Peer Review Comments

DOI: https://doi.org/10.1525/collabra.218
Article type: Original research report
Editor: Victoria Savalei
Article submitted: 18 December 2018
Editor decision: Accept submission
Revision submitted: 15 May 2019
Article accepted: 14 September 2019
Article published: 30 September 2019

Responses for Version 1

Reviewer B:

1) General comments and summary of recommendation
Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope. A checklist of things to you may want to consider is below:
- Are the methodologies used appropriate?
- Are any methodological weaknesses addressed?
- Is all statistical analysis sound?
- Does the conclusion (if present) reflect the argument, is it supported by data/facts?
- Is the article logically structured, succinct, and does the argument flow coherently?
- Are the references adequate and appropriate?

This paper presents a demonstration of how one could select viable targets for replication studies in a systematic fashion. The crux of the approach is to first identify studies that provide only weak statistical evidence for the original claim, in the form of a weak Bayes factor, and follow up with a qualitative assessment about the impact a replication study would have on our scientific knowledge. Studies that satisfy both quantitative and qualitative checks would then be ideal targets to replicate.

As far as I know, this is the first proposal about how to select replication targets that actually follows through and shows how it could be applied in practice. Well done. I think the overall rationale for the approach makes sense and I could see this possibly being a useful paper for practicing psychologists. The paper also includes reproducible data and code, which I think is great and deserves kudos. I also like the appendix as a quick primer for Bayes factors.
I have two primary comments that I hope would improve the manuscript, and some smaller ones that the authors may or may not want to address.

Primary comments:
1- I think the paper needs to provide more detail about the Bayesian analyses and possibly more guidance for potential users.

A search for “prior distribution” returns zero hits in the main body of the paper, and yet this is a crucial piece of information for interpreting the Bayes factors reported here. Looking at the provided script (yay open code), the priors used here seem to be the default ones provided by the Bayes factor package. Is there a principled reason for this choice vs other possible choices? I am not suggesting this choice is not appropriate, but I think some sort of justification should be in the paper.

Moreover, is this choice of prior being recommended as an integral part of this proposed procedure for choosing replication targets, or should others who take up this approach to selecting replication targets feel free to use a different prior for their selections? Some specific guidance for other users who aren’t as familiar with Bayes factors might improve the chance this procedure actually gets used by your average psychologist.

2- The paper “Making replication mainstream” by Zwaan et al. (2018), and its 30+ commentaries, generated a number of discussion points possibly relevant to this paper. Indeed, many commentaries specifically discussed how to select studies for replication and how to conduct replications that expand our theoretical knowledge beyond mere verification.

I won’t go through everything relevant from that compilation, but I will highlight a couple of specific commentaries here. Witte and Zenker stress (in an admittedly vague sense) the importance of replicating studies that have a chance to provide theoretically important advances to our knowledge. Hardwicke et al. and Coles et al. argue that to select studies for replication we should explicitly consider potential gains and costs, in addition to other factors such as their prior plausibility and evidence from the original study. Kuehberger and Schulte-Mecklenbeck argue against selecting replications studies at random and discuss potential biases that can emerge in the process of selecting studies to replicate.

Obviously, the fact that others have previously proposed ways to select studies for replication is no knock against this paper. One reason being that this paper actually goes on to implement its procedure using real data and provides in-depth discussion on individual cases, which other proposals don’t do. Nonetheless, it might be fruitful to discuss the overlap between the ideas proposed in this paper and those mentioned in Zwaan et al. and its
commentaries, if only to help readers situate this work more clearly in the current state of thinking on the topic.

Smaller points to chew on:

-Page 2: The mere use of p-values and NHST does not necessarily lead to publication bias. One needs to selectively publish based on the outcome of the test to induce such a bias, which of course could be done with many inferential methods (including BF).

-Page 2: A reference to the American Statistical Association’s (2016) statement on p-values might be relevant in the discussion about their misuse in the intro.

-Page 3: The reference to registered reports specifically mentions specifically Chris Chambers. A reference to one of his works would be appropriate there for those who do not know him.

