THE CHALLENGE OF SMALL (AND MEDIUM) *N* RESEARCH

Charles C. Ragin
Department of Sociology
and
Department of Political Science
University of Arizona
Tucson, AZ 85718

http://www.fsqca.com
http://www.compasss.org
http://www.u.arizona.edu/~cragin

This lecture is based on the following papers:

http://www.compasss.org/RaginDayOne.PDF
http://www.compasss.org/RaginSonnett2004.PDF

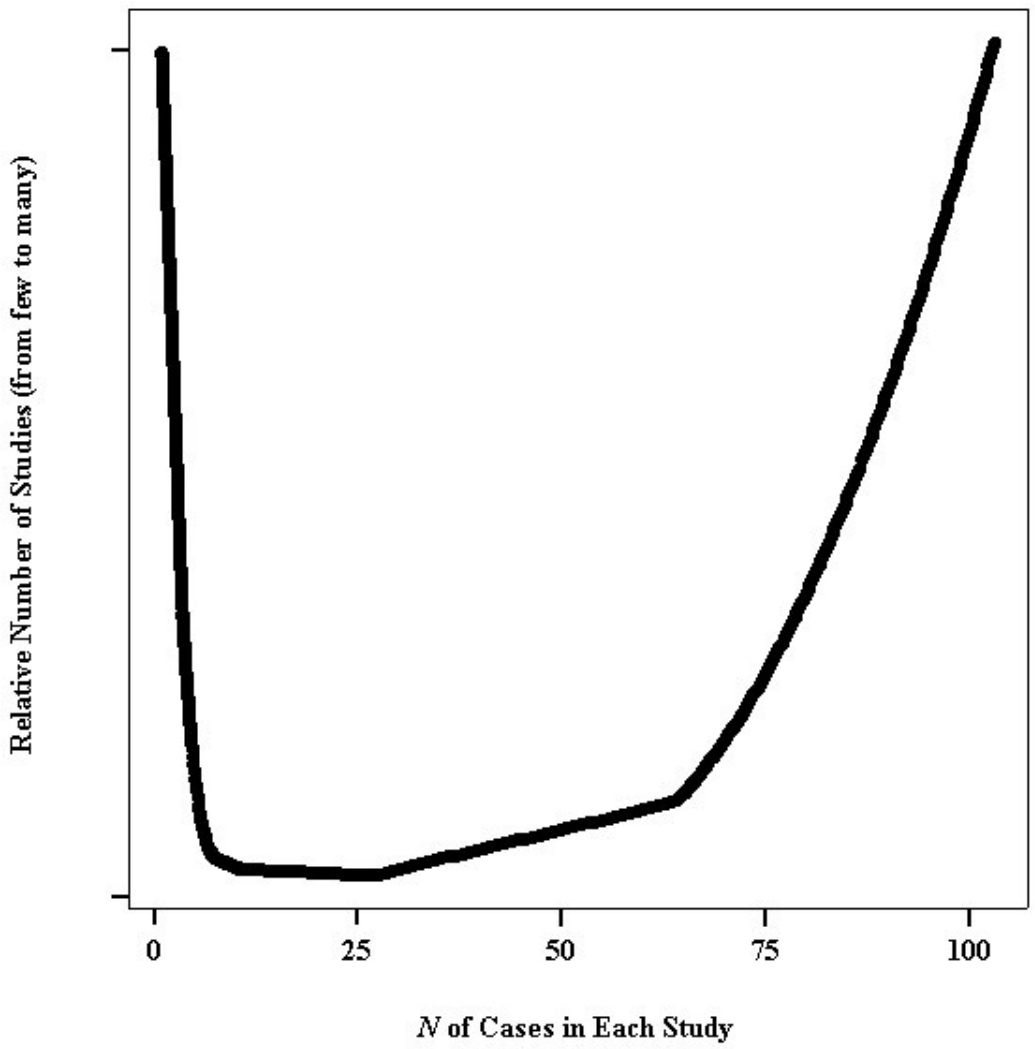


Figure 1.1 Plot of Relative Number of Studies against *N* of Cases in Each Study

Why the U-Shaped Pattern?

