

# **Fuzzy-Set Policy Research**

Charles C. Ragin  
Department of Sociology  
University of Arizona  
Tucson, AZ 85721 USA

and

Peer C. Fiss  
Department of Management & Organization  
University of Southern California  
Los Angeles, CA 90089 USA

## Overview

In the domain of policy research, the question of choosing the “right” methodology takes on a particular importance, as methodologies provide the foundations both for constituting evidence and for crafting representations of social phenomena (i.e., 'results') from this evidence. Whenever research is conducted with the goal of contributing to policy debate, methodology is thus not merely an academic concern, but also a political issue; different methodologies may produce qualitatively different findings, even when the definition of what constitutes evidence is the same.

While the choice of methodology thus plays a central role in determining the range of possible results, most policy-oriented social research today is dominated by conventional multivariate regression analysis. The prime concern of this specific methodology is assessing the relative importance of competing independent variables. For example: Which variable has the strongest impact on life chances: education, test scores, or family background? Furthermore, framing multivariate analyses as competition between variables dovetails with everyday forms of ideological opposition. That is, there is frequently a direct connection between conventional quantitative methodology and ideological debate because competing variables in multivariate analyses are usually linked to different ideological positions.

In this book, we offer an alternative to the conventional multivariate analysis of policy-relevant social data. Instead of asking, "What is the net effect of each independent variable (e.g., test scores) on the outcome?" we ask, "What combinations of causally relevant conditions are linked to the outcome?" In this different view, causal conditions do not compete with each other; rather, they combine in different ways to produce the outcome. This alternate approach allows for the possibility that there may be many paths to the same outcome, and it does not force the incremental effect of each causal variable on the outcome (e.g., on the odds of avoiding poverty) to be the same for each case. In essence, it is a diversity-oriented understanding of the connections between causal conditions and outcomes, because it views cases configurationally, i.e. in terms of the different ways they combine causally relevant conditions.

The diversity-oriented techniques we use here are set-theoretic in nature and build upon the case-oriented techniques that the first author originally presented in The Comparative Method (University of California Press, 1987) and then extended in Fuzzy-Set Social Science (University of Chicago Press, 2000). In the latter work, Ragin demonstrated how to identify the multiple paths to an outcome using set-theoretic methods. By viewing cases configurationally and

causes conjuncturally, it is possible to allow for much greater diversity and heterogeneity, and researchers can address nuanced questions about causal conditions. For example, instead of asking, "What is the net impact of test scores on poverty status?" they can ask, "Under what conditions is there a connection between low test scores and poverty?" This specific question can be answered by examining the different paths to poverty and pinpointing those that include low test scores as part of the mix of causally relevant conditions. Nuanced findings, in turn, are more directly relevant to policy makers and policy discourse; interventions can be tailored to specific cases, and identifying different paths to an outcome allows for greater flexibility in choosing the intervention with the greatest impact.

As an example of how diversity-oriented, set-theoretic methods may offer an alternative to policy analysts, we focus on the current controversy surrounding the role of test scores as predictors of life chances. The debate over test scores is a specifically appropriate case because of this debate's centrality to current issues in education and educational reform, and, in turn, the centrality of education to U.S. social and economic policy.

## **Chapter 1: Bringing Diversity-Oriented Methods to Policy Research**

Chapter 1 provides an overview of the book, especially the policy and methodological issues it addresses. We explain our interest in extending a methodology originally developed for macrosocial, comparative research to policy-relevant, individual-level data. We clarify our primary objective, which is to free large-N quantitative research from the restrictive, homogenizing assumptions that are integral to the most widely used analytic techniques. These assumptions are often at odds with the goals of policy analysts. For example, analysts are often more interested in "under what conditions" a variable influences an outcome than in its "average, net effect" on an outcome across all cases. Conventional quantitative techniques can answer the latter question with precision. They are less adept when it comes to the former. Thus, they can answer the question: Which explains more of the variation in poverty, test scores or parental income? But they have difficulty addressing the question: Under what conditions does having a low test score contribute to poverty?

We continue by arguing that the deficiencies of conventional multivariate analysis interact with ideological ideas about which variables are relevant to an outcome. Discussions about whether statistical models are under-specified or not usually hinge on differences in ideologies about the nature of social relations, and

thus the appropriateness of certain variables. Frequently, the results of such debates are differing camps of researchers, each pointing to their specific models as evidence for the correctness of their view. We argue that this undermines the potential value of social research to policy discourse, especially in such politically charged arenas as education. As researchers focus almost exclusively on the *competition* between variables, they fail to consider how different factors may work *together*. A related consequence is that they overlook the possibility that there may be many different paths to the same outcome, involving different combinations of causally relevant conditions, and these paths may differ by both race and sex. For example, the finding that test scores have a significant net effect on life chances does not help us understand *how* they have this effect, in what contexts, or in combination with what additional factors. We conclude by suggesting that a more textured understanding of the connection between test scores and life chances is possible only by abandoning the competition between variables and focusing on the diverse ways causal conditions and outcomes may be linked.

