

## Author response to reviews of

Manuscript JMLD.2022-0042

## Low prevalence of *a priori* power analyses in motor behavior research

submitted to *Journal of Motor Learning and Development*

---

**RC: Reviewer Comment**   AR: Author Response   ☐ Manuscript text

Dear Dr. Zelaznik,

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.

We would like to draw your attention to a change in the revised manuscript based on an error we recently discovered. We found that an article from 2017 was included in the sample due to a faulty DOI error. This article included a power analysis so our reported numbers in the revised manuscript have changed slightly for most analyses. We have also updated Figure 1 to reflect the removal of this paper.

Thank you again for considering our work and overseeing the review process.

### 1. Editor comments

**RC:** *I have one suggestion of my own. As you examined three journals, but you do not present any “effect of journal” reporting. I think you should. Think about it. I won’t demand this revision.*

**AR:** Thank you for this suggestion. We have added an analysis comparing the proportion of studies with a power analysis in each journal to the paper in an Exploratory analysis section. The following have been added to the manuscript:

Page 11, Lines 173-175:

Exploratory analysis

We used a 3-sample proportion test to compare the proportion of studies including a power analysis in the three journals surveyed.

Page 15, Lines 221-226:

Exploratory analysis

Difference among journals in proportion of studies with a power analysis

Studies published in Human Movement Science included a power analysis (59/271) 18% of the time, the Journal of Motor Learning and Development (9/128) 10% of the time, and the Journal of Motor Behavior (16/152) 7% of the time. The 3-samples proportions test was significant,  $\chi^2 = 13.52$ ,  $df = 2$ ,  $p = .001$ .

## 2. Reviewer 1 comments

**RC:** *I have thoroughly read this manuscript and am familiar enough to make comment. For transparency, I should also note that I am familiar with the preprint version of this manuscript.*

*Currently, the manuscript is strong and so my comments are relatively minor and focus on readability and clarity. For organizational purposes, I present my comments in order of appearance within the manuscript.*

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

**RC:** *Page 3, line 44. The authors might consider adding a statement indicating that the purpose of a pilot study should not be to determine estimates of the true effect, and any estimates reported from the underpowered sample are unreliable.*

**AR:** We generally agree with this comment for the reasons contained in the manuscript. Since we are arguing that estimates from previous studies and even meta-analyses are not the ideal input for a power analysis, we decided to refrain from criticizing pilot studies more forcefully than we already do.

**RC:** *Page 4, line 65. Should "a power analysis" be "an a priori power analysis"?*

**AR:** Unfortunately, it was rarely specified whether the power analyses that were reported were conducted a priori. We anticipated this before beginning and chose not to restrict our coding to explicitly a priori analyses.

**RC:** *The authors do an excellent job throughout the manuscript with "hand holding" and explaining the implication of earlier research and issues with power.*

**AR:** Thank you.

**RC:** *Page 5, lines 91-93. This might be a point for the discussion, but one reason for a lack of difference here is that some people perform "a post-hoc a priori power analysis". That is, people might report a priori power analyses that fit their final sample size, rather than recruit a sufficient sample size to fulfill their true a priori power analysis. I know that is cynical. But, I hypothesize it happens sometimes, given resource constraints, incentives to publish, journal pressures, etc. Could this partly explain some of the results?*

**AR:** We do not think Reviewer 1 is being cynical by suspecting some of the power analyses were conducted after the fact. We think this happens too. We have added the paragraph below to address the myriad selection effects that may be involved in the data generating process that created our sample.

Pages 18-19, Lines 305-311

Selection bias makes it challenging to interpret the impact of power analyses on positive results and sample sizes. It is possible that power analyses are sometimes added to a study to defend a null result or a small sample size. It is also possible that researchers who conduct power analyses choose to study larger effects, leading to smaller sample sizes. We encourage readers to focus on the overall low frequency of power analysis reporting, small average sample sizes, and high rates of positive results as a description of the modern motor behavior literature.

**RC:** *Table 1. "Calculate average per group". Does this mean the number per group? I apologize if I missed it, but I am not sure what average means in this context.*

**AR:** We only extracted one value, so in case of an unbalanced design we took the average number per group.

**RC:** *Figure 2 is excellent.*

**AR:** Thank you.

**RC:** *Page 16, lines 225-227. I think also including the percentages of each type of analysis in the discussion would be useful: For example, 63% of studies that reported a power analysis based their effect size assumption on a previous study (n, %), a pilot study (n, %), or on effect size benchmarks (n, %).*

**AR:** Thank you for this suggestion. We have added these values. Please see Page 16, lines 240-241.

**RC:** *Page 16, line 251. Maybe more explicitly explain what censored means here? For example "... such that many small effects or null results never made it into the pages of a journal or the scientific record due to publication biases..."*

**AR:** We have added the following footnote to clarify our usage of censored:

Page 17

We use the term "censored" here to mean that some experiments fail to enter the published literature for results-dependent reasons, commonly called publication bias or reporting bias.

