



Introduction to the Special Issue: Case-Based Approaches to the Analysis of Quantitative Data

Barry Cooper^a and Judith Glaesser^a

^aSchool of Education, Durham University

Within the social sciences generally, the conventional approach to the analysis of survey data remains variable-based, employing some member of the regression family. Such methods address the effect of one or more supposedly “independent” variables on some outcome. Individual cases, the carriers of the variables, usually remain in the background, as do, often, underlying causal mechanisms and processes. It is variables that act, having their effects on a dependent outcome variable. In the typical multivariate study, the purpose is to report the net effect of each independent variable. The underlying mathematics is matrix algebra and the typical model additive. Notwithstanding the use of interaction terms and the development of multi-level modelling and related techniques, causal homogeneity is still often assumed across cases (an assumption whose realism was questioned by Ralph Turner as long ago as 1948). Over the past 30 years, a number of authors have published important critiques of the assumptions of this form of variable analysis. Abbott (2001), Byrne (2002), Freedman (1991), Lieberson (1985), Pawson (1989) and Ragin (1987, 2000, 2008), amongst others, have contributed much to our understanding of its limitations.

Over a long period, various alternative approaches to variable analysis, all sharing the goal of focusing on the case as some kind of configuration of features, have been developed (Byrne & Ragin, 2009). These include cluster analysis (Bailey, 1994), sequence analysis (Abbott, 2001) and, arising from comparative political analysis, Ragin’s development of set theoretic methods in his Qualitative Comparative Analysis (QCA) in both its crisp and fuzzy set variants (Ragin, 1987, 2000, 2008). QCA is designed to address conjunctural causation and/or prediction; it allows a researcher to model processes rigorously whose outcomes are the result of the conjunction of several factors. QCA understands causal relations in terms of sufficiency and/or necessity rather than in terms of the more of x , the more of y , linear models that underlie regression. In the application of QCA’s holistic approach the case, rather than disappearing from sight, is retained, existing as a configuration of conjoined factors (Ragin, 1987, 2000, 2008). Many forms of case-based research are associated with qualitative work with small to medium sized datasets. However, cluster analysis is typically used with large datasets and QCA, though it developed via the analysis of small to medium datasets, has been used by Ragin and others with large survey datasets (e.g. Amoroso and Ragin, 1999; Cooper, 2005; Cooper & Glaesser, 2008, 2010; Eliason et al. 2008; Glaesser, 2008; Ragin, 2006; Ragin & Bradshaw, 1991).

A vigorous debate has arisen about the merits of QCA and, in particular, about the extent to which it differs from more sophisticated forms of regression (see, e.g., the exchange between Seawright, Ragin and others in 2005 in *Studies in Comparative International Development*). It has also been argued that QCA, like regression

approaches, does not, in itself, address causation as it is understood by philosophical realists (Pawson, 2008). The need to combine the study of regularities, whether correlational or configurational, with the study of generative mechanisms has become a growing concern for some users of regression methods (Sørensen, 2005) and their critics (George and Bennett, 2005). In this context, there has been discussion of how cases might be best selected as part of a move from cross-case to within-case analysis aimed at developing knowledge of generative causation (e.g. Seawright & Gerring, 2008). There has also been considerable attention paid to the problem of limited diversity in social data and the problems it causes for causal analysis, with Ragin and others arguing for the use of counterfactual reasoning as a way forward (Ragin & Sonnett, 2004; Ragin, 2008). There has also been discussion of how QCA might be combined with other methods (e.g. Bail, 2008). The six contributions to this special issue, five of which employ QCA either on its own or alongside or in combination with other methods, reflect these ongoing debates.

Dymnicki and Henry's paper employs cluster analysis, reanalysing American data from an earlier study of delinquent pathways in urban contexts. They illustrate the value of newer clustering techniques by comparing the results of clustering four waves of family variables with growth mixture models with the results of clustering the same data with a more traditional two-step approach. The results of the two approaches differ dramatically, reminding us of the need to employ considerable judgement in interpreting the results of this case-based technique, given the variety of forms in which it exists and the choices between and within algorithms that must be made by the analyst. They also show how their chosen approach can identify interesting groups of outliers, arguing that detailed analysis of such cases can provide additional insight into the developmental careers of individuals.

Fischer's paper addresses several problems that arise in applying case-based approaches to large micro-level datasets. It is usually the case that the larger the dataset, the less access the researcher will have to in-depth knowledge of each case. Since advocates of QCA often stress the importance of such in-depth knowledge, this appears to be a potential limitation in applying QCA to large datasets. Fischer's response is to apply, in a first stage, social network analysis to data on actors at the micro-level in order to create, at the meso-level, a smaller number of cases of networks of actors whose characteristics can be understood well enough to enable QCA to be used at a second stage, where the networks themselves are treated as the cases. He then demonstrates the value of this approach in an analysis of the degrees of political conflict that arose around various reforms in Switzerland.