## **Chapter 2: Policy Context: Test Scores and Life Chances**

In chapter 2, we examine a specific case of ideological contest over the interpretation of data: the debate over The Bell Curve (1994) and the role of test scores in determining life chances. We closely scrutinize the merits of this debate and how it is limited by the methods used to examine the reasons for poverty. We then replicate and contrast the analyses of both The Bell Curve and its most prominent refutation, Inequality by Design (1996), showing how both approaches are limited by the causal reasoning they apply.

As an empirical context for the application of diversity-oriented methods, the ongoing debate over the effect of test scores on poverty has an intuitive appeal. Clearly, understanding the reasons why some people end up in poverty why others do not has direct implications for those interested in public policy. Furthermore, within this debate on poverty, few works have created as much controversy as the publication of The Bell Curve. The critiques of this book were varied and many, but they usually tended to focus upon the idea that the model presented by Herrnstein and Murray was in some way or another underspecified, i.e. lacking important explanatory variables. We pay special attention to the response to H&M formulated by Fischer et al. in Inequality By Design, which presents probably the most important and widely acclaimed critique of The Bell Curve.

We begin by providing an overview of the Bell Curve debate and the different lines of attack in examining the findings of H&M. After doing so, we offer a complete re-analysis and comparison of both book's findings, using the original data from the National Longitudinal Survey of Youths as well as data provided by Fisher et al. The key contrast of our comparison is between the "parsimony reigns supreme" strategy of H&M and the "everything but the kitchen sink" approach of Fischer et al. We show how neither approach is satisfactory because they both suffer from the restrictive assumptions built into a common analytic technique: logistic regression. Essentially, in both analyses variables compete to explain variation in the outcome variable, and the goal of the analysis is to estimate the unique contribution of each independent variable across all cases. In the Bell Curve model, the independent variables are few and the results simple; the test score variable wins the competition to explain variation in the odds of living in poverty. By contrast, the variables in the Inequality by Design model are many and the findings complex. As a result, the test score variable withers in the face of an onslaught of competing variables. However, even the complex models of Fisher et al. offer cannot directly address the diverse links between a variety of individual-level characteristics, such as household formation decisions, and the odds of living in poverty.

By replicating and contrasting the analyses of both H&M and Fisher et al., we thus show how both approaches are limited by the causal reasoning they apply. While the more complex model proposed by Fisher et al. performs somewhat better than the model of H&M, neither of these logistic regression models is particularly successful at either correctly classifying those actually in poverty or at explaining variation in poverty outcomes. We find that the overall predictive power of both models is severely limited and differs depending on race and gender. This sets the stage for the next chapter, where we lay out how diversity-oriented research provides an alternative to the classic quantitative approach by abandoning the usual competition among explanatory variables for a method that allows to examine how combinations of attributes lead to different outcomes.

### **Chapter 3: Analytical Foundations: Set-Theoretic Versus Associational Reasoning**

In chapter 3, we accomplish two things. First, we show the key difference between set-theoretic and correlational thinking, and how set-theoretic methods are better suited to the arguments of social theory and public policy. Second, we

show how fuzzy set methods are an effective way of operationalizing set-theoretic statements, and how they overcome two frequent objections to set-theoretic research.

The key to understanding the logic of a diversity-oriented approach is to grasp the contrast between associational and set-theoretic relationships. In associational relationships, the focus is on correlations between variables. With set-theoretic relationships, on the other hand, the main concern is whether or not one set is a subset of another. For example, when a social scientist states that “democracies don't go to war against each other,” he or she is essentially claiming that the set of wars does not include instances in which the primary antagonists are both governed by democratic regimes. Set-theoretic methods focus on these sets specified in verbal statements, and in fact, almost all social scientific theory is implicitly or explicitly formulated in terms of sets.

Correlational methods, which constitute the core of conventional quantitative social science today, typically look at relationships between variables across all cases at once—across the entire analytic space defined by the relevant variables.

Consider another example: Is the set of Black females lacking high school education a rough subset of Black females living in poverty? If so, then we may claim, for Black females at least, that lacking a high school education may be a sufficient (though not necessary) condition for living in poverty. Of course, there may be many high-school educated Black females living in poverty, and in fact the vast majority of Black females in poverty may be high-school educated. Nevertheless, these facts do not challenge the set-theoretic relation between lacking a high school education and living in poverty. From an associational viewpoint, however, the fact that most Black females living in poverty are high-school educated would greatly weaken the association between the two variables (being high-school educated and living in poverty) and thus undermine the presumed value of the observation that most Black females lacking high school education are living in poverty.

