Book Review/Compte rendu

Stephen L. Morgan and **Christopher Winship**, *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Analytical Methods for Social Research. New York: Cambridge University Press, 2007, 328pp., $US 27.99 paper (978-0-521-67193-4), $US 75.00 hardcover (978-0-521-85615-7).

Among Sir R.A. Fisher's many seminal contributions to statistics was the fundamental technique of randomization in experimental design. In the 1950s, the great British statistician, himself a smoker and consultant to the tobacco industry, notoriously maintained that there was no good evidence that tobacco smoking causes lung cancer. One of Fisher's arguments was that it is possible that the association of tobacco smoking with lung cancer in observational data is due to a common genetic cause. (Fisher made important contributions to genetics as well as to statistics.) Even without the benefit of hindsight, Fisher's position seems perverse, but the more general difficulty of inferring causation from observational data is a real and continuing problem in epidemiology — witness, for example, the recent controversy over the efficacy and safety of hormone-replacement therapy for postmenopausal women. As every student of basic statistics knows (or should know), "Correlation does not imply causation."

Observational data are no less prominent in the social sciences than in epidemiology, and the issues that we address, for example in the area of public policy, are probably even more difficult to disentangle: Does capital punishment decrease homicide? Does the availability of legal abortion decrease violent crime? Do social-welfare programs raise the standard of living of the poor?

The difficulty of drawing causal conclusions from observational data has been understood for a long time, as has the basic strategy of controlling statistically for potentially confounding variables (for example, by matching, stratification, or regression analysis). Because it is always possible that a confounding prior cause has not been identified and observed, this strategy has a fundamental limitation not shared by randomized comparative experiments.

The last thirty years or so, however, have witnessed increased interest in causal inference for observational data: work in statistics by Rubin, Rosenthal, and others on the "counterfactual" approach to causal inference (see, e.g., Holland 1986, for a relatively compact account); Pearl's (2000) influential work (surprisingly, Pearl is a computer scientist) on strategies for drawing causal conclusions from statistical data; and work in econometrics by Heckman and others (e.g., Heckman, 2000 for a very general summary) on selection bias in program evaluation.

Morgan and Winship masterfully integrate the various streams of research on inferring causation from observational data, demonstrating their essential similarity and revealing how they differ. Their book is unified by the counterfactual or "potential outcomes" approach to causation, and by use of Pearl's directed graphs to explicate key points.

To focus ideas, consider a current issue: To address the relatively poor achievement and high drop-out rate among black students in Toronto schools, the Toronto District School Board has just approved plans for a "black-focused" high school. Suppose that, in several years, the black students who attend this school have a lower drop-out rate and higher average standardized test scores than their counterparts in other public schools. We could imagine, however, that each of the students attending the black-focused school could instead have attended a regular public high school, as would have been the case had the black-focused school not been established. Similarly, each black student in a regular public high school could, at least in principle, have attended the black-focused school.

Let y_i represent the ith student's score on the Ontario Secondary School Literacy Test, administered to students in grade 10; y_i^1 represent the score that the student would obtain were he or she to attend the black-focused school; and y_i^0 the score that the student would obtain were he or she to attend another public high school. For students *actually* in the black-focused school, the observed score $y_i = y_i^1$, and y_i^0 is an unrealized potential outcome, hence counterfactual. Likewise, for black students in other public high schools, $y_i = y_i^0$ and y_i^1 is counterfactual. Because in both cases one of the two potential outcomes is unobserved, it is impossible to observe individual causal effects such as $y_i^1 - y_i^0$.

Under certain circumstances, however, it is possible to estimate various kinds of average causal effects. Suppose, for example, that more students want to attend the black-focused high school than can be accommodated, and that those admitted to the school are selected by lottery. Let the indicator variable D_i=1 for a student selected to attend the black-focused school, and D_i= 0 for a student who applies for admission but is not selected. One can then estimate the average causal effect $E(y_i^1 - y_i^0) = E(y_i^1) - E(y_i^0)$ for students who volunteer for the school by comparing the average performance of students in the black-focused school with that of their unselected peers, that is, $(\bar{y}|D = 1) - (\bar{y}|D = 0)$. This is the potential-outcomes account of causal inference in a randomized comparative experiment.

Suppose, however, that students are admitted to the black-focused school on a first-come-first-served basis, or alternatively that all applicants are accepted. We could try to estimate $E(y_i^1 - y_i^0)$ by comparing the average test scores of black students in the black-focused school with those in other public high schools, but this attempt will fail if students in the two settings differ in relevant respects, as is likely because the students attending the black-focused school are self-selected. This is the essential conundrum of causal inference from observational data.

Most of Morgan and Winship's book deals with this setting: an observational study in which the goal is to estimate various kinds of average causal effects of a factor with two levels, generically termed "treatment" (such as attendance at the black-focused high school) and "control" (attendance at a regular public high school). The authors present relatively new work on a variety of strategies for dealing with the kinds of confounding that can occur in observational data, including (among others) new methods for stratifying and matching observations in the treatment and control groups, including the use of "propensity scores," which are the estimated probabilities of membership in the treatment and control groups; methods based on regression analysis, in which potentially confounding variables are controlled statistically; the use of instrumental variables estimators; and attempts to establish bounds for causal effects based on relatively weak assumptions about the causal process generating the data. All of these methods are explained with cleverly contrived "data" that illustrate how the methods work, when they are likely to be successful, and how they can fail.

Counterfactuals and Causal Inference is not a manual for applying the methods discussed, and the reader interested in using these methods will have to consult the statistical literature to which the authors refer. Moreover, there is no real data in the book — a weakness, in my opinion — and relatively little discussion of issues of causal inference in real social research. The latter is particularly a pity, because where the authors do address the research literature — for example in a relatively brief treatment of Coleman's controversial research on the effects of Catholic schooling in the United States on student achievement — their commentary is invariably insightful and illuminating. Occasionally, as well, the authors treat a subject so briefly that the exposition becomes opaque. In a chapter on mechanisms and causal explanation, for example, Morgan and Winship state that, "For … a mechanism to count as a sufficiently deep explanation, its causal pathways must be finely enough articulated that it meets whatever standard of bottoming out is maintained in the relevant field of study," a claim that skates dangerously close to tautology. Likewise, I doubt whether the discussions of models for time-series and panel data will be accessible to those not already familiar with the subject.

Despite these minor reservations, Morgan and Winship have written an important, wide-ranging, careful, and original introduction to the modern literature on causal inference in nonexperimental social research.

References

Heckman, James J. 2000. Causal parameters and policy analysis in economics: A twentieth century retrospective. *The Quarterly Journal of Economics* 115:45–97.

Holland, Paul W. 1986. Statistics and causal inference. *Journal of the American Statistical Association* 81:945–70.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

McMaster University John Fox

John Fox has written many articles and books on social statistics and quantitative methods. He has two books forthcoming in 2008 with Sage Publications: *Applied Regression Analysis and Generalized Linear Models*, *Second Edition*, and *A Mathematical Primer for Social Statistics*. jfox@mcmaster.ca