Leknes Affective Brain Lab
Department of Psychology
University of Oslo
Forskningsveien 3A, 0373 Oslo

molly.carlyle@psykologi.uio.no

27th September 2022

PCI Registered Report: Stage 1 Resubmission, Round 2

Dear Recommender Chris Chambers and PCI Review Panel,

Please find submitted the revised Stage 1 manuscript (round 2) '*Does childhood adversity alter opioid drug reward? A conceptual replication in outpatients before surgery*'.

We hope that we have now fully addressed the comments on the manuscript and analysis plan, and our point-by-point response is provided below. We have made particular effort to address Zoltan Dienes' point 3 and 5. For point 3 regarding the addition of the Bayesian analyses, the Bayesian estimates are still included in the manuscript as we still see value in these in addition to the frequentist analyses. We have included greater detail on the reasons and specific interpretations for these, however we can remove these analyses if the Reviewers do not agree with our approach. We have also changed the author order in the manuscript.

We hope that the manuscript is now ready for In Principle Acceptance, and again would like to thank the Reviewers for taking the time to provide insightful suggestions for our Stage 1 manuscript and analysis plan.

Kind regards,

Dr Molly Carlyle

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

Round #2

Author's Reply:

---

by Chris Chambers, 25 Aug 2022 13:53
Manuscript: **https://osf.io/xr2vb/?view_only=4238d2ee3d654c4f908a94efea82a027** version v2

Minor Revision
The three reviewers who assessed your initial submission have now evaluated the revised manuscript, and the good news is that we are getting close to Stage 1 acceptance. You will find some remaining methodological points to address in two of the reviews, including a key point about streamlining the analysis (and consequentially the logical chain of inference), and the suggestion to remove exploratory analyses from the Stage 1 manuscript (with which I agree).

I will consult swiftly with Zoltan Dienes concerning your further revised submission to ensure that his points have been adequately addressed (especially his points 3 and 5, which are most important).

Reviewer 1
Reviewed by anonymous reviewer, 20 Aug 2022 15:58
The authors provided a thoughtful consideration of, and response to, all of the concerns raised.

> *Thank you. We appreciate the time you have taken to review and help improve our Stage 1 manuscript.*

Reviewer 2
Reviewed by Yuki Yamada, 13 Aug 2022 05:49
I would like to thank the authors for revising the manuscript based on the review comments. My opinion is that IPA could be granted for this proposed revised plan.

> *Thank you for taking the time to review our Stage 1 manuscript, we appreciate your previous suggestions and have also now addressed your additional points below.*

The following points are minor and should be confirmed by the recommender:

1. In multiple regression equations, β usually represents the partial regression coefficient, and x etc. would represent predictor variables. Perhaps the brackets themselves may represent the predictor, but \(\hat{Y} \) also contains a bracketed name, which can be confusing, so I think it would be better to write it in the least misleading way possible.

   Thank you for raising this, we have now made adjustments in line with your suggestion, as follows:

Page 8: "~~$\hat{Y}$(post-drug score) = β^0 + β^1(CTQ) + β^2(age) + β^3(sex) + β^4(opioid) + β^5(weight) + β^6(surgery) + β^7(pre-drug score) + ϵ~~ $\hat{Y}_i = \beta_0 + \beta_1 CTQ_i + \beta_2 Age_i + \beta_3 Sex_i + \beta_4 Opioid_i + \beta_5 Weight_i + \beta_6 Surgery_i + \beta_7 Pre\text{-}drug\ score_i + \epsilon_i$,

where $\hat{Y}$ is post-drug score."

2. Since there was no cleaned manuscript, many typos, etc. may be present. It is recommended that the cleaned manuscript file be checked by multiple third-party eyes in the final version before IPA if possible.

   This is a good point. Since we have created a new v3 manuscript document to address these comments, the manuscript is considerably cleaner. We have made sure to read through this version carefully to check for typos.

