March 1, 2017

Duane F. Alwin
Sociological Methodology
Pennsylvania State University

Dear Editors:

Please find the responses to reviewers below for submission SMX-15-0022.R1.

Thank you.

Reviewer 1:

Thank you for your continued support of this method. We appreciate your concerns regarding the importance of intermediate solutions and counterfactual expectations compared to complex (and even parsimonious) solutions. In this paper, we employ our method on complex solutions for ease of interpretation for the naïve QCA user. Our method, however, *does* allow researchers to consider intermediate solutions, as does the package (with a few specifications).

To demonstrate the applicability of our method to various types of solutions, we have added two endnotes to the manuscript (3 and 4) to address similarities between complex and intermediate solutions in our data, before and after applying the method. In our case, the results for the complex versus an intermediate solution sets are identical. Not only were the cases covered and solutions similar, but so were the assessments of robustness before and after applying the recommended consistency scores and configurational N cutoffs. For example, before applying the robustness assessment, the complex solution for our data exhibit $\sim$95% randomness (which, after applying the recommendations from the method, drops to about 9 percent randomness). Applying the method to the intermediate solution results in $\sim$96% randomness (which drops to, again, $\sim$9% randomness after applying recommendations from our method). The method and package are built to work with whichever type of solution the researcher specifies.

Reviewer 2:

We appreciate Reviewer 2's comments. We have been having trouble with the software ourselves, as the reviewer correctly assumes, due to the packages on which ours depends. The package has been updated and is more stable than previous iterations because pack-

ages on which it depends have no conflicts. We are surprised, however, that the reviewer did not look to our theoretical justification in the previous response, which was the majority of our response (the software was a minor point).

Reviewer 3:

Thank you for the constructive comments. As for the first hypothesis test you mentioned, an additional paper could look at out-of-sample prediction of causal configurations, such as splitting the data, simulating from the predicted causal configurations, and seeing if it matches data not included in the original sample (also called cross-validation). Given the small N is most QCA studies, this would be difficult to achieve practically in most cases. Another option is to simulate a causal configuration process and apply the artificial data to QCA, to see if QCA would reproduce it. In a perfect QCA world, no error would exist in the data and the process would be recreated perfectly. So, we would only be informative in terms of how much error is tolerated with QCA, but this is ultimately done with sensitivity analysis; the only way a perfect-world QCA result could be misconstrued is by errors in measurement or stochastic variables not determined by a nonstochastic process. Those papers do that, and sensitivity analysis is certainly an additional test one could do. A null hypothesis test would only be informative for a tolerance of error, since the theory of QCA is one of deterministic processes with a little error here and there. The sensitivity analyses do not have a null or baseline as of yet, but could derive a configuration-specific baseline model trivially with the number of variables involved (i.e. the number of possible combinations of causal conditions). But actually trying to determine a tolerance for error is much more involved. Despite this method's noncomprehensive nature of the problems it solves, we believe that it improves substantially upon the former, ad-hoc method.

On the second hypothesis test: the theory of QCA relies on intimately linked causal configurations that do not operate independently – in fact, each result consists of a collection of necessarily-linked causal conditions. Since the configurations are deterministically linked, it does not fit in with the paradigm of QCA. One can argue for or against this, but we want to remain relevant to the researchers involved in this research paradigm, and improve their use of this method. This paper is really intended to be the first p-value-esque test for QCA. Another point is that sensitivity analysis is a useful approach for determining the tolerance of QCA in practice, either through the causal configuration or causal conditions. Determining what is exactly a null model is in that context is not immediately clear to us.

In regards to power analysis, this is a good point we had considered and mentioned in the paper, though using less elegant language. We included a function in the software which lists cases excluded in the final analysis after applying configurational N and consistency score recommendations from the method. We have warned against using this method as a strict p-value type of threshold, with this exact consideration in mind. Ultimately, it will require an intimate knowledge of the cases analyzed if very few configurations hold enough cases to warrant a 'robust' result. We also added more language in the paper for users to take care to observe how configurations will change with increasing configurational N thresholds. Thank you for this suggestion and we believe it will be mighty helpful for researchers wondering about which configurations they might miss in the context of improving robustness.

Sincerely,
Authors