

When less is more: selection problems in large-N and small-N cross-national comparisons

Ebbinghaus, Bernhard

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Ebbinghaus, B. (2006). When less is more: selection problems in large-N and small-N cross-national comparisons. In K.-S. Rehberg (Ed.), *Soziale Ungleichheit, kulturelle Unterschiede: Verhandlungen des 32. Kongresses der Deutschen Gesellschaft für Soziologie in München. Teilbd. 1 und 2* (pp. 4013-4021). Frankfurt am Main: Campus Verl. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-142348>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

When Less is More: Selection Problems in Large-N and Small-N Cross-national Comparisons¹

Bernhard Ebbinghaus

In recent years, important debates over epistemological and methodological issues in comparative cross-national research arose. Proponents of a statistical perspective criticize the use of small-N comparison (Lieberson 1991; King/Keohane/Verba 1994; Goldthorpe 1997), while various comparativists have responded to this criticism, emphasizing the importance of case-oriented studies and shortcomings of quantitative approaches (Mahoney/Rueschemeyer 2003; Brady/Collier 2004). A major issue has been the attributed ›selection bias‹ of qualitative comparative studies (Geddes 1990), in this view, the selection of cases based on their positive outcome leads to false inference. However, as I will show, there are also selection problems with large-N comparisons. In this paper, I want to revisit the problem of case selection in comparative research: How do we select our cases for comparative cross-national analysis? I will focus on cross-national comparison, be it *cross-sectional* statistical analysis or *cross-case* logical comparison, referring to ›cases‹ in the sense of politically defined macro-social units, that is independent national states.

In this paper, I exclude within case analyses that inform much of comparative small-N studies (Mahoney 2003) as well as comparative multi-level analyses, which use micro-level data and control for cross-national institutional variations by using macro variables (Przeworski/Sprague 1986). I also pass over a discussion of Galton's problems posed by increased transnationalisation (Ebbinghaus 1998). In the following, I first review some of the main selection problems and other shortcomings in quantitative large-N statistical analysis. In a second step, I rebut some of the voiced criticism and point at the advantages of qualitative small-N comparison. As empirical illustration, I will use some examples from cross-national studies of modern welfare states, limiting the set of countries mainly to the OECD countries for pragmatic reasons.

¹ Gekürzte Version des auf Englisch erschienen Beitrags: Bernhard Ebbinghaus (2005), »When Less is More: Selection Problems in Large-N and Small-N Cross-national Comparisons«, *International Sociology*, vol. 20, no. 2, pp. 133–152.

1. Selection Problems of Large-N Comparisons

1.1 The Illusion of Random Sampling

Statistical inferences based on (random) samples of the population are well established, yet it remains doubtful whether we can speak of a »random sample« in cross-national comparison. First of all, it is often difficult to define the »population at risk«, that is the full set of all cases due to divergent (often politically loaded) definitions of what constitutes a nation-state, a region or other social entity. While descriptive statistics should be applied to the description of a given population, inference statistics should be confined to inferences from a random sample to the population of cases. Yet »many economists and political scientists will even perform statistical tests that assume a random sample, when the units of analysis at their disposal mount up to the complete population.« (Pennings/Keman/Kleinnijenhuis 1999: 82). Several rationales are given for the »attraction« of statistical tests: (1) the non-random country sample is taken as the best approximation of the full universe of cases, (2) it helps dealing with (random) measurement errors, (3) that there is an infinity of models that could be applied to the same evidence, and (4), a »metaphysical« claim that a selected set represents a sample from the unknown universe of past, present and future cases. All these claims are questionable since the cases at hand are themselves preselected not least by historical processes.

In practice, variable-oriented researchers choose a set of cases for largely pragmatic reasons: cultural familiarity, similar socio-economic development and – last but not least – availability of data. Instead of a random sample, cross-national studies are commonly dealing with a (non-random) sample from a *categorical* set. Moreover, data availability is unequally distributed across potential cases: macro-indicators and time-series are more likely to be available for the more advanced economies, larger societies and long-term democracies, in particular if they are members of international organizations such as the OECD or Eurostat that collect and disseminate such data-series.

