Small N’s and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases*

STANLEY LIEBERSON, Harvard University

Abstract

An increasing number of studies, particularly in the area of comparative and historical research, are using the method of agreement and method of difference proposed by Mill (1872) to infer causality based on a small number of cases. This article examines the logic of the assumptions implicit in such studies. For example, the research must assume: (1) a deterministic approach rather than a probabilistic one, (2) no errors in measurement, (3) the existence of only one cause, and (4) the absence of interaction effects. These assumptions are normally inappropriate, since they contradict a realistic appraisal of most social processes, but are mandatory if we follow Mill’s causal analyses based on small N’s. Research should not attempt employment of such methods in small-N cases without a more rigorous justification of heroic assumptions and a guard against possible distortions.

This article evaluates a sociological approach that is gaining in usage, especially in researching historical and comparative problems. Namely, we will consider the causal inferences drawn when little more than a handful of nations or organizations — sometimes even less — are compared with respect to the forces driving a societal outcome, such as a political development or an organizational characteristic. 1 Application of this method to a small number of cases is not new to sociology, being in one form or another a variant of the method of analytic induction, described by Znaniecki (1934) and analyzed succinctly by Robinson (1951) and Turner (1953). 2 The conclusions rely on a formalized internal logic derived from Mill’s (1872) method of agreement and his method of difference (see the discussion of Mill in Nichols [1986]). The formal rigor of this type of analysis sets it off from other small-case procedures that also imply causality.

* I am indebted to William Alonso, Rogers Brubaker, John Campbell, William Kruskal, and Peter V. Marsden for stimulating discussions or comments on this topic. A modified version of this article will be reprinted in What Is a Case, edited by Charles C. Ragan and Howard S. Becker, Cambridge University Press, 1992. Direct correspondence to Stanley Lieberson, Department of Sociology, Harvard University, Cambridge, MA 02138.

such as in *Street Corner Society* (Whyte 1943) or in the development of the urban structure and growth model of Burgess (1925). It is also different from case studies that seek to point out merely that a given phenomenon *exists* in some setting, as opposed to analyzing its causes. The comments are, however, to some degree relevant in evaluating the Boolean method proposed by Ragin (1987) for dealing with somewhat larger samples used in comparative and historical research. Moreover, although the analysis is stimulated by recent developments in macrohistorical research, it is pertinent to a wide variety of other studies that use Mill’s logic with a small number of cases.

There are numerous instances of these small-N studies, which apply Mill’s methods to infer causality. For example, Moore (1966) uses them in his analysis of the role of various classes and peasantry leading to three different outcomes in seven different nations. Skocpol’s (1979) influential study of social revolutions in a small number of countries also uses these methods and attributes their application to de Tocqueville and Bloch. Katzenelson (1985) explains “the puzzle of variation between English and American patterns of working-class formation” (260) by eliminating any common features shared by the two nations, searching for a distinctive difference that must account for these developments. Likewise, Stepan (1985) employs similarities and differences between Argentina, Brazil, Chile, and Uruguay to infer relations between “the power of the state and the power of civil society” (318). Orloff and Skocpol (1984) compare welfare developments in the U.S. and Britain around the turn of the century, eliminating major conventional theories of welfare-state development because the independent variables are considered the same in both nations. Only one variable is found to differ between the two nations and causal significance is therewith assigned.

More generally, we are dealing with any study that attempts to infer the existence or absence of causality through the correlation between the dependent variable and an independent variable across a small number of cases. When the cases differ in outcomes, then it is assumed that the causal variable will have a correlation of ±1.0 with the dependent variable, with all other correlations having other values. When the cases have the same outcome, then it is assumed that the causal variable will be the same for all cases, with each of the other variables having some variability across cases. Even the classic study by Weber (1930) of the role of the Protestant ethic in the development of capitalism can be used as an illustration of the method of difference. A capitalistic system is observed to develop in only one part of the world. Weber then asks what variable is found in the Protestant West but nowhere else. All other variables, which the countries share, are eliminated as the necessary causal variable. One can readily understand why the logic is appealing despite Mill’s (1872) warning that it is inappropriate for social science problems. He specifically illustrates the inapplicability of the method of difference by describing differences between nations, and he observes the weakness of the method of agreement for the typical social science problem as well. As Tilly (1984) observes, the presence of more than one common causal variable makes it impossible for the latter method to sort out the actual cause.

