Cambridge University Press
978-1-107-06507-9 - Counterfactuals and Causal Inference: Methods and Principles for Social Research: Second Edition
Stephen L. Morgan and Christopher Winship
Excerpt
More information

Part I: Causality and Empirical Research in the Social Sciences

Cambridge University Press
978-1-107-06507-9 - Counterfactuals and Causal Inference: Methods and Principles for Social Research: Second Edition
Stephen L. Morgan and Christopher Winship
Excerpt
More information

Chapter 1

Introduction

Do charter schools increase the test scores of elementary school students? If so, how large are the gains in comparison to those that could be realized by implementing alternative educational reforms? Does obtaining a college degree increase an individual's labor market earnings? If so, is this particular effect large relative to the earnings gains that could be achieved only through on-the-job training? Did the use of a butterfly ballot in some Florida counties in the 2000 presidential election cost Al Gore votes? If so, was the number of miscast votes sufficiently large to have altered the election outcome?

At their core, these types of questions are simple cause-and-effect questions of the form, Does X cause Y? If X causes Y, how large is the effect of X on Y? Is the size of this effect large relative to the effects of other causes of Y?

Simple cause-and-effect questions are the motivation for much research in the social, demographic, and health sciences, even though definitive answers to cause-and-effect questions may not always be possible to formulate given the constraints that researchers face in collecting data and evaluating alternative explanations. Even so, there is reason for optimism about our current and future abilities to effectively address cause-and-effect questions. Over the past four decades, a counterfactual model of causality has been developed and refined, and as a result a unified framework for the prosecution of causal questions is now available. With this book, we aim to convince more social scientists to apply this model to the core empirical questions of the social sciences and to applied research questions of public importance.

In this introductory chapter, we provide a skeletal précis of the main features of the potential outcome model, which is a core piece of the more general counterfactual approach to observational data analysis that we present in this book. We then offer a brief and selective history of causal analysis in quantitatively oriented observational social science. We develop some background on the examples that we will draw on throughout the book, and we conclude with an introduction to directed graphs for systems of causal relationships.

1.1 The Potential Outcome Model of Causal Inference

With its origins in early work on experimental design by Neyman (1990[1923], 1935), Fisher (1935), Cochran and Cox (1950), Kempthorne (1952), and Cox (1958), the potential outcome model of causal inference was formalized in a series of papers by the statistician Donald Rubin (1974, 1977, 1978, 1980a, 1981, 1986, 1990). The name "potential outcome" is a reference to the potential yields from Neyman's work in agricultural statistics (see Gelman and Meng 2004; Rubin 2005). The model also has roots in the economics literature (Roy 1951; Quandt 1972), with important subsequent work by James Heckman (see Heckman 1974, 1978, 1979, 1989, 1992), Charles Manski (1995), and others. The model is now dominant in both statistics and economics, and it is being used with increasing frequency across the basic and applied social and health sciences.

The core of the potential outcome model is simple. Suppose that each individual in a population of interest can be exposed to two alternative states of a cause. Each state is characterized by a distinct set of conditions, exposure to which potentially affects an outcome of interest, such as labor market earnings or scores on a standardized mathematics test. If the outcome is earnings, the population of interest could be adults between the ages of 30 and 50, and the two states could be whether or not an individual has obtained a college degree. Alternatively, if the outcome is a mathematics test score, the population of interest could be high school seniors, and the two states could be whether or not a student has taken a course in trigonometry. For the potential outcome model, these alternative causal states are referred to as alternative treatments. When only two treatments are considered, they are referred to as treatment and control. Throughout this book, we will conform to this convention.

The key assumption of the model is that each individual in the population of interest has a potential outcome under each treatment state, even though each individual can be observed in only one treatment state at any point in time. For example, for the causal effect of having a college degree rather than only a high school diploma on subsequent earnings, adults who have completed only high school diplomas have theoretical what-if earnings under the state "have a college degree," and adults who have completed college degrees have theoretical what-if earnings under the state "have only a high school diploma." These what-if potential outcomes are counterfactual in the sense that they exist in theory but are not observed.

