Title
Translating Quantitative Methods of Qualitative Researchers: The Case of Selection Bias

Journal
American Political Science Review, 89(2)

Author
Collier, David

Publication Date
1995-06-01

Peer reviewed

TRANSLATING QUANTITATIVE METHODS FOR QUALITATIVE RESEARCHERS: THE CASE OF SELECTION BIAS

DAVID COLLIER *University of California, Berkeley*

King, Keohane, and Verba's *Designing Social Inquiry: Scientific Inference in Qualitative Research* (KKV) is an ambitious attempt to translate for the qualitative researcher a series of insights derived from quantitative methodology. The authors build on the basic framework of regression analysis—which has been enriched over the past two decades with innovations introduced by econometricians and statisticians—to make recommendations about how qualitative analysts should confront a variety of methodological problems. KKV are strongly committed to the premise that the underlying logic of quantitative and qualitative research is basically the same (p. ix). At the same time, they are attentive to the specific dilemmas that arise in actually carrying out qualitative research, and they provide many useful examples and employ clear, nontechnical language. This is a book that moves the discussion forward, and therefore merits close attention.

Selection bias is one of the important topics into which quantitative methodologists have recently offered significant new insights. Hence, the assessment of KKV's treatment of this topic[1] provides a useful window for evaluating their effort to transpose complex issues of quantitative method to the sphere of qualitative research. This is also an interesting topic to address because KKV are centrally concerned with selection bias resulting from deliberate selection by the investigator. Their recommendations are consequently of special importance: if their diagnosis is correct, a small improvement in methodological self-awareness can yield a large improvement in scholarship. Finally, KKV's recommendations merit examination precisely because they are quite emphatic. Given their emphatic character, readers may desire assurance that they are, in fact, receiving sound advice.

The question of how to situate the problem of selection bias in relation to a spectrum of other methodological and theoretical issues is not an easy one. At one pole, in discussions of selection bias in quantitative *sociology*, one finds an influential article suggesting that the impact of selection bias is not as serious as has been believed, that efforts to introduce statistical corrections for selection bias may create more problems than they solve, and that among the many problems of quantitative analysis, this one does not merit special attention (Stolzenberg and Relles 1990). In the present context, the appropriate point of entry into the problem is different. First, in the examination of selection bias in qualitative *political research*, it is much more difficult to assess its precise impact, so that conclusions about its importance are inevitably more tentative. Second, KKV's recommendations about selection bias are centrally concerned with deliberate selection by the investigator. Hence, relevant corrections do not involve statistical procedures, but rather basic choices about case selection that are relatively easy to achieve. Third, one of KKV's central goals is to explore the interrelations among a series of different methodological problems. They are not singling out selection bias as a paramount problem: in their view, it is one of many.

I conclude that KKV offer useful recommendations regarding selection bias. Yet they subsume under this term various issues with which qualitative researchers may already be familiar, but under different labels. Obviously, in a complex field it is common to find that a given phenomenon is named in different ways. For example, what was probably the first paper ever published on selection bias referred to it as a problem of spurious correlation (Berkson 1946, 51). Nevertheless, the overlap of labels raises the question of whether KKV's methodological insights really offer something new to the qualitative analyst.

In fact, some of the important recommendations offered by KKV can just as well be viewed not as insights derived from advanced quantitative methods, but rather as part of a long-standing effort to encourage qualitative scholars to be more methodologically and theoretically aware of which cases they are analyzing. This self-consciousness can also be encouraged by insistently posing a question that, according to the traditional lore of the comparative politics subfield, should often be asked at doctoral dissertation defenses: "What is this a case of?"