We can easily appreciate the goal of making causal inferences based on a small number of cases. If data were available with the appropriate depth and
Small N's and Big Conclusions

Lieberson, Stanley, *Small 'N's' and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases*, Social Forces, 70:2 (1991:Dec.) p.307

detail for a large number of cases, obviously the researcher would not be working with these few cases (assuming a minimal time-energy cost). Since the data are not available, or the time-energy cost is too great, one can only approach these efforts with considerable sympathy for their objective. We address three questions: (1) What are the assumptions underlying these studies? (2) Are these assumptions reasonable? (3) What can we do to improve such studies in those instances when they might be appropriate forms of inquiry?

However, there is a special difficulty in comprehending the weakness of these applications and the gross assumptions that are necessarily made; namely, the methods are typically applied to topics for which the reader has no experiential grasp. Readers cannot easily see the unreasonableness of a conclusion about macrosocietal events in a distant time and place. To display the nature of this logic, this article draws illustrations from hypothetical applications to automobile accidents and the evaluation of airlines.

Probabilistic and Deterministic Perspectives

Let us start by distinguishing between causal propositions that are deterministic as contrasted with those that are probabilistic. The former posits that a given factor, when present, will lead to a specified outcome. The latter is more modest in its causal claim, positing that a given factor, when present, will increase the likelihood of a specified outcome. When we say, "If \( X_1 \), then \( Y \)," we are making a deterministic statement. When we say, "The presence of \( X_1 \) increases the likelihood or frequency of \( Y \)," we are making a probabilistic statement. Obviously, if given the choice, deterministic statements are more appealing. They are cleaner, simpler, and more easily disproved than probabilistic ones. One negative case (\( Y \)'s absence in the presence of \( X_1 \)) would quickly eliminate a deterministic statement.

Alas, a probabilistic approach is often necessary to evaluate the evidence for a given theoretical perspective, even if we think in deterministic terms. This occurs for a variety of reasons, not the least being measurement errors — a serious problem in the social sciences. The existence of a measurement error means that a given data set may deviate somewhat from a hypothesized pattern without invalidating the hypothesis. In addition to this technical matter, there is an additional problem: complex multivariate causal patterns operate in the social world, such that a given outcome can occur because of the presence of more than one independent variable and, moreover, may not occur at times because the influence of one independent variable is outweighed by other influences working in the opposite direction. Under such circumstances, the influence of \( X_1 \) is only approximate (even without measurement errors), unless one can consider all the other independent variables, through controls or otherwise.

Furthermore, we often do not know or cannot measure all the factors that we think will influence \( Y \). As a consequence, we are again obliged to relinquish a deterministic measurement of the influence of \( X_1 \) on \( Y \), even if we are prepared to make a deterministic statement about its influence. There are yet other reasons for reverting to a probabilistic rather than a deterministic approach,
namely, the role of chance in affecting outcomes. Beyond consideration here is
the question of whether chance per se exists or is simply a residual label
referring to our ignorance about additional influences and/or inadequate
measures for the variables under scrutiny. In either case, some form or another
of indeterminacy is clearly useful to employ in the physical sciences, let alone
in the social sciences (see examples in Lieberson [1985:94-97]). Any of these
factors would lead to probabilistic statements of outcomes rather than deter-
nomistic ones.

