**Dear Dan,**

**Thanks for the comments from both reviewers and you, and especially nice to get the review from an intended user in addition.**

**We are grateful for the extra push to provide an option in the software to compute the required sample size for a specific design, and have been able to implement it in what we think is a very nice way (see Figure 6). Since our plot function already computed power for each sample size in a range specified by the user, it made sense to allow them to specify a desired power level, and highlight the n for each effect in the plot. We are very happy with this additional functionality and expect it will fill a need.**

* **We have added a hat to sample statistics, as requested (and left the verbal descriptions (e.g., ‘in the population’) so these differences should now be clear.**
* **In box 1 we now refer to the pooled standard deviation.**
* **We added in the text which order cell means are entered (i.e., a1\_b1, a1\_b2, a2\_b1, a2\_b2, a3\_b1, a3\_b2).**
* **We added a sentence on the number of simulations people should use for the ANOVA\_power function – the 100,000 used in our paper is indeed overkill for practical purposes, we now advise 1000 to test the simulation settings, and 10.000 for the definite power analysis.**
* **We checked all references, and added missing information.**
* **We now separate the discussion of the R code and Shiny app more clearly.**
* **We have updated the Shiny app and incorporated all suggestions.**
* **We fixed all typos.**

**The only comment we could not incorporate is renaming the package SuperPower (instead of Superpower), the package is already on CRAN, and R is case sensitive, so it can not be changed.**

**Finally, we link to the online manual, and the 2 locations for the shiny apps, in the text.**

**We hope the manuscript is now ready for submission. Below this letter, we provide detailed responses to the reviewers.**

**Sincerely,**

**Daniel Lakens & Aaron Caldwell**

**07-Apr-2020**

**Dear Dr. Lakens:**

**Thank you for submitting your General Article (AMPPS-19-0062.R1) entitled "Simulation-Based Power-Analysis for Factorial ANOVA Designs" to Advances in Methods and Practices in Psychological Science (AMPPS). The manuscript has now been reviewed, and the reviewer comments appear at the end of this letter. In addition to the original two reviewers, I sent the manuscript to one of our volunteers who agreed to evaluate the usability of tutorials.**

**Both of the reviewers who saw the initial version were pleased with the thoroughness of the revision. I agree that you did an excellent job of addressing the primary concerns they raised. Based on my assessment and these reviews, I am accepting the manuscript pending a minor revision that address the reviewer comments (and my comments below). Most of these comments involve minor clarifications and corrections designed to make it easier for readers to use the R code and Shiny app.**

**There is one more substantive suggestion that I do feel would greatly enhance the usefulness of the code and package. As noted in the previous stage of review, it would be ideal to add an option to calculate the required sample size for a given level of power. The new "usability" reviewer explicitly mentioned that, and it came up in the previous round as well. Many of our readers might want to be able to calculate the sample size needed to achieve a given level of power for an assumed population effect size. That goal might well be the most common one for power analysis. If it is possible to implement that feature without too much trouble, I'd strongly encourage you to do it.**

**Done, see below.**

**I also agree with reviewer 2 that it is important to distinguish population parameters and sample statistics. Doing so could be as simple as adding a "hat" when you're talking about the sample estimate rather than the population parameter (e.g., hat-mu rather than mu).**

**Done, see below.**

**Below, along with the reviewer comments, I've listed a few minor things I noticed when reading the manuscript.**

**When you submit your revision, please include include a letter detailing your point-by-point responses to each reviewer comment and indicating how you changed the manuscript to address them. I will handle the revised manuscript without sending it for further external review. You can submit your revision by logging into https://eur02.safelinks.protection.outlook.com/?url=http%3A%2F%2Fmc.manuscriptcentral.com%2Fampps&amp;data=02%7C01%7C%7C28e01f6d185f4eb58f4008d7db0dda5e%7Ccc7df24760ce4a0f9d75704cf60efc64%7C1%7C0%7C637218724860488886&amp;sdata=%2FANXQrG8fX9I%2B0vjkYsELnQvmUxecp%2BGUI8l9JVJp90%3D&amp;reserved=0 and entering your Author Center, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.**

