

Peer Review Comments

Article: Nalborczyk, L., et al. (2019). Pragmatism should Not be a Substitute for Statistical Literacy, a Commentary on Albers, Kiers, and Van Ravenzwaaij (2018). *Collabra: Psychology*, 5(1): 13. DOI: <https://doi.org/10.1525/collabra.197>

Article type: Original Research Report

Editor: Victoria Savalei

Article submitted: 27 September 2018

Editor decision: Accept Submission

Revision submitted: 17 January 2019

Article accepted: 03 February 2019

Article published: 26 March 2019

Responses for Version 1

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?:

Disclaimer: as first author of the paper to which the current manuscript is a response, I clearly cannot act as an independent peer. I do, however, hope that my comments below are helpful to the editor in reaching a decision.

A couple of weeks before receiving this review request, the authors already kindly pointed me to the preprint they put online. I have discussed this preprint with my co-authors, Don van Ravenzwaaij and Henk Kiers, and have read the submitted manuscript.

Although the authors make some interesting point, much of the manuscript seems to be a response to things that weren't actually in our paper. I get the impression that our discussion section was overlooked.

For us, one of the main messages is the opening paragraph of our discussion section: "In the present paper, we have demonstrated by means of various examples that confidence intervals and credible

intervals, in various practical situations, are very similar and will lead to the same conclusions for many practical purposes when relatively uninformative priors are used.”

The authors do not really respond to this, except by stressing that there are theoretical differences in how to interpret Confidence Intervals (CIs) and Credible Intervals (CrIs). We never said these differences are absent, nor did we comment on whether it is advisable to use uninformative priors. We claim that if two procedures consistently yield virtually the same numerical results, their interpretations are automatically virtually interchangeable.

Indeed, it might not be a best practice to use an uninformative prior (especially in the case of abundant prior information). However, it is a common practice. The statistical program JASP has gained a lot of popularity in the past couple of years, and this program (so far) only uses uninformative/flat priors.

The near-equivalence in numerical outcomes, and thus interpretation is not only useful to the JASP-using scientists who might lack the technical skills to use more sophisticated software. It is also useful when reading published scientific papers. When, for instance, reading a paper discussing the results of a ‘standard’ frequentist linear regression, it is possible to interpret the presented CI as a CrI for a flat prior, with all the benefits that come with the Bayesian approach. The “just do Bayesian”-suggestion in the submitted manuscript, is of no help when reading published literature.

In the second paragraph of our discussion section, we indicate that CrI and CI do differ substantially when (1) there is a restriction on the range; and/or (2) informative priors are used. We also mentioned the differences that occur when using (ridiculously) small samples. The list of situations given in Section 2.5 of the submitted manuscript thus is not new: we already reflected on these topics in our paper.

Figure 2 shows a case where the CI and CrI indeed differ. However, as this example is based on $n = 6$ numbers, it is an example of the above-mentioned small sample case. For such cases, we never claimed that CI and CrI are virtually different.

On p.9 the authors write “In contrast to what Albers et al. (2018) advocate, we thus suggest that the researcher interested in the Bayesian interpretation should use and report Bayesian credible intervals.” This is not in contrast to what we wrote, but in line with what we wrote. We end our manuscript with the statement, clearly in favour of Bayesian methods, that “posterior distributions can ideally work towards the enhancement of science.” This is not the only place in the submitted manuscript where the authors introduce or exaggerate differences between the original paper and this reply. Another example is in the abstract of the submitted manuscript, where it is written that “We argue that this heuristic [...] does not generalise well to more complex situations”. Again, this is 100% in line with what we already wrote in our manuscript (e.g. the sentence “Depending on the research context, vastly different intervals can be obtained if one chooses a specific informative prior”).

All in all, a large part of this manuscript is not a response to our paper and the differences in advices between both papers are much smaller than suggested. In our paper we clearly indicate that, in general, the Bayesian approach is advisable. There are situations, however, where this approach cannot be used, e.g. when the researcher lacks the technical skills to use e.g. R or specific Bayesian software, or when reading already published results. In such cases, and only in case of the specific circumstances outlined in our paper, it is useful to realise the numerical similarities between CI and CrI.

