

Author response to reviews of

Manuscript JMLD.2022-0061

On the reproducibility of power analyses in motor behavior research

submitted to *Journal of Motor Learning and Development*

RC: Reviewer Comment AR: Author Response ☐ Manuscript text

Dear Dr. Immink,

Thank you overseeing the review of our manuscript for publication at the *Journal of Motor Learning and Development*. We appreciate the positive and helpful comments we received from this review process. We believe these comments have helped to strengthen our manuscript.

Based on comments from both Reviewers, as well as yourself, we have made some changes to the manuscript. Below we provide our point-by-point response (in normal font) to the Reviewers' comments (in **bold-italic**). All substantial changes in the revised manuscript are included in this response letter (in a text box) and appear in **blue** text in the revised manuscript. Thank you again for considering our work and overseeing the review process.

1. Editor comments

RC: *Reviewer 2 has acknowledged that they became aware of authorship and of the preprint version of this manuscript. JMLD intends to conduct blinded reviews. Based on the usefulness of Reviewer 2's comments and transparency in acknowledging knowledge of authorship, I am of the opinion that their awareness of authorship did not limit or compromise the peer-review process. If you have concerns or feel differently about this, please contact me for further discussion.*

AR: We have no concerns about Reviewer 2 being aware of our identities or the peer-review being compromised. We feel this might be a regular issue for those who take advantage of preprint servers.

RC: *Please direct your attention to your figures. As per author guidelines, color figures are acceptable. Do note that some readers might have some form of color weakness resulting in difficulty in distinguishing certain colors - like purple/pink. As a protan, I thought the figures used shades of blue until I read Reviewer 1's comments. Consider using palette that is more color weakness friendly.*

AR: We have updated the figures to use the color friendly Okabe Ito palette.

2. Reviewer #1

RC: *In this manuscript, the authors tested the reproducibility of power analyses reported in 84 articles from motor behavior journals. Strictly based on the information reported in the manuscript, the authors were able to*

reproduce 7% of the power analyses. When assumed plausible values were used to fill the missing parameters, reproducibility increased to 43%.

The entire manuscript is extremely clear and well organized. The information reported is of great interest to our field. In the introduction, the authors remind us of the importance of power analyses and their underuse. The results show low reproducibility of power analyses. In the discussion, the authors propose ways to remedy underpowered motor behavior research.

I have only minor suggestions that the authors might decide to consider if they find them useful.

AR: Thank you for the positive comments about our manuscript.

RC: In the introduction or discussion, could the authors introduce/discuss power analyses in the context of more complex analyses such as mixed-effects models (this thread may provide some references: https://twitter.com/CForestier_PhD/status/1556554485111062530) and mediation analyses (e.g., Fritz & Mackinnon, 2007; Schoemann et al., 2017)?

AR: This is an important issue as mixed-effects models and mediation analyses will hopefully train some traction in the motor learning literature. However, neither of these types of analyses appeared in our sample. We therefore decided to keep the paper focused on the relevant, and more common to our field, types of analyses that we aimed to reproduce. Given the added complexity of such analyses, we think mixed-effects models and power requires its own paper. Instead, we have provided some information that refers the interested reader to R packages that can be used for these more complex analyses. This section now reads as:

On page 5, lines 76-85:

Choice of software to conduct a power analysis is also important, as different designs may require different software. For instance, G*Power cannot, accurately calculate power for mixed factorial designs that include three or more levels of the within-subjects factor. While other packages, such as `Superpower`, can handle this more complex design, there are many possible designs that will require simulation-based approaches and likely consultation with a statistician. For example, power analysis for mixed-effects models can be conducted via simulation with the R package `faux` (DeBruine et al., 2021), and power analysis for mediation analyses can be conducted with the package `powerMediation` (Qiu, 2021). Each of these common pitfalls can result in conducting an underpowered study, or (less likely) an inefficient study.

RC: In the figures, the "light purple" looks more like pastel purple to me, and the "pink" looks like a light purple.

AR: We have updated the figures to use the color friendly Okabe Ito palette.

RC: Could the authors add a CRediT Authorship Contribution Statement (Brand et al., 2015)?

AR: We have added this to the manuscript. It was removed in the original submission due to blinding requirements of the journal.

RC: Could the authors indicate that the unrefereed pre-publication version of this article can be found on <https://doi.org/10.51224/SRXIV.184>

AR: We have included a link to the preprint version of the article at the end of the manuscript.