1.2 The Stratified »Sample« Problem

In contrast to a population census, the cases of a macro-comparison are heterogeneous social units, violating the homogeneity assumption of inferential statistics. Given the large cross-national differences in population and economic resources, it may be misleading to analyze each case as equally important. The most common »fix« to this problem is to apply *statistical controls* for measuring (otherwise unobserved) heterogeneity, including population or resource variables as controls in a

regression model. However, adding control variables will reduce the degree of freedom considerably in cases where cross-country analysis are based on few countries and might limit the explanatory power of other independent variables in case of multicollinearity.

The alternative strategy sometimes undertaken is to reduce heterogeneity by selecting countries with similar conditions (e.g. advanced industrial countries) or by agnostically assuming similarity due to membership in a particular dataset (e.g. OECD). Both strategies have major repercussions in descriptive and inference statistics. Hence, there are two problems that cross-sectional analysis faces: (1) the population of macro-units is very heterogeneous, contradicting the homogeneity assumption, and (2) any selection hereof be it for size or categorical membership, will have considerable impact on dependent and independent variables (*selection bias*). These problems cannot be remedied by better (or larger) sampling, but are problems resulting from the historical contingency of real world macro-social units.

1.3 The Historical Contingency Problem

When selection is based on a regional (e.g. Southern Europe), categorical (e.g. Social democratic welfare states) or organizational membership (e.g. EU), the population of real world cases is ›biased‹ as a result of historical and political contingencies. Choosing the members of a political organization (e.g. EU) implies ›self-selection‹ by historical or political processes since accession is contingent on fulfilling admission criteria, political decisions by the applicant, and the admission by the international organization. Even when we use regional or categorical sets or for that matter all existing countries, the universe of cases is ›biased‹. Today's observable national states are a highly contingent set of (surviving and redesigned) cases. Yet all observable cases represent only a ›limited diversity‹ (Ragin 1987) compared to the many more logical possibilities.

1.4 The Path-Dependency Problem

In the search for increasing degrees of freedom, quantitative researchers have embraced time-series and more recently cross-sectional (or ›pooled‹) time-series analysis (Beck 2001). Yet general linear models applied in quantitative studies face a problem related to path dependence. Stanley Lieberson called this problem asymmetrical forms of causation (Lieberson 1985: Ch. 4). This comes very close to the increasing returns of a path dependent phenomenon (Arthur 1994). Hence, the

selection of an appropriate time window is an additional important consideration in comparative research.

2. Deliberate Selection in Small-N Case Study

2.1 Comparison and Single Case Studies

For many critics, a single case study is of very limited value to comparative analysis since it is neither directly comparative, nor does it allow to generalize beyond the case. Yet depending on the selection, in-depth study of a single case can in fact make a contribution to our general knowledge, it can »make a case« (Gerring 2001): an *extreme case* that clarifies the outlier of previous statistical analysis; a *typical case* that stands for a larger set of countries; a *crucial case* that approaches most clearly the paradigmatic case of a particular theory; a *counterfactual case* that is a theoretical comparison of what might have happened with what actually did. The case study purpose can thus be manifold: a particular case can confirm, disprove, alter or generate a theory (Lijphart 1971; Collier 1991). Certainly, the scientific process cannot stop here, the amended or new theory should then be applied to other cases; the single case study then plays only a part in a larger collective enterprise that may well require to be more interdisciplinary and international (Smelser 2003).

2.2 »Too Many Variables, too Few Cases« Problem

The most common criticism of the comparative method is the »many variables, small N problem« (Lijphart 1971: 686; Goldthorpe 1997): there are just too few cases to allow the testing of all potentially relevant variables. In statistical language, we face a »degree of freedom« problem, though this implies that we wish to test a large number of variables at the same time. However, as I would argue, comparison may serve a different function, namely to test a proposition with a few given variables, and this makes the too-many-variables problem less pertinent. Hence, one particular case suffices to disconfirm a theory unless we face severe measurement problems on dependent or independent variables (see below).