Formalizing this conceptualization for a two-state treatment, the potential outcomes of each individual are defined as the true values of the outcome of interest that would result from exposure to the alternative causal states. The potential outcomes of each individual i are y_i^1 and y_i^0, where the superscript 1 signifies the treatment state and the superscript 0 signifies the control state. Because both y_i^1 and y_i^0 exist in theory for each individual, an individual-level causal effect can be defined as some contrast between y_i^1 and y_i^0, usually the simple difference $y_i^1 - y_i^0$. Because it is impossible to observe both y_i^1 and y_i^0 for any individual, causal effects cannot be observed or directly calculated at the individual level.[1]

[1]The only generally effective strategy for estimating individual-level causal effects is a crossover design, in which individuals are exposed to two alternative treatments in succession and with enough

By necessity, a researcher must analyze an observed outcome variable Y that takes on values y_i for each individual i that are equal to y_i^1 for those in the treatment state and y_i^0 for those in the control state. We usually refer to those in the treatment state as the treatment group and those in the control state as the control group.[2] Accordingly, y_i^0 is an unobservable counterfactual outcome for each individual i in the treatment group, and y_i^1 is an unobservable counterfactual outcome for each individual i in the control group.

In the potential outcome modeling tradition, attention is focused on estimating various average causal effects, by analysis of the values y_i, for groups of individuals defined by specific characteristics. To do so effectively, the process by which individuals of different types are exposed to the cause of interest must be modeled. Doing so involves introducing defendable assumptions that allow for the estimation of the average unobservable counterfactual values for specific groups of individuals. If the assumptions are defendable, and a suitable method for constructing an average contrast from the data is chosen, then the resulting average difference in the values of y_i can be given a causal interpretation.

The potential outcome model is one core piece of the more general counterfactual approach to causal analysis that we will present in this book. Another core piece, which we will introduce at the end of this chapter, is the directed graph approach to causal analysis, most closely associated with the work of the computer scientist Judea Pearl. We will use Pearl's work extensively in our presentation, drawing on his 2000 book, *Causality: Models, Reasoning, and Inference* (2nd edition, 2009), as well as related literature.

A counterfactual account of causation also exists in philosophy, which began with the seminal 1973 article by David Lewis, titled "Causation." It is related to the counterfactual model of causal analysis that we will present in this book, but the philosophical version, as implied by the title of Lewis' original article, aims to be a general model of causation.[3] As noted by the philosopher James Woodward in his 2003 book,

---

time elapsed in between exposures such that the effects of the cause have had time to dissipate (see Rothman, Greenland, and Lash 2008). Obviously, such a design can be attempted only when a researcher has control over the allocation of the treatments and only when the treatment effects are sufficiently ephemeral. These conditions rarely exist for the causal questions that interest social scientists.

[2] We assume that, for observational data analysis, an underlying causal exposure mechanism exists in the population, and thus the distribution of individuals across the treatment and control states exists separately from the observation and sampling process. Accordingly, the treatment and control groups exist in the population, even though we typically observe only samples of them in the observed data. We will not require that the labels "treatment group" and "control group" refer only to the observed treatment and control groups.

[3] In this tradition, causation is defined with reference to counterfactual dependence (or, as is sometimes written, the "ancestral" to counterfactual dependence). Accordingly, and at the risk of a great deal of oversimplification, the counterfactual account in philosophy maintains that (in most cases) it is proper to declare that, for events c and e, c causes e if (1) c and e both occur and (2) if c had not occurred and all else remained the same, then e would not have occurred. The primary challenge of the approach is to define the counterfactual scenario in which c does not occur (which Lewis did by imagining a limited "divergence miracle" that prevents c from occurring in a closest possible hypothetical world where all else is the same except that c does not occur). The approach differs substantially from the regularity-based theories of causality that dominated metaphysics through the 1960s, based on relations of entailment from covering law models. For a collection of essays in

Making Things Happen: A Theory of Causal Explanation, the counterfactual account of causation championed by Lewis and his students has not been influenced to any substantial degree by the potential outcomes version of counterfactual modeling used by statisticians, social scientists, and other empirical researchers. However, Woodward and other philosophers have engaged the directed graph approach to causality with considerable energy in the past decade. We will discuss the broader philosophical literature in Chapters 2, 10, and 13, as it does have some implications for social science practice and the pursuit of explanation more generally.

1.2 Causal Analysis and Observational Social Science

The challenges of using observational data to justify causal claims are considerable. In this section, we present a selective history of the literature on these challenges, focusing on the varied usage of experimental language in observational social science. We will also consider the growth of survey research and the shift toward outcome-equation-based motivations of causal analysis that led to the widespread usage of regression estimators. Many useful discussions of these developments exist, and our presentation here is not meant to be complete.[4] We review only the literature that is relevant for explaining the connections between the counterfactual approach and other traditions of quantitatively oriented analysis that are of interest to us here.