We did not make statements that go beyond these specific circumstances. Hence, the message of the current submission - that such statements are ill-advised - is not in conflict with our paper.

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):

Nice figures.

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:

No ethical approval is necessary.

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:

The article is well written

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?:

Thank you for inviting me to review "Pragmatism should not be a substitute for statistical literacy, a commentary on Albers, Kiers, and van Ravenzwaaij (2018)." I take great interest in statistical pragmatism and statistical literacy, hypothesis testing, and how statistical inference shapes scientific practice. In preparing this review, I read both the target article and the submitted commentary. I do want to disclose that my own prior belief in this debate slightly favors pragmatism, and I have published in the past on the pragmatic value of NHST and frequentist statistics. I believe, however, that I am still able to provide a neutral review of this submission.

The submitted commentary has a number of strengths. The simulations are accessible, useful, and clear (with one exception noted below). The paper is nicely structured and mostly well-written. I think *Collabra* is a great venue for commentary submissions like this. I strongly agree with the authors that researchers should choose the method – and resulting interpretation – that is appropriate for their goals and research question. Still, I believe the manuscript needs work before it can be accepted for publication. I detail my concerns and remaining questions below.

First, my biggest concern with this submission is what, exactly, the authors are trying to claim about Albers et al.'s conclusions. I cannot figure out whether the present submission is written show that Albers et al.'s claims are **WRONG** and should be ignored by the scientific community, or whether the present authors simply **DISAGREE** with Albers et al.'s suggestions and provide a more nuanced conclusion. The former argument is obviously a much stronger claim to make, and I'm not sure the commentary provides enough evidence to support it, yet the authors provide an unqualified quote by Berger (2006) to this end. Therefore, I would suggest the authors make it clearer in the manuscript that Albers et al. may not necessarily be wrong, but rather, that their claims do not generalize to more complex and nuanced (yet still realistic) cases. It would be nice to know whether the authors explicitly disagree with Albers et al.'s simplest point: that frequentist and Bayesian CI yield similar results for an estimated mean in a generally unbounded unimodal distribution.

Second, I struggle with the argument (or at least, the implication) that statistical pragmatism has no place in research. How should an everyday researcher (or journalist, or policy-maker), who cannot be bothered by the frequentist vs. Bayesian debate, approach the many off-the-shelf analyses that are available to them? Given that the present authors suggest choosing the analysis (frequentist CI or Bayesian CI) that best fits their needs, how might we empower clinicians and bench scientists to do this? I think the manuscript would be stronger (and have wider readership) if it spent a bit more time making some practical recommendations on this topic. Statistical literacy is well and good, and I would never argue against encouraging it. However, simply appealing to more statistical knowledge and complexity can feel overly idealistic and like cheap talk.

Third, it strikes me in Figure 2 that the Bayesian credible intervals plotted are simply, on average, wider than the frequentist intervals. As the credible intervals stretch upward to .60 on the y-axis, it strikes me as obvious that they would also stretch downward to include 0. Can the authors explain in the manuscript why this occurs (or why it must be so)? A skeptic may argue here that if a slightly more restrictive assumption is built into the Bayesian credible intervals to make them narrower, and thus more precise, that they will begin to occasionally miss the population value (as do the fre-

quentist intervals). My opinion is that we need a bit more setup and explanation to believe that the simulation isn't being built in a way that demands support for the authors' argument.

Minor points and a few typos:

The authors raise what appears to me to be a straw man that we might consider reporting a credible interval in addition to a confidence interval and then dismissing this suggestion out of hand. Perhaps this point can simply be removed.

I'm not sure the authors' section 2.1 is entirely fair to the frequentist approach; it is certainly the case that Bayesian analyses can also be run using faulty or nonsensical assumptions (prior selection, distribution shape, etc). I think the section title (conditioning on nonsense) could better communicate the point the authors are trying to make.

p. 10: "...more variable that it would happen"

p. 11: "...incredible large variety" (I would just drop the editorializing, if not slightly humorous, use of the word incredible)

p. 12 "...adopt here lead to the conclusion"

Thank you for this interesting paper, and I wish the authors the best in their future research!

