Abstract draft

Maximum 350-word limit.

The abstract should include the purpose of the research; data science methods; any results and conclusions (or results expected by the time of presentation).

Package review for longitudinal casual inference studies in R

In recent years, there has been an increased interest in estimating causal effects in observational studies with time-dependent variables. However, implementing causal inference using longitudinal data remains challenging due to the complexity of the advanced statistical methods for interested researchers coming from a variety of backgrounds. The purpose of this study is to provide a review of the features and functionalities of three software packages developed for longitudinal causal inference in *R*programming language: *ipw, gfoRumla,* and *ltmle*. We start with a brief introduction to the method of focus of the three packages, which are g-computation, inverse probability weights and longitudinal targeted maximum likelihood estimation, respectively. We simulated two longitudinal data sets with continuous and binary outcomes, respectively. Simultaneously, another data set with censored continuous outcome was simulated. We apply each package of interest to the data sets, aiming to estimate the treatment effect for data with the continuous outcome and the odds ratio for data with binary outcome. In the context of the simulated data set, we explore the key features of each package and highlight their specialties as well as limitations. We focused specifically on i) required input data format, ii) statistical outputs, iii) accommodated models and data. If the package does not allow for the desired calculation, we provide an alternative tutorial to demonstrate the manual calculations that cannot be achieved using the package alone. We conclude by making suggestions for package application in longitudinal causal inference studies and future package development for researchers.

It’s increasingly common to design observational studies to estimate causal effects. It’s an efficient and viable design over traditional randomized controlled trial to assess long term efficacy and safety. Well-developed and well-documented statistical software has 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*. For this review, 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 key features of each package were assessed. We focused 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.

Advance causal inference analysis in R: A software Review

Estimation of causal effects using observational data continues to grow in popularity in many fields. In recent years, advanced causal estimation methods such as the targeted maximum likelihood estimation, have been developed to handle complex study design and data. However, these newer methods have been out of reach for practitioners due to their complexity and the difficulty in implementation. In this project, the student and supervisor will complete a review of three statistical software packages for implementing advanced causal inference methods in R programming language: tmle, ltmle, and gfoRmula. We will apply each package to a pediatric rheumatology data and review key features of each package. We will focus specifically on i) support references and tutorial papers, ii) required input data format, iii) numeric and graphic outputs, iv) model diagnostic and fitting features, and (v) a list of accommodated statistical models and data. Working together with the supervisor, the student will draft a report documenting the review results and provide recommendations to guide practitioners in choosing an appropriate package based on the planned causal analysis.