1.2.1 Experimental Language in Observational Social Science

Although the common definition of the word experiment is broad, in the social sciences it is most closely associated with randomized experimental designs, such as the double-blind clinical trials that have revolutionized the biomedical sciences and the routine small-scale experiments that psychology professors perform on their own students.[5]

philosophy on counterfactuals and causation, see Collins, Hall, and Paul (2004). For a penetrating examination of the counterfactual model in philosophy and its rivals, see Paul and Hall (2013). The counterfactual model that we consider in this book is close to what Paul and Hall label the “causal model” approach to causation, which they consider one of four variants of counterfactual modeling.

[4]To gain a more complete appreciation of the expansive literature on causality in the social sciences, see, for sociology, Barringer, Leahey, and Eliason (2013), Berk (1988, 2004, 2008), Blossfeld (2009), Bollen (1989), Bollen and Pearl (2013), Firebaugh (2008), Fox (2008), Gangl (2010), Goldthorpe (2007), Harding and Seefeldt (2013), Lieberson (1985), Marini and Singer (1988), Morgan (2013), Rohwer (2010), Singer and Marini (1987), Smith (1990, 2003, 2013), Sobel (1995, 1996, 2000), and Treiman (2009). For economics, see Angrist and Krueger (1999), Angrist and Pischke (2009, 2010), Heckman (2000, 2005, 2008b, 2010), Imbens and Wooldridge (2009), Keane (2010), Lee (2005), Manski (1994, 1995, 2003), Moffitt (2003), Pratt and Schlaifer (1984), and Rosenzweig and Wolpin (2000). For political science, see Brady and Collier (2010), Druckman, Kuklinski, and Lupia (2011), Dunning (2012), Gerber and Green (2012), Gerring (2007), Goertz and Mahoney (2012), King, Keohane, and Verba (1994), Morton and Williams (2010), and Sekhon (2009). For applied evaluation and policy analysis, see Shadish, Cook, and Campbell (2001), and especially for education research, see Murnane and Willett (2011)

[5]The *Oxford English Dictionary* provides the scientific definition of experiment: “An action or operation undertaken in order to discover something unknown, to test a hypothesis, or establish or illustrate some known truth” and also provides source references from as early as 1362.

Randomized experiments have their origins in the work of statistician Ronald A. Fisher during the 1920s, which then diffused throughout various research communities via his widely read 1935 book, *The Design of Experiments*.

Statisticians David Cox and Nancy Reid (2000) offer a definition of an experiment that focuses on the investigator's deliberate control and that allows for a clear juxtaposition with an observational study:

> The word *experiment* is used in a quite precise sense to mean an investigation where the system under study is under the control of the investigator. This means that the individuals or material investigated, the nature of the treatments or manipulations under study and the measurement procedures used are all selected, in their important features at least, by the investigator.
>
> By contrast in an observational study some of these features, and in particular the allocation of individuals to treatment groups, are outside the investigator's control. (Cox and Reid 2000:1)

We will maintain this basic distinction throughout this book. We will argue in this section that the potential outcome model of causality that we introduced in the last section is valuable precisely because it helps researchers to stipulate assumptions, evaluate alternative data analysis techniques, and think carefully about the process of causal exposure. Its success is a direct result of the language of potential outcomes, which permits the analyst to conceptualize observational studies as if they were experimental designs controlled by someone other than the researcher – quite often, the subjects of the research. In this section, we offer a brief discussion of other important attempts to use experimental language in observational social science and that succeeded to varying degrees.

Samuel A. Stouffer, the sociologist and pioneering public opinion survey analyst, argued that "the progress of social science depends on the development of limited theories – of considerable but still limited generality – from which prediction can be made to new concrete instances" (Stouffer 1962[1948]:5). Stouffer argued that, when testing alternative ideas, "it is essential that we always keep in mind the model of a controlled experiment, even if in practice we may have to deviate from an ideal model" (Stouffer 1950:356). He followed this practice over his career, from his 1930 dissertation that compared experimental with case study methods of investigating attitudes, to his leadership of the team that produced *The American Soldier* during World War II (see Stouffer 1949), and in his 1955 classic *Communism, Conformity, and Civil Liberties.*

On his death, and in celebration of a posthumous collection of his essays, Stouffer was praised for his career of survey research and attendant explanatory success. The demographer Philip Hauser noted that Stouffer "had a hand in major developments in virtually every aspect of the sample survey – sampling procedures, problem definition, questionnaire design, field and operating procedures, and analytic methods" (Hauser 1962:333). Arnold Rose (1962:720) declared, "Probably no sociologist was so ingenious in manipulating data statistically to determine whether one hypothesis or another could be considered as verified." And Herbert Hyman portrayed Stouffer's method of tabular analysis in charming detail:

> While the vitality with which he attacked a table had to be observed in action, the characteristic strategy he employed was so calculating that one can sense it from reading the many printed examples. . . . Multivariate analysis for him was almost a way of life. Starting with a simple cross-tabulation, the relationship observed was elaborated by the introduction of a third variable or test factor, leading to a clarification of the original relationship. . . . But there was a special flavor to the way Sam handled it. With him, the love of a table was undying. Three variables weren't enough. Four, five, six, even seven variables were introduced, until that simple thing of beauty, that original little table, became one of those monstrous creatures at the first sight of which a timid student would fall out of love with our profession forever. (Hyman 1962:324–25)

Stouffer's method was to conceive of the experiment that he wished he could have conducted and then to work backwards by stratifying a sample of the population of interest into subgroups until he felt comfortable that the remaining differences in the outcome could no longer be easily attributed to systematic differences within the subgroups. He never lost sight of the population of interest, and he appears to have always regarded his straightforward conclusions as the best among plausible answers. Thus, as he said, "Though we cannot always design neat experiments when we want to, we can at least keep the experimental model in front of our eyes and behave cautiously" (Stouffer 1950:359).

Not all attempts to incorporate experimental language into observational social science were as well received. Most notably in sociology, F. Stuart Chapin had earlier argued explicitly for an experimental orientation to nearly all of sociological research, but while turning the definition of an experiment in a direction that agitated others. For Chapin, a valid experiment did not require that the researcher obtain control over the treatment to be evaluated, only that observation of a causal process be conducted in controlled conditions (see Chapin 1932, 1947). He thus considered what he called "ex post facto experiments" to be the solution to the inferential problems of the social sciences, and he advocated matching designs to select subsets of seemingly equivalent individuals from those who were and were not exposed to the treatment of interest. In so doing, however, he proposed to ignore the incomparable, unmatched individuals, thereby losing sight of the population that Stouffer, the survey analyst, always kept in the foreground.

Chapin thereby ran afoul of emergent techniques of statistical inference, and he suffered attacks from his natural allies in quantitative analysis. The statistician Oscar Kempthorne, whose 1952 book *The Design and Analysis of Experiments* would later become a classic, dismissed Chapin's work completely. In a review of Chapin's 1947 book, *Experimental Designs in Sociological Research*, Kempthorne wrote:

> The usage of the word "experimental design" is well established by now to mean a plan for performing a comparative experiment. This implies that various treatments are actually applied by the investigator and are not just treatments that happened to have been applied to particular units for some reason, known or unknown, before the "experiment" was planned. This condition rules out practically all of the experiments and experimental designs discussed by the author. (Kempthorne 1948:491)

Chapin's colleagues in sociology and demography were often just as unforgiving. Nathan Keyfitz (1948:260), for example, chastised Chapin for ignoring the population of interest and accused him of using terms such as "experimental design" merely to "lend the support of their prestige."

In spite of the backlash against Chapin, in the end he has a recognizable legacy in observational data analysis. The matching techniques he advocated will be discussed in Chapter 5. They have been reborn in the new literature, in part because the population of interest has been brought back to the foreground. But there is an even more direct legacy. Many of Chapin's so-called experiments were soon taken up, elaborated, and analyzed by the psychologist Donald T. Campbell and his colleagues under the milder and more general name of "quasi-experiments."[6]

The first widely read presentation of Campbell's perspective emerged in 1963 (see Campbell and Stanley 1966[1963]), in which quasi-experiments were discussed alongside randomized and fully controlled experimental trials, with an evaluation of their relative strengths and weaknesses in alternative settings. In the subsequent decade, Campbell's work with his colleagues moved closer toward observational research, culminating in the volume by Cook and Campbell (1979), *Quasi-Experimentation: Design & Analysis Issues for Field Settings*, wherein a whole menu of quasi-experiments was described and analyzed: from the sort of ex post case-control matching studies advocated by Chapin (but relabeled more generally as nonequivalent group designs) to novel proposals for regression discontinuity and interrupted time series designs (which we will discuss in Chapter 11). For Cook and Campbell, the term quasi-experiment refers to "experiments that have treatments, outcome measures, and experimental units, but do not use random assignment to create the comparisons from which treatment-caused change is inferred" (Cook and Campbell 1979:6).[7] And, rather than advocate for a reorientation of a whole discipline as Chapin had, they pitched the approach as a guide for field studies, especially program evaluation studies of controlled interventions. Nonetheless, the ideas were widely influential throughout the social sciences, as they succeeded in bringing a tamed experimental language to the foreground in a way that permitted broad assessments of the strengths and weaknesses of alternative study designs and data analysis techniques.