Reviewer 3
Reviewed by Zoltan Dienes, 25 Aug 2022 13:21
The authors have addressed many of my points. There remain a few issues to resolve, the last one listed being most important.

Thank you for your helpful insights on our manuscript and we are pleased we addressed many of your initial points. We have now also attempted to address your remaining points, and taken particular care with the final point.

1. " if any of the two tests were significant (p>.01 for the Shapiro-Wilk and p>.05 for the Kolmogorov-Smirnov)" The ">"s should be "<"s.

   Thank you for spotting this! We have now corrected these to <.

2. "Outliers for the CTQ scores were assessed using boxplots" State how outlier is defined.

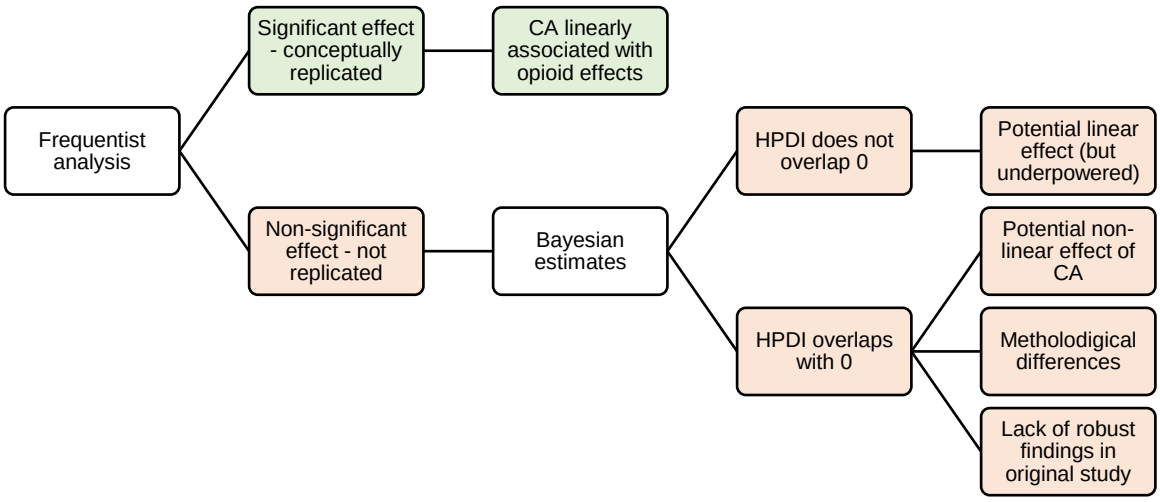
   Thank you for this point, we have now clarified this information in the manuscript, as follows:

   Page 8: "Outliers (defined as responses >3 standard deviations from the mean) for the CTQ scores were assessed using boxplots."

3. For the Bayesian analysis, why specifically 89% CIs? Why "AND HDIs"? But the bigger point is I don't know what role the Bayesian analyses play in the planned inference. What would count as the Bayesian analysis "concurring" with the frequentist one? A CI/HDI doesn't in itself allow rejecting or accepting H1 or H2. In fact the posterior distribution is guaranteed to give 100% probability to the claim the relevant effect exists. I suggest pick one analysis and stick with it.

   Thank you for this important point. The purpose of the inclusion of the Bayesian estimates was not to reject or accept the hypotheses, but as a complementary method to ensure we fully addressed the previous overarching reviewer point ***"to take additional steps to maximise bias control and rigour for a Level 2 study (e.g. via conservative statistical threshold; recruitment of a blinded analyst; robustness testing, multiverse/specification analysis, or other approach)."***

