Methods to address correlated exposure and outcome error for failure time outcomes

Pamela Shaw

Associate Professor of Biostatistics
E-mail: shawp@pennmedicine.upenn.edu

Abstract: Electronic health records (EHR) data are increasingly used in medical research, but these data are often subject to measurement error. These errors, if not addressed, can potentially bias results in association analyses. Methodology to address covariate measurement error has been well developed; however, methods to address errors in time-to-event outcomes are relatively underdeveloped. We will consider methods to address errors in both the covariate and time-to-event outcome that are potentially correlated. We develop an extension to the popular regression calibration method for this setting. Regression calibration has been shown to perform well for settings with covariate measurement error, but it is known that this method is generally biased for nonlinear regression models, such as the Cox model for time-to-event outcomes. Thus, we additionally propose raking estimators. Raking is a standard method in survey sampling that makes use of auxiliary information on the population to improve upon the simple Horvitz-Thompson estimator applied to a subset of data (e.g. the validation subset). Raking estimators are consistent when based on a consistent estimating equation for the validation subset. We demonstrate through numerical studies the relative performance of the regression calibration and raking estimators. We will discuss the choice of the auxiliary variable and aspects of the underlying estimation problem that affect the degree of improvement that the raking estimator will have over the simpler, biased regression calibration approach. We consider the relative performance under varying levels of signal, covariance, and censoring. We further illustrate the methods with a real data example using observational EHR data on HIV outcomes from the Vanderbilt Comprehensive Care Clinic.