Only when we shift from deductive proposition testing to induction do we face the small-N problem. Indeed, there are always too many potential, nontrivial independent variables that could be considered. It is indeed true that Mill's *inductive* cannon of logic cannot be used to infer causal relationships other than about necessary conditions, except in the unlikely case that we could indeed control all other

relevant variables by selection of appropriate cases. Thus, the comparative method, in combination with the appropriate selection of cases, can provide a means to test propositions deduced from a given theory or may help to eliminate some competing hypotheses. Here, the number of cases is not the limiting factor for the variables, but the conflict between predicted and actual outcome.

2.3 The Selection by Outcome Bias

Another criticism voiced against small-N studies is the »selection bias« problem, that is a violation of a methodological rule from statistical analysis (Geddes 1990; King/Keohane/Verba 1994): when we deliberately exclude cases with high (or low) values of the dependent variables, the statistical results are biased compared to the full set of observations. Barbara Geddes points at two problems:

»The first (...) involves jumping to the conclusion that any characteristic that the selected cases share is a cause. The other involves assuming that a relationship (or the absence of a relationship) between variables within the selected set of cases reflect relationships in the entire population of cases« (1990: 132f.).

Geddes criticism, that by adding more observations to an extensive bivariate analysis the linear relationship might change seems unfair to the work of those who study the social processes in a few cases intensively and make no claim beyond the cases at hand.

While the large-N researchers seem to hope that the more cases they study the more ground they have to *generalize* their findings beyond the observed cases, qualitative comparative methods usually lead to more complex findings when the number of cases is increased. With each new case we learn more about whether a universal relationship holds or whether we have to add further conditional factors to its applicability. The basic selection questions, which categorical set is the right one depends on the availability of cases, our in-depth knowledge about them, their comparability and their theoretical relevance.

The more serious challenge is the first problem mentioned by Geddes, the problem of causality. According to Mill's logic of agreement, we would need a most dissimilar country design (MDSO) when all selected cases show the same outcome, that is, all independent variables should differ in all but the crucial variable. By selecting on the outcome, however, we can only establish *necessary* causes, this is claim that all cases share the same preconditions and that further (unclear) causal mechanisms are at stake. The selection by outcome allows us not only to find some potential necessary conditions but it provides a means to eliminate all those potential variables that are always (or in most cases) not congruent with a positive outcome.

Again the flag of »too few cases for too many variables« is waved. It will probably be impossible to single out the one causal necessary condition, and some of the congruent observations may be trivial and some may be spurious factors. Again, theoretical considerations and additional evidence will be needed to distinguish between true, trivial and spurious necessary conditions. However, the most important contribution of selection by outcome is that this method can help to eliminate some non-necessary conditions in a first step. For those who refute that finding necessary conditions is enough to establish causality, this is certainly not enough.

2.4 Deterministic Causality

A fundamental difference in epistemological and ontological positions turns around the controversy over deterministic vs. probabilistic causality (Lieberson 1991; Goldthorpe 1997). Small-N comparison, using Mill's logic or Boolean algebra, are criticized as »deterministic« since already one case can falsify a proposition, thus ignoring the possibility of chance events, the frequency of (dis)confirming cases, and likelihood of measurement errors. In contrast, a probabilistic model would entail an ontological view that social processes are stochastic, the statistical law of large numbers, and the natural distribution modelling of measurement errors. Stanley Lieberson (1991; 1994), among other critics of the »deterministic« comparative logic, argues for a probabilistic model given the complexity of social processes. When there are complex multivariate causes at work, social science is unable to measure, control or model *all* relevant factors and interactions, particularly not in a small-N research design. Certainly in larger small-N studies, using Boolean comparative methods, we may be able to go beyond Mill's one factor perspective and also include interactive configurational analysis (Ragin 1987).

