1 Integrative Multi-Method Research

The social sciences are in the middle of a boom in multi-method research. An increasing number of books and articles combine techniques from different methodological families within a single study. In conjunction with this trend in application, there has been a marked increase in methodological debate regarding the merits of and best practices for multi-method designs.

This book advances the proposition that well-designed and well-executed multi-method research has inferential advantages over research relying on a single method. I argue that multi-method research can test assumptions that are generally untested in single-method research, thereby transforming key issues of descriptive and causal inference from matters of speculative assertion into points of empirical debate. Yet in order to realize these advantages, multi-method research must be designed from the start with a clear focus on testing assumptions – a priority that informs decisions about case selection, statistical analysis, and the substantive targeting of qualitative inquiry.

While multi-method research has potential advantages for diverse goals, including concept formation and refinement (e.g. Pearce *et al.* 2003), description (Campbell and Fiske 1959; Eid and Diener 2006), and applied policy evaluation (Smith and Lewis 1982; Greene, Caracelli, and Graham 1989), this book focuses on designing multi-method research for causal inference. This emphasis is not intended as a slight against the other families of goals just listed. After all, these goals are deeply interrelated. Good conceptualization and description in particular are essential components of successful causal inference; causal claims, in turn, are routinely central to work in normative theory

and policy evaluation. Instead, causal inference is emphasized because it has been the primary focus of sustained debate regarding multi-method social science, with some scholars arguing that multi-method work has no advantages for this goal vis-à-vis single-method designs (Beck 2006; Ahmed and Sil 2009; Kuehn and Rohlfing 2009; Beck 2010). Hence, showing how multi-method research can improve causal inference is more urgent than demonstrating its less-contested role in other domains.[1]

Multi-method designs are not a panacea for the challenges of causal inference. Even the best designs will leave some issues unaddressed, and some common designs arguably have little advantage in comparison with single-method designs. Nonetheless, when carefully constructed and executed, multi-method research can make a major contribution to the social sciences. This book is intended to help scholars design, execute, and evaluate such research.

1.1 The Multi-Method Boom

Multi-method research involves combining data-gathering and -analyzing techniques from two or more methodological traditions. Examples of multi-method research thus include studies that combine survey data with laboratory experiments, focus groups with participant observation, statistical analysis of a corpus of text with careful qualitative interpretation of a selected handful of texts, and so forth. Out of the wide range of possible multi-method combinations, this book focuses on designs that combine quantitative and qualitative methods[2] in support of a single causal inference. This focus is adopted in part because it has been the central theme of most debates regarding multi-method designs, and in part because the very significant differences between the families of methods being combined render particularly vivid both the challenges and the advantages of multi-method approaches.

Issues of how best to combine methods to form the basis for a causal inference have become a central concern in the social sciences

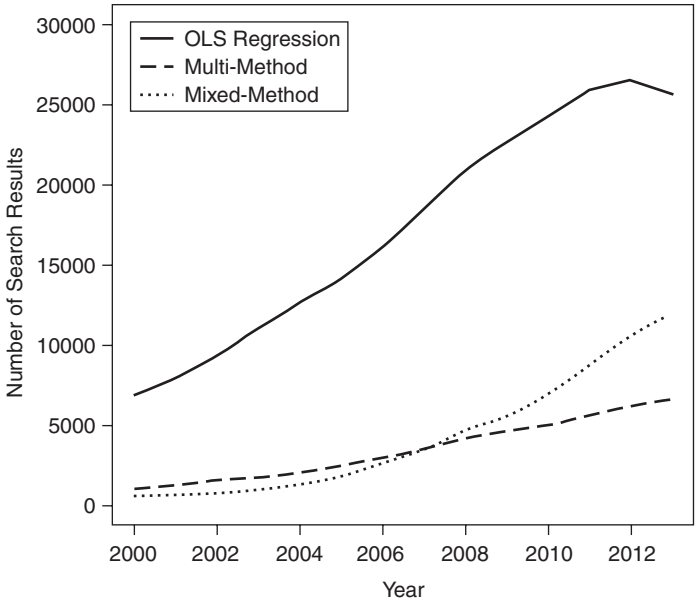

Figure 1.1 Google Scholar Search Results Related to Multi-Method Research.