1. The nature of the discipline, including training, publishing, invisible colleges, and so on. Researchers tend to use methods they learn in graduate school, where training typically is bifurcated.

2. The underdevelopment of methods for medium-sized Ns.

3. The difficulty of knowing a large number of cases in an in-depth manner.

4. The difficulty of keeping track of (N)(N-1)/2 paired comparisons.

5. The difficulty of considering 2^k logically possible combinations of conditions (relevant to counterfactual analysis), where k is the number of causal conditions. (This expands to 3^k - 1 if the researcher considers subsets of k.)

Some Assertions:

1. Social scientists seek generalizations. They are interested in constructing statements about general patterns.

2. Cross-case analysis is central to the process of constructing generalizations. It is not a necessary ingredient, but is a very common way of arriving at general statements.

3. The results of cross-case analysis, especially in isolation from other knowledge, can be very misleading. The spurious correlation is the best known example of the limitations of cross-case analysis.

4. Causal processes are most visible at the level of the single case.

5. The best way to address the limitations of cross-case analysis is by complementing it with within-case analysis. If possible, it is good to balance cross-case and within-case analysis in social research.

This goal of balancing within and cross-case analysis is a central motivation behind “configurational comparative research” (and QCA).

Configurational Comparative Research: The Middle Path

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
Goals	Case study researchers focus on the problem of making sense of a very small number of cases, usually one and rarely more than three, selected because they are substantively or theoretically important in some way. The key concern is the representation of the case.	Comparative researchers study substantively or theoretically defined categories of cases (usually five to 50 or more), with the goal making sense of both individual cases and sets of similar cases, using cross-case analysis to inform within case analysis, and vice versa.	Variable-oriented research seeks to document general cross-case relationships between variables characterizing a large population of generic observations. The key focus is on the relative conformity of cross-case relationships with theoretically based models.

CCR seeks limited generalizations using small to moderate-sized *N*s.

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
Populations	The case-study researcher's answer to "What is my case a case of?" may change throughout the course of the investigation, as the investigator learns more about the phenomenon in question and refines his or her guiding concepts and analytic schemes. The fact that a single case can be defined in multiple ways is usually seen as a strength, making the case "rich."	In comparative research, the investigator constructs a carefully delimited set of cases, using theoretical and substantive knowledge as guides. The boundary around this set is initially flexible and becomes more fixed as the research proceeds, through the interaction of ideas and evidence. Concept formation and empirical categorization go hand-in-hand.	In variable-oriented research, cases and populations are typically seen as given. The ideal-typic case (or "observation") is the survey respondent. Macrolevel cases such as countries are treated in the same generic manner. The key issue is how to derive a representative sample from the abundant supply of "given" observations.

In CCR populations are constructed, not given.

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
Role of theory	Case-study researchers use in-depth study of cases to advance theory. Thus, they often choose cases that are anomalous in some way from the viewpoint of current theory. A case study is successful even if it succeeds in showing only that existing theory is inadequate. Thus, case selection is critically important.	Existing theory is rarely well-formulated enough to provide explicit hypotheses in comparative research. The primary theoretical objective of comparative research is not theory testing, but concept formation, elaboration, and refinement, and also theory development. Sharpening the definition of the set of relevant cases is often an important theoretical advance in itself.	In variable-oriented research, it is often presumed that researchers have well-defined theories and well-formulated hypotheses at their disposal from the very outset of their research. Theory testing is the centerpiece of social research. The ideal variable-oriented investigation adjudicates between competing theories.

In CCR, vague theory is refined and elaborated, not formally tested.