Most fundamentally, there may be unmeasured chance events and unaccountable stochastic processes that intervene in the real life social processes and can thus only be captured by probabilistic statements. »Except for probabilistic situations that approach 1 or 0 (in other words, those that are almost deterministic), studies based on a small number of cases have difficulty in evaluating probabilistic theories« (Lieberson 1991: 310). However, this criticism does not hold for necessary conditions, which should be present in all those cases that lead to an outcome partly due to intervening chance events. We may not be able to study the stochastic process itself by »deterministic« comparative methods; yet necessary *auxiliary* conditions could still be analyzed. Furthermore, stochastic processes could be studied by *within-case* analyses (for instance, analyzing several events over time) the results of which could then be analyzed across cases using comparative methods.

2.5 Frequency and Historical Contingency

A way to circumvent the hard deterministic test, frequency considerations can be used in the qualitative comparative analysis, particularly in medium-N studies. »Rather than impose absolute standards in all investigations (...), researchers also can make inferences about sufficiency using probabilistic methods« (Ragin 2000: 109). Indeed, Charles Ragin's fuzzy set approach allows for assessing probabilistic propositions of »quasi-sufficiency, applying formal statistical (binominal) test with predefined benchmarks. Already few confirming cases may thus be enough to pass a statistical test. Similarly Bayesian estimators of the confidence in a hypothesis could be used. As Douglas Dion suggests: »in determining whether a necessary condition is true, we must compare the hypothesis that the necessary condition is at work to some alternative that the necessary condition is not at work« (1998: 134). Using for instance a prior probability of 50 percent that the alternative hypothesis is correct, only five confirming cases are needed to reach a 95 percent confidence that a condition is indeed necessary (Dion 1998).

Finally, we also face the problem of »limited diversity« (Ragin 1987) due to historical contingency. Many master variables »band« together, thus the universalist Nordic welfare states are all »smaller« export-oriented, Protestant and homogenous societies, with relatively successful left parties and powerful unions. Moreover, not all logically possible configurations of master variables have occurred (thus far) and therefore we cannot test alternative combinations with real world observations. This »limited diversity« has major implications for Boolean configurational analysis since the outcome of non-existing combinations is unknown. Boolean analysis can be used to reveal the implicit simplifying assumptions about non-existing cases (Ragin 1987: 104–113). Historical contingency also sets some limits for frequency considerations in fuzzy set analysis. If we set our frequency level for example to four out of five cases, it would depend on historical nation-building contingency whether enough cases could be found to meet the benchmark. Nevertheless, frequency related qualitative comparison allows researchers to evaluate the probability *within* particular configurations and thus circumvent the pitfall of assuming linear relationship across cases that are asymmetrically distributed due to historically contingent nation-building.

3. Conclusion

The two juxtaposed research approaches certainly have both advantages and disadvantages. The analysis of selection problems in comparative studies indicates several

pitfalls, some of which are peculiar to macro-level analysis whether quantitative or qualitative. Most fundamentally, the universe of macro-social units from which we choose our cases is highly contingent, and itself an outcome of historical and political processes. These contingencies are not always well understood and unfortunately not often explicitly reflected by cross-national studies. The danger is that quantitative researchers want to analyze the available data and test their hypotheses too quickly, instead of explicitly discussing why and how their cases came to be in their »sample«. Availability of data and past research leads them to a »let's take what we have« approach in case selection. This stands in contrast to the explicit deliberate selection of cases following the qualitative comparative method that seeks to choose cases with a view to particular theory-relevant configurations. Cases are selected because they are most similar or most dissimilar, with important consequences for the kind of causal inferences that can be tested.

It is problematic that quantitative studies use inference statistics as if the cases are a quasi-random sample, but instead it is usually a relatively stratified set of cases. Any grouping of national states is a historically contingent set, implying considerable problems for cross-national statistical inference. The practice of steering around these issues by introducing statistical control variables seems a rather unsatisfying research strategy. This is not to rule out the use of quantitative methods altogether, but rather to call for more modesty in its use and implications. Statistical analysis can play a role in data exploration, yet there remain major obstacles to inference from any given »sample« of national states to the (implied) universe of all macro-social units.

