February 27, 2017

Duane F. Alwin
Sociological Methodology
Pennsylvania State University

Dear Editors:

Please find the responses to reviewers below for submission SMX-15-0022. Please also note that, at

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 a footnote to the manuscript to address differences 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 minimum number of cases per solution. For example, before applying the robustness assessment, the complex solution for our data exhibit 95% randomness (which, after applying the recommendations from the method, drops to 9 percent randomness). Running the method on the intermediate solution results in 96% randomness (which drops to, again, 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. We believe this package is more stable than the previous iterations given that the packages it depends on now have no conflicts. We are surprised, however, that the reviewer didn't look into our theoretical justification we wrote as a response, which was the meat of the response (the software was a minor point).

Reviewer 3:

Thank you for the constructive comments. 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, separated by a logical "OR". This makes the second null hypothesis test (while completely reasonable to frequentists like us...) nonsensical within this paradigm.

The first hypothesis test would be difficult to do without substantial extension of the method, and is ultimately beyond our scope. 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. But, this is ultimately done with sensitivity analysis – one of QCA's great weaknesses is its utter reliance on humans to make sense of the results, and it can taint the final analysis. The test itself determines whether it will return an answer at all, not which answers it will: it's certainly not solving every problem of QCA, but we believe that it improves substantially upon the former, ad-hoc method.

In regards to power analysis, this is a good point we had considered and mentioned in the paper, though using less elegant language. One of the many things we disagree with about the QCA founder(s) is the attitude towards data: they intend to include every case with relevance, and thus power is of no object: not enough cases observed that have been drawn from a configuration means that the configuration simply does not predict the outcome. So, that's one perspective on one part of power.

We still need to deal with those cases excluded from QCA that are removed from increasing configurational N. We have included in the software that lists which configurations that are excluded in the final analysis due to a lack of cases, along with their configurational n thresholds. 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. Thanks for the 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