Chart

Description automatically generated

How to use the app?

1. Convert your data set as a CSV file. Make sure to name the outcome as Y and the binary treatment as A
2. Make sure you installed the package “Shiny” in R
3. In R session, run shiny::runApp(PATH), where PATH is the path of the folder
   1. Eg: shiny::runApp("~/Desktop/HuangGroup/cvtmle\_plasmode/Code/REFINE2")
4. Set the “Path of the data” to your data path
5. Set 4 models and other parameters
   1. Models should in format of R formula: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/formula.html
6. Click the “Run” button

What does this do?

In this study, we found relative performance of efficient estimators to differ substantially from previous simulation studies and we attribute this to the structure of our specific molecular epidemiologic data at hand. As noted previously (Balzer et al) it is unlikely that universal recommendations can be made as to which estimators will best balance efficiency and robustness to finite-sample properties in every instance. Instead, estimators should be tested against the analyst’s specific data and estimation goals. To this end, we developed an offline application that enables analyst to easily input their data, generate plasmode simulations of known structure and effect size, and compare the performance of different estimators and libraries. The current version of the Realistic Evaluations of Finite sample INference using Efficient Estimators (REFINE2) application is freely available to download at: XXXX and currently supports estimation of ATEs using the estimators and learning libraries implemented in this study. Future updates will include additional estimators, customizable SuperLearner libraries, additional estimands (e.g. risk / rate ratios), and imputation approaches for missing data.

This tool is mainly used to check the appropriateness of the effect estimation on the average treatment effect (ATE). After one obtains an estimate, the program simulates multiple datasets using the linear model with the estimated parameter. We then use the same model to perform estimation on the simulated datasets. If it gives too much bias or less than the nominal coverage, then it might not be a good method given the covariate structure presented in the dataset.

Q&A:

1. How the plasmode are run when you open the program (re-estimated every time you hit run? So there may be different estimates each time).
   1. There is a place we can set seed. If we fix the seed then the result will change everytime we click run. Right now the seed is hard-coded.
2. How the (true) fixed effect size is determined.
   1. The true fixed effect size is determined by performing the selected estimation method using the full data.