Estimating and testing vaccine sieve effects using machine learning

Benkeser, David; Gilbert, Peter; Carone, Marco

JASA ACS Reproducibility Initiative - Author Contributions Checklist Form

Data

Abstract

The data included with this submission includes a simulated version both vaccine efficacy trials. The data are similar in structure to the real data used in the analysis.

Availability

Due to existing privacy policies, it is difficult to provide the data used for these analyses. However, it is theoretically possible to gain access to the RV144 data by applying for it through a concept proposal. We encourage interested readers to contact the authors for more information.

Description

The simulated data can be found as part of an R package that can be downloaded from GitHub (https://github.com/benkeser/survtmle) or CRAN (https://CRAN.R-project.org/package=survtmle). The data are available in .RData format and documentation for the data is available as part of the distributed R package documentation.

Code

Abstract

Code to implement the methods developed in the paper is available in an R package that can be downloaded from GitHub (https://github.com/benkeser/survtmle) or CRAN (https://CRAN.R-project.org/package=survtmle). Additional code used to perform the analysis is included in a separate GitHub repository (https://github.com/benkeser/sieveml). We have also included the code used to perform the simulation study.

Description

The code used to implement both analyses has been made available through an R package and code tutorial on GitHub under an MIT License. The tutorial presents simplified versions of the analyses performed in the paper for the sake of computational time. However, the tutorial also includes the additional functions that were used to perform the full analysis if users are interested. The full analysis of the RTS,S/AS01 data was extremely computationally intensive and was implemented using a parallelized shell script that is specific to the host institution's computing platform. Please contact the authors for this code.

The code to execute the simulation study and summarize the results is also included. The full simulation study is fairly computationally intensive, but for simplicity, we included a sequential execution of the simulation code. We also have a parallelized version, which was run to obtain the results in the manuscript, available upon request.

Instructions for Use

Reproducibility

Using simulated data, the GitHub code tutorial reproduces the results of the RV144 sieve analysis that are included in the text, as well as Figure 1 from the analysis. The default implementation in the GitHub tutorial uses a simplified Super Learner library, but also includes details on how the full library could be implemented. The tutorial also illustrates the code used to obtain the multiply outputed estimates for the RTS,S/AS01 analysis using 10 simulated data sets. The code analyzes these data sets one by one and illustrates the code that was used to combine the data sets to obtain the multiply outputed estimates.

Replication

The R package documentation includes several other examples for implementing the code using simulated data.