**IMPORTANT: Your original files are available to you when you upload your revised manuscript. Please delete any redundant or outdated files before completing the submission.**

**Once again, thank you for submitting your manuscript to AMPPS and I look forward to receiving your revision. I think this well written tutorial (along with the R code and Shiny app) will provide a useful resource for our readers.**

**Best,**

**Dan**

**--------**

**Daniel J. Simons, Editor**

**Advances in Methods and Practices in Psychological Science (AMPPS) Psychology Department, University of Illinois ampps.editor@gmail.com**

**EDITOR ADDITIONAL COMMENTS**

**• I'd suggest adding a table listing other power analysis packages to the paper as a helpful resource. You could refer to it at the end of paragraph 2.**

**We have instead added a footnote to an online spreadsheet that lists and compared power analysis software. By linking to an online spreadsheet, we can keep it up to date as software packages continue to develop.**

**• minor suggestion: name the package SuperPower rather than Superpower to better highlight the "power" part of it.**

**Regrettably, the package is already on CRAN, so this is not possible (and R is case sensitive).**

**• Typo: "(be entering a “w” or a “b”)."**

**• Typo in shiny screenshot: "specficied"**

**• "which answers the question whether there are any differences between group means" --> technically, it answers the question of whether there are any differences AMONG the group means. "Between" suggests pairwise differences, but the omnibus ANOVA doesn't necessarily suggest any particular pairwise differences. "Among" might lead to less misunderstanding.**

**• It might be good to point out that the power for an N=80 within design is equivalent to an 80\*conditionLevels between design when the correlation is zero (and higher when it is >0. You do the latter, but don't make the point about equivalent power when correlation=0).**

**We added:**

**“In our example a within design would require three times less participants as a between subjects design with three conditions, but would achieve practically the same statistical power even when the three measurements are not correlated.”**

**• typo: "population standard violations varied extremely across conditions" - "standard deviations"**

**Reviewer(s)' Comments to Author:**

**Reviewer: 3**

**Comments to the Author**

**In this manuscript, Lakens & Caldwell introduces an R package and Shiny app that calculates the statistical power for factorial ANOVA as well as t-tests. This is a very useful tutorial as, at least to my knowledge, current alternatives are either unable to handle ANOVA designs with more than 2 factors or are overly convoluted. Even for one-way ANOVAs, many software require additional calculations (such as specific effect sizes) that can vary widely between software, and have confusing defaults (as stated by the authors). Superpower provides an easy and intuitive alternative, especially for ANOVA designs with more than 2 factors. It took me about 2 hours to go through the paper and try out the code.**

**Comments:**

**As the authors pointed out, one of the purposes of carrying out an a-priori power analysis is to ensure that the study is adequately powered, in this case for an ANOVA. I think there should be more explicit mention or even an example on how to obtain the sample size required for a certain power given. The plot\_power function provides one way of doing so, but I feel that explicitly stating that this could be a method for obtaining sample size requirements would be useful especially for readers who are skimming through it. In addition, although the main goal of the paper is for power analysis, I think it would be useful to a wider audience if it could be extended to include sample size calculations.**

**We agree. Thanks for the suggestion – it is incorporated (see Figure 6).**

**Below are comments more specific to the usability of the tutorial:**

**(a) all necessary materials are available**

**- This might be obvious to most, but I think it would be if the authors explicitly stated the name of the R package (Superpower).**

**- The links to the Shiny apps are not immediately apparent. The authors provide a link to the manual, which contains the links to the apps, but this is not immediately apparent and confused me for a while.**

**(b) any provided code runs successfully**

**- Yes, all provided code ran successfully.**

**(c) step-by-step instructions are clear and unambiguous**

**- For the app, I would have liked more specific instructions for the definition of “Factor & level labels”. I think it is important to mention that the levels of a factor have to be defined before defining the next factor. Perhaps a specific example would be useful. Also, there is a typo (‘young’ instead of ‘yound’)**