Many issues that underlie this question and that are highly relevant to KKV's discussion of selection bias have previously been raised in discussions of the comparative method, that is, the branch of methodology concerned with the systematic, qualitative analysis of relatively small numbers of cases (a "small N").[2] In assessing methodological claims about selection bias, it is useful to take these earlier discussions as a base line. With regard to issues of case selection, they include the ongoing evaluation of J. S. Mill's methods of experimental inquiry, the related distinction between "most similar" and "most different" systems designs, and a new perspective on case selection in small-N studies arising from counterfactual analysis. Regarding the problem of applying concepts and indicators across diverse contexts, an issue that arises in KKV's discussion, relevant insights from work on comparative method include the traditional concern with "conceptual stretching" and the use of "system-specific," as opposed to "common," indicators. Regarding the issue of generalization, relevant insights include the argument that it may at times be appropriate for scholars to limit severely the scope of generalizations from a given set

of cases, an argument with important implications for what it means to think of these cases as being selected from a larger population.

The following discussion devotes central attention to the relationship between KKV's arguments and these familiar issues in the field of comparative method. After presenting an overview of selection bias as a methodological problem, along with some initial caveats, I shall consider the potential strategy of avoiding selection bias through random sampling, examine KKV's treatment of selection bias in descriptive inference, and then explore two issues that arise in their discussion of causal inference. Finally, I shall consider how qualitative researchers understand the role of generalization in social inquiry and the implications for selection bias. Although the discussion generally supports the thesis of a convergence in the logic of quantitative and qualitative methods, it is evident that qualitative researchers at times have different priorities in designing research.

Overview of Selection Bias

Selection bias is commonly understood as occurring when the *non*random selection of cases results in inferences, based on the resulting sample, that are not statistically representative of the population. The focus of the present discussion is on selection bias deriving from deliberate selection by the investigator.[3] A common problem arising from such selection is that it may overrepresent cases at one or the other end of the distribution on a key variable. When this specifically involves the selection of cases that fall above or below a particular value on the distribution of that variable, it is referred to as a form of *truncation*.[4]

The statistical insight crucial to understanding the consequences of such selection is the observation that selecting cases so as to constrain variation toward high or low values of the *dependent* variable tends to reduce the slope estimate produced by regression analysis, whereas an equivalent mode of selection on the *explanatory* variable does not have this effect. If, for example, the analyst selects a sample that is truncated to include only cases that have higher scores on the dependent variable, the sample will tend to overrepresent cases *above* the regression line that is derived from the full data set. This mode of selection therefore gives disproportionate weight in the calculation of the slope to cases for which factors *in addition to* the principal explanatory variable play an important role in producing higher scores on the dependent variable (or lower scores, in the case of a negative relationship). As a consequence, unless the investigator can identify missing variables that explain the position of these cases above the regression line, the bivariate relationship within this subset of selected cases will appear to be *weaker* than in the larger set of cases from which they are selected. A corresponding effect occurs if selection is biased toward the lower end of the dependent variable. By contrast, if selection is biased toward the higher or lower end of the *explanatory* variable, then, as long as the underlying relationship is linear, the expected value of the slope will not change (although it may vary in particular data sets).

This asymmetry is the basis for warnings about the hazards of "selecting on the dependent variable." This expression refers not only to the deliberate selection of cases according to their scores on this variable but to any mode of selection correlated with the dependent variable (i.e., tending to select cases that have higher, or lower, values on that variable) once the effect of the explanatory variables is removed (pp. 138–39). If such a correlation exists, causal inferences will tend to be biased.

Initial Caveats

Selection bias deriving from truncation is in some respects *less* serious—and in other respects *more* serious—than might initially appear to be the case. First, as KKV note, on average it will lead analysts to underestimate the strength of causal effects, and they suggest that estimates derived from the sample may be understood as a "lower bound" in relation to the true causal effect (pp. 130, 139). In qualitative research, where the inductive character of the analysis may entail a kind of ad hoc hypothesizing that can lead to "overfitting" the data, this kind of constraint might be a useful corrective.