This distinction is more than merely an academic one. Rather, it is em-
bedded in our daily thinking. Suppose we examine the influence of alcohol on
automobile accidents. Even if we believe there is such an influence, we would
still expect some sober drivers to have chargeable accidents and not all drunk
drivers to experience accidents. If we find that some sober drivers did cause
accidents and some drunk drivers did not, these deviations would not lead us
to reject automatically the proposition that drunkenness causes automobile
accidents. Rather, we would look at a set of data and ask if the probability or
frequency of accidents was greater for drunk drivers than sober ones. Why is
this so? Even if taking a deterministic view, we would expect several factors to
influence the likelihood of an accident, alcohol being only one of them. Indeed,
we would expect an interaction effect for drunkenness, such that one drunk
driver might run a red light in a busy intersection and have an accident
whereas another driver might be fortunate to enter the intersection when the
light was green. To be sure, we might want to take some of these additional
factors into account, and we would then expect the influence of drinking to be
more sharply displayed. But it is unlikely that we could isolate alcohol's
influence from all the additional conditions that either prevent drinking from
causating an accident or lead a sober driver to have an accident. The net effect is
that we would not totally reject our idea about alcohol and driving if we
compare a drunk driver with a sober one and find the latter has had an accident
and the former has not.

Despite these facts, small-N studies operate in a deterministic manner,
avoiding probabilistic thinking either in their theory or in their empirical
applications. As one distinguished proponent of the small-N approach says, "In
contrast to the probabilistic techniques of statistical analysis — techniques that
are used when there are very large numbers of cases and continuously
quantified variables to analyze — comparative historical analyses proceed
through logical juxtapositions of aspects of small numbers of cases. They
attempt to identify invariant causal configurations that necessarily (rather than
probably) combine to account for outcomes of interest" (Skocpol 1984:378). One
good reason for this disposition is the following principle: 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.

Let us draw an analogy with flying a given airline. Suppose a rude
employee is encountered, luggage is lost, or the plane is delayed. After such an
experience, one could decide to use a different airline. However, one would
know that although airlines may differ in their training programs, employee
relations, morale, luggage practices, airplane maintenance, and other factors
affecting their desirability, a very small number of experiences is insufficient to evaluate airlines with great confidence. If airlines differ, it is in the frequency of unpleasant experiences rather than in the ability of one airline to have consistently polite employees, never to lose luggage, or to avoid all mechanical problems. Based on a small number of experiences, one may decide to shun a certain airline, and the decision is not totally wrong, since the probabilities of such experiences in any given small number of events is indeed influenced by the underlying distribution of practices in different airlines. However, conclusions drawn on the basis of such practices are often wrong. Passengers with a small number of experiences will draw very different conclusions about the relative desirability of various airlines. These conclusions are often wrong because a small number of cases is an inadequate basis for generalizing about the process under study. Thus, if we actually knew the underlying probabilities for each airline, it would be possible to calculate how often the wrong decision will occur based on a small number of experiences. The consumer errors are really of no great consequence since making decisions on the basis of a small number of events enables the flyer to respond in some positive way to what can otherwise be a frustrating experience. Such thinking, however, is not innocuous for the research problems under consideration here; it will frequently lead to erroneous conclusions about the forces operating in society. Moreover, other samples based on a small number of different cases — when contradicting the first sample, and this is almost certain to occur — will create even more complicated sets of distortions as the researcher attempts to use deterministic models to account for all the results. In my judgment, this is not a step forward.

Briefly, in most social research situations it is unlikely that the requirements of a deterministic theory will be met. When these conditions are not met, the empirical consequences of deterministic and probabilistic theories are similar in the sense that both will have to accept deviations: the former because of errors in measurement and controls, the latter both because of those reasons and because the theory itself incorporates some degree of indeterminacy (due to inherent problems in either the measurement or knowledge of all variables or due to some inherent indeterminacy in the phenomenon).

The implications of these deviations are frequently seen in social research. For example, it is very difficult to reject a major theory because it does not appear to operate in some specific setting. One is wary of concluding that Weber was wrong because of a single deviation in some insufficiently understood time or place. Such a conclusion would be no different from deciding that an accident caused by a sober driver disproves the notion that drinking causes automobile accidents.

