HSR Health Services Research

DOI: 10.1111/j.1475-6773.2011.01243.x
COMMENTARY

TECHNICAL REPORT
R-373
April 2011

Statistics and Causality: Separated to Reunite—Commentary on Bryan Dowd's "Separated at Birth"

Judea Pearl

Introduction: Tensions and Confusions

Bryan Dowd (2010) should be commended for laying before us the historical roots of the tensions between statisticians and econometricians which, until today, perpetuate the myth that causal inference is somehow confusing, enigmatic, or controversial. While modern analysis has proven this myth baseless, it is often the historical accounts that put things in the proper perspective.

I see the tension between statistics and economics or, more generally, between statistics and causality, to be rooted in a more fundamental schism than the one portrayed in Dowd's account. Moreover, and contrary to Dowd's narrative, I believe that the schism was justified, necessary, and not sufficiently emphasized. In fact, it was only after the distinction between statistical and causal concepts was made crisp and formal through new mathematical notation that a productive symbiosis has emerged which now benefits both paradigms.

Dowd's account portrays the schism as a product of unfortunate circumstances that could have been avoided, if only the players were more aware of each other work. Economists, we are told, developed causal inference techniques that yield regressional estimates of causal effects (e.g., IV, confounding-control) and, since regression is a proud invention of statistics, there was no reason for statisticians to shun causal analysis as strongly as they did. If they did oppose structural equations, instrumental variables, and observational studies, it must have been due to an unfortunate rhetorical distinction or, perhaps, a fluke in the history of science.

Upon reading Dowd's account on how close statisticians were to develop causal inference techniques by themselves, readers might be tempted to conclude that the distinction between causal and statistical inference is perhaps unwarranted, and that the former is but a nuance of the latter. This would be a mistake and would not cohere with the lesson I draw from the history of causal analysis.

While there has been indeed a century of tension and misunderstanding between statisticians and econometricians regarding causation, the tension was justified: economists were prepared to posit untestable structural assumptions and derive their consequences, but statisticians were not. As a result, economists developed IV estimators and methods of controlling for confounders—all based on theoretical, untestable assumptions—while statisticians were alienated by those assumptions and found refuge in Fisher's controlled randomized trial (CRT), where the only assumptions needed were those concerning the nature of randomization.

Dowd recognizes this basic difference but attempts to minimize its importance, arguing that statistical analysis too is laden with untested assumptions, for example, that randomization balances "unobserved confounders," or that certain measurements were taken under identical conditions, or that a certain error is Gaussian, or that Bayesian priors have certain values, or that an experimental group is "representative" of a target population.

There are, however, fundamental differences between the assumptions that underlie statistical studies and those needed for causal inference in observational studies. The first difference is that most of the assumptions in conventional statistical studies, while untested perhaps in a given study, are testable in principle, given sufficiently large sample and sufficiently refined measurements. Causal assumptions, in contrast, cannot be tested even in principle, unless one resorts to experimental control. This difference stands out in Bayesian analysis. Although the priors that Bayesians commonly assign to statistical parameters are untested quantities, the sensitivity to these priors tends to diminish with increasing sample size. In contrast, sensitivity to prior causal assumptions, say that treatment does not change gender, remains substantial regardless of sample size.

The second, and perhaps deeper difference between statistical and causal information is that the latter cannot be expressed in probability

---

Address correspondence to Judea Pearl, Ph.D., Professor, Computer Science Department, University of California, Los Angeles, CA 90095-1596; e-mail: judea@cs.ucla.edu

calculus—the standard language of statistical analysis. Any mathematical approach to causal analysis must acquire new notation—probability calculus is insufficient. (Skeptics are invited to write down a mathematical expression for the English sentence: "The rooster crow does not cause the sun to rise.")