**RC:** *I hope that the authors find these comments to be useful.*

**AR:** We appreciate Reviewer 1's feedback, which has helped to improve our manuscript.

### 3. Reviewer comments

**RC:** *I would like to declare upfront that I am aware of the authors of this manuscript. I say this since it seems the journal operates a double-blind review procedure. I have seen this work presented and had already read the preprint of the manuscript prior to conducting the review.*

*I commend the authors on an important and rigorously conducted study. This topic has been gathering more and more attention in the broad movement science field and this specific focus on three motor behavior journals expands on previous analyses and draws more attention to the issues. It is also promising that the Journal of Motor Learning & Development considers it for publication since self-acknowledgement*

*of these issues by the entities with the power to make changes (journals) will hopefully be impactful going forward. I appreciate that the authors have shared their materials and code on the OSF. Since I do not code with R, I have not checked the code but the fact that it is publicly shared already gives me confidence in the transparency and robustness of the work. In my review, I did not detect any specific errors or issues in the numerical results. I have made a few specific comments for the authors below for their consideration but overall, I have no major concerns or comments on the manuscript. One thing that could be addressed more in the discussion is how to tackle this issue. The authors stop at the enforcement of journal guidelines, but I am not sure that will be enough for the reasons I outline below. Perhaps the authors can expand further on that section to give more direction for others in the field.*

AR: We would like to thank Reviewer 2 for their feedback during the review process, which has helped to improve the quality of the manuscript.

RC: *Lines 47-49: It might also be worth noting that using an effect size estimate from a meta-analysis, even when accurate, results in a power analysis for the “average” effect in the literature, which may or may not be what the researcher really wants to be able to reliably detect.*

AR: This is a great point. We have added the following text to expand on this:

Page 4, Lines 52-55

Targeting the estimated mean effect will only provide the desired power if the effect is average or above average in the specific context being investigated. Therefore, even if researchers are privy to a precise and unbiased estimate of the mean effect they are studying, they still might choose a different effect to target in their power analysis.

RC: *Line 91-93: I am curious what the thought behind this hypothesis was, especially considering that no specific direction of the effect was stated. The authors write that the field is typically underpowered, so why would the presence of a power analysis result in smaller sample sizes? Another aspect here is that this assumes that the power analyses are correctly performed, with a justifiable effect size of interest and not “gamed” or adjusted to reach a more feasible sample size. The recent article on power analyses in the journal Gait and Posture found that there was no clear change in sample sizes after the journal introduced a requirement for articles to report their power analyses, which the authors also suggested might be due to “power analysis hacking” to achieve more feasible sample sizes.*

AR: We did not specify a direction because of the myriad factors that impact our access to the sample. In our sample, anything from “power analysis hacking” to researchers choosing different research questions that could be addressed with smaller samples might have led to smaller sample sizes being associated with reporting a power analysis. We can imagine a context in which we would predict that power analyses lead to larger samples: if we randomly assigned researchers to conduct a motor behaviour experiment with or without conducting an appropriate power analysis, granted them unlimited time and resources, and required them to complete the experiment. Instead, due to selection bias we decided to avoid making any confirmatory predictions and instead focused on describing the literature.

RC: *Lines 121-122: Did a hypothesis have to be explicitly stated (i.e., at the end of the introduction) or did the presence of statistical hypothesis testing suffice?*

AR: We extracted a quote of each hypothesis, so it had to be explicitly stated. That said, the hypothesis did not need to be precisely stated. For example, if the authors wrote “We tested whether our [independent variable] had an effect on our [dependent variable]” we would consider that a hypothesis because it implies a test of the

null hypothesis.

**RC:** *Line 130: Eight people is quite a large group for data extraction. Was agreement in screening and extraction assessed? In what ways was the consistency of this process ensured?*

**AR:** During piloting the research team extracted data from 40 articles. There were 4 disagreements out of 120 items. We describe this detail in our preregistration, but can add it to the manuscript if necessary. Beyond this initial assessment, we relied on the fact each coded value was agreed on by two researchers, whether initially or after a third researcher resolved the conflict. As a final layer of oversight, the researchers flagged articles they found challenging, prompting the intervention of the lead author. The challenging papers were used as learning opportunities and therefore received the most attention.

**RC:** *Line 223-224: I would disagree with this conclusion slightly. The results suggest that the reporting of power analyses is not yet widely adopted. It is not possible, with the current data, to determine whether or not researchers use power analyses in planning their studies. Indeed, many ethics review boards require this, so there may be some percentage of the articles that did not report a power analysis had actually previously performed one.*

**AR:** We did not intend for this conclusion to come across so strongly. We have changed our wording to the following:

Page 15, Line 237

Our results cause us to wonder whether motor behavior research has not yet widely adopted power analyses to inform study design.

**RC:** *Lines 252-257: This finding is similar to the Gait and Posture paper – the guidelines explicitly require sample size justification, but the majority of articles don't report one. Stricter enforcement and perhaps reviewer guidelines would perhaps be beneficial. The broader problem here though is that this will also encourage post-hoc power analyses if authors have not actually performed an a priori power analyses. So while such enforcement might increase the percentage of papers with a power analyses, at least a portion of these will have been conducted after the study was conducted. A stricter requirement might be for journals to require evidence of an a priori calculation via ethics application, trial registration or preregistration.*

**AR:** We completely agree. We have added the following text to capture this idea:

Pages 17-18, Lines 281-285

Even if power analysis reporting becomes the norm, it may remain unclear whether power analyses were conducted during the design of the study or at a later stage. Therefore, the most credible confirmatory studies will have preregistered power analyses. A bold journal might even require it.

**RC:** *Line 288: check reference format.*

**AR:** Thanks for catching this. We have fixed the issue.