Suppose, for example, there is a single deviation among a small number of cases or a modest number of deviations among a larger number of cases. What are the consequences for the deterministic theory under consideration? If the deterministic theory is univariate, then either one variable or one specific combination of variables (an interaction) causes a given outcome. The theory can be rejected with a single deviation if we are confident that there are no measurement errors (a nontrivial consideration for either statistical or qualitative descriptions) and that there are no other possible causes of the dependent variable. As for a multivariate deterministic theory, where more than one
variable or combination of variables could account for the consequence, the
theory can be rejected with a single deviation if we are confident that there are
no measurement errors — as above — and if we know and consider all other
factors hypothesized to affect the outcome.

The formal procedures used in the small-N comparative, historical, and
organizational analyses under consideration are all deterministic in their
conception. Indeed, small-N studies cannot operate effectively under probabilistic
assumptions because they would require much larger N’s to have any
meaningful results. This ineffectiveness becomes clear when we analyze the
reasoning used in small-N studies based on Mill’s methods.

Mill’s Methods

As Skocpol (1986) observes, the key issue is the applicability of Mill’s method
of agreement and method of difference to small-N data. Nichols (1986) agrees
but shows how the application of this logic makes a variety of assumptions that
are unrealistic for social processes and lead to serious distortions. I will build on
and modify this important critique.

Let us start with the method of difference, which deals with situations in
which the dependent variable (outcome) is not the same for all cases. Here the
researcher examines all possible independent variables that might influence this
outcome, looking for a pattern where only one of the independent variables is
perfectly correlated with the dependent variable, whereas the other independent
variables are either constant (regardless of shifts in the dependent variable) or
imperfectly correlated with the dependent variable. Examples of this might be
where $X_i$ is constant in all cases or varies between cases in a manner different
from the dependent variable. This method is applied even with two cases, as
long as only one of the independent variables differs and the others are constant
across the cases (Orloff & Skocpol 1984). Table 1 illustrates this type of analysis.
For simplicity, let us assume that all the independent variables as well as the
explicandum are dichotomies with “yes” and “no” indicating the presence or
absence of the attribute under consideration. To illustrate my points as clearly
as possible, I have used an illustration based on automobile accidents. The logic
used follows that of Mill’s methods and is identical to that employed in
deterministic studies of macrophenomena.

Applying the method of difference to the hypothetical data in Table 1, we
would conclude that the auto accident is caused by $X_2$ because in one case a car
entered the intersection whereas in the other case no car entered. We would
also conclude that the accident is not caused by drunk driving or the running of
a red light because the variables (respectively $X_1$ and $X_3$) are the same for both
drivers yet only one had an accident. Such conclusions are reached only by
making a very demanding assumption that is rarely examined. The method’s
logic assumes no interaction effects are operating (i.e., that the influence of each
independent variable on $Y$ is unaffected by the level of some other independent
variable). The procedure cannot deal with interaction effects; it cannot distin-
guish between the influence of inebriation or running a red light from another
constant, such as the benign fact that both drivers were not exceeding the speed
TABLE 1: Application of the Method of Difference

<table>
<thead>
<tr>
<th>Accident</th>
<th>Drunk Driving</th>
<th>Car Entering from Right-Hand Direction</th>
<th>Driver Speeding</th>
<th>Runs a Red Light</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y)</td>
<td>(X₁)</td>
<td>(X₂)</td>
<td>(X₃)</td>
<td>(X₄)</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

limit. Since $X₁$ and $X₄$ are constant, under this logic it would follow that neither inebriation nor running a red light was related to the accident. The procedure does not empirically or logically eliminate interaction effects. Rather, it arbitrarily assumes that they do not operate and that therefore constants cannot influence the dependent variable.⁶ Unless interactions are automatically ruled out a priori, this means that the results in Table 1 (and all other small-$N$ applications of the method) fail to provide any determination of the influence of variables $X₁$, $X₃$, and $X₄$ on the phenomenon under consideration.⁷ Consider another example: ten people apply for a job; five are black and five are white. One of the five blacks and all the five whites are hired. Applying the method of difference, one would conclude that race did not affect employment. Rather, it would have to be some variable that separates all the employed persons from the four who did not get a job. Using a small $N$ with the method of difference, it is not possible to examine interaction effects or multiple causes. Their absence is assumed.