This notational requirement, which economists tried to circumvent using structural equations, was unacceptable to statisticians who insisted that all empirical information be expressed in contingency tables, probability functions, or covariance matrices. In particular, Fisher's generation of statisticians could not accept the ambiguities associated with structural equations. Indeed, if we examine Dowd's first equation

$$Y_i = X_i\beta + T_i\beta_T + u_i \tag{1}$$

from a statistical perspective, the ambiguities are overwhelming. How is one to distinguish this equation from a regression equation, in which the error is automatically orthogonal to T_i? What empirical information is conveyed by this equation, if any? Does β_T have a causal interpretation? Does this interpretation vary with the statistics of u_i or with the existence of a confounder W?

The ambiguities associated with interpreting this seemingly innocent equation became, to no exaggeration, the greatest confusion of the 20th century. Paul Holland (1995, p. 54), for example, writes: "I am speaking, of course, about the equation: $\{y = a + bx + \varepsilon\}$. What does it mean? The only meaning I have ever determined for such an equation is that it is a shorthand way of describing the conditional distribution of $\{y\}$ given $\{x\}$." Today we know, of course, that the structural interpretation of equation (1) has nothing to do with the conditional distribution of $\{y\}$ given $\{x\}$; rather, it conveys causal and counterfactual information that is orthogonal to the statistical properties of $\{x\}$ and $\{y\}$ (see Pearl 2009, Chapter 7). But such an understanding was not to be expected from traditional statisticians who, even as late as the 1990s, considered structural equations to be "meaningless" (Wermuth 1992). Naturally, statisticians gravitated to Fisher's experiments for as long as they could and, when mathematical analysis of causal relations became necessary, they invented the Neyman–Rubin "potential outcome" notation (Rubin 1974) and continued to oppose structural equations as a threat to principled science (Rubin 2004, 2009, 2010; Sobel 2008), not recognizing that the two languages are in fact equivalent (Pearl 2009, pp. 98–102).

David Freedman, another staunch critics of structural equations found them not only ambiguous but utterly "self-contradictory" (Freedman 1987, p. 114). Today we understand that Freedman's error was to interpret structural equations as ordinary equations when in fact they are nonalgebraic—they

change meaning under legitimate algebraic operations such as moving terms from one side of the equation to the other. What Freedman and others failed to realize is that the equality sign in equation (1) stands not for algebraic equality but for an assignment operator (: =) which Nature invokes to assign values to Y_i, based on the current values of X_i, T_i, and u_i (Pearl 2009, p. 138).

Remarkably, the recent flair-up of interest in instrumental variables and other structurally based identification methods (Angrist and Pischke 2010) does not reflect statisticians' acceptance of structural equations as legitimate carriers of scientific knowledge; rather, it reflects an uncritical reliance on the semblance between quasi-experiment (e.g., instrumental variables) and certain properties (e.g., balance) of Fisher's randomized experiment—the supreme compass of statistical right and wrong. Members of the so-called "experimentalist" camp in econometrics still refuse to recognize structural equation modeling for what it is—a transparent, formal language for causal and counterfactual information, logically equivalent to the opaque jargon of structureless models,[1] within which powerful identification methods could be both justified and derived (Heckman 2010; Keane 2010; Leamer 2010; Nevo and Whinston 2010).

The Fruits of Reunification

I have elaborated on this century-old tension to convince readers that a reconciliation between statisticians and economists was inconceivable in the days of Fisher and Wright; it had to wait for the logic of structural equations to be explicated, axiomatized, and put on firm mathematical grounds. Most important, it had to wait until a new algebra was developed for causal analysis, be it in structural, graphical, or counterfactual form.

Today, we know precisely which assumptions are causal, which are statistical, and how the two interact to produce meaningful inferences from both observational and experimental studies. Today we also understand that it is important to keep the two types of information apart, cast in notational distinction, so as to enable investigators to trace back the theoretical assumptions that enable causal conclusions to be derived from data.

