Commentary on Bryan Dowd's Paper "Separated at Birth: Statisticians, Social Scientists, and Causality in Health Services Research"

A. James O'Malley

This issue of the journal contains a very insightful and informative paper by Bryan Dowd. Although health services research is a relatively new field, there are valuable lessons to be learned from an historical review of critical junctures involving its core methodological disciplines, including statistics, economics, and sociology. Extensive use of observational data and need for causally defensible results makes health services research an ideal stage for comparing disciplines.

This commentary begins by addressing some of the major points in Dowd (2011, this issue) followed by a review of how causality has featured in my own work. The commentary concludes by describing causality from the operational subjective statistics viewpoint, which complements Dowd's paper because it has a different philosophical foundation than the frequentist and Bayesian paradigms and a different viewpoint of causality.

GENERAL COMMENTS

Dowd's paper examines why statisticians and social scientists (economists, sociologists) have tended to view causality differently and use different approaches to causal inference. Upon absorbing the arguments made in the paper, the reader might wonder whether it is all that surprising that the different methods have to a large extent remained the property of the fields in which they were developed. One might also just ask how different are the methods and are they related?

At the time randomized trials were discovered (Fisher 1926), statisticians typically worked in agriculture or other industries where treatments could be assigned. In the social sciences, practical or ethical reasons generally ruled out

randomized experiments and so randomized experiments held limited utility (these days the Internet provides more opportunities). Instrumental variables provided a pathway to causality that bypassed the need to externally manipulate and randomly set treatments; thus, they (Wright 1928) represented a major breakthrough for social scientists, although this took some time to be realized. Because statisticians and social scientists encountered different problems, it is not surprising that discipline-specific methods resulted. Furthermore, the tendency for substantive fields to develop their own jargon and in some cases methods, rather than rely on statisticians, might in part be due to the historical affiliation of statisticians with mathematics departments, which may have limited their exposure or incentive to work on applied problems.

Statisticians traditionally have had a preference for working on methods with empirically testable assumptions, whereas social scientists have relied much more on theoretical models for justifying assumptions used in statistical analyses. As a consequence, the uptake of structural equation models and instrumental variable methods among statisticians was slow, while the practice of conducting databased sensitivity analysis to evaluate the robustness of results to model assumptions was less emphasized in the social sciences. Yet the principal methods for causal inference do not seem far apart when (as noted by Dowd) one considers that randomization is the “perfect” instrument.

One of Dowd's major points is that it is the use of different philosophical principles and language concerning causality, rather than technical aspects of methods themselves, that led to the “separation.” An unfortunate consequence of statisticians' (particularly those working on clinical trials) unfamiliarity with instrumental variables is that intention-to-treat analyses may have been used when an analysis of the treatment received (the therapeutic effect of treatment) using instrumental variable methods would have been enlightening.

Today times are changing and stronger ties are now being formed between disciplines. The principal stratification framework of Frangakis and Rubin (2002) is a case in point. Principal stratification includes instrumental variable methods as a special case of a more general methodology for the situation where one or more intermediary variables (e.g., compliance status), which are possibly only partially observed, are available to define strata for which causal effects can be evaluated (Barnard et al. 2003). Robins and colleagues have developed a variety of novel methods for causal inference

---

Address correspondence to A. James O'Malley, Ph.D., Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115-5899; e-mail: omalley@hcp.med.harvard.edu.

with a particular focus on longitudinal and survival analysis (e.g., Robins et al. 1992; Robins and Greenland 1994; van der Laan and Robins 2002). Therefore, I wonder whether boxes with linkages between the “RCT side” and the “correlation side” could be added to Figure 2 in Dowd’s article. For example, a box “studies with failed randomization” with inward links from the randomized trials and structural equation modeling or instrumental variables boxes seems appropriate.

PERSONAL EXPERIENCE

For the past 9 years, I have been a statistician in a Department of Health Care Policy—a department also composed of health economists, medical sociologists, physicians, and general health policy researchers. The mission of the department is to conduct and publish research that assists policy makers to improve the quality and financing of health care in the United States.

