

**SPR 14TH ANNUAL MEETING
PRECONFERENCE WORKSHOP #3
May 30, 2006
8:30 AM – 5:00 PM**

Title Missing-data methods for causal inference

Presenter Joseph L. Schafer
Department of Statistics and The Methodology Center
The Pennsylvania State University

Description Causal inference in nonrandomized studies and broken randomized experiments can be characterized as a problem of missing data. Consider a simple comparison of participants who received a new intervention (Treatment) and those received care as usual (Control). Each participant has two potential outcomes--the response under Treatment and the response under Control--and we may define a causal effect as the average difference between these two outcomes in the population. Of course, only one of these two potential outcomes is seen for any participant, and making inferences about an average causal effect becomes a missing-data problem. Estimates of average causal effects can be constructed by many of the same methodological tools developed in recent years for handling missing values including direct maximum likelihood and Bayesian estimation, multiple imputation, and weighted estimating equations. Special care is needed in these settings, however, due to the high rates of missing information and high degree of sensitivity to model misspecification.

Objectives This one-day workshop is designed for methodologically-oriented researchers in the social, behavioral and medical sciences in search of practical tools for causal inference from experimental and observational data.

Participants will first be introduced to the potential-outcomes framework of causal inference as espoused by Holland (1986), Rosenbaum (2002), Rubin (2005) and others, in contrast to the regression-based approaches (e.g. structural equations modeling) which have historically dominated the social sciences. This will be followed by a broad overview of the available techniques for handling missing values--weighting, imputation, likelihood-based methods, etc.--with discussion on how they apply to the problem of estimating an average causal effect. These missing-data methods will be related to other methods proposed for causal inference, including stratification and matching based on the propensity score (Rosenbaum and Rubin, 1983), weighted estimating equations (Robins, Rotnitzky and Zhao, 1994), and the new "doubly robust" methods that combine propensity scores with regression predictions (Rotnitzky, Robins & Scharfstein, 1998; van der Laan & Robins, 2003). These methods will be compared and contrasted with respect to the underlying assumptions, performance characteristics (bias, efficiency, robustness) availability of software and ease of implementation.

#####

Gelman, A. and Meng, X.L. (Eds.) (2004) Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives. New York: Wiley.

Holland, P.W. (1986) Statistics and causal inference. Journal of the American Statistical Association, 81, 945--970.

Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. Journal of the American Statistical Association, 89, 846-866.

Rosenbaum, P.R. (2002) Observational Studies, Second edition, New York: Springer.

Rosenbaum, P.R. and Rubin, D.B. (1983) The central role of the propensity score in observational studies for causal effects. Biometrika, 70, 41-55.

Rotnitzky, A., Robins, J.M. and Scharfstein, D.O. (1998) Semiparametric regression for repeated outcomes with ignorable nonresponse. Journal of the American Statistical Association, 93, 1321-1339.

Rubin, D.B. (2005) Causal inference using potential outcomes: design, modeling, decisions. Journal of the American Statistical Association, 100, 322-331.

van der Laan, M.J. and Robins, J.M. (2003) Unified Methods for Censored Longitudinal Data and Causality. New York: Springer.