Note also how this method has a certain limited generality unless one assumes, a priori, that only one variable causes the phenomenon under study. For variables that are constant, it is impossible to rule out their influence under different levels simply because there are no measurements. From Table 1, for example, we know that an accident occurs although $X₃$ is constant. Even ignoring the question of interaction effects, it is impossible to conclude that $X₃$ does not cause accidents unless one assumes there is only one cause of accidents. In this case, and this asymmetry is common in small-$N$ studies, we only know about situations where drivers are not speeding. Note again the assumptions that are introduced: if there is any generality to the results, it means that only a single causal variable is operating; otherwise, under the logic used in such studies, the influence of constants are not really taken into account in the method of difference.⁸ These assumptions have considerable bearing on the generality of such small-$N$ comparative studies.

In Table 2, we have a new situation in which two drivers both experience accidents. As before, the two drivers are drunk, both cars run red lights, and again in only one instance was another car appropriately entering the intersection. This time, however, the second person was driving at a high speed whereas the first driver was not. Intuitively, it is not unreasonable to assume that high-speed driving could affect the chances of an accident, by the car failing to make a turn or by causing a skid. At any rate, since both drivers have accidents, the logic generated by Mill’s method of agreement is applied, where
presumably the causal variable is isolated by being the only constant across the two instances, whereas all of the other attributes vary. However, notice what happens under that logic. The previous cause, \(X_2\), is now eliminated since it varies between two drivers who both have accidents. Previously, \(X_1\) and \(X_4\) could not have caused an accident, but they are now the only two contenders as a possible cause. Since only one driver is speeding in this case and both drivers have accidents, it follows that the addition of this factor could not have caused an accident — an extraordinary conclusion, too. What has gone wrong? This example shows how Mill’s method cannot work when more than one causal variable is a determinant and there are a small number of cases. Comparison between the two tables shows how volatile the conclusions are about whether variables cause or do not cause accidents. Every fact is identical about the first driver in both cases, but the fact that the second driver was speeding and therefore had an accident completely alters our understanding of what caused the first driver to have an accident. Another shortcoming to such data analyses is that the conclusions are extremely volatile if it turns out that a multideterministic model is appropriate. Moreover, with a small-\(N\) study, although it is possible to obtain data that would lead one to reject the assumption of a single-variable deterministic model (assuming no measurement error), it is impossible for the data to provide reasonable assurance that a single-variable deterministic model is correct, even if the observed data fit such a model.

These comparisons suggest more than the inability of Mill’s methods to be used in a small number of cases to deal with a multivariate set of causes. As Nichols (1986) points out, Mill had intended these methods as “certain only where we are sure we have been able to correctly and exhaustively analyze all possible causal factors” (172). Nichols goes on to observe that Mill rejects these methods when causality is complex or when more than one cause is operating. Beyond these considerations, important as they are, the above analysis also shows how exceptionally vulnerable the procedure is to the exclusion of relevant variables. In case 2, had we excluded \(X_4\), inebriation would have emerged as the causal factor. Large-\(N\) studies also face the potential danger that omission of variables will radically alter the observed relations, but the susceptibility to spurious findings is much greater here.

Suppose a researcher has a sufficient number of cases such that there are several drivers who have accidents and several who do not. Would this
facilitate interpreting the deterministic model based on a small number of cases? In my opinion, it is unlikely. If drinking increases the probability of an accident but does not always lead to one, and if sobriety does not necessarily enable a driver to avoid causing an accident, then it follows that some drunk drivers will not experience accidents and some sober drivers will. Under the circumstances, there will be no agreement for these variables among all drivers experiencing an accident, and there will be no agreement among those not experiencing an accident. This lack of correspondence means that neither of Mill’s methods will work. A difference in the frequency of accidents linked to drinking will emerge, but this is ruled out (and more or less has to be) in the deterministic practices involving small-N studies. Multicausal probabilistic statements are simply unmanageable with the procedures under consideration.

