Taking Causality Seriously: Propensity Score Methodology Applied to Estimate the Effects of Marketing Interventions

Donald B. Rubin

Department of Statistics, Harvard University, Cambridge MA
rubin@stat.harvard.edu

Propensity score methods were proposed by Rosenbaum and Rubin (1983, Biometrika) as central tools to help assess the causal effects of interventions. Since their introduction two decades ago, they have found wide application in a variety of areas, including medical research, economics, epidemiology, and education, especially in those situations where randomized experiments are either difficult to perform, or raise ethical questions, or would require extensive delays before answers could be obtained. Rubin (1997, Annals of Internal Medicine) provides an introduction to some of the essential ideas. In the past few years, the number of published applications using propensity score methods to evaluate medical and epidemiological interventions has increased dramatically. Rubin (2003, Erlbaum) provides a summary, which is already out of date.

Nevertheless, thus far, there have been few applications of propensity score methods to evaluate marketing interventions (e.g., advertising, promotions), where the tradition is to use generallly inappropriate techniques, which focus on the prediction of an outcome from an indicator for the intervention and background characteristics (such as least-squares regression, data mining, etc.). With these techniques, an estimated parameter in the model is used to estimate some global "causal" effect. This practice can generate grossly incorrect answers that can be self-perpetuating: polishing the Ferraris rather than the Jeeps "causes" them to continue to win more races than the Jeeps ¡=¿ visiting the high-prescribing doctors rather than the low-prescribing doctors "causes" them to continue to write more prescriptions.

This presentation will take "causality" seriously, not just as a casual concept implying some predictive association in a data set, and will show why propensity score methods are superior in practice to the standard predictive approaches for estimating causal effects. The results of our approach are estimates of individual-level causal effects, which can be used as building blocks for more complex components, such as response curves. We will also show how the standard predictive approaches can have important supplemental roles to play, both for refining estimates of individual-level causal effect estimates and for assessing how these causal effects might vary as a function of background information, both important uses for situations when targeting an audience and/or allocating resources are critical objectives.

The first step in a propensity score analysis is to estimate the individual scores, and there are various ways to do this in practice, the most common

N. Lavrač et al. (Eds.): ECML 2003, LNAI 2837, pp. 16–22, 2003.

being logisitic regression. However, other techniques, such as probit regression or discriminant analysis are also possible, as are the robust methods based on the t-family of long tailed distributions. Other possible methods include highly non-linear methods such as CART or neural nets. A critical feature of estimating propensity scores is that diagnosing the adequacy of the resulting fit is very straightforward, and in fact guides what the next steps in a full propensity score analysis should be. This diagnosing takes place without access to the outcome variables (e.g., sales, number of prescriptions) so that that objectivity of the analysis is maintained. In some cases, the conclusion of the diagnostic phase must be that inferring causality from the data set at hand is impossible without relying on heroic and implausible assumptions, and this can be very valuable information, information that is not directly available from traditional approaches.

Marketing applications from the practice of AnaBus, Inc. will also be presented. AnaBus currently has a Small Business Innovative Research Grant from the US NIH to implement essential software to allow the implementation of the full propensity score approach to estimating the effects of interventions. Other examples will also be presented if time permits, for instance, an application from the current litigation in the US on the effects of cigarette smoking (Rubin, 2002, Health Services Outcomes Research).

An extensive reference list from the author is included. These references are divided into five categories. First, general articles on inference for causal effects not having a focus on matching or propensity scores. Second, articles that focus on matching methods before the formulation of propensity score methods – some of these would now be characterized as examples of propensity score matching. Third, articles that address propensity score methods explicitly, either theoretically or through applications. Fourth, articles that document, by analysis and/or by simlulation, the superiority of propensity-based methods, especially when used in combination with model-based adjustments, over model-based methods alone. And fifth, introductions and reviews of propensity score methods. The easiest place for a reader to start is with the last collection of articles.

Such a reference list is obviously very idiosyncratic and is not meant to imply that only the author has done good work in this area. Paul Rosenbaum, for example, has been an extremely active and creative contributor for many years, and his text book "Observational Studies" is truly excellent. As another example, Rajeev Deheija and Sadek Wahba's 1999 article in the Journal of the American Statistical Association had been very influential, especially in economics.

