~ measured before as well as after opioid administration.”

We have also included these in the regression equation below.

5. For clarity, specify the full regression equation that will be used.

We have now added the following regression equation into the manuscript (analysis page 8):

“The regression equation for these analyses were:

$$\hat{Y}(\text{post-drug score}) = \beta^0 + \beta^1(\text{CTQ}) + \beta^2(\text{age}) + \beta^3(\text{sex}) + \beta^4(\text{opioid}) + \beta^5(\text{weight}) + \beta^6(\text{surgery}) + \beta^7(\text{pre-drug scores}) + \epsilon$$

Surgery type was categorical and dummy coded, where a regression coefficient was obtained for each level of the variable. Pre-drug scores in the regression equation were only relevant for feeling good.”

6. A lower-powered back up analysis is suggested by collapsing change scores into three categories. This gives another shot at the cherry. I suggest deleting this analysis.

Thank you for this suggestion, we have now removed this analysis from the manuscript.

7. Subjective effects will be measured in three different ways (feeling good, liking, feeling high). This gives three shots at getting the effect. I suggest averaging these ratings together (or else adjusting familywise error rate). Averaging will increase the reliability of the measure and give more power to detect a given raw effect size (i.e. difference in ratings units).

It is important for comparison and generalisation of these findings to keep these outcomes separate. Although both may reflect positive drug effects, they do not operate exactly the same way (as observed in the original Carlyle et al., paper, where ratings are different between ‘drug liking’ and ‘euphoria’). Also, as mentioned we are now removing the ‘High’ subscale since it does not function or translate in the same way in a Norwegian population, so there are two main tests.

Thus, from a psychopharmacological perspective there are important nuances related to the outcomes ‘feeling good’ and ‘liking effects’ that we would like to assess separately. While these measures are typically significantly correlated the association

is fairly weak (e.g. Kendall's $\tau = .148$). We now specify how we will interpret a significant finding on none, one or both of the replication tests, and adjust for multiple testing using the Holm-Bonferroni correction.

However, we do appreciate that using single-items may be less reliable, and welcome the suggestion of increasing power to detect differences. We will discuss the similarities between these outcomes in light of the results within the discussion section of the paper.

8. Determine what difference in rating units would be just meaningful, given the purpose to which the study could be used. How many units of feeling high is enough to care about? Put another way, a previous study found the bottom limit if the 95% CI for euphoria was 7 units on a 100 point scale. This corresponds to 0.7 units on a 10 point scale. Is this still enough to care about? (See p 10 here: <https://psyarxiv.com/yc7s5/>). If so, the fact that it is the bottom of a CI could be used to indicate it is roughly the lower limit of what is plausible; and if it is an effect one would care about, it is a minimal meaningful effect size that is just plausible. That means it is appropriate to be the effect size used for a power analysis. Note when converting from a raw to a standard effect size, take into account if the DV is averaged, which will increase the standardized effect size for a given raw effect size.

This is an interesting point that we have given a lot of consideration. Because the study is not assessing a clinically-meaningful measure (e.g., the use of opioids) but rather a subjective measure that may have a relationship with later use (liking the effects of opioids). We know that subjective measures such as liking are predictive of drug use behaviour among people with drug experience or dependency (e.g., <https://doi.org/10.1016/j.pain.2012.07.035>), but we do not know what this means clinically in generally healthy populations - including in pre-operative contexts such as the current study. We consider any significant link between childhood trauma and positive responses to opioid analgesic important because it may relate to cumulative risk for misuse.

More broadly, this is a relatively unexplored area for subjective drug effects. To our knowledge, only one study has attempted to define clinically meaningful differences in ratings of feeling high among opioid-experienced, non-dependent men (<https://doi.org/10.1007/s11136-011-0012-7>). This study reported the lowest clinically-relevant changes between 5.73-8.54 (out of 100) in relation to drug use; the difference scores in the study we attempt to replicate here were larger than this suggested minimum clinically relevant effect.

Of course we will still report and comment on the magnitude and robustness of the effects, and discuss the potential clinical implications.

Minor point from Introduction: Why would a reduction in mu-opioid receptor density create heightened reward sensitivity (as it is associated with a reduced analgesic response to the drug)?

Thank you for raising this point. We agree that the link between density and reward is not clear and we understand why this was raised by the reviewers. We have now removed the two sentences on opioid receptor density, since the link is not necessary

for the aims of this study. We have also adjusted the end of the paragraph on page 3 to ensure flow, as follows:

~~“Heightened reward responses among animals with early adversity were also associated to reductions in mu-opioid receptor density, and a reduced analgesic response to the drug (8, 10). Reduced mu-opioid receptor expression after early life stress has also been reported for mice (11), and preliminary positron emission tomography (PET) evidence also linked reductions in resting mu-opioid receptor availability to insecure childhood attachment styles. There are several potential mechanisms to help explain this heightened reward response after early adversity. Panksepp (12) proposed that opioid drugs may mimic the pleasure experienced from caring social bonds by binding to the mu-opioid attachment circuitry, and that exposure to adverse social factors (such as isolation) may increase the desirability of opioids. Accordingly, this may be one explanation for an and that this enhanced pleasure response to opioids will may be greater among for those with limited early experiences of stable-caring social bonds in childhood. However, support for this theory has scarcely been translated from preclinical findings to humans.”~~

Zoltan Dienes