in light of recent growth in the application of such methods. As a rough measure of the development of multi-method techniques, consider the number of Google Scholar search hits for the terms "multi-method" and "mixed-method." Figure 1.1 shows the number of scholarly texts that contain these two search texts for each year between 2000 and 2013. By way of comparison, the figure also includes the trend for the phrase "OLS regression" (without quotation marks) during the same period.[3] All three search categories show increasing numbers of hits over time – a result that no doubt reflects a combination of increasing digitalization, an increase in the number of venues for scholarly publication, and trends in methodological usage. Relative comparisons are, perhaps, most informative. During this period, the number of references to multi-method research grew from 14% of the references to OLS regression, to 26%. References to mixed-method research grew far more impressively, from 8% to 47% of the number of search results for OLS regression. These suggestive data, combined with qualitative indicators such as the publication of textbooks (e.g.

Greene 2007; Teddlie and Tashakkori 2009; Creswell and Plano Clark 2011) and the launching of a specialized journal, *The Journal of Mixed Methods Research*, specifically devoted to the methodology of multi-method research, indicate quickly and substantially growing scholarly interest in, attention to, and application of research designs combining qualitative and quantitative strategies.

1.2 Integration, not Triangulation

While multi-method research is evidently experiencing a surge of popularity, there are reasons to worry about whether multi-method applications are in fact producing more grounded, justified, and persuasive inferences than studies using a single method. These concerns arise from the "triangulation" framework (Webb *et al.* 1966; Jick 1979; Tarrow 1995) that has for some decades served as the prototypical research design for multi-method social science. Simply put, triangulation designs involve asking the same question of causal inference using two different methods, and checking that the same substantive conclusions are produced by both. The metaphor is to the geometric technique of estimating distance by measuring the angle of sight toward an object from two different vantage points.

Triangulation in the social sciences has major flaws. One is well-known and widely discussed: what conclusion should be drawn when the two methods produce different findings? Unfortunately, the list of intellectually plausible responses to such an outcome is all but unbounded. For example, a scholar could reasonably conclude that both findings are correct but capture different aspects of the phenomenon of interest; that one method or the other displayed fundamental limitations in the analysis; that the divergence provides evidence against the credibility of the assumptions involved in both methods; that the outcome involves a contrast in terms of scope or relevant populations for the two methods; or that the failure to triangulate simply leaves the inference in a state of uncertainty. In decades of writing about triangulation, no definitive

guidance has yet emerged about how to respond to divergent findings between the multiple methods in a study.

In my view, this frequently discussed limitation of triangulation designs is a product of a deeper incoherence in their conception. What, in fact, does it mean to say that a qualitative and a quantitative and/or experimental study ask and answer the same research question? In a very general sense, the answer is, perhaps, clear: if both methods produce causal inferences in which a particular concept is important, then there is some degree of overlap in the results. Yet because qualitative and statistical approaches produce results that are different in kind, it is only possible to assess such convergence very abstractly. If a variable has an effect estimate that is positive and statistically significant in some analysis, and a related variable also seems to play an enabling causal role in an analysis of sequences of events within a single case, then that variable has positive support from both methods. This requires essentially ignoring the magnitude of the quantitative effect estimate, as well as qualitative inferences about sequence and context that go along with the discovery of a role for this one variable – effectively, everything but the sign of the inference must be disregarded. At any more detailed level of analysis, it becomes difficult or impossible to decide whether qualitative and statistical results correspond.

For example, if a scholar does a cross-national statistical analysis relating various hypothesized explanatory variables to the outcome of civil war, and then conducts case-study analysis of the dynamics of civil war or its absence in some of the cases used in the statistical part of the study (see, for example, Fearon and Laitin 2003, 2008), then it is clearly true that we have a quantitative and a qualitative study of the outcome of civil war. If one finds "positive" results on related variables from the statistical analysis and the case studies, then there is an overlap, and the triangulation design has made a connection.