References

General Causal Inference Papers

(1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." Journal of Educational Psychology, 66, 5, pp. 688-701.

(1977). "Assignment to Treatment Group on the Basis of a Covariate." Journal of Educational Statistics, 2, 1, pp. 1-26. Printer's correction note 3, p. 384.

(1977). "Assignment to Treatment Group on the Basis of a Covariate." Journal of Educational Statistics, 2, 1, pp. 1-26. Printer's correction note 3, p. 384.

(1978). "Bayesian Inference for Causal Effects: The Role of Randomization." The Annals of Statistics, 7, 1, pp. 34-58.

(1983). "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome." The Journal of the Royal Statistical Society, Series B, 45, 2, pp. 212-218. (With P.R. Rosenbaum).

(1983). "On Lord's Paradox." Principles of Modern Psychological Measurement: A Festschrift for Frederick Lord, Wainer and Messick (eds.). Erlbaum, pp. 3-25. (With P.W. Holland).

(1984). "Estimating the Effects Caused by Treatments." Discussion of "On the Nature and Discovery of Structure" by Pratt and Schlaifer. Journal of the American Statistical Association, 79, pp. 26-28. (With P.R. Rosenbaum).

(1984). "William G. Cochran's Contributions to the Design, Analysis, and Evaluation of Observational Studies." W.G. Cochran's Impact on Statistics, Rao and Sedransk (eds.). New York: Wiley, pp. 37-69.

(1986). "Which Ifs Have Causal Answers?" Discussion of Holland's "Statistics and Causal Inference." Journal of the American Statistical Association, 81, pp. 961-962.

(1988). "Causal Inference in Retrospective Studies." Evaluation Review, pp. 203-231. (With P.W. Holland).

(1990). "Formal Modes of Statistical Inference for Causal Effects." Journal of Statistical Planning and Inference, 25, pp. 279-292.

(1990). "Neyman (1923) and Causal Inference in Experiments and Observational Studies." Statistical Science, 5, 4, pp. 472-480.

(1991). "Dose-Response Estimands: A Comment on Efron and Feldman." Journal of the American Statistical Association, 86, 413, pp. 22-24.

(1994). "Intention-to-Treat Analysis and the Goals of Clinical Trials." Clinical Pharmacology and Therapeutics, 87, 1, pp. 6-15. (With L.B. Sheiner).

(1996). "Identification of Causal Effects Using Instrumental Variables." Journal of the American Statistical Association, 91, 434, as Applications Invited Discussion Article with discussion and rejoinder, pp. 444-472. (With J.D. Angrist and G.W. Imbens).

(1997). "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance." The Annals of Statistics, 25, 1, pp. 305-327. (With G. Imbens).

(1997). "Estimating Outcome Distributions for Compliers in Instrumental Variables Models." Review of Economic Studies, 64, pp. 555-574. (With G. Imbens).

(1998). "More Powerful Randomization-Based p-values in Double-Blind Trials with Noncompliance." Statistics in Medicine, 17, pp. 371-385, with discussion by D.R. Cox, pp. 387-389.

(1999). "Addressing Complications of Intention-To-Treat Analysis in the Combined Presence of All-or-None Treatment-Noncompliance and Subsequent Missing Outcomes." Biometrika, 86, 2, pp. 366-379. (With C. Frangakis).

(1999). "Causal Inquiry in Longitudinal Observational Studies," Discussion of 'Estimation of the Causal Effect of a Time-varying Exposure on the Marginal Mean of a Repeated Binary Outcome' by J. Robins, S. Greenland and F-C. Hu. Journal of the American Statistical Association, 94, 447, pp. 702-703. (With C.E. Frangakis).

(1999). "Teaching Causal Inference in Experiments and Observational Studies." Proceedings of the Section on Statistical Education of the American Statistical Association, pp. 126-131.

(2000). "Statistical Issues in the Estimation of the Causal Effects of Smoking Due to the Conduct of the Tobacco Industry." Chapter 16 in Statistical Science in the Courtroom, J. Gastwirth (ed.). New York: Springer-Verlag, pp. 321-351.