One way of thinking about this small-N methodology is to visualize a very small sample taken from a larger population. Let us say we have a small sample of nations or of political developments drawn randomly from the universe of nations or the universe of political developments. What is the likelihood that the application of Mill’s methods to this small sample will reproduce the patterns observed for the larger universe? In my estimation, rarely do we encounter large-N studies where all of the relevant causal variables are determined, where there are no measurement errors, and where all cases are found as neatly as is assumed with small-N studies. Yet this is exactly what the small-N study assumes in order to obtain any conclusions, to wit, if all cases were equally well known, rather than limited to the small sample, the patterns observed would be duplicated with essentially no exceptions in the universe. Is this reasonable? Also, how often in large-N studies would restrictions to a deterministic univariate theory make sense?

This type of analysis cannot guard against the influence of chance associations. Indeed, the assumption is that chance cannot operate to generate the observed data. Because it is relatively easy to develop a theoretical fit for small-N data, researchers are unable to guard against a small-N version of the ad hoc curve fitting that can be employed in large-N studies (see the discussion of Taylor’s theorem in Lieberson [1985:93]). Ironically, small-N deterministic analyses actually have the same goal as some types of large-scale empirical research, namely, explaining all the variance. The former is just another version of this large-scale research, subject to the same dangers (Lieberson 1985), as well as to special ones because of demanding assumptions necessary when using small N’s.

**Theoretical Concerns**

Two implications follow from this review; one is theoretical and the other deals with empirical procedures. Dealing with the theoretical questions first, we see that the small-N applications of Mill’s methods cannot be casually used with all macrocultural data sets. The methods require very strong assumptions: a deterministic set of forces, the existence of only one cause, the absence of interaction effects, confidence that all possible causes are measured, the absence
of measurement errors, and the assumption that the same clean pattern would occur if data were obtained for all relevant cases.

At the very least, users must recognize that these assumptions are mandatory in this procedure. The issue then becomes: Under what conditions is it reasonable to make these assumptions ("reasonable" in the sense that they have a strong likelihood of being correct)? The empirical data themselves cannot be used to test whether or not the assumptions are correct; for example, the empirical data gathered in the typical small-N study cannot indicate whether a univariate deterministic cause is operating or if there are no interaction effects. Theories of large-scale organizations, "qualitative" or not, must direct themselves to these questions before the data analyses begin. Moreover, the theories have to develop ways of thinking about these problems so the researcher can decide if they are reasonable. Certainly, the Boolean method proposed by Ragin (1987) is a step in the right direction, although it does require a relatively larger N than the type of small-N studies under consideration here.10

The Quality of Qualitative Data

It should be clear how critical it is that we take extraordinary care in the design and measurement of the variables in small-N studies, whether or not the study is "qualitative." Care is always appropriate, but the impact of error or imprecision is even greater when the number of cases is small. The deterministic model used in these studies requires error-free measurement. The choice of cases for study is itself critical, requiring considerable thought about the appropriate procedure for choosing them. Presumably, these are self-evident facts to practitioners of this approach, and the intense scrutiny of a small number of cases should mean exceptional care with the descriptions.

However, exceptionally rigorous practices are necessary to avoid some methodological pitfalls. If a small number of cases are selected using reasonably rigorous criteria, then it makes a great difference whether or not the outcomes are the same in each case. If they are the same, then the method of agreement is used such that a solution occurs only if one variable is constant in all cases; if different, then the only solution occurs when all but one of the variables are constant across all cases. This approach is nothing more than a repetition of procedures dating back to Mill. Less obvious, as far as I can tell, are the implications this approach has for the delineation of each independent variable. If an independent variable consists of nominal categories, there should be little difficulty delineating them, since presumably trained observers would agree on the classification of each measure. The researchers use the same checks as they would in any large-scale study (e.g., content analysis). But if the independent variable is even ordinal, there is a certain arbitrariness in the way an ordered variable is dichotomized or otherwise divided (polytomized).