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):

Sufficient; no comments.

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.:

Some suggestions listed above.

Editor Decision for Version 1

Editor: Victoria Savalei

Affiliation: University of British Columbia, CA

Editor decision: Revisions Required

Decision date: 21 December 2018

Dear Mr Ladislav Nalborczyk,

After review, we have reached a decision regarding your submission to *Collabra: Psychology*, "Pragmatism should not be a substitute for statistical literacy, a commentary on Albers, Kiers, and van Ravenzwaaij (2018)". Our decision is to request revisions of the manuscript prior to acceptance for publication.

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:

The first reviewer is one of the original authors. While obviously that introduces some bias, I do agree with most of his comments. At the same time, I do find your contribution valuable. Therefore, I invite you to submit a revision. Below are the suggested lines of revision:

Please make sure you do not misstate what the original authors actually argued or advocated. From myself, I can add that ironically, the tone of this commentary is in line with how these types of debates are characterized in the original paper, i.e., "harsh" or combative. I don't see any need for this. I don't think the divide is as wide as you paint it to be. Perhaps that is my bias, but I do think arguments for Bayesian statistics will be more appealing if they are made without exaggerating the advantages or the divide. The second reviewer's comments can also help in this line of revision.

I would like to see this commentary shortened somewhat. Upon my reading of it, I found it to be too elaborate and repetitive for a commentary; the text could be tightened. The introduction of so many short subsections also seems unnecessary. Finally, I did not find section 2.3 useful, but I will leave it up to you.

Please address the reviewers' comments about the simulation presented in Figure 2.

On p. 5, you give an example where it is a priori known that a parameter cannot be negative. Yes, it does seem ridiculous to assume the $U[-\infty, \infty]$ prior on such a parameter; however, nothing actually goes wrong if one does, and in reasonable size samples it will hardly ever matter. Furthermore, if one is working with small samples, constrained estimation under the frequentist paradigm is possible. Surely the authors know this. This example is not convincing as written. (As an aside, I work with structural equation models, where a negative variance estimate is indicative of model misspecification, and use of Bayesian or constrained estimation methods actually hides this useful piece of information...such nonsensical estimates can be useful!)

On p. 11, you argue that "contrary to what the authors postulate", the situation where CIs and credible intervals are numerically equivalent is not common. It is not difficult to see that what the original authors meant was that most research articles use these simple methods, whereas you enumerate situations is by simply listing types, and then counting how many have this equivalence, without taking their frequency of occurrence/use into account. This misinterpretation creates a disagreement where there isn't one necessarily.

Thank you for interesting commentary. I plan to assess your revision myself and not send it out for further review.

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 January 21st. 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: Ladislav Nalborczyk

Affiliation: University of Grenoble Alpes, FR; Ghent University, BE

Revision submitted: 17 January 2019

Dear Pr. Savalei,

Herewith, I submit a revised version of our commentary for publication in *Collabra: Psychology*.

The responses to reviewers have been uploaded as Supplementary Materials.

Sincerely,

Ladislav Nalborczyk

Attached document:

<https://s3-eu-west-1.amazonaws.com/ubiquity-partner-network/ucp/journal/collabra/197-2774-1-SP.pdf>

Editor Decision for Version 2

Editor: Victoria Savalei

Affiliation: University of British Columbia, CA

Editor decision: Accept Submission

Decision date: 03 February 2019

Dear Mr Ladislav Nalborczyk,

We have reached a decision regarding your submission to *Collabra: Psychology*, "Pragmatism should not be a substitute for statistical literacy, a commentary on Albers, Kiers, and van Ravenzwaaij (2018)", and are happy to accept your submission for publication, pending the completion of copyediting and formatting processes.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point. The accepted submission will now undergo final copyediting. You will be contacted once this is complete to answer any queries that may have arisen during copyediting and to allow a final chance to edit the files prior to typesetting. If you wish to view your submission during this time, you can log in via the journal website.

Kind regards

Victoria Savalei

v.savalei@ubc.ca