Second, the basic asymmetry that calls for warnings about selecting on the dependent variable applies to the slope, but *not* the correlation. The correlation is a "symmetric" measure, and constraining variance on the explanatory variable *can* affect the correlation, potentially making researchers even more vulnerable to selection bias. In quantitative research the slope is more widely used for causal inference than the correlation, in part for this reason. Yet for some specific kinds of causal analysis, the correlation (or the standardized slope, which is a closely related coefficient) is more appropriate (Achen 1982, 74–76). To the extent that it is, the warning about selection bias must be extended to the problem of selecting on the explanatory variable. It remains a topic for further investigation whether the intuitive assessment of causal relationships by qualitative researchers should be understood as more nearly analogous to the slope or the correlation—and hence whether they should be concerned with selection on the independent variable, as well as the dependent variable.

Third, advice about the distinctive problems associated with the dependent variable should be qualified in another sense as well. Selecting on the explanatory variable can affect the slope estimate under some circumstances.[5] With a bivariate linear relationship, sampling toward the high or low end of the explanatory variable does not affect the expected value of the slope. However, if the underlying relationship is nonlinear, selecting different parts of the distribution on the explanatory variable can yield different slope estimates. This is *not* due to selection

bias, but it *does* involve the more general issue that case selection can influence findings. Hence, this more generic form of the problem does not arise only with selection on the dependent variable.

The Option of Random Sampling

Given that selection bias is conventionally understood as deriving from the nonrandom selection of cases, one option for the qualitative researcher might be to engage in random sampling and thereby (hopefully) avoid the whole problem. Yet KKV suggest that random sampling is not the solution.

In the statistical literature, it has been traditional to treat selection bias and sampling error that results from random selection as separate issues.[6] By contrast, KKV argue that sampling error can produce selection bias. They offer a hypothetical small-N example of three cases that have high, medium, and low scores on the dependent variable, from among which only two cases will be selected. They discuss three alternative selection rules and point out that of the three combinations of cases that can result, two will constitute a sample that is biased either toward the high or low end of the dependent variable. They conclude that "since random selection of observations is equivalent to a random choice of one of these three possible selection rules, *random selection of units in this small-n example will produce selection bias with two-thirds probability!*" (p. 126, italics added).

KKV's statement that sampling error produces selection bias with two-thirds probability entails a usage that will surprise many readers familiar with standard statistical terminology.[7] However, their goal here is to point out that with a small N and only *one* random sample, the same kind of error that is associated with selection bias is quite likely to occur. Despite the presumed virtues of random sampling, with a very small N it can produce the same kind of error that is identified with selection bias deriving from a truncated sample. Hence, they argue that the investigator is much better off engaging in a carefully planned form of nonrandom selection (pp. 125–26). This is good advice.

Descriptive Inference

KKV's discussion of selection bias in descriptive inference includes two interesting examples that will be examined here. The first concerns case studies of deterrence in international relations that focus primarily on deterrence failure (pp. 134–35). This focus has an important consequence: inferring an overall success rate of deterrence from such case studies is a big mistake. Achen and Snidal, who are cited by KKV, give the example of a prominent social scientist who used a study of 12 cases of conventional deterrence to reach the surprising conclusion that conventional deterrence fails 83.3% of the time, i.e., 10 out of 12 cases (1989, 162). Yet it does not require a deep knowledge of modern regression analysis to grasp this problem, which can readily be understood within the conventional framework of the comparative method. Thus, it is essential for researchers to keep in mind how and why the cases were originally selected, and scholars following any variant of what J. S. Mill called the "method of agreement" (1974, 388–90) should not use the cases thereby selected as a basis for descriptive generalizations concerning the dependent variable.

In another example concerned with descriptive inference, KKV consider a hypothetical assessment of support for the Liberal party in New York State, based on votes for candidates endorsed by the Liberals in elections for the State Assembly. Because the Liberals do not endorse candidates in many districts where they believe they will lose, such a study would provide an inadequate assessment of support for the party. Thus, KKV argue, descriptive inferences derived from this study would suffer from selection bias (p. 135).