Because of the collaborative culture in my department, recognition of the value of considering and ultimately using methods developed in other disciplines is the norm, not the exception. In this environment, a statistician who is unfamiliar with instrumental variables, structural equation modeling, or the use of theoretical economic and sociological models to motivate and supplement empirical analysis quickly adds these to his/her repertoire. This contrasts with the scenario in the study-section depicted in Dowd’s paper! Conversely, the economists, sociologists, and physician researchers in my department become acquainted with a wider range of statistical models, methods, and practices (e.g., hierarchical, mixed-membership, and multivariate models; Bayesian analysis; experimental design; and methods for evaluating the sensitivity of results to model assumptions) than they might have otherwise.

Because most of my collaborative research projects involve observational data, concerns about selection of patients (or health providers and health plans) into treatments are paramount. In formulating an analysis plan, consideration is given to all methods of analysis with the choice ultimately depending on what is most appropriate for the problem at hand. If a strong instrumental variable is available (i.e., one backed by strong theoretical arguments), this would suggest using an instrumental variables analysis. If inundated with predictor variables but lacking a good instrument, then propensity scores methods would become more attractive. If interaction(s) involving the treatment were thought to be present or there was interest in testing for such effects, then multiple regression analysis, possibly in the

context of an instrumental variable or propensity score analysis, would be considered. If more than one of these approaches can be applied to a problem, we might perform each and compare the results in order to evaluate the sensitivity to results of the differing assumptions. If large differences are found, we try to provide explanations. This often involves further empirical analysis, perhaps using simulation experiments, as well as theoretical models based on the insights of social scientists and physicians, all of which can lead to a more enlightened analysis.

Often the randomized trial data we use have their own difficulties. For example, randomized clinical trials in mental health are notorious for treatment noncompliance and subsequent nonresponse. To overcome these problems, we have used principal stratification methods to move beyond intention-to-treat analysis (O'Malley and Normand 2005). Surveys are another valuable source of data for health policy and health services research. An integral step in designing a survey is the removal of redundant items, while in designing reports based on survey data it is often desirable to combine items into meaningful scales. Latent variable models such as factor analysis (traditionally the domain of sociology and psychology) are often used to study the correlation structure of the survey items and in so doing help accomplish these objectives (O'Malley et al. 2005).

Some problems we encounter present new challenges for causal inference. A prime example is social network analysis. Network data are characterized by interdependence among individuals. For example, the same individual may influence and be influenced by multiple other individuals. Thus, the data embody complex dependence structures that provide serious difficulties for statistical inference. For example, if individual A influences individual B and individual C influences individual A, then A's treatment (including influence from C) may affect B's outcome. This violates the stable unit treatment value assumption (SUTVA), a condition for identifying causal effects (Rubin 1980). Sobel (2006) and Hudgens and Halloran (2008) investigate identifying causal effects when SUTVA is violated, but to my knowledge this has not been addressed generally in the context of social network data.

THE OPERATIONAL SUBJECTIVIST VIEWPOINT

To provide yet another perspective on causality, I now consider the viewpoint of the operational subjective statistician, a statistical paradigm founded by Bruno de Finetti (see de Finetti 1972, 1975) and staunchly followed today by a

small community of statisticians. Observable quantities are all that matter in operational subjective statistics with models being of our opinions, not some supposed unobservable probabilistic generating structure for the observables. Under this paradigm, it is operationally meaningless to formulate inferences about quantities that you will never observe (e.g., model parameters, the causal effect of one variable on another) or to use approaches reliant on assumptions that cannot be observed as true or false (F. Lad, personal communication). However, a theoretical model (e.g., social science theorem) may be inherent in your asserted opinions about the observable quantities of interest.

Operational subjective statistics is predictive in nature, conditioning on whatever data you have already observed to obtain your opinion for the values of yet-to-be observed observations. For example, in a clinical setting, one might compute the predictive probabilities of various treatment outcomes for a patient conditional on outcome and treatment for past patients under the testable assumption that the sequence of outcomes is exchangeable; by exchangeability it is meant that the probability of the outcomes is invariant to the order in which they are observed (see Lad 1996, Sections 3.8–3.12, for a thorough treatment of exchangeability). Unlike the frequentist and Bayesian approaches, inferences are targeted solely toward future values, not the parameters of hypothetical distributions from which the observed data are assumed to have been generated. A policy maker following the operational subjectivist paradigm might frame policy decisions on their predictions of the next patient's (or a range of patients') clinical outcome of treatment, evaluated conditional on each treatment of interest. The concept of manipulating one variable (e.g., treatment) while holding others (e.g., covariates) fixed helps define the quantities of interest; to be observable (and thus of interest to policy makers) the quantities must correspond to patients—treatment combinations that exist.

