Response to reviewers

We thank both reviewers for the insightful discussions. Beyond their overall encouragement, they have identified terminological and conceptual errors on our parts and given additional methodological advice for which we are very grateful. We note the reviewers were largely in agreement with regards to their overall evaluation and particular criticisms. We have found it straight-forward to include all of their advice, and think this has greatly improved the manuscript.

In particular, reviewer #1 has pointed out that we employed the term parameter not only in the sense of e.g. a population parameter with which inference is concerned, but also incorrectly to refer to (confounding) covariates ("nuisance parameter"). In the reworked version of the manuscript, the term parameter is only employed in the sense of a population parameter to be estimated. We instead employ the term "nuisance variable" to refer to that which we had previously incorrectly called "nuisance parameter". We have also corrected the other confusingly written or wrong uses of statistical terminology and concepts, such as those concerning variance in parameter estimation via multiple regression (which reviewer #2 has also pointed out).

Reviewer #2 (Jake Westfall) inquires about quantitative estimates of the failure rates of this approach. We have amended the methods section with a short simulation, where we show that in the described procedure, in realistic contexts, failures (e.g. failures to reject confounded samples when a significant effect is only observed due to confounding) occur more often than correct rejections (e.g. correctly rejecting a sample for confound imbalance).

We have also rephrased multiple sections for style, shortened the discussion of collinearity, extended the discussions of the error of "accepting" the null, further emphasized the importance of providing descriptive statistics (which both reviewers have repeatedly asked for), reworked and shortened the confusingly written section on employing multiple regression to control for confounds (including a reference to Westfall & Yarkoni, 2016, which appeared just days after our initial submission). Finally, we have included a simulation that demonstrates the impact of the problem.

We again wish to thank the reviewers for their fair and constructive criticism.