3. Reviewer #2

RC: *I would like to declare upfront that I am aware of the authors of this manuscript, as well as the preprint of the manuscript.*

In this manuscript, a follow-up analysis to a previous manuscript is reported in which the authors investigate the reproducibility of reported power analyses found in the previous manuscript. Similar to the previous manuscript, this is a very important topic for our field to address and one that has received too little attention. I commend the authors for their thorough work on this and also the journal for considering it for publication. Generally, I have found no major issues with the manuscript. It is very clearly written and will be understandable to authors not specifically familiar with all of the nuances of power analyses. The procedures and results are outlined clearly. I have a few minor suggestions below for the authors to consider.

AR: Thank you for the positive assessment and comments.

RC: *Lines 96-101: I am not sure journals requiring the reporting of power analyses necessarily improves the reliability of the literature. In many cases, a good power analysis may not have been conducted a priori and this enforcement will lead to the reporting of post hoc power calculations presented as a priori power calculations. An intervention such as this would only be likely to improve reliability in the case of preregistration and registered reports.*

AR: This is a good point (and one we had discussed during this project). We have updated this part of the manuscript to clarify our position that the most direct benefits of including reproducible power analyses are likely to be seen in pre-registered studies or Registered Reports. We include a caveat that even a post-hoc power calculation presented as an *a priori* one is not completely without value; it allows the reader to know what power the study had for a given effect size target and informed readers can judge whether that target is plausible.

On pages 5-6, lines 100-103:

However, the largest benefits to increased usage of reproducible power analyses would likely be seen in preregistered studies or Registered Reports. Otherwise, power analyses may be conducted post-hoc, limiting (but not eliminating) their usefulness.

RC: *Line 150: Since an alpha of 0.05 and power of 0.8 are so common, I wonder why these were not used as potential defaults when the information was missing?*

AR: This is a good question. We did not guess alphas other than .05. For one study (Mani 2021), we assumed alpha = 0.05 to be able to reproduce the analysis. Two studies did not report power (Li 2019 and Bartonek 2021) and for both studies more parameters were missing so assuming power was 80% would not have helped. We have updated this sentence to improve the clarity of our approach.

On page 7, lines 151-152:

All plausible analyses were attempted, but effect size, power, and alphas other than .05 were not guessed.

RC: *Lines 283-300: This section seems to repeat some tips given elsewhere in the article. I would suggest removing the common G*Power errors part from the introduction and rather leave them here in the discussion since, to me, they fit more logically here as potential pitfalls to avoid in the future. Just a suggestion.*

AR: Thank you for this suggestion, which we have followed.

RC: *General comment on the discussion: For me personally, I find that the discussion repeats the results a bit too much, making the text longer and also making the conclusions and recommendations a bit "crowded out" in the text. This is a stylistic point and doesn't need to be changed, but I would prefer more concise text in which the key conclusions and recommendations would stand out more clearly.*

AR: We agree with this assessment and thank you for the suggestion. We have removed the repetition so the key conclusions and recommendations stand out more clearly.

RC: *Line 325: Check reference format.*

AR: Thanks for catching this.

RC: *Related to my previous point, I recommend the authors consider making a table or figure with their key recommendations included, simply to provide an obvious, accessible list of the key points.*

AR: Than you for the recommendation. We have updated the manuscript to include a table (see page 20).

RC: *There is one issue that feel has gone without much mention (except for a few sentences in the limitations). The reporting and reproducibility of power analyses is addressed, and also the accuracy in terms of agreement between power analysis and hypothesis being addressed. However, little mention is made of the validity or appropriateness of the effect size of interest chosen. This is another topic entirely but I do feel that one major issue in both the conducting and reporting of power analyses in this field is that most researchers struggle to define and explain things like the minimum effect size of interest, smallest worthwhile effects and the like. In many cases, this leads researchers to use defaults that can be part of accurately reported and reproducible power analyses which are nevertheless inappropriate and unlikely to lead to studies of high informational value. I invite the authors to consider expanding on this issue in the discussion to some degree.*

AR: We agree that this is an important issue. We have added some additional discussion in the revised manuscript. In particular, we give advice to avoid using benchmarks altogether and focus on raw differences.

On page 22, lines 365-371:

In the meantime, it is important for researchers to think carefully about the specific effects they are investigating and not rely on effect size benchmarks to inform their power analyses. In fact, the benchmarks recommended by Cohen (1988) and used in G*Power change depending on the type of analysis, rendering them inconsistent and illogical for use in sample size planning (Correll et al., 2020). Instead, researchers should think about raw differences they would not want to miss to help arrive at a smallest effect of interest.