Within the tradition of work on the comparative method, qualitative researchers might more readily understand this example as raising issues linked to the problem of conceptual stretching and also as a complex measurement problem that may call for system-specific indicators. Central issues here are the definition of a political party and the problem of measuring party support. If one accepts a Sartori-type definition, according to which parties are political groups that present candidates in elections to public office (1976, 64), then the Liberals are not acting as a party in those districts where they do not present candidates. This doubtless construes the definition too narrowly, but it is appropriate to emphasize that in multiparty, as opposed to two-party, systems, the practice of not running candidates in selected districts is more common, and the question of what is and is not a party is often more complex. Some conceptual reflection would seem essential here. Second, analysts who use elections as a source of data on political support would normally devote close attention to which candidates run in a particular year, because alternative candidates for a given party will generate different profiles of support. Hence, it is *not* the case that elections usually provide a straightforward measure of party support. If elections are, nonetheless, used to measure party support, some other system-specific measure should be used in the districts where the Liberals do not run candidates.

Thus, one could argue that in the case of the Liberal party in New York, using the popular vote to measure party support will produce descriptive inferences that suffer from selection bias. Alternatively, this can be viewed as a problem of conceptualization and of developing system-specific measures, topics not covered in KKV's book but that are familiar themes of comparative method.

Mild Versus Extreme Selection Bias

An important element in KKV's discussion of selection bias in causal inference is their distinction between a "milder" form of selection bias, that results

from merely constraining variation on the dependent variable, and what they refer to as an "extreme" form that results from selecting only *one value* on the dependent variable—according to them, a grave mistake (p. 130). This strategy of selecting only one value may be adopted by scholars who are analyzing an outcome of exceptional interest (e.g., deterrence failure, revolutions, or high growth rates) and who wish to focus only on this outcome, out of a belief that they will thereby achieve greater insight into the phenomenon itself and into its causes. Alternatively, they may be dealing with an outcome about which previous theories, conceptualizations, measurement procedures, and empirical studies provide limited insight. Hence, they may be convinced that a carefully contextualized analysis of one or a few cases of the outcome will be more analytically productive than a broader study that compares cases of its occurrence and nonoccurrence.

This distinction between mild and extreme selection bias conflates two distinct issues with which KKV are concerned. The first is the core issue of investigator-induced bias, involving the fact that the greater the constraint on the variance of the dependent variable, the more severe the bias in inference is likely to be. The second issue is that at the outer limit, when variance on the dependent variable disappears and the investigator focuses on only one outcome on that variable, a shift to a different kind of research design has occurred. Where there is no variance, selection bias certainly may be present in that the sample may well overrepresent cases for which factors other than the main explanatory variable play an important role in accounting for their higher scores on the dependent variable. But that outcome can more usefully be treated as a different issue from the switch in research design.

Selecting One Value on the Dependent Variable

Some of the strongest criticisms regarding selection bias have been leveled against studies that focus on a single outcome on the dependent variable, which KKV characterize as an extreme form of selection bias. In such studies, according to them, "nothing whatsoever can be learned about the causes of the dependent variable without taking into account other instances when the dependent variable takes on other values" (p. 129). They suggest that the need for variation on the dependent variable "seems so obvious that we would think it hardly needs to be mentioned" and that research designs lacking such variation "are easy to deal with: avoid them!" (pp. 129, 130).

On the one hand, given that they advance a definition of "causal effect" that requires the observation of at least two different values on the dependent variable (pp. 81–82), within their own framework their position can be seen as making sense. Yet other perspectives on this question are available. In the field of comparative method, a traditional way of thinking about this design is in terms of J. S. Mill's method of agreement, a perspective that KKV note (p. 134) but do not develop. This label is used by Mill because all cases under investigation "agree" on the dependent variable. Many authors have examined the strengths and weaknesses of this design, and a standard view, expressed by Mill himself, is that this design fails to provide a positive demonstration of causation, and rather should be viewed as a "method of elimination," which can exclude causal factors if they are consistently not present when a given outcome occurs (1974, 392). As Jervis has suggested, this design may serve to assess the necessary conditions of a given outcome, or, to put it more precisely, to eliminate some hypothesized necessary conditions (1989, 194). In this sense, KKV's assertion that this type of design makes it "impossible to evaluate any individual causal effect" (p. 134) seems incomplete: it can serve to *eliminate* some hypothesized causes, which can be a useful first step in causal analysis.