**- Perhaps it would be better to have a separate section for the app rather than combining it with the section that first introduces Superpower in R. Or, shift it to the end of the section. It was a little confusing/jumpy when Superpower was introduced in R, then the Shiny app was referenced for a bit, then it jumped back to R. For example, in the text it mentions that “for a visual confirmation of the input a figure is created..” (pg 4, line 4), but this is referring to only the Shiny app although the text preceding and following refers to the R package.**

**The r code can also produce a figure when adding ‘plot = TRUE’, which we now added to the example. Furthermore, we have moved the Shiny app screenshot below, to the location here it is discussed. This should prevent the ‘jumping back and forth’.**

**- Pg 4, line 44: my output does not exactly match the output in the manuscript. Instead of “p\_cheerful\_sad”, I have “p\_condition\_cheerful\_condition\_sad”.**

**True – we shortened these labels in the submission because of layout issues. We can use long labels, and ask the copy editors to make sure it is printed well.**

**- Under the ‘Power for interactions’ section (pg 7), it would be helpful to give an example of how to specify a 3x2 design in R. It is not clear how the means should be specified in a 3x2 design (or for any factorial design with more than 2 factors and 2 levels). I had to refer to the Shiny app to discover that in a 3x2 design, the means are specified in the following manner:**

**a1\_b1, a1\_b2, a2\_b1, a2\_b2, a3\_b1, a3\_b2**

**rather than**

**a1\_b1, a2\_b1, a3\_b1, a1\_b2, a2\_b2, a3\_b2**

**We added a 3x2 design for the new section on the power plot, to add this information on how means are entered to the manuscript.**

**(d) prerequisites specified in the tutorial are accurate and realistic**

**- There was no mention of prerequisites, but I believe that it is obvious to those who use R that the package first needs to be installed and loaded.**

**(e) the tutorial teaches the skill(s) it promises to teach**

**- The tutorial sought to demonstrate how to conduct power analyses, especially for factorial ANOVA designs (for which there are no satisfactory software currently). Two easy-to-use and easy-to-understand solutions were provided in the form on an R package and a Shiny app.**

**That’s good to hear. The preprint has been online for a while, and is downloaded more than 1000 times and is cited 7 times. We are receiving regular feedback from users and will continue to incorporate this in the future.**

**Reviewer: 2**

**Comments to the Author**

**Overall, the authors were responsive to my comments on the previous version of the manuscript. Moreover, as already outlined in my previous review, I think the software described in this paper is useful, mostly because it allows to calculate (or estimate) power for a variety of ANOVA designs as function of population means, standard deviations, and correlations under H1. Thus, except for a few minor issues that are easily corrected (see below), I think the revision is now ready for publication.**

**However, I still doubt that it is a good idea to use the same notation for (population) parameters and (sample) statistics. I am afraid that this will cause confusion and misconceptions in many readers. For example, it might suggest to researchers that it is possible calculate power for an effect size observed in the sample (“observed power”). This of course has nothing to do with power in the statistical sense because power always refers to specific parameters (or population effect sizes) that define H1. However, I leave it to the editor to decide whether different notations for parameters and statistics are necessary.**

**We have incorporated the suggestion to use a hat to specify sample statistics.**

**Minor issues:**

**P. 2, left column, 31: delete one “with”**

**P. 2., right column: Figure 1: Cohen’s d is labelled differently in the figure (x-axis) and in the text**

**P. 2., right column, Title of Figure 1: According to the rule specified in Footnote 1, “alternative hypothesis assuming d = 0.5” should read “alternative hypothesis assuming d in the population = 0.5”**

**P. 2., right column, Title of Figure 2: According to the rule specified in Footnote 1, “alternative hypothesis assuming partial eta-squared = 0.0588” should read “alternative hypothesis assuming partial eta-squared in the population = 0.0588”**

**P. 3, left column, 13-15: “The goal of an a-priori power analysis is to determine the sample size to observe a p value smaller than the chosen alpha level WITH A PREDETERMINED PROBABILITY (i.e., THE DESIRED POWER), given an assumption ABOUT the true POPULATION effect size”**

**P. 3, let column, 18: “the sample size” is redundant here, please delete**

**P. 3, left column, 26: According to the rule specified in Footnote 1, “true effect size is” should read “true population effect size is”**