(2000). "The Utility of Counterfactuals for Causal Inference." Comment on A.P. Dawid, "Causal Inference Without Counterfactuals". Journal of the American Statistical Association, 95, 450, pp. 435-438.

(2000). "Causal Inference in Clinical and Epidemiological Studies via Potential Outcomes: Concepts and Analytic Approaches." Annual Review of Public Health, 21, pp. 121-145. (With R.J.A. Little).

(2000). "Statistical Inference for Causal Effects in Epidemiological Studies Via Potential Outcomes." Proceedings of the XL Scientific Meeting of the Italian Statistical Society, Florence, Italy, April 26-28, 2000, pp. 419-430.

(2001). "Estimating The Causal Effects of Smoking." Statistics in Medicine, 20, pp. 1395-1414.

(2001). "Self-Experimentation for Causal Effects." Comment on 'Surprises From Self-Experimentation: Sleep, Mood, and Weight', by S. Roberts. Chance, 14, 2, pp. 16-17.

(2001). "Estimating the Effect of Unearned Income on Labor Supply, Earnings, Savings and Consumption: Evidence from a Survey of Lottery Players." American Economic Review, 19, pp. 778-794. (With G.W. Imbens and B. Sacerdote).

(2002). "Statistical Assumptions in the Estimation of the Causal Effects of Smoking Due to the Conduct of the Tobacco Industry." [CD-ROM] In Social Science Methodology in the New Millennium. Proceedings of the Fifth International Conference on Logic and Methodology (J. Blasius, J. Hox, E. de Leeuw and P. Schmidt, eds.), October 6, 2000, Cologne, Germany. Opladen, FRG: Leske + Budrich. P023003.

(2002). "School Choice in NY City: A Bayesian Analysis of an Imperfect Randomized Experiment", with discussion and rejoinder. Case Studies in Bayesian Statistics, Vol. V. New York: Springer-Verlag. C. Gatsonis, B. Carlin and A. Carriquiry (eds.), pp. 3-97. (With Barnard, J., Frangakis, C. and Hill, J.)

(2002). "Principal Stratification in Causal Inference." Biometrics, 58, 1. pp. 21-29. (With C. Frangakis).

(2002). "Clustered Encouragement Designs with Individual Noncompliance: Bayesian Inference with Randomization, and Application to Advance Directive Forms." With discussion and rejoinder, Biostatistics, 3, 2, pp. 147-177. (With C.E. Frangakis and X.-H. Zhou.)

(2002). "Discussion of 'Estimation of Intervention Effects with Noncompliance: Alternative Model Specification,' by Booil Jo". Journal of Educational and Behavioral Statistics, 27, 4, pp. 411-415. (With F. Mealli.)

(2003). "Assumptions Allowing the Estimation of Direct Causal Effects: Discussion of 'Healthy, Wealthy, and Wise? Tests for Direct Causal Paths Between Health and Socioeconomic Status' by Adams et al.'". Journal of Econometrics, 112, pp. 79-87. (With F. Mealli.)

(2003). "A Principal Stratification Approach to Broken Randomized Experiments: A Case Study of Vouchers in New York City." Journal of the American Statistical Association, 98, 462, with discussion and rejoinder. (With J. Barnard, C. Frangakis, and J. Hill.)

(2003). "Assumptions When Analyzing Randomized Experiments with Noncompliance and Missing Outcomes." To appear in Health Services Outcome Research Methodology. (With F. Mealli.)

(2003). "Hypothesis: A Single Clinical Trial Plus Causal Evidence of Effectiveness is Sufficient for Drug Approval." Clinical Pharmacology and Therapeutics, 73, pp. 481-490. (With C. Peck and L.B. Sheiner.)

(2003). "Teaching Statistical Inference for Causal Effects in Experiments and Observational Studies." To appear in The Journal of Educational and Behavioral Statistics.

Matching Methods, Pre-propensity Score Paper

(1973). "Matching to Remove Bias in Observational Studies." Biometrics, 29, 1, pp. 159-183. Printer's correction note 30, p. 728.