-Page 4 discusses a study by Hoekstra et al that does something relevant in the context of this paper, but what they did or why it is relevant is not elaborated upon. I am left wondering what exactly did they do and how is it relevant?

-Page 5: The paper might provide a reference for the claim that classic p-value based tests are asymmetric in the conclusions one can draw. The reference I like is Dienes (2011).

-Page 12: Post-hoc power is essentially a nonsense metric (see Hoenig and Heisey, 2001; O’Keefe, 2007). I’d strongly recommend the paper avoid hinging any of its arguments on it.

-The correct interpretation of the Bayes factor is not always given. Example: “Through the lens of the Bayes factor, scenario 1 presents ambiguous evidence: BF = .94 (i.e., the null and the alternative hypotheses are about equally likely, given the data).” As the paper points out in the appendix, the BF refers to probabilities of data, not hypotheses. Please be careful not to accidentally switch the conditional.

-Some language is ambiguous or confusing to me. A few examples stand out: +“Conceptual and direct replications can be used in partnership to great effect to firmly establish studies.” What does it mean to establish a study? +“This QRP [uncorrected alpha inflation] … is responsible for the overestimation of the importance of many study results.” Surely the paper does not mean to imply that a significant p-value determines the importance of a study.
“If an originally reported effect was a type 1 error, a direct replication can be an effective tool in reinforcing results.” In what way does a replication reinforce a type 1 error?

What role does the possibility of publication bias play in the selection of targets for replication? Presumably this bias weakens the actual evidence in the original studies and would create a larger subset of potential targets than if we were to ignore the bias. E.g., the analysis in Etz and Vandekerckhove (2016) suggests that most of the original studies included in the RPP provide only weak evidence when publication bias is taken into account, indicating a larger set of possible replication targets than the proposed analysis would provide.

Citations:


O'Keefe, D. J. (2007). Brief report: post hoc power, observed power, a priori power, retrospective power, prospective power, achieved power: sorting out appropriate uses of statistical power analyses. Communication methods and measures, 1(4), 291-299.


2) Figures/tables/data availability:
Please comment on the author’s use of tables, charts, figures, if relevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Collabra here):
Good. Data and code are available via OSF.

3) Ethical approval:
If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant
authority present? Where humans have participated in research, informed consent should also be declared. If not, please detail where you think a further ethics approval/statement/follow-up is required.: N/A

4) Language:
Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process.: Language is generally good. I highlight a couple of confusing phrases in my review.

Reviewer C:

1) General comments and summary of recommendation
Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope. A checklist of things to you may want to consider is below:
- Are the methodologies used appropriate?
- Are any methodological weaknesses addressed?
- Is all statistical analysis sound?
- Does the conclusion (if present) reflect the argument, is it supported by data/facts?
- Is the article logically structured, succinct, and does the argument flow coherently?
- Are the references adequate and appropriate?: See attachment

2) Figures/tables/data availability:
Please comment on the author’s use of tables, charts, figures, if relevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Collabra here).: See attachment

3) Ethical approval:
If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant authority present? Where humans have participated in research, informed consent should also be declared. If not, please detail where you think a further ethics approval/statement/follow-up is required.: See attachment

4) Language:
Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process.

See attachment

Editor Decision for Version 1

Editor: Victoria Savalei
Affiliation: University of British Columbia
Editor decision: Revisions required
Decision date: 15 March 2019

Dear Ms Sarahanne Miranda Field,

After review, we have reached a decision regarding your submission to Collabra: Psychology, "When and Why to Replicate: As Easy as 1, 2, 3?". Our decision is to request revisions of the manuscript.

The full review information should be included at the bottom of this email. There may also be a copy of the manuscript file with reviewer comments available once you have accessed the submission account. A summary of the requested edits from the editorial team can be found below. Please consider these points and revise the file accordingly:

Editorial Revision Requests:
--

In my reading, the paper could be made more concise--particularly in the beginning, I would encourage the authors to get to the point faster, given how much has been written on the replicability crisis. However, throughout, you could work on the writing to make it more concise (even at the level of the sentence structure). Finally, please submit a double-spaced file next time around, so that length can be gauged more easily.