To simplify the point, let us consider dichotomies. The method of agreement will only work if all the cases for one causal variable fall in the same category and if no other variable has such uniformity. This means that the cutoffs are critical. The same effect holds for the method of difference, but here the results must be uniform for all but one variable, with the one critical exception being
associated with differences in the dependent variable. Under the circumstances, the delineation of the dichotomies or polytomies is critical and has to be done as rigorously as possible since the boundaries will influence the results enormously. This means that rigor is mandatory when locating the variables if they are nominal and even more so when they are ordinal (e.g., careful driver vs. careless driver).

With the method of difference, where there is an inverse relationship between the number of cases and the importance of finding all but one variable constant across cases, researchers have to guard against using such broad categories as to make it relatively easy for cases to fall into the same rubric. With the method of agreement, where it is vital that all but one variable be different across the cases, the danger is in constructing narrow categories within each variable so that it will be relatively hard for cases to fall into the same rubric. In short, because of the subtle pressure to obtain only one variable that is homogeneous (in the case of agreement in the dependent variable) or only one variable that is heterogeneous (in the case of disagreement in the dependent variable), one must also guard against the bracketing of attributes in the former case and the decomposition in the latter. For these methods to work at all, researchers must introduce formal criteria for these decisions that can be followed in advance of a given research project. To my knowledge, these criteria do not currently exist. (It would be an interesting study in both the sociology of knowledge and research methodology to examine whether the breadth of categories in recent studies is related to which method the study involves.)

Because of the small N's and the reasoning Mill's methods require, it is vital to include all possible causal variables. Yet this will tend to lead to inconclusive results if carried out in a serious way, since the method of agreement will probably turn up with more than one variable that is constant for all the cases and, likewise, the method of difference will have more than one independent variable that is associated with the difference in the dependent variable. Suppose, for example, we find that a drunk driver does not experience an automobile accident but the sober one does. In this case, using the small-N methods practiced in historical sociology, the investigator is in danger of concluding that sobriety causes automobile accidents or at the very least is the cause in the observed situation. At best, and only if the correct causal factor is included, the study will conclude that either sobriety or some other factor causes automobile accidents. At worst, if the correct causal factor is excluded, sobriety will be the cause. So there is a dilemma here; a clean result will tend to occur only with a modest number of independent variables, but this very step is likely to increase the chances of an erroneous conclusion.

Also, the relationship between the independent variables and the dependent variable is distorted if the cases are selected so as to have agreement or disagreement with respect to the dependent variable (rather than simply sampling from all of the cases). It can be shown that sampling in order to obtain a certain distribution with respect to the dependent variable distorts the explicandum's association with the independent variables (unless the ratio of odds is used). Obviously, not all cases are equally good since the quality of the data presumably varies between them, as does the researcher's access and knowledge about the relevant information. However, this distortion is beyond
that problem and makes it even more difficult to assume that a small sample by one researcher and another small sample by a second researcher can be combined to generate a more accurate model of the forces under consideration.

Conclusions

A number of assumptions made in small-N macrocomparative studies are not only very demanding, but to my knowledge they are neither normally made explicit nor seriously examined. Yet they entail assumptions that are usually indefensible in social research. This leads to a certain curiosity. One possibility is that these assumptions occur because they are the only way of proceeding with such data sets, not because the investigators commonly believe they are correct. In that circumstance, the same assumptions will collapse when studies based on large N's are attempted. Another possibility is that such assumptions are appropriate for certain subject matters, such as those involving major institutions or nations. If that is the case, then a very important step is missing, since these assumptions are rarely justified with empirical data based on a larger number of cases. (That is, as a set of tests, sampling an extremely small number of cases of the usual sort from large macrosocietal data sets, it should be possible to show that the same conclusions would occur with Mill's methods as they would by studying the universe of cases.) At the moment, however, it appears that Mill's procedures cannot be applied to small-N studies. There are strong grounds for questioning the assumptions essential to causal analyses generated by such procedures.

