

COMMENT: LIMITED DIVERSITY AND THE UNRELIABILITY OF QCA

*Jason Seawright**

*Northwestern University

Corresponding Author: Jason Seawright, j-seawright@northwestern.edu

DOI: 10.1177/0081175014542082

Simulations have to date played a limited but meaningful role in discussions of qualitative comparative analysis (QCA). Hug (2013), for example, used simulations to show that measurement error can lead QCA to badly misrepresent the relationships in the data, and Lucas and Szatrowski (this volume, pp. 1–79) use similar techniques to demonstrate (among a wealth of other results) that findings of causal asymmetry in QCA are a methodological artifact of setting consistency thresholds greater than .5. Marx's (2010) simulations showed that QCA can distinguish between purely random data and data with a meaningful structure as long as the analyst pays attention to the proportion of combinations of conditions for which cases show contradictory outcomes, and Ragin and Rihoux (2004:23) supportively discussed a hypothetical simulation in which QCA was applied to two random subsamples from a given initial sample. Thus, simulation has been used by both critics and advocates of QCA; if QCA works in some context, simulations can provide supporting evidence and help identify the relevant context.

At the same time, existing simulation studies address only parts of what we need to know about QCA: Hug's (2013) work addressed measurement error, Marx's (2010) simulation looked only at purely random data, and Ragin and Rihoux (2004) were interested only in discussing the logical relation between subsample QCA results and the results for the sample as a whole. Although these studies advance methodological knowledge of QCA in important ways, they do not answer the overarching question of whether QCA is ever a reliable data analytic tool.

Lucas and Szatrowski use a series of fascinating simulation studies to address exactly this issue, showing time and again that QCA results fail to replicate the relationships among variables in the genuine data-generating process. However, for most of these simulations, data are generated using a logit model. These results provide the valuable insight that QCA fails, often quite badly, when asked to analyze data that are

generated through some process other than Boolean truth functions. However, QCA supporters may complain with some justification that the simulations simply do not take their ontological commitments seriously. It seems fitting to focus on the one issue that is most widely agreed to be important for QCA: limited diversity, or the problem of determining the outcome for unobserved combinations of independent variables.

To isolate this issue of limited diversity from other possible challenges, the simulation reported here adopts standard QCA assumptions essentially across the board. That is to say, it involves deterministic causal relations, with a strictly Boolean functional form, that are constant across all cases, among variables measured without error. Furthermore, the direction of causality is taken to be known in advance. This is carried out by calculating the complete truth table implied by the Boolean function:

$$Y = V_1 + V_2 * v_3 * V_4 + V_3 * V_5. \quad (1)$$

The truth table also includes three variables, which are not in the Boolean function that serves as the data-generating process and which are therefore causally irrelevant by design: V_6 , V_7 , and V_8 . The key test in the simulation is to determine whether it is likely that QCA will identify the three genuinely relevant combinations of independent variables, while not including the three irrelevant variables in the solution.

For a given iteration of the simulation, a random sample with replacement of a given size is drawn from that truth table. This means that, in these data, the true data-generating process for Y is exactly the one given in the equation above. Furthermore, there is no measurement error, and there are no contradictions. The only messy issue in the data is limited diversity, a problem that QCA advocates and critics all agree is widespread in practice.

Each simulated data set is then subjected to Boolean minimization, implemented using the eqmcc procedure within the QCA library in R (Thiem and Dusa 2013). All unobserved combinations of conditions are treated as logical remainders; this is equivalent to asking QCA to produce the most parsimonious solution. Because the data contain no contradictions, consistency parameters are irrelevant. For computational reasons, the simulation adopts the conditional approach of setting a minimum number of cases per retained condition at 2 unless that deletes most or all of the data, in which case the fallback threshold of 1 is used. This pragmatic solution seems in keeping with guidance from QCA