1.2.2 "The Age of Regression"

Even though the quasi-experiment tradition swept through the program evaluation community and gained many readers elsewhere, it lost out in the core social science disciplines to regression-equation-based motivations of observational data analysis,

[6] In his first publication on quasi-experiments, Campbell (1957) aligned himself with Stouffer's perspective on the utility of experimental language, and in particular Stouffer (1950). Chapin is treated roughly by Campbell and Stanley (1966[1963]:70), even though his ex post facto design is identified as "one of the most extended efforts toward quasi-experimental design."

[7] Notice that Cook and Campbell's definition of quasi-experiments here is, in fact, consistent with the definition of an experiment laid out by Cox and Reid, which we cited earlier in this section. For that definition of an experiment, control is essential but randomization is not. The text of Cook and Campbell (1979) equivocates somewhat on these issues, but it is clear that their intent was to discuss controlled experiments for which randomization is infeasible and which they then label quasi-experiments.

under the influence at first of researchers who promoted regression modeling from a path-modeling orientation. In sociology, Hubert Blalock and Otis Dudley Duncan are usually credited with introducing the techniques, first via Blalock's 1964[1961] book *Causal Inferences in Nonexperimental Research* and then later via Duncan's 1966 article, "Path Analysis: Sociological Examples," which was published as the lead article in that year's *American Journal of Sociology*.[8] In both presentations, caution was stressed. Blalock discussed carefully the differences between randomized experiments and observational survey research. Duncan stated explicitly in his abstract that "path analysis focuses on the problem of interpretation and does not purport to be a method for discovering causes," and he concluded his article with a long quotation from Sewall Wright attesting to the same point.

A confluence of developments then pushed path models toward widespread usage and then basic regression modeling toward near complete dominance of observational research in some areas of social science.[9] In sociology, the most important impetus was the immediate substantive payoff to the techniques. *The American Occupational Structure*, which Duncan cowrote with Peter Blau and published in 1967, offered new decompositions of the putative causal effects of parental background and individuals' own characteristics on later educational and occupational attainment. By pushing social stratification research into new terrain, their book transformed a core subfield of the discipline of sociology, leading to major theoretical and methodological redirections of many existing lines of scholarship.[10]

Researchers seemed to then ignore many of the cautionary statements of Blalock, Duncan, and others. In their defense, it should be noted that Blalock's guidance was confusing at times. When introducing regression equations in his 1961 book, specified as $Y_i = a + bX_i + e_i$, where X is the causal variable of interest and Y is the outcome variable of interest, Blalock stated the matter correctly and clearly:

> What if there existed a major determinant of Y, not explicitly contained in the regression equation, which was in fact correlated with some of the

[8]Goldberger (1972) and Heckman (2000) offer a history of usage in economics, which begins before the history we offer for sociology. The biologist Sewall Wright (1925, 1934) deserves credit for the earliest developments (see Bollen and Pearl 2013; Pearl 2009).

[9]Regression estimation of systems of linear causal models with observed variables should be regarded as a restricted form of the more general structural equation modeling approach that Blalock, Duncan, and others introduced into sociology (see Bollen and Pearl 2013 for an explanation, especially their debunking of the myth "SEM and Regression Are Essentially Equivalent"). In these early and very influential pieces, Blalock and Duncan considered only linear causal models with observed variables. Duncan (1966:7) clarified: "As a statistical technique, therefore, neither path analysis nor the Blalock-Simon procedure adds anything to conventional regression analysis as applied recursively to generate a system of equations. As a pattern of interpretation, however, path analysis is invaluable in making explicit the rationale for a set of regression calculations." The literature moved quickly from the late 1960s to consider overidentified models (see Duncan, Haller, and Portes 1968; Hauser and Goldberger 1971), after which a general latent variable structural equation model was developed (see Bollen 1989). Nonetheless, the pattern of practice that elevated regression to its dominant position, we maintain, was shaped by these early pieces, as well as later work on interpreting the regression coefficients estimated for basic linear path models with observed variables (e.g., Alwin and Hauser 1975).

[10]For example, compare the methods (and substantive motivations) in Sewell (1964), with its non-parametric table standardization techniques, to Sewell, Haller, and Portes (1969), with its path model of the entire stratification process.