I am also in agreement with the reviewers that a glaring omission from the paper is any mention of the prior distribution on the effect sizes under H1. Related, consider Figure 1 from the perspective of a frequentist. If this were the relationship between any two psychological constructs, all researchers would agree that we have a problem of discriminant validity. Maybe the solution is to illustrate how priors can be useful, or perhaps it is to broaden your perspective so that you appeal to a wider audience and note that study selection can also occur using p-values? Right now, you appear to sell Bayes Factors short, while also appearing overly dogmatic about the choice between them versus p-values.

I look forward to your revisions!
--
To access your submission account, follow the below instructions:
1) login to the journal webpage with username and password
2) click on the submission title
3) click 'Review' menu option
4) download Reviewed file and make revisions based on review feedback
5) upload the edited file
6) Click the 'notify editor' icon and email the confirmation of re-submission and any relevant comments to the journal.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing; therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please do contact us.

Please could you have the revisions submitted by April 15th. If you cannot make this deadline, please let us know as early as possible.

Kind regards,

Victoria Savalei
v.savalei@ubc.ca

Author’s Response to Review Comments for Version 1

Author: Sarahanne Field
Affiliation: Rijksuniversiteit Groningen
Revision submitted: 15 May 2019

Dear Dr. Savalei and the editorial office staff,

I believe you will find the 'When and Why to Replicate: As Easy as 1, 2, 3?' resubmission complete. I admit I find the resubmission pages a little confusing, so please let me know if I have done something wrong with the revision submission. It seemed a little simple and fast to me.

There are two documents - a PDF of the revised manuscript as well as a response to reviews letter. Hopefully the revision makes its way to your desk in good order!

Warm regards,

Sarahanne

Attached document:
Responses for Version 2

Reviewer A:

1) General comments and summary of recommendation
Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope.
A checklist of things to you may want to consider is below:
- Are the methodologies used appropriate?
- Are any methodological weaknesses addressed?
- Is all statistical analysis sound?
- Does the conclusion (if present) reflect the argument, is it supported by data/facts?
- Is the article logically structured, succinct, and does the argument flow coherently?
- Are the references adequate and appropriate?:

My previous comments have been well addressed. The details of the prior distribution are there now, and the details of its specification and recommendations for readers seem good. The varying interpretations of Bayes factors have also been cleaned up and look good. Overall, this is a nice paper. Congrats to the authors!

One small comment: The reference to Hardwicke et al mentioned in the cover letter text appears to have gone missing from the newly added section at the bottom of page three.

2) Figures/tables/data availability:
Please comment on the author’s use of tables, charts, figures, if relevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Collabra here).
Good.

3) Ethical approval:
If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant authority present? Where humans have participated in research, informed consent should also be declared.
If not, please detail where you think a further ethics approval/statement/follow-up is required.: NA.
4) Language:
Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process.:
Yes, jargon is explained.

Editor Decision for Version 2

Editor: Victoria Savalei
Affiliation: University of British Columbia
Editor decision: Accept submission
Decision date: 14 September 2019

Dear Ms Sarahanne Miranda Field,

I apologize for the long review cycle; we were unable, in the end, to obtain two reviews of the revised manuscript; however, based on the one review and my own reading, I'm happy to accept the manuscript.

However, please re-submit the final file after the following checks and minor changes have been done. Please check/confirm:

- that all references are included (see Reviewer A's comment about a reference that appears in the letter but not in the article-did you mean for this to happen?)

- that the document is rendered properly (e.g., I see weird letters in the pdf on p. 23, fourth line of text from above);

- that the text in Appendix A is not copyrighted by the original journal that published it (i.e., the article/journal is open access?).

- notation in Appendix A. For example, on p. 26, "delta" appears as a word rather than a symbol; in Equation (2), the integration appears over big delta (\Delta) rather than \delta.

After the final file is submitted, the document will go through copyediting, and you will contacted if there are any questions.
The review information should be included in this email.

Kind regards,

Victoria Savalei
UBC
v.savalei@ubc.ca