A second perspective on this design becomes relevant if analysts compare cases that are matched on the dependent variable but are extremely different from one another in other respects, in which case this can be called a "most different systems" design (Przeworski and Teune 1970). One of the merits of such a design is that the challenge of distilling a common set of explanatory factors out of this diversity can push scholars to discover new explanations that might not have emerged from the analysis of a more homogeneous set of cases (Collier 1993, 112).

A third perspective is found in Fearon's discussion of counterfactuals as a means of testing hypotheses within the framework of small-N analysis (1990, 179–80). He suggests that one can make "methodological sense" of designs with no variance on the dependent variable by recognizing that scholars can employ counterfactual analysis to introduce variance, and he goes on to present a detailed discussion of how such counterfactual analysis can be carried out.

Given these three alternative perspectives, KKV's claim that "nothing whatsoever can be learned about the causes of the dependent variable" if it does not vary within a given study would seem to be excessively limiting.

Two final observations may be made about this type of design. First, it appears that studies lacking variation on the dependent variable may be less common than scholars concerned with selection bias have sometimes implied, and studies that appear to lack it may have it after all. Michael Porter's (1990) book on industrial competitiveness, analyzed by KKV, is a case in point. The authors argue that Porter focuses on 10 nations that share a common outcome on the dependent variable of "competitive advantage," a research design that "made it impossible to evaluate any causal effect" (134). However, as Porter repeatedly points out, a central concern of his study is with explaining success and failure not at the level of nations, but rather at the level of firms and industrial sectors (e.g., pp. 28–29, 33, 69, 577, 735), of

which he considers both successful and unsuccessful cases. At this level, it is incorrect to state that he lacks variance on the dependent variable. He may or may not take full advantage of the variance that is present in his study, but that is a different issue. Studies can doubtless be found in which such variance is completely lacking. Yet on closer inspection, one may at times discover some variation after all. In fact, due to a scholarly instinct for "variation seeking," analysts may have a strong tendency to find some variation on the dependent variable.

The other observation concerns the real trade-offs between these different designs. If little is known about a given outcome, then the close analysis of one or two cases of its occurrence may be more productive than a broader study, focused on positive and negative cases, in which the researcher never becomes sufficiently familiar with the phenomenon under investigation to make good choices about conceptualization and measurement, which in turn can lead to conclusions of dubious validity. On the other hand, by not utilizing the comparative perspective provided by the examination of negative cases, the researcher gives up a lot. In general, it is productive to build in a comparison of contrasting outcomes.

Samples, Populations, and the Role of Generalization

Another area in which issues of selection bias intersect with discussions of comparative method concerns the relationship among samples, populations, and the issue of generalization. Discussions of selection bias by definition presume the existence of a larger set of cases, from among which the cases under analysis have in some sense been chosen. Indeed, the claim that one is selecting cases that tend toward the high end of the dependent variable is not meaningful apart from the identification of a larger set of cases that define a range for this variable. Although in some domains of research the definition of the population is clear, in many domains, as KKV (p. 125) and others have noted, it may not be clearly specified, or its definition may be a matter of debate.

In qualitative comparative studies, a central issue in the definition of the population is a fundamental ambivalence about the process of generalization to additional cases. On the one hand, the generalization of empirical findings from an initial set of cases is a basic priority of social science research, and findings that cannot be generalized are routinely considered less important. On the other hand, over the past couple of decades a concern with sensitivity to context has been stimulated by a diverse spectrum of authors.[8] This concern has led many analysts to conclude that even important theories may sometimes apply only to limited domains. If the cases under study in fact constitute the full set of theoretically relevant cases, then an issue of selection bias in relation to a larger set of cases does not arise.[9]