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
Conception of outcomes	Case-study researchers often select cases specifically because of their uncommon or anomalous outcomes. The usual goal is to resolve the anomaly in a theoretically progressive way, based on in-depth knowledge of the selected case(s). Sometimes there is no sharp separation of causal conditions and outcomes, for an outcome may seem inherent in the constitution of the case.	Comparative researchers often begin by intentionally selecting cases that do not differ greatly from each other with respect to the outcome that is being investigated; they are all "positive cases." The constitution and analysis of the positive cases is usually a prerequisite for the specification of relevant negative cases--if they can be reasonably identified.	Variable-oriented researchers are advised to direct their attention to "dependent variables" that display a healthy range of variation across a systematic sample of cases drawn from a large population. Usually, the more fine-grained this variation, the better. Outcomes that do not vary across cases cannot be studied because there is no variation to explain.

CCR focuses on specific outcomes. Positive cases are easier to define than negative cases.

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
Causation	Case-study researchers examine causation holistically, in terms of a convergence of structures, actors, and events. They are also centrally concerned with sequences and timing of events, with an eye toward turning points and path dependence.	Comparative researchers usually look at causation in terms of multiple pathways. Positive cases often can be classified according to the general path each traveled to reach the outcome. Each path, in turn, can be seen as involving a different combination of relevant causal conditions.	Variable-oriented researchers assess the relative importance of competing independent variables in order to test theory. The key focus is on the relative importance of causal variables across cases, not on how they come together or combine in any single case. A single causal model is derived that applies equally to all cases.

CCR assumes that causation is complex, often involving multiple combinations of conditions sufficient for an outcome. (INUS causation)

The Greatest Challenge of Case-Oriented Research

In an article in *Studies in Comparative International Development*, Christopher Achen, a well-known quantitative researcher, notes:

> Few social scientists dispute the need to combine qualitative and quantitative methods and evidence in the profession. The question is how. As . . . [many] scholars have said, first-rate social science theorizing seems to integrate the two in ways we do not fully understand. For example, contemporary case-study methods are difficult to explicate in conventional statistical theory, and yet they are frequently quite powerful and successful in ways that no statistical methods could match. *An important clue is that they often carry out an implicit comparison against known background relationships*, most obviously so in single-case studies (Ragin 2000:206). But what is the precise inferential logic of this step and why is it so successful? *No one knows*. (*Italics added*)

Olav Stokke's Truth Table for Causes of Successful Shaming in International Regimes

Advice (A)	Commitment (C)	Shadow (S)	Inconvenience (I)	Reverberation (R)	Success (Y)
1	0	1	1	1	1
1	0	0	1	0	0
1	0	0	1	1	0
0	0	0	1	0	0
1	1	1	1	1	1
1	1	1	1	0	0
1	1	1	0	0	1
1	0	0	0	0	1

1. Advice (A): Whether the shamers can substantiate their criticism with reference to explicit recommendations of the regime's scientific advisory body.
2. Commitment (C): Whether the target behavior explicitly violates a conservation measure adopted by the regime's decision-making body.
3. Shadow of the future (S): Perceived need of the target of shaming to strike new deals under the regime--such beneficial deals are likely to be jeopardized if criticism is ignored.
4. Inconvenience (I): The inconvenience (to the target of shaming) of the behavioral change that the shamers are trying to prompt.
5. Reverberation (R): The domestic political costs to the target of shaming for not complying (i.e., for being scandalized as a culprit).

HOW STOKKE'S EVIDENCE IS TYPICAL

- The number of cases (10 cases; 8 configurations) is more than a handful, but still small enough to permit familiarity with each case.

- From the viewpoint of conventional quantitative social science, however, the number of cases is very small relative to the number of causal conditions (5). This ratio essentially eliminates the possibility of any form of multivariate statistical analysis.

- If the cases are viewed configurationally, then the prospects seem even more discouraging, for there are 2^5 logically possible combinations of five causal conditions. We have empirical evidence on only eight of the 32 combinations.

- This pattern of **limited diversity** is characteristic of comparative research and, more generally, of research on naturally occurring social and political phenomena.

- Causal combinations without cases are potential counterfactual cases.

SIMPLE EXAMPLE OF LIMITED DIVERSITY

Strong Unions (U)	Strong Left Parties (L)	Generous Welfare State (G)	N of Cases
Yes	Yes	Yes	6
Yes	No	No	8
No	No	No	5
No	Yes	????	0 (they don't exist)

Is it strong left parties (L) that cause generous welfare states (G) or is it the combination of strong unions and strong left parties (L*U) that causes generous welfare states (G)?