We have revised the strategy in accordance with your comment, as follows: The frequentist analysis will provide a significance threshold to conceptually replicate the significant findings reported in the previous study (Carlyle et al., 2021). However, possible reasons for non-significant results could be because (1) our data is limited by the sample size and power to detect small effects, (2) the effect of childhood adversity is non-linear (e.g., there is a 'threshold effect' that is only impactful when childhood adversity is more severe – as the prior Carlyle et al. study only recruited people with either no or high childhood adversity), or (3) a true non-significant effect. While this additional analysis will not be used as confirmatory, the posterior distributions derived from the Bayesian analysis could serve as a sensitivity analysis, resolving some remaining uncertainty about the possibility of a true (linear) effect of childhood adversity on subjective opioid effects. The Bayesian analysis is also able to incorporate prior expectations of the distribution of data by defining priors for the outcomes (mu and sigma) which are informed by the original study (Carlyle et al., 2021). The specific interpretations we will draw are summarised by:



The Bayesian analyses will help to inform whether there is a linear effect of childhood adversity in the case that we are not powered to detect it. If the Bayesian estimates do not support this, this may either indicate a potential non-linear effect of childhood adversity, or a true null effect. These additional insights are important for planning future research that assesses the impact of childhood adversity, i.e., by indicating whether studies are better served recruiting across different levels of adversity (linear relationship) or to identify and pre-screen those only with very high adversity scores.

We realise the addition of these Bayesian estimates are only valuable in the case of non-significant findings, and we have specified that Bayesian estimates will only be calculated in the case of non-significant results. We have also specified that we will not use the Bayesian analyses to accept or reject the hypotheses, but only to help guide further research.

In line with the diagram, the specific interpretations we will draw from the Bayesian estimates are:

1: If the 95% HDPI does not overlap with 0, this may support that a linear effect of childhood adversity in line with the alternative hypothesis (H1/H2) is plausible, but we were not powered to detect it. We will also have the mean posterior estimate to indicate the most plausible beta value, which could be interpreted as a tentative effect size. This linear effect should be examined in a higher-powered study.

2: If the 95% HDPI does overlap with 0, this will provide further support for the null hypothesis and could indicate that there may not be a linear relationship between CTQ scores and subjective opioid effects. Given the previous findings reported a significant effect in people with high childhood adversity, this could indicate a 'threshold (non-linear) effect', where only exposure to high levels of childhood adversity leads to altered subjective effects (e.g., only if it was severe, but not in the case of moderate or low exposures). Other possible interpretations could include the methodological differences between this and the previous study, or lack of robustness in the previous study. We will however not overinterpret this, and still highlight the need for further research – possibly with recommendations for study design.

We have specified this in the manuscript as follows:

Analysis page 9: "Bayesian posterior probabilities were calculated ~~to assess the robustness of the~~ for non-significant ~~findings~~results, using the 'rstan' (31) and 'rethinking' packages (32). Quadratic approximation was used to calculate the posterior probabilities [$outcome_i \sim Normal(\mu_i, \sigma)$] for the centered linear relationships with CTQ score [$\mu_i = \alpha + \beta(CTQ_i - \bar{x})$]. Priors were constructed using the means, standard deviations and slopes from the previous study (12) and were tested using prior predictive simulations, with drug liking: $\alpha \sim Normal(30,15)$, $\sigma \sim Uniform(0,20)$ and $\beta \sim Normal(0,1)$; and feeling good (measured as euphoria in the previous study): $\alpha \sim Normal(20,10)$, $\sigma \sim Uniform(0,10)$ and $\beta \sim Normal(0,1)$. The posterior mean ~~, 89% credible interval,~~ and ~~89~~95% highest posterior density intervals (HPDI; the narrowest interval containing ~~89~~95% 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 t~~The Bayesian ~~analysis~~ estimates ~~did not concur with the frequentist analysis, potential reasons for the lack of robustness were discussed~~were interpreted as supporting a potential effect to be examined in a higher-powered study if the 95% HPDI for beta did not overlap with 0, and the most plausible beta given by the posterior mean."