This mathematical distinction has given rise to symbiotic methodology that scored a rather impressive record of achievements in the past two decades, many of which are routinely used in the health sciences (Greenland, Pearl, and Robins 1999; Robins 2001; Petersen, Sinisi, and van der Laan 2006; VanderWeele and Robins 2007; Glymour and Greenland 2008; Shrier and

Platt 2008; Hafeman and Schwartz 2009; Schisterman, Cole, and Platt 2009; VanderWeele 2009; Ojha 2010). I would mention a few highlights of this methodology to give readers a taste of the excitement currently sweeping the landscape of symbiotic causal analysis.

What emerged from the structural-counterfactual symbiosis is a methodology based on nonparametric structural equation models—a natural generalization of those used by econometricians and social scientists in the 1950–1960s, yet cast in new mathematical underpinnings, liberated from parametric blindfolds that have obscured the causal content of structural equations. This nonparametric framework, which I dubbed structural causal models (SCM), has enabled the development of several inferential tools, including the following:

1. Tools for explicating and enumerating the causal assumptions embodied in structural equation models as well as the assumptions that support each causal claim in the analysis (Pearl 2004, 2009, p. 101).
2. Methods of identifying the testable implications (if any) of the assumptions in (1), and ways of testing the testable implications of the assumptions behind each derived claim (Verma and Pearl 1990; Pearl 2004; Kyono 2010).
3. Methods of deciding, before taking any data, what measurements ought to be taken, whether one set of measurements is as good as to another, and which measurements tend to bias our estimates of the target quantities (Pearl 1993, 1995; Tian and Pearl 2002; Pearl and Paz 2010).
4. Methods for devising critical statistical tests by which two competing theories can be distinguished (Pearl and Paz 2010).
5. Methods of deciding mathematically if the causal relationships of interest are estimable from nonexperimental or quasi-experimental data and, if not, what additional assumptions, measurements, or experiments would render them estimable (Pearl 1993, 1995; Tian and Pearl 2002; Shpitser and Pearl 2006; Kyono 2010).
6. Methods of recognizing and generating equivalent models (Verma and Pearl 1990; Pearl 2009, p. 19; Ali, Richardson, and Spirtes 2009).
7. Methods of recognizing instrumental variables and auxiliary instrumental variables in structural models (Pearl 2009, pp. 257–8; Brito and Pearl 2002).
8. Generalization of structural equation models to categorical data and nonlinear interactions (Pearl, 1995; Pearl 2009, Chapter 7).

9. A solution to the so-called "mediation problem" (Baron and Kenny 1986), taking the form of estimable formulas for direct and indirect effects that are applicable to both continuous and categorical variables, and to linear as well as nonlinear interactions (Pearl 2001, 2010a, b).
10. A formal solution to the "external validity" problem (Campbell and Stanley 1963), which uses knowledge about differences and commonalities between populations to decide whether and how causal effects in one population can be inferred from experimental studies on another (Pearl and Bareinboim 2010).

These tools and results owe their developments to new mathematical notation in which statistical and causal relationships are kept apart, each governed by its own calculus, and each contributing its appropriate share to the inference sought.

CONCLUSIONS

Causal and statistical information are two different species that do not and should not be mixed. The latter deals with probabilistic relationships among observed variables; the former deals with hypothetical relationships in new situations. These relationships should be kept apart by notational distinctions and be governed by separate calculi. Once the mathematical distinction is accomplished, symbiotic analysis can benefit both causal and statistical inferences.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This research was supported in parts by NIH grant 1R01 LM009961-01, NSF grant IIS-0914211, and ONR grant N000-14-09-1-0665.

Disclosures: None.

Disclaimers: None.

NOTE

1. This opaqueness, and the blind reliance on, the RCT paradigm has already resulted in several blunders, among them, misguided advice on the choice

of covariates (Pearl, 2010c), paradoxical definitions of direct and indirect effects (Pearl, 2010a), and inadequate definitions of "surrogate endpoints" (Pearl and Bareinboim, 2010).