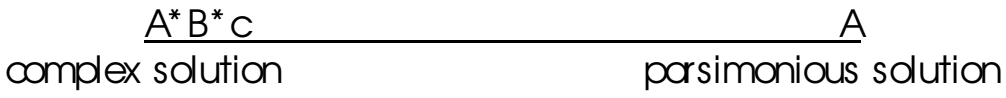
From a correlational viewpoint, having a strong left party (L) is perfectly correlated with having a generous welfare state (G). A parsimonious explanation has been achieved.

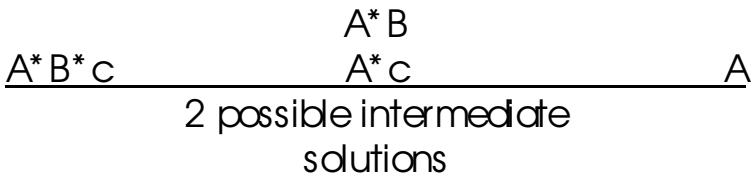
From a case-oriented perspective, however, all instances of generous welfare state share two causally relevant conditions (strong left parties and strong unions) and none of the negative cases display this combination. This pattern suggests a more complex explanation.

Limited Diversity in a Truth Table with Four Causal Conditions

A	B	C	D	Y
no	no	no	no	no
no	no	no	yes	?
no	no	yes	no	?
no	no	yes	yes	?
no	yes	no	no	no
no	yes	no	yes	no
no	yes	yes	no	?
no	yes	yes	yes	no
yes	no	no	no	?
yes	no	no	yes	?
yes	no	yes	no	?
yes	no	yes	yes	?
yes	yes	no	no	yes
yes	yes	no	yes	yes
yes	yes	yes	no	?
yes	yes	yes	yes	?

PARSIMONY VERSUS COMPLEXITY (HYPOTHETICAL DATA)





At the left end of the continuum is the complex solution; the right end shows the parsimonious solution. The complex solution is a subset of the parsimonious solution.

Assume theoretical and substantive knowledge indicates that it is the presence of these four conditions (A, B, C, D) and not their absence (a, b, c, d) that should be linked to the outcome (Y). This knowledge defines A*B*C as an **easy** counterfactual, yielding solution A*B; it defines A*b*c as a **difficult** counterfactual. (This second counterfactual is what is required to produce A*c as a solution.)

Olav Stokke's Truth Table for Causes of Successful Shaming in International Regimes

Advice (A)	Commitment (C)	Shadow (S)	Inconvenience (I)	Reverberation (R)	Success (Y)
1	0	1	1	1	1
1	0	0	1	0	0
1	0	0	1	1	0
0	0	0	1	0	0
1	1	1	1	1	1
1	1	1	1	0	0
1	1	1	0	0	1
1	0	0	0	0	1

1. Advice (A): Whether the shamers can substantiate their criticism with reference to explicit recommendations of the regime' s scientific advisory body.
2. Commitment (C): Whether the target behavior explicitly violates a conservation measure adopted by the regime' s decision-making body.
3. Shadow of the future (S): Perceived need of the target of shaming to strike new deals under the regime--such beneficial deals are likely to be jeopardized if criticism is ignored.
4. Inconvenience (I): The inconvenience (to the target of shaming) of the behavioral change that the shamers are trying to prompt.
5. Reverberation (R): The domestic political costs to the target of shaming for not complying (i.e., for being scandalized as a culprit).

PARSIMONY VERSUS COMPLEXITY IN STOKKE'S EVIDENCE

A·cs·i·r +	
A·CS·i·r +	i +
A·S·I·R	S·R
complex	parsimonious

A·cs·i·r +		
A·CS·i·r +	A·i +	i +
A·S·I·R	A·S·R	S·R
	intermediate	

In the complex solution, none of the combinations without cases is used as a counterfactual case. In the parsimonious solution, any combination without cases that yields a simpler solution is incorporated into the solution (i.e., both easy and difficult counterfactuals). The assumptions are: A, C, S, i, R. These assumptions yield the intermediate solution.