Within the framework of a single piece of research, a critical issue is the appropriate balance between a legitimate process of delimiting the scope of findings and a degree of particularism that excessively limits the contribution of the study. A further issue of balance arises when other analysts become interested in the findings of a given study and wish to extend them to additional cases. On the one hand, these analysts should be alert to the limitations on the scope of claims that the original author sought to impose. On the other hand, from a different theoretical or comparative perspective these other analysts might make a different decision about the appropriate scope and seek to extend the analysis to additional cases. Hence, for them a problem of selection bias could arise that was not an issue for the original author. This kind of shift can occur in any sphere of research. However, it may have special importance in areas of qualitative comparative research in which investigators are particularly concerned about imposing constraints on the scope of their findings.

KKV's arguments about selection bias are most usefully understood as pushing qualitative researchers to think about a spectrum of selection issues. These include: (1) the core problem of selection bias that has been illuminated by advanced quantitative methods, that is, the specific impact on causal inference of certain kinds of deliberate case selection; (2) other issues already familiar to many qualitative researchers, including broader questions of case selection and their implications for various approaches to descriptive and causal analysis; and (3) additional areas in which the priorities of quantitative and qualitative researchers may sometimes be quite different—as with the issue of selecting matched versus contrasting cases and the implications for selection bias of severe restrictions on claims about scope.

These points of convergence lend support to KKV's claim that the underlying logics of quantitative and qualitative research are similar. The convergence also underscores the fact that some of their important recommendations do *not* provide qualitative researchers with new methodological insights. Finally, the divergences remind us that these two traditions sometimes make different choices about underlying trade-offs entailed in the design of research.

From the perspective of qualitative researchers, the core concern that should emerge out of these discussions can again be expressed in terms of the question, "What is this a case of?" If the debate on selection bias stimulates qualitative researchers to address this question more frequently and successfully, it will have accomplished a lot.

Notes

This work grows out of an analysis of selection bias in the field of comparative politics that I am carrying out jointly with James Mahoney. Christopher Achen, Henry Brady, Gary King, Mark Lichbach, and Laura Stoker provided exception-

ally helpful comments on earlier drafts. Valuable suggestions were also made by Jake Bowers, Ruth Collier, Neil Fligstein, David Freedman, Lynn Gayle, Lincoln Moses, Michael Pretes, Thomas Romer and Mark Turner. My work on this analysis at the Center for Advanced Study in the Behavioral Sciences was supported by National Science Foundation Grant No. SBR-9022192.

1. Chapter 4 discusses the overall problem of selecting cases, or observations (as KKV call them), and one third of it is specifically concerned with selection bias.

2. Most of these issues were raised in such "classic" statements on comparative method as Bendix 1963; Lijphart 1971; Przeworski and Teune 1970; Sartori 1970; and Smelser 1976.

3. On other specific contexts in which selection bias arises, see Achen 1986; Geddes 1990, 145–48; King 1989; and Przeworski and Limongi 1992, 1993.

4. Truncation can take other forms as well; see p. 142 and Moses 1968.

5. KKV make a parallel point on p. 137.

6. See, for example, the definition of sampling error in the classic *Dictionary of Statistical Terms* prepared for the International Statistical Institute (Kendall and Buckland 1960, 255–56).

7. This formulation is stated in such a way that it appears to overlook a key theoretical idea about sampling. With truncated samples, the expected value of the estimate is biased, whereas with random samples, the expected value of the estimate is unbiased in that, if the sample is drawn a sufficient number of times, the average value of the estimates provided by the samples will be equal to the parameter one is estimating. Thus in their example it would be more helpful to say that there is a two-thirds probability that *any one sample* will contain this kind of error.

8. See Geertz 1973; Przeworski and Teune 1970; Ragin 1987; and Skocpol and Somers 1980; see also Walker and Cohen's (1985) discussion of "scope statements."

9. Moses (1968, 197) and Stolzenberg and Relles (1990, 407–08) likewise argue that problems of selection bias depend on the definition of the relevant population.