

Text to consider adding to manuscript reviews for journals. Explanations in boldface, suggested text not in boldface. Please add practices, ideas to this document! - alex.holcombe@sydney.edu.au YOU DON'T NEED TO CITE OR ATTRIBUTE ANY OF THESE SENTENCES/PARAGRAPHS WHEN USED IN A REVIEW (CC-ZERO) BUT IF IN SOME OTHER DOCUMENT, PLEASE CITE (CC-BY)

In my reviews, I preface some of the below with this header: I INCLUDE THE BELOW RECOMMENDATIONS, IN SOME FORM, IN ALL MY REVIEWS

REDUCING P-HACKING, MAKING P-VALUES INTERPRETABLE

In accordance with the movement towards greater transparency in reporting of studies, I suggest the authors add something like "We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study." to their paper, in line with http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2160588 and the recent Transparency and Openness guidelines being adopted by many journals. <http://centerforopenscience.org/top/> .

Especially in cases like this where there are a lot of possible interactions that might occur and Bonferroni can really kill one's statistical power, I recommend (for future studies) preregistering one's hypotheses and analyses so that the post-hoc tests don't have to be post-hoc.

REGISTRATION STANDARD

I recommend (for future studies) registering one's hypotheses and analyses before conducting the study so that the post-hoc tests don't have to be post-hoc. Registration also eliminates the presumption of many readers that researcher degrees of freedom makes the p-values uninterpretable - the reader doesn't know how many statistics were tried, in an attempt to make results significant.

ABSENCE OF STUDY REGISTRATION REDUCES CREDIBILITY

I note that this study was not preregistered, despite that it appears that the study was thought out in advance very carefully, making the lack of preregistration a missed opportunity. In this section, I argue for the importance of preregistration in the hopes of influencing the researchers and journal editors' future decisions about how they do their work.

Unfortunately, p-values are only meaningful if the number of analysis decisions made contingent on the data is known, and that can only be assured if a study is preregistered (e.g. Davis et al., 2017; Wasserstein & Lazar, 2016). Due to the existence of researchers who p-hack wittingly (as admitted in surveys) or unwittingly (recall Feynman's "The first principle is that you must not fool yourself – and you are the easiest person to fool"), individual results unfortunately have lowered credibility if not preregistered. Researchers should preregister their analysis choices, including statistical tests, outlier exclusion criteria, data transformations, etc. Otherwise, readers without personal experience of the internal practices of the lab will consider the results to not be particularly credible. A priori power analysis can also help.

Theory validation suffers from the same problem. Because many authors admit to Hypothesising After the Results are Known (HARKing), confidence in the conclusions of a paper are undermined by the lack of preregistration of the detailed predictions of theory.

Non-preregistered work is often excellent of course. Sadly, the p-values don't mean as much, but sometimes high-precision measurement can make a convincing case despite the reader not being able to be confident in how many comparisons were done.

I am aware that these facts are not familiar to many researchers. But all researchers should be aware of the replication crisis by now, and with the evidence suggesting that fewer than 50% of results in multiple sciences replicate in replication studies, readers will naturally be skeptical unless authors can show (such as by preregistering) why their results are more credible than the average historical study.

REGISTRATION REQUIREMENT OF THE HELSINKI DECLARATION

The authors indicated that their protocol followed the recommendations of the Declaration of Helsinki. However, recent versions (e.g. 2013) of the Declaration of Helsinki say that "Every research study involving human subjects must be registered in a publicly accessible database before recruitment of the first subject.", and they state that the most recent version supersedes previous versions. <http://www.wma.net/en/30publications/10policies/b3/> If the authors wish to state that they complied with all aspects of the DoH except preregistration, they could say that.

POSTING THE RAW DATA

Posting the raw data is important to improve the reproducibility of science, and also accelerates progress. Indicating that the data is available "upon request" has formerly been the usual practice but has not proven effective - studies have shown that most requests for data are not answered. You may have noticed that several journals, such as Psychological Science, have adopted open data badges as well as open materials and preregistered badges <https://osf.io/tvyxz/wiki/home/> . Posting the data at the Open Science Framework is a good choice for psychology. Xenodo is good for extremely large files.

TOO MUCH NHST

Need some boilerplate about how authors rely too much on rejecting null hypothesis when actually you are theoretically more interested in effect sizes, or sometimes some positive prediction of your theory.

REPRODUCIBILITY

Posting of the raw data and the code needed to reproduce the experiment would also be a good idea.

Also, I don't like papers where the (anonymized) data are not made available. But I know that's not the norm (yet).

Stronger steps: those who have pledged to make open practices a pre-condition for a review. <https://opennessinitiative.org/>

BEING A RESPONSIBLE AND ETHICAL REVIEWER

see the elements of the oath here: <http://f1000research.com/articles/3-271/v1>

REFUSING TO REVIEW FOR ELSEVIER

Example text:

<https://docs.google.com/document/d/1gDbQ7PpwljXSpJkQ9-Oqqm8mDTjhlW884xQaNdFWz14/edit>

TRANSPARENCY, ACCOUNTABILITY IN REVIEWING

Signed (I sign all my reviews),
Alex Holcombe

Note put at the end of each review:

To improve the transparency of peer review and editorial decisions, I avoid entering anything in boxes like the "Confidential Comments to the Associate Editor" provided here. Also I am concerned that such boxes can easily be abused with "stealth rejections" and/or unsupported insinuations.

JUDGMENTS OF IMPORTANCE ARE UNRELIABLE

I know that the policy of this journal is to reject articles based on not being important or ground-breaking enough, but I'm against making decisions on that basis currently, because I think academia currently prizes that way to much at the expense of weighting rigor of methodology and open practices. I have a longer essay about this which I sometimes include in reviews but I can't find it now. So to improve things at the current dysfunctional margin of the political economy of science, I refuse to answer the PNAS Review questions that were included about this issue. I am sorry if this comes at the expense of some authors, a kind of collateral damage in the service of improving science, which may unfortunately happen a lot with my somewhat counter-cultural behavior.