REFERENCES

Ali, R., T. Richardson, and P. Spirtes. 2009. "Markov Equivalence for Ancestral Graphs." *The Annals of Statistics* 37: 2808–37.

Angrist, J., and J.-S. Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives* 24: 3–30.

Baron, R., and D. Kenny. 1986. "The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51: 1173–82.

Brito, C., and J. Pearl. 2002. "Generalized Instrumental Variables." In *Uncertainty in Artificial Intelligence, Proceedings of the Eighteenth Conference*, edited by A. Darwiche and N. Friedman, pp. 85–93. San Francisco: Morgan Kaufmann.

Campbell, D., and J. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research.* Chicago: Wadsworth Publishing.

Dowd, B. 2010. "Separated at Birth: Statisticians, Social Scientists, and Causality in Health Services Research." *Health Services Research* 46(2): 397–420.

Freedman, D. 1987. "As Others See Us: A Case Study in Path Analysis (with Discussion)." *Journal of Educational Statistics* 12: 101–223.

Glymour, M., and S. Greenland. 2008. "Causal Diagrams." In *Modern Epidemiology.* 3d Edition, edited by K. Rothman, S. Greenland, and T. Lash, pp. 183–209. Philadelphia, PA: Lippincott Williams & Wilkins.

Greenland, S., J. Pearl, and J. Robins. 1999. "Causal Diagrams for Epidemiologic Research." *Epidemiology* 10: 37–48.

Hafeman, D., and S. Schwartz. 2009. "Opening the Black Box: A Motivation for the Assessment of Mediation." *International Journal of Epidemiology* 3: 838–45.

Heckman, J. 2010. "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy." *Journal of Economic Literature* 48: 356–98.

Holland, P. 1995. "Some Reflections on Freedman's Critiques." *Foundations of Science* 1: 50–7.

Keane, M. 2010. "A Structural Perspective on the Experimentalist School." *Journal of Economic Perspectives* 24: 47–58.

Kyono, T. 2010. *Commentator: A Front-End User-Interface Module for Graphical and Structural Equation Modeling.* Technical Report R-364, Department of Computer Science, University of California, Los Angeles, CA. Available at http://ftp.cs.ucla.edu/pub/stat_ser/r364.pdf

Leamer, E. 2010. "Tantalus on the Road to Asymptopia." *Journal of Economic Perspectives* 24: 31–46.

Nevo, A., and M. Whinston. 2010. "Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference." *Journal of Economic Perspectives* 24: 69–82.

Ojha, R. 2010. "A Cautionary Note on the Inclusion of Strong Predictors of the Exposure in Propensity Score Analysis." *Journal of Epidemiology and Community Health*, published October 15. E-Letter Reply to, S. Cousens, J. Hargreaves, C. Bonell, B. Armstrong, J. Thomas, B. R. Kirkwood, and R. Hayes, "Alternatives to Randomisation in the Evaluation of Public-Health Interventions: Statistical Analysis and Causal Inference." *Journal of Epidemiology and Community Health* Online First, Published July 13, 2010, doi:10.1136/jech.2008.082610.

Pearl, J. 1993. "From Bayesian Networks to Causal Networks." In *Proceedings of the Adaptive Computing and Information Processing Seminar*. Brunel Conference Centre, London. See also *Statistical Science*, 8(3): 266–9.

———. 1995. "Causal Diagrams for Empirical Research." *Biometrika* 82: 669–710.

———. 2001. "Direct and Indirect Effects." In *Uncertainty in Artificial Intelligence, Proceedings of the Seventeenth Conference*, edited by J. S. Breese and D. Koller, pp. 411–20. San Francisco, CA: Morgan Kaufmann.

———. 2004. "Robustness of Causal Claims." In *Proceedings of the Twentieth Conference Uncertainty in Artificial Intelligence*, edited by M. Chickering and J. Halpern, pp. 446–53. Arlington, VA: AUAI Press.