**P. 3, left column, 41: “be” should read “by”**

**P. 5, right column, 39: delete “when”**

**P. 7, right column, Box 2: The notation in Equation (8) differs from the notation in the text.**

**P. 10-11: Many references are incomplete (i.e., lack volume numbers, page numbers, and doi’s)**

**Reviewer: 1**

**Comments to the Author**

**AMPPS-19-0062.R1**

**This is a very good revision that clearly addressed many of my suggestions in the previous review. My focus this round is more on user experience with the Shiny and R package to enhance usability.**

**As a small point of clarification in Box 1, I would suggest referring to the pooled standard deviation rather than the standard deviation as this the most common approach to calculation of Cohen’s d (although Grisson and Kim’s Effect Sizes book does address some limitations of that approach). If you do make explicit reference to the pooled standard deviation, it would add a nice link to ANOVA-type statistics to show that the square root of MSw/in is equivalent to the pooled standard deviation in this case.**

**Thanks, we have done so.**

**I ran the included code several times with various permutations. One persistent issue is the speed of processing for the monte carlo power approaches. The paper does note that more simulations take longer but I think that a warning for less advanced users would be useful. I recognized right off the bat that the 100k sims in the first example might take a good bit of time. However, novice users might simply think the code isn’t working. On this note, I tried the example using 100 sims (1.4 seconds), 1000 sims (12.6 seconds), 10,000 sims (2.1 minutes), and 100,000 sims (26.1 minutes). That said, it is likely better to use an example that won’t make people wait for so long to get an outcome. The Shiny app does well to show the progress but if people are using R they may get frustrated.**

**On a related note, some guidance as to how many sims should be used would be helpful. I admittedly, am just guessing here but given norms for bootstrapping (usually around 2,00), I expect that 100k is overkill. In a quick test, I designed around 64 per group. 100k gave power of 80.97, 10k gave power of 80.59, 2k had power of 79.9, whereas exact power came to 80.14. Those are all pretty accurate (if we take exact as the “right” answer here) so I expect that there is not much gained from going bigger.**

**Thanks, this is a very good point. Indeed, the 100k simulations are overkill (although we used this many for our publication anyway to have high accuracy for all simulations). We now use 1000 simulations in the example code, note this number of simulations should be completed within 15 seconds, and recommend to try out the simulation with 1000 simulations, and perform a larger simulation for the final power analysis. We write:**

**“In the code below 1000 simulations are performed, which should take approximately 15 seconds and yields reasonably accurate results when trying out the power analysis. For most designs increasing the number of simulations to 10000, which would take a few minutes to complete, should give accurate enough results for most practical purposes.”**

**It would be useful to have sample sizes in the output and a multiple reminders about what the sample size refers to. You do note sample size per group but (and I know this from shortcomings of my own package), without explicit reporting in the output regarding what the sample sizes both per groups and overall, users get confused and cite the wrong sample size particularly since the sample size is not reported in the pdf report.**

**Thanks for sharing this experience. We have added the total sample sizes for the entire study where relevant - e.g., (for a total of 240 participants), (for a total of 160 participants). We have noted the feature request to report the total sample size in the Superpower output, which is not yet added to the package, but we will strongly consider adding this in a future update. The online manual will also clearly explain how the sample size per condition would translate into a total sample size.**

**It would be useful to be more explicit on the location of the Shiny app up front. The Shiny does work well but there are other versions online and that may lead to some confusion for authors (particularly since this one https://eur02.safelinks.protection.outlook.com/?url=https%3A%2F%2Farcstats.io%2Fshiny%2Fanova-power%2F&amp;data=02%7C01%7C%7C28e01f6d185f4eb58f4008d7db0dda5e%7Ccc7df24760ce4a0f9d75704cf60efc64%7C1%7C0%7C637218724860488886&amp;sdata=KhS8Yq6MZUiqXc4ZhijzLcJMTdYJkxpZquC3cTqyn2o%3D&amp;reserved=0) throws an error.**

**We link directly to the app within the manuscript now.**

**Chris Aberson**