At first it might appear that the subjective statistics paradigm allows you to say anything you like. However, to encourage honesty and accuracy in asserting your opinions, proper scoring rules (in the context of decision problems these are equivalent to utility functions) are used to quantify the relative quality of different states of uncertain knowledge (Lad 1996, Chapter 6). Specifically, your joint distribution for the variables of interest is “scored” against the ultimate observation of them. If you state values that do not cohere with your actual state of knowledge, you can expect to score poorly in the sense that if you were to place bets in accordance with your stated values on how the observations were to come out, you could expect to lose money. Thus, while the operational subjectivist approach does not assume that an absolute

truth exists nor requires that your opinion be the same as someone else's, it utilizes tools for ensuring coherent specification and updating of knowledge.

Operational subjective statistics provides an interesting alternative viewpoint of the role of causality. However, despite its pure and philosophically sound principles, it is rarely used in practice. The primary reason is that substantial effort is required to implement the approach and computational barriers become prohibitive in smaller dimensions than with frequentist or Bayesian methods. However, the ever-increasing level of computing power will lessen the computational barriers and so these methods may be utilized more in the future, especially in situations where expert opinion is needed to supplement experimental evidence in decision making.

To conclude, I hope and expect that Bryan Dowd's excellent paper and the accompanying commentaries will result in a robust and useful discussion regarding the estimation of causal effects in health services research.

REFERENCES

- Barnard, J., C. E. Frangakis, J. L. Hill, and D. B. Rubin. 2003. "Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City." *Journal of the American Statistical Association* 98: 299–311.
- Dowd, B. (2011, this issue). "Separated at Birth: Statisticians, Social Scientists and Causality in Health Services Research." *Health Services Research*.
- de Finetti, B. 1972. *Probability, Induction, and Statistics*. New York: Wiley.
- . 1975. *Theory of Probability: A Critical Introductory Treatment*, Vol. 2. New York: Wiley.
- Fisher, R. 1926. "The Arrangement of Field Experiments." *Journal of the Ministry of Agriculture* 33: 500–13.
- Frangakis, C. E., and D. B. Rubin. 2002. "Principal Stratification in Causal Inference." *Biometrics* 58: 21–9.
- Hudgens, M. G., and M. E. Halloran. 2008. "Toward Causal Inference with Interference." *Journal of the American Statistical Association* 103: 832–42.
- van der Laan, M., and J. M. Robins. 2002. *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer.
- Lad, F. 1996. *Operational Subjectivist Statistical Methods*. New York: Wiley.
- O'Malley, A. J., and S-L. T. Normand. 2005. "Likelihood Methods for Treatment Noncompliance and Subsequent Nonresponse in Randomized Trials." *Biometrics* 61: 325–34.
- O'Malley, A. J., A. M. Zaslavsky, R. D. Hays, K. A. Hepner, S. Keller, and P. D. Cleary. 2005. "Exploratory Factor Analyses of the CAHPS® Hospital Pilot Survey Responses across and within Medical, Surgical, and Obstetric Services." *Health Services Research* 40: 2078–95.

- Robins, J. M., D. Blevins, G. Ritter, and M. Wulfson. 1992. "G-Estimation of the Effect of Prophylaxis Therapy for *Pneumocystis carinii* Pneumonia on the Survival of AIDS Patients." *Epidemiology* 3: 319–36.
- Robins, J. M., and S. Greenland. 1994. "Adjusting for Differential Rates of Prophylaxis Therapy for PCP in High versus Low-Dose AZT Treatment Arms in an AIDS Randomized Trial." *Journal of the American Statistical Association* 89: 737–49.
- Rubin, D. B. 1980. "Discussion of Randomization Analysis of Experimental Data: The Fisher Randomization Test." *Journal of the American Statistical Association* 75: 591–3.
- Sobel, M. E. 2006. "What do Randomized Studies of Housing Mobility Demonstrate?" *Journal of the American Statistical Association* 101: 1398–407.
- Wright, P. 1928. *The Tariff on Animal and Vegetable Oils*. Appendix B. New York: Macmillan.