(1976). "Multivariate Matching Methods that are Equal Percent Bias Reducing, I: Some Examples." Biometrics, 32, 1, pp. 109-120. Printer's correction note p. 955.

(1976). "Multivariate Matching Methods that are Equal Percent Bias Reducing, II: Maximums on Bias Reduction for Fixed Sample Sizes." Biometrics, 32, 1, pp. 121-132. Printer's correction note p. 955.

(1980). "Bias Reduction Using Mahalanobis' Metric Matching." Biometrics, 36, 2, pp. 295-298. Printer's Correction p. 296 ((5,10) = 75%).

Propensity Score Techniques and Applications

(1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects." Biometrika, 70, pp. 41-55. (With P. Rosenbaum).

(1984). "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." Journal of the American Statistical Association, 79, pp. 516-524. (with P.R. Rosenbaum).

(1985). "The Use of Propensity Scores in Applied Bayesian Inference." Bayesian Statistics, 2, Bernardo, DeGroot, Lindley and Smith (eds.). North Holland, pp. 463-472.

(1985). "Constructing a Control Group Using Multivariate Matched Sampling Incorporating the Propensity Score." The American Statistician, 39, pp. 33-38. (With P.R. Rosenbaum).

(1985). "The Bias Due to Incomplete Matching." Biometrics, 41, pp. 103-116. (With P.R. Rosenbaum).

(1992). "Projecting from Advance Data Using Propensity Modelling". The Journal of Business and Economics Statistics, 10, 2, pp. 117-131. (With J.C. Czajka, S.M., Hirabayashi, and R.J.A. Little).

(1992). "Affinely Invariant Matching Methods with Ellipsoidal Distributions." The Annals of Statistics, 20, 2, pp. 1079-93. (With N. Thomas).

(1992). "Characterizing the Effect of Matching Using Linear Propensity Score Methods with Normal Covariates." Biometrika, 79, 4, pp. 797-809. (With N. Thomas).

(1995). "In Utero Exposure to Phenobarbital and Intelligence Deficits in Adult Men." The Journal of the American Medical Association, 274, 19, pp. 1518-1525. (With J. Reinisch, S. Sanders and E. Mortensen.)

(1996). "Matching Using Estimated Propensity Scores: Relating Theory to Practice." Biometrics, 52, pp. 249-264. (With N. Thomas).

(1999). "On Estimating the Causal Effects of Do Not Resuscitate Orders." Medical Care, 37, 8, pp. 722-726. (With M. McIntosh).

(1999). "The Design of the New York School Choice Scholarship Program Evaluation". Research Designs: Inspired by the Work of Donald Campbell, L. Bickman (ed.). Thousand Oaks, CA: Sage. Chapter 7, pp. 155-180. (With J. Hill and N. Thomas).

(2000). "Estimation and Use of Propensity Scores with Incomplete Data." Journal of the American Statistical Association, 95, 451, pp. 749-759. (With R. D'Agostino, Jr.).

(2002). "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." Health Services & Outcomes Research Methodology, 2, pp. 169-188, 2001.

Matching & Regression Better than Regression Alone

(1973). "The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies." Biometrics, 29, 1, pp. 184-203.

(1973). "Controlling Bias in Observational Studies: A Review." Sankhya - A, 35, 4, pp. 417-446. (With W.G. Cochran).

(1979). "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies." The Journal of the American Statistical Association, 74, 366, pp. 318-328.

(2000). "Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates." Journal of the American Statistical Association, 95, 450, pp. 573-585. (With N. Thomas.)

Propensity Score Reviews

(1997). "Estimating Causal Effects From Large Data Sets Using Propensity Scores." Annals of Internal Medicine, 127, 8(II), pp. 757-763.

(1998). "Estimation from Nonrandomized Treatment Comparisons Using Subclassification on Propensity Scores." Nonrandomized Comparative Clinical Studies, U. Abel and A. Koch (eds.) Dusseldorf: Symposion Publishing, pp. 85-100.

(2003). "Estimating Treatment Effects From Nonrandomized Studies Using Subclassification on Propensity Scores." To appear in Festschrift for Ralph Rosnow. Erlbaum publishers.