Yet if the question is taken more seriously, this connection becomes far more tenuous. The easiest way to see the problem is to ask what it means to decide whether the quantitative and qualitative component of a multi-method design agree. For example, in Fearon and Laitin's logit

analysis of civil wars, the logged percentage of a state's territory that is mountainous has a coefficient of 0.219 – an estimate that is significantly different from zero (Fearon and Laitin 2003). What kind of finding from a case-study analysis of Colombia would confirm this estimate? What would contradict it? Colombia is, of course, a mountainous country with a history of endemic civil war. Geographically, the relevant point is that three major Andean mountain ranges run through the center of Colombia, running the length of the country and housing most of its heavily populated areas, as well as its capital city. Furthermore, in terms of civil war and related forms of political violence, Colombia has a long and regrettable record, including a substantial number of wars in the nineteenth century, as well as La Violencia between 1948 and 1958, and the current civil war, which has been ongoing at one level of intensity or another since the 1960s (LeGrand 2003). So far, so good (for the quantitative finding, if not for Colombia), one might suppose.

With a closer examination, though, the apparent agreement between case study and statistical result becomes mired in dilemmas. At some historical moments and for certain components of the conflict, the mountainous regions of Colombia have in fact been host to key events in the civil war. For example, an early event in Colombia's ongoing conflict involved a military attack on a de facto autonomous republic of leftists in the high Andean community of Marquetalia (Arenas 1972). In a very different phase of the conflict, the M-19 revolutionary movement was based in mountainous regions of Colombia and carried out several of its most famous actions in the city of Bogotá – itself in the mountainous region (Duran *et al.* 2008). A great many other examples could also be provided, of course.

At the same time, the mountainous areas near the national capital at Bogotá have often been among the safest areas and the zones with strongest state presence in the country. Furthermore, mountain-free regions in the southeast and along the Caribbean coast have also served as important areas of refuge for anti- and non-state armed actors, thereby facilitating and prolonging the civil war. For instance, through the 1990s the FARC guerrillas had their primary bases in the jungle

regions of southeastern Colombia. Thus, case-study consideration of Colombia would have to suggest that mountainous terrain plays a supporting causal role at some points and for certain actors in the Colombian conflict, but is not a relevant consideration at other points. Further research would likely illuminate how and why terrain matters for some aspects of Colombian political violence but not others.

How does this line of inquiry compare with the estimated coefficient of 0.219? It seems that the best one can say is that the two results do not obviously contradict each other; but nor are they clearly mutually supportive. Does Colombian history show too much of a role for terrain in light of a statistical coefficient that is significant but substantively moderate? After all, the current civil war began in part because anti-state actors had created refuges for themselves in the mountains; a conceivable counterfactual is that less rugged terrain would have prevented these key actors from organizing in the first place. On the other hand, perhaps the case suggests that the coefficient is too large; armed actors have at times found the jungles and other regions as hospitable a refuge as the mountains, so various forms of difficult terrain may be substitutes in a way that the statistical results fail to demonstrate. Or perhaps these competing considerations are just what the value of 0.219 implies?

I think it is in fact impossible to decide whether the case study and the logit coefficient agree. Furthermore, this is not a problem specific to Colombia; any case study would face similar issues of interpretation. The fundamental problem is that the qualitative and quantitative methods are not in fact asking the same question, even though they focus on the same topic. The statistical analysis estimates the difference in conditional probability of civil war associated with a given contrast in geographies across countries, while the case study asks whether there is within-case evidence consistent with the proposition that terrain features helped enable or prevent various components of a particular civil war. These *questions* are fundamentally different, and so it is essentially useless to ask whether the *answers* are the same.[4]

The practical result of a triangulation perspective on multi-method research is the proliferation of studies in which scholars effectively carry out two separate analyses sharing a broad topic and theoretical orientation but with no serious intellectual interaction at any level of detail. Factors that "matter" to some extent in both studies are emphasized, while the actual meaning of inferences drawn in each study, as well as any contrasts in results, is neglected. Such loose forms of multi-method design often do not substantially persuade their target audience. More seriously, there is no obvious reason why they should. I cannot find either a serious argument – beyond the flawed triangulation metaphor – or an empirical demonstration that causal inference is more likely to succeed when it is done twice with different tools and non-comparable results within a single book or article.