Much of the criticism of the small-N approach seems to be derived from a misleading analogy based on statistical methodological reasoning and considerable misunderstandings about the purpose of cross-case comparison. The comparative method cannot be reduced to the exercise of logical truth-table analysis only, but is built upon the insights from *within*-case analysis such as detailed historical study of the underlying social processes that stand behind macro-variables. A probabilistic perspective can be partially integrated through frequency benchmarks, while stochastic processes can be assessed through within-case-analysis. Moreover, the small-N configurational analysis is better suited to take into account the historical and political contingency of macro-social units, as it allows for configurational analysis, while the selection of cases is informed by the strategy to control for much real world diversity and similarity. Given the strong supporting evidence of path dependent developments, comparative analysis that engages in a dialogue between within-case analysis and configurational comparison seem to me to be more appropriate than those cross-national statistical analyses that assume homogeneity, independence and representativeness of its cases.

References

- Arthur, W. Brian (1994), *Increasing Returns and Path Dependence in the Economy*, Ann Arbor.
- Beck, Nathaniel (2001), »Time-Series-Cross-Section Data: What Have We Learned in the Past Few Years?«, *Annual Review of Political Science*, vol. 4, pp. 271–293.
- Brady, Henry E./Collier, David (eds.) (2004), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Lanham.
- Collier, David (1991), »The Comparative Method: Two Decades of Change«, in: Rustow, Dankwart A./Erickson, Kenneth P. (Hg.), *Comparative Political Dynamics: Global Research Perspectives*, New York, pp. 7–31.
- Dion, Douglas (1998), »Evidence and Inference in the Comparative Case Study«, *Comparative Politics*, vol. 30, pp. 127–145.
- Ebbinghaus, Bernhard (1998), »Europe Through the Looking-Glass: Comparison and Multi-Level Perspectives«, *Acta Sociologica*, vol. 41, H. 4, pp. 301–313.
- Geddes, Barbara (1990), »How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics«, in: Stimson, James (ed.), *Political Analysis*, Ann Arbor, S. 131–150.
- Gerring, John (2001), *Social Science Methodology: A Critical Framework*. Cambridge.
- Goldthorpe, John H. (1997), »Current Issues in Comparative Macrosociology: A Debate on Methodological Issues«, *Comparative Social Research*, vol. 16, S. 1–26.
- King, Gary/Keohane, Robert O./Verba, Sidney (1994), *Designing Social Inquiry, Scientific Inference in Qualitative Research*, Princeton.
- Lieberson, Stanley (1985), *Making It Count: The Improvement of Social Research and Theory*, Berkeley.
- Lieberson, Stanley (1991), »Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on Small Number of Cases«, *Social Forces*, vol. 70, H. 2, pp. 307–320.
- Lieberson, Stanley (1994), »More on the Uneasy Case for Using Mill-Type Methods in Small-N Comparative Studies«, *Social Forces*, vol. 72, H. 4, pp. 1225–1237.
- Lijphart, Arend (1971), »Comparative Politics and the Comparative Method«, *American Political Science Review*, vol. 65, pp. 682–693.
- Mahoney, James (2003), »Strategies of Causal Assessment in Comparative Historical Analysis.« in: Mahoney, James/Rueschemeyer, Dietrich (eds.), *Comparative Historical Analysis in the Social Sciences*, New York, pp. 337–372.
- Mahoney, James/Rueschemeyer, Dietrich (ed.) (2003), *Comparative Historical Analysis in the Social Sciences*, New York.
- Pennings, Paul/Keman, Hans/Kleinnijenhuis, Jan (1999), *Doing Research in Political Science: An Introduction to Comparative Methods and Statistics*, London.
- Przeworski, Adam/Sprague, John (1986), *Paper Stones. A History of Electoral Socialism*, Chicago.
- Ragin, Charles C. (1987), *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*, Berkeley.
- Ragin, Charles C. (2000), *Fuzzy-Set Social Science*, Chicago.
- Smelser, Neil J. (2003), »On Comparative Analysis, Interdisciplinarity and Internationalization in Sociology, *International Sociology*, vol. 18, H. 4, pp. 643–657.