Combination A·S·I·R:

1. Causal conditions S and R cannot be removed because they appear in the corresponding parsimonious term at the other end of the continuum.
2. The support of the regime's the scientific advisory body (A) is certainly linked to the success of shaming. This causal condition should be retained.
3. The fact that it is inconvenient for the targets of shaming to change their behavior (I) does *not* promote successful shaming. Thus, inconvenience (I) can be dropped from the combination A·S·I·R because inconvenience of behavioral change to the target of shaming is not central to the success of A·S·R in generating conformity.

The intermediate combination is **A·S·R**.

Combination A·C·S·i·r:

1. Condition i (the behavioral change is not inconvenient) cannot be dropped because it appears in the corresponding parsimonious term.
2. Condition A (support from the regime's scientific advisory board) should remain because this condition is clearly linked to the success of shaming.
3. Condition C (the offending behavior clearly violates a prior commitment) also should not be dropped, for this too is something that should only contribute to the success of shaming.
4. Condition S (the violator will need to strike future deals with the regime) is also a factor that should only promote successful shaming.
5. Condition r (absence of domestic reverberations for being shamed) can be removed. Clearly, the presence of domestic reverberation (R) would promote successful shaming.

The intermediate combination is **A·C·S·i**.

Combination A·c·s·i·r:

1. Condition i must be retained because it appears in the corresponding parsimonious term.
2. Condition A is retained as well, for the reasons stated previously.
3. Condition r (absence of domestic reverberations) can be removed, as it was from the previous combination, for the same reason provided.
4. Condition c (absence of violation of a commitment) can be removed, for surely these instances of successful shaming would still have been successful if there had been an explicit violation of a commitment (C).
5. Condition s (absence of a need to strike future deals with the regime) can be safely removed because only its presence (S) should contribute to the success of shaming.

The intermediate term is **A·i**.

These three intermediate terms can be joined into a single equation:

$$\mathbf{A{\cdot}S{\cdot}R + A{\cdot}C{\cdot}S{\cdot}i + A{\cdot}i \longrightarrow Y}$$

which can then be simplified to:

$$\mathbf{A{\cdot}S{\cdot}R + A{\cdot}i \longrightarrow Y}$$

because the term A·C·S·i is a subset of the term A·i and is thus logically redundant. (All cases of A·C·S·i are also cases of A·i.) These results indicate that there are two paths to successful shaming: (1) support from the regime's scientific advisory body (A) combined with the need to strike future deals (S) and domestic reverberations for being shamed (R), and (2) support from the regime's scientific advisory body (A) combined with the fact that the behavioral change is not inconvenient (i).

CONCLUSIONS

1. CCR (configurational comparative research) poses a number of important challenges for researchers. These challenges are at the core of social scientific methodology, and include issues regarding the role of theory, the construction of populations, the nature of causation, and so on.

2. In CCR, the issue of limited diversity is especially apparent and salient. Limited diversity is a **characteristic** feature of naturally occurring social phenomena (i.e., non-experimental data). The resolution of the problem of limited diversity involves the use of **counterfactual analysis** in some way.

3. In CCR, the resolution of the problem of limited diversity is **knowledge and theory dependent**. "How" this happens in case-oriented research (Achen's query) is through the incorporation of **"easy" counterfactuals**.

4. In order to define "easy" counterfactuals, researchers apply their substantive and theoretical knowledge to the "remainder" combinations in QCA. In practice, this allows them to craft an intermediate solution, situated between the "most complex" and "most parsimonious" QCA (configurational) solutions. The use of background knowledge in case-oriented research is made explicit through QCA.

5. In quantitative research, the problem is also addressed through assumptions. However, these assumptions (e.g., linearity and additivity) are usually invisible to users. They are rarely examined or challenged, even when they are known to be unreasonable.