The contrast between set-theoretic and correlational thinking is striking, and, as we show in this chapter, it is important for formulating policy responses.

Until now, almost all large-N policy research has been correlational in nature. However, set-theoretic statements and arguments permeate and constitute both social science theory and policy analysis. We show that the set-theoretic approach provides a much better way to connect data analysis to theory and thus to policy discourse. For example, we can ask: Which is more useful to policy discourse, the knowledge that there is a very modest positive relationship between test scores and poverty status, or the knowledge that high test scores seem to offer some sort of immunity from poverty?

After establishing the key differences between set-theoretic and associational approaches, we describe the set-theoretic methods used to re-analyze the Bell Curve data. Central to these methods is the use of fuzzy sets to permit graded memberships in sets, with membership scores ranging from 0 (full nonmembership) to 1 (full membership). We show that our approach allows us to address the two objections most often raised against set-theoretic analysis. The first objection is that set-theoretic methods are “all or nothing”—that one exception disproves the rule. In response, we show how to use benchmarks and probabilistic criteria to evaluate qualified set-theoretic relationships. The second objection is that set-theoretic methods apply only to categorical variables, while the phenomena that interest social scientists mostly vary by level or degree. In response, we show how fuzzy sets, which allow for varying degrees of membership in sets, solve this problem and in fact often provide better and substantially meaningful measures of social phenomena. Taken together, we demonstrate in chapter 3 that key set-theoretic operations (e.g. assessing subset relations) are possible with fuzzy sets, and that fuzzy sets offer an excellent way for addressing qualified set-theoretic arguments.

#### **Chapter 4: Fuzzy-Set Methods for Studying Causal Conditions**

In this chapter, we apply our set-theoretic approach to studying causation. As we have argued in the previous chapter, most current research uses a correlational approach to causation. We now continue to present an entirely different approach to the analysis of causal conditions. Instead of asking, "What is the net effect of each independent variable on the outcome?" we ask, "What combinations of causally relevant conditions are linked to the outcome?" In this view, causal conditions do not compete with each other; rather, they combine in different ways to produce the outcome. The key question thus concerns not separate, independent effects, but how causal conditions come together and with what consequences. This alternate approach allows for the possibility that there may be many and diverse paths to the same outcome, and it does not force the incremental effect of each causal variable to be the same for each case. In essence, it is a diversity-oriented understanding of the connection between causal conditions and outcomes because it views cases configurationally, in terms of the different ways they may combine causally relevant conditions.

In the present chapter, we also demonstrate procedures for analyzing and logically simplifying the different combinations of conditions that display high concentrations of the outcome. By viewing cases configurationally and causes

conjuncturally, it is possible to identify the multiple pathways to an outcome and thereby allow for greater diversity and heterogeneity among cases than is possible using conventional methods. We begin by showing how to measure the membership of cases in combinations of characteristics, using fuzzy-set methods. We then show how to test "all possible paths" implied in a specific set of causal conditions and how to use logical rules to simplify the paths that pass our test of set-theoretic consistency. Finally, we demonstrate how to assess the relative importance of paths. Our demonstration of these procedures uses NLSY data on White males, focusing on the set-theoretic links between three causal conditions (high test scores, high parental income, and college education) and one outcome (the set of cases not in poverty). In sum, chapter 4 shows that combinations of causally relevant conditions can be conceived and studied as paths, and suggests that this approach has much more to offer to policy discourse than analysis that focuses on the net effects of independent variables, conceived as adversaries attached to competing theoretical perspectives.

## **Chapter 5: Fuzzy-Set Analysis I: Causal Conjunctures**

In this chapter, we present our central findings concerning the combinations of casual conditions linked to poverty for four groups: Black females, Black males, White females, and White males. We begin this chapter by reconstructing the main causal variables of Inequality By Design as fuzzy sets, focusing especially on education, test scores, parental income, marital status, and children, as well as two outcome variables: the fuzzy set of individuals "in poverty" and its negation, the fuzzy set of people "not in poverty." Our reconstructed variables offer a major advancement over the variables that have so far been used. Specifically, we improve H&M's and Fisher et al.'s studies by directly tying variable values to substantively meaningful thresholds. For example, in analyzing the effect of parental income on the odds of being in poverty, Fisher et al. treat poverty as a dichotomous outcome; a respondent is either in or out of poverty, as defined by the official poverty threshold corrected for household size. However, such a measure of poverty is inherently problematic. Using a binary outcome measure places families whose income is just barely above the poverty level in the same category as families whose income is far above the poverty threshold, such as upper-middle class or even upper-class families. Our fuzzy set of membership in the set of households in poverty avoids this problem of conflating such outcomes. Rather than using a binary measure, we construct a fuzzy set of membership in the set of poor