We have also updated this in the study design table:
Page 11: "The Bayesian intervals will inform on non-significant findings. If the 95% HDPIs do not overlap 0, this may indicate we were not powered to detect the effect, and a higher-powered study should confirm this. If the intervals do overlap with 0, this may provide more support for the null."

For why specifically 89%: *The original choice was partly due to wanting to avoid confusion with confidence intervals and to signal the arbitrariness of the cut-offs (per Richard McEllrath's Statistical Rethinking). We have now however changed these thresholds to 95% HDPI - both to be more conservative and because there is an intuitive relationship with the standard deviation (± 2SD) in the case of a normal posterior distribution. We will use the HPDIs which readers may not confuse with Confidence intervals, and that also works well in the case of non-normal posteriors.*

For ***Why "AND HDIs"****? We've now stuck to only HPDIs as our interval of interest as they are the narrowest interval containing 95% of the specified probability mass.*

4. To keep things clean, don't list exploratory analyses at this stage.

   We have now removed the exploratory analyses for the Stage 1 manuscript for clarity.

5. Most importantly, past relevant work found small to medium effect sizes, and the current study calculates power for small to medium effect sizes. That means the study is not powered to detect *all* plausible effects of interest. Thus a non-significant result would not count against the hypotheses of an effect being there. The authors cannot change N, so they should temper their conclusions such that a non-significant result just means reserve judgment.

   Thank you for this point, we will ensure that we do not over interpret a null effect, and have now made this more explicit in manuscript and the PCI-RR Study Design table.

   In the analyses section, we have removed that we would interpret the null effect in the context of methodological differences (as we will discuss these irrespective of significant/non-significant effects), and specified that the null effects will not be interpreted as no effect being there:

   Page 8: "For non-significant findings or significant associations in the opposite direction than hypothesised, we concluded that the conceptual replication was unsuccessful. ~~, and~~ Because the study was only powered to detect small-medium effect sizes, ~~this~~ any null effect ~~was~~ was not interpreted as support for no effect. ~~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).~~"

   For the Design table:
   Interpretation section Page 11:

   > *"H1 and H2 will be rejected if we find no effect, or significant effects in the opposite direction. However, ~~Any null or opposite effect~~ because the study sample size is limited and we are only~~will be interpreted as down to differences in: (i) the context of opioid use (e.g., surgery compared with a research study or recreational use), (ii) motivations for use (e.g., people who would take part in acute drug study vs surgery sample)~~ powered to detect medium-large effects, we will not conclude this as support for- the null effect, but rather that we are not powered to reliably detect smaller effects ($f^2 < .05$).*

We will also interpret any findings in line with the different opioid drugs, doses, and route of administration, in addition to the amount of variation with CTQ scores, and study context (hospital vs research study).”

Theory section Page 11:

“We are not powered to support the null hypothesis (that ~~If the current outcomes do not support this theory, it is possible that this may indicate~~ childhood adversity ~~may~~ is not ~~be considered~~ a risk factor for persistent use of medically prescribed opioids). In the case of null findings, we can only tentatively discuss ~~. It may highlight~~ the potential ~~importance~~ role of methodological differences, limited statistical power, or a non-linear effect of adversity.

We will also broadly explore potential challenges in generalising laboratory-based research to naturalistic settings, which is important when considering these studies for policy.

~~However, given the methodological differences and limitations in power we would only cautiously interpret any null effect.~~”

We have checked through the manuscript to make sure that we do not make reference to strong statements of the null effect. We have also added this as a proposed discussion point for the Stage 2 manuscript:

Page 10:

- *Describing the role of effect sizes in the findings (particularly in the case of non-significant findings), and that we are not powered to detect all plausible effect sizes. For this reason, we could not provide support for the null hypothesis (that there is no effect). We will discuss reasons for why the effect may be smaller than expected, for example that there may be proportionally less people in the severe range of childhood adversity. ~~This will be particularly important if the hypothesis is fully rejected.~~*