As matters now stand, it appears that the methodological needs are generating the theory rather than vice versa. Put bluntly, application of Mill's methods to small-N situations does not allow for probabilistic theories, interaction effects, measurement errors, or even the presence of more than one cause. For example, in the application shown above, the method cannot consider the possibility that more than one factor causes automobile accidents or that there is an interaction effect between two variables. Indeed, if two drivers are drunk but one does not have an accident, the procedure would conclude that the state of inebriation could not have been a cause of the accident that did occur.

The automobile accident example should clearly show that the special deterministic logic does not operate in that instance. Perhaps one may counter that nations and major institutions are neither persons nor roulette wheels; surely, their determination is less haphazard and therefore deterministic thinking is appropriate for these cases. Hence, one might argue, the points made are true for automobile accidents but not for major social institutions or other macrosocietal phenomena. This sounds plausible, but is it true? It turns out that many deep and profound processes are also somewhat haphazard, not so easily relegated to a simple determinism. Elsewhere, I have cited a wide variety of important phenomena that appear to involve chance processes or those that are best viewed that way. These include race riots, disease, subatomic physics, molecules of gas, star systems, geology, and biological evolution (Lieberson 1985:94-99, 225-27). One must take a very cautious stance about whether the
Small N’s and Big Conclusions / 319

deterministic mechanisms implied by the methods used in these small-N studies cause institutional and macrosocietal events. At the very least, advocates of such studies should learn how to estimate whether the probabilistic level is sufficiently high enough so that a quasi-deterministic model will not do excessive damage.

Notes

1. This is different from historical or comparative analyses based on larger numbers, as, for example, in Isaac and Griffin (1989).
2. A brief history of earlier applications of this reasoning is given by Znanieck (1934:236-38).
3. Following Marini and Singer (1988), by “cause” and “causal,” they distinguish “causation from association, recognizing that causes are responsible for producing effects, whereas noncausal associations are not. Although causal terminology has been imprecise and has waxed and waned in popularity . . . ., the ideas of agency and productivity which it conveys have continued to be viewed as distinctive and important in social science” (347).
4. For my purpose at this point, it is not vital to define small, modest, or larger.
5. Needless to say, determination of measurement error should not be made on the basis of whether deviations occur — all the more reason to expect rigorous procedures in both qualitative and quantitative analyses.
6. One cannot argue, by the way, that a new variable combining being drunk and running a red light could serve as a substitute for unmeasured interactions. This is because there would be no way of distinguishing such a combination from other combinations, such as not speeding and running a red light, or for that matter from a grand variable that includes all the constants and the red light variable.
7. Were there to be a larger number of cases in the chart, perhaps 100 where Y is “yes” in 60 cases and “no” in 40 cases, and where the presence or absence of X2 is always in the form shown above, and the other variables vary in a random way, there would be considerable confidence in the very same conclusion that is questionable with a small N.
8. In fairness, of course, the influence is tested if the constant is at a level where it is believed to affect the dependent variable.
9. This ignores the added problem when the small sample is not a random one but rather a selective set of cases.
10. For the most part, I would say that his approach is, however, a deterministic one. Particularly relevant is his treatment of contradictions. The emphasis is primarily on finding additional variables that resolve the contradictions and/or on changing the delineation of the dependent variable. However, he does consider a type of statistical solution as well.
11. The Boolean methods proposed by Ragin (1987) advance our ability to deal with some of these problems, although they require a larger number of cases than are often used in these attempts to apply Mill’s methods.
12. As for the former, Turner (1953) observes that the method of analytic induction is “ill-equipped to cope” (609) with multiple causes.

References


320 / Social Forces 70:2, December 1991


Skocpol, Theda. 1979. States and Social Revolutions: A Comparative Analysis of France, Russia, and China. Cambridge University Press.


