Advance causal inference analysis in R: A software Review

Abstract

It is increasingly common to design observational studies to estimate causal effects. With more longitudinal data sets available to researchers, it is an efficient and viable design over the traditional randomized controlled trial to assess long-term efficacy and safety. Well-developed and well-documented statistical software have been thought to reduce the barrier in the use of statistical methods in medical and public health research. The purpose of this study is to conduct a software review for time-dependent causal inference methods, marginal structure models, g-computation and longitudinal targeted maximum likelihood estimation. Specifically, we identified three packages in R programming language for this review, *ipw, gfoRumla,* and *ltmle*. We simulated three two-visit time-dependent treatment data sets with a continuous outcome, a binary outcome, and a continuous outcome with right censoring. All three packages were applied to the three simulated data sets to estimate the average treatment effect, and the key features of each package were assessed. The focus of this review was specifically on i) required input data format, ii) statistical outputs, iii) accommodated models and data. We provided suggestions and comments on the implementation of each package, and general recommendations for software development of advanced statistical methods.