Multi-method research can be much more powerful than the triangulation design. In this book, I advocate a contrasting family of multi-method designs, which I will refer to as integrative multi-method research. Integrative designs are multi-method designs in which two or more methods are carefully combined to support a single, unified causal inference. With such a design, one method will produce the final inference, and the other is used to design, test, refine, or bolster the analysis producing that inference.[5]

For example, the scholar may use a regression-type model to produce the final causal inference, drawing on case-study research to test and adjust key assumptions about measurement, omitted variables, causal interactions, and pathways. Here, the final product is a causal effect estimate drawn from a statistical coefficient, as in a purely quantitative study. Yet that coefficient comes from a model designed, refined, and tested in light of serious qualitative analysis. Both methods contribute substantially to the overall inference, even though the final product comes from one and not the other. Integrative designs may also ultimately rely on case-study methods to produce a final causal inference. In such designs, quantitative analysis is often used to test especially important, sensitive, or elusive steps in the case study's causal chain connecting the initial cause to the outcome of interest; if these

analyses support the existence of a causal effect at key points, then the overall claim regarding the set of connections from cause to effect is strengthened, and the case study inference becomes more robust.

The central idea of integrative multi-method research is to use each method for what it is especially good at, and to minimize inferential weaknesses by using other methods to test, revise, or justify assumptions. This is in contrast to triangulation designs, in which whole inferences – fully dependent on both the strengths and the weaknesses of each method – are carried out and then somehow compared and combined. Integrative designs thus use multiple modes of inference to substitute strengths for weaknesses, whereas triangulation designs compare whole causal inferences whose individual strengths and weaknesses are the same as those of single-method designs. For this reason, a well-constructed integrative multi-method design will yield a more robust and higher-quality causal inference than either a triangulation design or a single-method design using only one approach. Furthermore, with integrative designs the issue of how to reconcile very different kinds of inferences from quantitative and qualitative analysis does not arise, because these methods are deliberately used from the very beginning to ask different kinds of questions about separate issues related to the causal inference. While the final causal findings from an integrative study will depend on the answers to all of these questions, they are not a direct combination of those answers; one method influences the final inference by shaping the way the other is used, while the second method generates the final causal inference.

Much existing work on multi-method designs emphasizes research sequences, differentiating designs by whether qualitative and quantitative components are simultaneous or sequential and – if the design is sequential – also by which method is used at each stage (Tashakkori and Teddlie 1998: Chapter 3; Teddlie and Tashakkori 2009: Chapter 7; Cresswell and Plano Clark 2011: Chapter 3). The integrative approach requires a sequential design – if qualitative and quantitative research are carried out in parallel, then it is extremely difficult for

the findings of one component to inform design decisions involved in the other. In an integrative research sequence, one method provides an initial summary of current knowledge about a problem of causal inference. An alternative method then tests the assumptions behind that initial summary, ideally discovering new material that can be incorporated into an improved version of the first analysis. In turn, the assumptions behind that improved analysis can be tested and improved with a new round of multi-method research. Thus, in contrast to the linear sequences or parallel designs typical of triangulation-based research, integrative multi-method designs involve an indefinite cycle of discovery and refinement.

1.3 The Origins of Proportional Representation: An Integrative Example

Integrative multi-method research designs can be implemented within a single study, perhaps by a single author, but they can also play out across debates among scholars; integrative designs are a wonderful tool for evaluating and critiquing others' research, as well as for strengthening one's own causal inferences. For example, beginning with Boix's (1999) essay on the origins of proportional representation (as opposed to single-member district) electoral systems, the *American Political Science Review* hosted a decade-long debate about the politics of electoral institutional choice. Boix's initiating argument draws on classic theories in comparative politics and a simple game-theoretic model to argue that proportional representation results when the electoral left is powerful and the right is politically fragmented; a regression analysis serves as the primary empirical test of the hypothesis.

Cusack *et al.* (2007) argue, against Boix, that proportional representation is instead caused by an elective affinity between that system and politico-economic arrangements in which workers' skill formation is coordinated between employers and representatives of labor. This argument is tested using a triangulation multi-method design; somewhat methodologically loose qualitative argumentation