Glaesser and Cooper's paper arises from their concern that the study of cross-case regularities in survey data, whether via regression methods or QCA, ideally needs to be combined with further in-depth study of cases in order to develop understanding of the causal processes by which configurations of factors do or do not lead to some outcome. They draw on an ongoing study of educational transitions in England and Germany that combines QCA-based analyses of large datasets with in-depth interviewing (e.g. Glaesser & Cooper, 2010; Cooper & Glaesser, 2011). They show how they have used their QCA models to select cases for interview with the specific intention of developing causal understanding, for a given configuration of factors, of why typical cases do achieve the outcome and why a minority of deviant cases do not.

Hellström, in an analysis of political parties' positions on European integration, focuses on the similarities and differences between regression and QCA. It has been argued by some that QCA shares more with regression than Ragin claims (Seawright, 2005) and, by others, that regression can be used to undertake analyses of sufficiency and necessity (Clark et al, 2006). Hellström addresses these claims by undertaking analyses employing both crisp and fuzzy set QCA alongside regression analyses. He uses forms of regression, and particular specifications of his models, that parallel his QCA analyses. He argues that, while QCA can provide the capacity to explore causal substitutability (i.e. multiple paths to a given outcome), the statistical elements associated with the regression modelling can provide robust indications of the probable validity of the QCA-based findings. He also discusses the problems that a limited sample size might create for this approach. His arguments here might be read in conjunction with the results, in the following paper, of Marx and Dusa's simulations.

Marx and Dusa, developing earlier work by Marx (2010), address the problems that arise when users of QCA (though the same problem applies in the context of other methods, of course) have a large number of causal

conditions in relation to the number of cases in their dataset. It has often been assumed, they argue, that if QCA models are ill-specified this will become clear as the analyst examines the indices of consistency with sufficiency and explanatory coverage that the QCA software provides. Using randomly generated data and simulation techniques they show that this cannot, in fact, be safely assumed. They show that additional assumptions are required, these concerning the ratio of causal conditions to numbers of cases. They use the results of their simulations to provide benchmark tables that users of QCA can consult to judge the number of cases that will be required in order to enter a given number of causal conditions in a QCA model without threatening the validity of subsequent analyses. Of course, alongside these considerations, the researcher will also want, following Ragin's advice, to think theoretically about their QCA results, especially when the ratio of cases to conditions puts their study on or near a borderline in one of Marx and Dusa's benchmark tables.

Thomson, like Glaesser and Cooper, applies QCA to large survey datasets. Here, she draws on some analyses from her ongoing research, with the empirical focus being on the ways in which parental interest in their children's education, in conjunction with other factors, is or is not sufficient for children to achieve various levels of mathematical achievement. Initially, she investigated configurations of sex, maternal interest and social class, aiming to find quasi-sufficient configurations for high or very high attainment in mathematics. She then made two changes to improve this model. One involved changing the level of mathematical achievement used as her outcome and the second the addition of a new factor, general ability, to her model. These changes allowed quasi-sufficiency to be achieved, but only at the cost of generating limited diversity in that some configurations came to be populated by small numbers of cases. She shows how she addressed this problem of limited diversity by employing Ragin's writing on counterfactual reasoning. Her detailed account of complex, intermediate and parsimonious solutions, i.e. three types of QCA model which treat configurations with few or no cases differently, also provides a useful background for understanding the choices made by Hellström in his contribution.

It can be seen that, taken as a group, these papers address a number of outstanding issues arising from attempts to apply case-based approaches to quantitative data. They are a small sample of the broader range of work that has arisen, and continues to arise, from the various critiques of correlation-based variable analysis and from attempts to develop new ways of establishing, and trying to understand, regularities in datasets. Ragin's work in particular has excited many, both positively and negatively! This sample of the range of responses to his long-term intervention in methodological debate provides yet more evidence of the effects of his work in social science.

References

- Abbott, A. (2001) *Time Matters*. Chicago: Chicago University Press.
- Amoroso, L. M & Ragin, C. C. (1999) 'Two approaches to understanding control of voluntary and involuntary job shifts among Germans and Foreigners from 1991 to 1996', *Quarterly Journal of Economic Research*, 68(2), 222-229.
- Bail, C. A. (2008) 'The configuration of symbolic boundaries against immigrants in Europe', *American Sociological Review*, 73, 37-59.
- Bailey, K. D. (1994). *Typologies and taxonomies. An introduction to classification techniques*. Thousand Oaks: Sage.
- Byrne, D. S. (2002) *Interpreting Quantitative Data*. London: Sage.
- Byrne, D. & Ragin, C.C. (2009) *The Sage Handbook of Case-Based Methods*. London: Sage.
- Cooper, B. (2005) 'Applying Ragin's crisp and fuzzy set QCA to large datasets: social class and educational achievement in the NCDS', *Sociological Research Online*, 10, 2.
<http://www.socresonline.org.uk/10/2/cooper1.html>