———. 2009. *Causality: Models, Reasoning, and Inference*. 2d Edition. New York: Cambridge University Press.

———. 2010a. "The Foundations of Causal Inference." *Sociological Methodology* 40: 75–149.

———. 2010b. *The Mediation Formula: A Guide to the Assessment of Causal Pathways in Non-Linear Models*. Technical Report R-363, Department of Computer Science, University of California, Los Angeles, CA. Available at //ftp.cs.ucla.edu/pub/stat_ser/r363.pdf. To appear in C. Berzuini, P. Dawid, and L. Bernardinelli (Eds.), *Statistical Causality*. Forthcoming, 2011.

———. 2010c. "On a Class of Bias-Amplifying Variables That Endanger Effect Estimates." In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*. AUAI, Corvallis, OR, 425–32. Available at http://ftp.cs.ucla.edu/pub/stat_ser/r356.pdf

Pearl, J., and E. Bareinboim. 2010. *Transportability across Studies: A Formal Approach*. Technical Report R-372, Department of Computer Science, University of California, Los Angeles, CA. Available at http://ftp.cs.ucla.edu/pub/stat ser/r372.pdf

Pearl, J., and A. Paz. 2010. "Confounding Equivalence in Causal Equivalence." In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, edited by P. Grunwald, and P. Spirtes, pp. 433–41. Corvallis, OR: AUAI.

Petersen, M., S. Sinisi, and M. van der Laan. 2006. "Estimation of Direct Causal Effects." *Epidemiology* 17: 276–84.

Robins, J. 2001. "Data, Design, and Background Knowledge in Etiologic Inference." *Epidemiology* 12: 313–20.

Rubin, D. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66: 688–701.

———. 2004. "Direct and Indirect Causal Effects via Potential Outcomes." *Scandinavian Journal of Statistics* 31: 161–70.

———. 2009. "Author's Reply: Should Observational Studies be Designed to Allow Lack of Balance in Covariate Distributions across Treatment Group?" *Statistics in Medicine* 28: 1420–3.

———. 2010. "Reflections Stimulated by the Comments of Shadish (2010) and West and Thoemmes (2010)." *Psychological Methods* 15: 38–46.

Schisterman, E., S. Cole, and R. Platt. 2009. "Overadjustment Bias and Unnecessary Adjustment in Epidemiologic Studies." *Epidemiology* 20: 488–95.

Shpitser, I., and J. Pearl. 2006. "Identification of Conditional Interventional Distributions." In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, edited by R. Dechter and T. Richardson, pp. 437–44. Corvallis, OR: AUAI Press.

Shrier, I., and R. Platt. 2008. "Reducing Bias through Directed Acyclic Graphs." *BMC Medical Research Methodology* 8: doi: 10.1186/1471–2288–8–70.

Sobel, M. 2008. "Identification of Causal Parameters in Randomized Studies with Mediating Variables." *Journal of Educational and Behavioral Statistics* 33: 230–1.

Tian, J., and J. Pearl. 2002. "A General Identification Condition for Causal Effects." In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, edited by R. Dechter, M. Kearns, and R. S. Sutton, pp. 567–73. Menlo Park, CA: AAAI Press/The MIT Press.

Wermuth, N. 1992. "On Block-Recursive Regression Equations." *Brazilian Journal of Probability and Statistics* 6: 1–56.

VanderWeele, T. 2009. "Marginal Structural Models for the Estimation of Direct and Indirect Effects." *Epidemiology* 20: 18–26.

VanderWeele, T., and J. Robins. 2007. "Four Types of Effect Modification: A Classification Based on Directed Acyclic Graphs." *Epidemiology* 18: 561–8.

Verma, T., and J. Pearl. 1990. "Equivalence and Synthesis of Causal Models." In *Uncertainty in Artificial Intelligence, Proceedings of the Sixth Conference*, edited by P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, pp. 255–68. Cambridge, MA: Elsevier Science Publishers, B.V.