households using a ratio of the official poverty level. In the resulting fuzzy set, we define those households with total family income at or below the poverty level as having full membership in the set of households in poverty. Conversely, we define families with an income of four times or more the poverty level as fully out of poverty, since four times the poverty level on average lies significantly above the median family income in the U.S. Families with an income between one and four times the poverty level are assigned varying degrees of partial membership in the set of those in poverty, thus allowing for a very fine-grained measure of poverty that at the same time relies on substantive knowledge about income levels for determining membership thresholds.

After constructing our variables, we conduct analyses to determine both necessary and sufficient conditions for being out of poverty. These analyses identify the different combinations of causal conditions linked to high concentrations of individuals in poverty or to high concentrations of individuals not in poverty. We then algebraically simplify these combinations to produce logical statements describing the multiple conjunctures of conditions linked to these two outcomes.

Our results contextualize and localize the connections between test scores and poverty for the four groups, showing important differences across categories of race and sex. For example, while having a high school education is a necessary condition for being out of poverty across all sub-groups, the effects of not having a low test score is only significant for Whites, while environmental factors such as high unemployment rate only show an effect for Blacks. Regarding sufficient conditions for being out of poverty, coming from a background with higher parental income itself is usually sufficient by itself for White respondents. For both black males and females, on the other hand, coming from a high parental income background itself is not sufficient, but only becomes so when combined with another characteristic such as having high test scores or being college educated. These findings are a marked improvement over previous research, as they offer a detailed insight into how causal conditions combine and interact to affect the life chances of differing social sub-groups.

## **Chapter 6: Fuzzy-Set Analysis II: Limited Diversity**

As we have stated before, conventional quantitative methods focus on the net effects of "independent" variables and overlook the fact that variables can also be seen as aspects of interpretable configurations. Viewed in a "configurational" light, however, the analysis of quantitative data is not so much

about variables and their separate, independent effects, as it is about different kinds of cases (defined by configurations of values on variables) and the outcomes these different kinds of cases experience. From a configurational point of view, each finding should be examined in terms of the kinds of cases it embraces. For example, suppose that both Black males with high test scores and White males with high test scores constitute clusters or concentrations of not-in-poverty individuals. It is important to examine both of these groups with respect to their other relevant characteristics. If the White males with high test scores are diverse in terms of their family background and educational variables, while the Blacks males with high test scores are college educated and have high-income parents, then it is hazardous to give the same interpretation to what appear to be parallel findings regarding the importance of high test scores. Viewing cases as configurations, it is apparent, in this hypothetical example, that the Black males with high test scores are configurationally uniform and that their concentration in the not-in-poverty set is somewhat "overdetermined." That is, their concentration in this set may have nothing to do with their high test scores, per se. The White males with high test scores, by contrast, are configurationally diverse and their concentration in the not-in-poverty set may in fact be connected to their high test scores and not to their other characteristics. The key point is that it important to examine configurationally defined differences, especially when comparing across groups and attempting to construct meaningful generalizations.

## **Chapter 7: Implications for Policy and Policy Research**

By contextualizing and localizing the effects of such variables as test scores and education on life chances, it is possible to go beyond global (and typically vague) statements about their effects. One of the simplest ways to address a debate regarding the interpretation of an effect is to locate it--identify the conditions for its impact on the outcome. Set-theoretic methods are especially well-suited for this task. By viewing cases as configurations and examining conjunctures of causal conditions, it is possible to bring greater texture and depth to the analysis of large-N data sets. With set-theoretic methods, this texture usually entails the specification of multiple paths to an outcome (e.g., in poverty versus not in poverty), where paths are defined in terms of combinations of causally relevant conditions. The implications of "paths" for policy are straightforward. If there are multiple paths to an outcome, more than likely some will be more open to policy intervention than others. For example, if a

common path to poverty for White females involves single motherhood, then policy efforts can be directed toward delaying pregnancy. Any methodology that allows the identification of paths provides policy makers with a range of choices regarding intervention points. By contrast, the usual results of conventional quantitative analysis indicate only the relative importance of different causal conditions across all cases--delineating, in effect, a single path or "formula" for all. Such findings often make it seem that the possibilities for intervention are limited, especially when causal variables that are less manipulable by policy makers appear to have strong effects on the outcome, outweighing the effects of the manipulable variables.