- Cooper, B. & Glaesser, J. (2008) 'How has educational expansion changed the necessary and sufficient conditions for achieving professional, managerial and technical class positions in Britain? A configurational analysis', *Sociological Research Online*, 13, 3. <http://www.socresonline.org.uk/13/3/2.html>
- Cooper, B. & Glaesser, J. (2010) 'Contrasting variable-analytic and case-based approaches to the analysis of survey datasets: exploring how achievement varies by ability across configurations of social class and sex', *Methodological Innovations Online* 5(1), 4-23.
- Cooper, B. & Glaesser, J. (2011) 'Using case-based approaches to analyse large datasets: a comparison of Ragin's fsQCA and fuzzy cluster analysis', *International Journal of Social Research Methodology* 14(1), 31-48.
- Clark, W. R., Gilligan, M. J., & Golder, M. (2006). 'A simple multivariate test for asymmetric hypotheses', *Political Analysis*, 14(3), 311-331.
- Eliason, S.R., Stryker, R. & Tranby, E. (2008) 'The welfare state, family policies, and women's labor force participation: combining fuzzy-set and statistical methods to access causal relations and estimate causal effects', in L. Kenworthy and A. Hicks (eds.) *Method and Substance in Macrocomparative Analysis*. London: Palgrave Macmillan.
- Freedman, D. (1991) 'Statistical models and shoe leather', *Sociological Methodology*. 21, 291-313.
- George, A. L. and Bennett, A. (2005): *Case studies and theory development in the social sciences*. Cambridge, Massachusetts: MIT Press.
- Glaesser, J. (2008) 'Just how flexible is the German selective secondary school system? A configurational analysis', *International Journal of Research & Method in Education* 31(2), 193-209.
- Glaesser, J. & Cooper, B. (2010) 'Selectivity and flexibility in the German secondary school system: a configurational analysis of recent data from the German socio-economic panel. *European Sociological Review*. doi: 10.1093/esr/jcq026
- Liebersohn, S. (1985) *Making it Count*. Berkeley: California University Press.
- Marx, A (2010) 'Crisp set Qualitative Comparative Analysis (csQCA) and model specification', *International Journal of Multiple Research Approaches*, 4(2), 138-158
- Pawson, R. (1989) *A Measure for Measures*. London: Routledge.
- Pawson, R. (2008) 'Causality for beginners', *ESRC/NCRM Research Methods Festival*. Available at: < <http://eprints.ncrm.ac.uk/245/> > 20/04/10
- Ragin, C.C. (1987) *The Comparative Method*. Berkeley: California University Press.
- Ragin, C.C. (2000) *Fuzzy Set Social Science*. Chicago: Chicago University Press.
- Ragin, C.C. (2006) 'The limitations of net effects thinking', in B. Rihoux and H. Grimm (eds.) *Innovative Comparative Methods for Political Analysis*. New York: Springer.
- Ragin, C.C. (2008) *Redesigning Social Inquiry*. Chicago: Chicago University Press.
- Ragin, C. C. and Bradshaw, Y. B. (1991) 'Statistical analysis of employment discrimination: a review and critique', *Research in Social Stratification and Mobility*, 10, 199-228.
- Ragin C. C. & Sonnett, J. (2004) 'Between complexity and parsimony: limited diversity, counterfactual cases, and comparative analysis', in S. Kropp and M. Minkenberg (eds), *Vergleichen in der Politikwissenschaft*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Seawright, J. (2005). 'Qualitative Comparative Analysis vis-à-vis regression (Symposium on Qualitative Comparative Analysis)', *Studies in Comparative International Development*, 40(1), 3-26.
- Seawright, J. and Gerring, J. (2008) 'Case selection techniques in case study research. A menu of qualitative and quantitative options', *Political Research Quarterly* 61(2), 294-308.

Sørensen, A.B. (2005) 'Theoretical mechanisms and the empirical study of social processes', in Hedström, P. & Swedberg, R. (2005) (eds.) *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.

Turner, Ralph H. (1948) 'Statistical logic in social research', *Sociology and Social Research*, 32, 697-704.

Biographies:

Barry Cooper is Emeritus Professor of Education at Durham University where he was, from 1998 to 2005, Director of Research in Education. He was from 2004-2007 co-editor of the *British Educational Research Journal*. His interests are in the sociology of education, especially social class, educational achievement and assessment, set-theoretic research methods and the evaluation of educational aid projects. His most recent book was, with Máiréad Dunne, *Assessing Children's Mathematical Knowledge: Social class, sex and problem-solving*. A new book, Cooper, Glaesser, Gomm and Hammersley's *Challenging the Qualitative-Quantitative Divide: Explorations in Case-focused Causal Analysis* will be published by Continuum in 2012.

Judith Glaesser is an ESRC Research Fellow in the School of Education at Durham University, where she will take the post of lecturer in education from October 2011. Her interests include sociology of education, inequality and meritocracy in education, and research methods, particularly QCA. She studied for a PhD at Konstanz University (published as *Soziale und individuelle Einflüsse auf den Erwerb von Bildungsabschlüssen*). Currently, with Barry Cooper, she is exploring the application of case-based methods to large datasets in comparing transitions in the English and German secondary school systems. A new book, Cooper, Glaesser, Gomm and Hammersley's *Challenging the Qualitative-Quantitative Divide: Explorations in Case-focused Causal Analysis* will be published by Continuum in 2012.