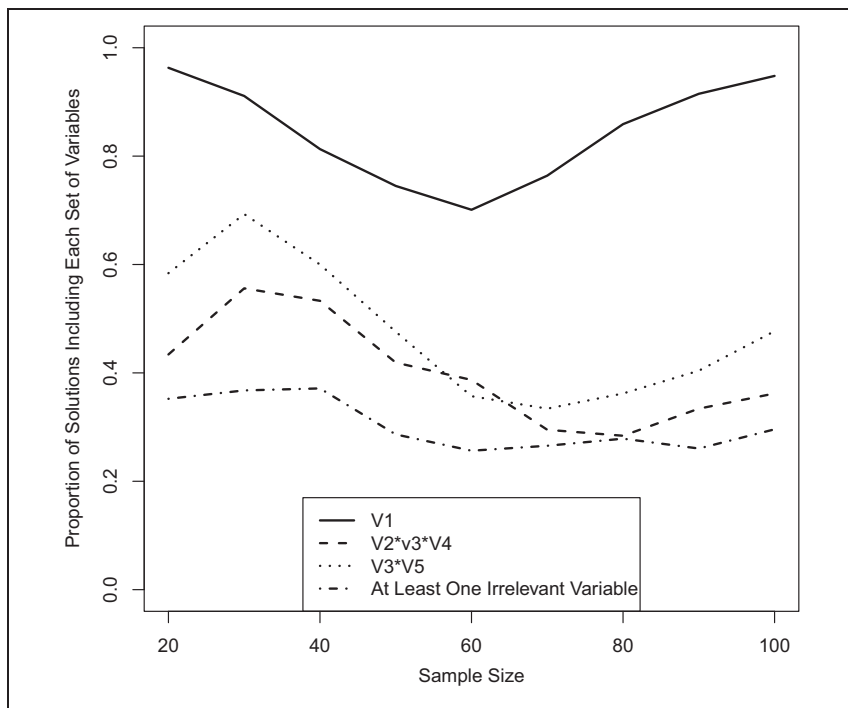


Figure 1. Qualitative comparative analysis success rates with limited diversity.

methodologists, who encourage a higher threshold but obviously do not hold scholars to that rule if the total number of cases is small.

Sample sizes vary between 20 and 100. For each sample size, 1,000 independent samples are drawn from the truth table. Figure 1 shows the results. The top line in the figure shows QCA's success rate at identifying the stand-alone sufficient cause, V_1 , as part of the final causal solution. Evidently the method works well for such simple causal patterns, although there is a dip in the middle of the plot, representing a tendency for medium-sized samples to overcomplexify the relationship of V_1 with Y .

The results are far less promising for the second two causal combinations, $V_2 * v_3 * V_4$ and $V_3 * V_5$. Success rates for these two combinations vary in a range from about 35% to about 70%, but mostly fall near 40%. Across the simulation as a whole, only 20% of analyses correctly identify both of these combinations. These unimpressive results are made all the more troubling by the fact that 30.4% of analyses include at least

one irrelevant variable in the final minimized solution. In other words, given the reality of limited diversity, QCA is in practice not particularly powerful at detecting genuine causal complexity, yet it is fairly prone to false-positive results.

This simple extension reinforces Lucas and Szatrowski's conclusions. As it stands, QCA seems to fail quite often, even under the most favorable conditions, if there is limited diversity—and one thing nearly all scholars agree on is that in the social sciences, there is always limited diversity.¹ Perhaps future developments within the QCA tradition will rectify this situation; until they do, however, scholars should consider turning to other methodological traditions when dealing with the kinds of causal structures with which QCA is intended to work.

Note

1. Further results are available from the author showing that classification and regression trees quite often outperform QCA even in this most favorable test.

References

- Hug, Simon. 2013. "Qualitative Comparative Analysis: How Inductive Use and Measurement Error Lead to Problematic Inference." *Political Analysis* 21(3): 252–65.
- Marx, Axel. 2010. "Crisp-set Qualitative Comparative Analysis (csQCA) and Model Specification: Benchmarks for Future csQCA Applications." *International Journal of Multiple Research Approaches* 4(2):138–58.
- Ragin, Charles C., and Benoit Rihoux. 2004. "Qualitative Comparative Analysis (QCA): State of the Art and Prospects." *Qualitative Methods* 2(2):3–13.
- Thiem, Alrik, & Adrian Dusa. 2013. *Qualitative Comparative Analysis with R: A User's Guide*. New York: Springer.

Author Biography

Jason Seawright is an assistant professor of political science at Northwestern University. He is the author of *Party-System Collapse: The Roots of Crisis in Peru and Venezuela* (Stanford University Press, 2012), and his work on multimethod research design and on representation has been published in *Perspectives on Politics*, *Political Analysis*, *Comparative Political Studies*, and several other journals and edited volumes.