

January 20, 2018

Dear Doctors Brown and Wood,

I have now received reports from four reviewers and Andrew Chang regarding your manuscript, “Which test not witch hunts: a diagnostic approach for conducting replication research,” (*Manuscript Number 2363; Discussion Paper Number 2017-77*). The reviewers were very favorable, with the majority recommending acceptance without revision.

Based on the feedback I have received, I am inviting you to submit a revision. Your revision should address the individual points raised by each of the reviewers. In addition, you should include a separate document that provides a point-by-point response to each of the reviewer’s comments. You are not required to do everything the reviewers say, but if you disagree with a recommendation, you should state your reason. A successful revision will include addressing the following:

1) As noted by the “first” Reviewer 3 (sorry about there being two Reviewer 3’s!), there is vagueness about the definition of replication that you are employing. Reviewer 3 suggests you engage with Duvendack, Palmer-Jones, and Reed (2014 and 2017) and Clemens (2017). I personally prefer Reed’s (2017) taxonomy of replications (Reed, W.R., 2017, Replication in labor economics, *IZA World of Labor*, 2017:413, doi: 10.15185/izawol.413). But whatever you think is best, you should try and place your notion of replication within a larger taxonomy of replications. Make clear whether you think your checklist is appropriate for all types of replications, or just for some types and, if the latter, explain why.

2) “First” Reviewer 3 also said it would be helpful to include “Don’ts” in your checklist. This is an excellent suggestion. Perhaps you could address the reviewer’s point by giving some general “Don’ts” that you have come across in your experience with replications, without listing specific examples.

3) Section 5 of your paper (pages 11f.) discusses heterogeneous effects in which you identify the trade-off between using smaller, subsamples at the cost of statistical power. However, you don’t offer much in the way of wisdom about how to make this trade-off. I recognize that there is no “right” approach to this, but it would be very valuable to other researchers to hear your thoughts on this subject.

4) When discussing propensity score matching and replications, please read and incorporate into your discussion Lampach and Morawetz (Lampach, N., Morawetz, U.B., 2016. Credibility of propensity score matching estimates. An example from Fair Trade certification of coffee producers. *Appl. Econ.* 48, 4227–4237. doi:10.1080/00036846.2016.1153795).

5) Finally, as noted by Reviewer 2, the use of “which test” in the title, while catchy, is somewhat misleading because your paper is about a lot more than tests. Accordingly, you might want to consider changing your title.

A revision that satisfactorily addresses the points above, as well as responding to the individual reviewers’ concerns, is likely to result in a favourable publication decision.

Thank you for submitting your manuscript to *Economics: The Open-Access, Open-Assessment E-Journal*. I look forward to receiving your revision.

Sincerely,

W. Robert Reed  
Co-Editor, *Economics E-Journal*