This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

Part 1: Data

☐ This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Abstract

We analyze two datasets: **CD4 Data** from the Multi-Center AIDS Cohort Study, containing 2,376 CD4+ cell count measurements from 369 HIV-infected men over 8.5 years, with 1–12 measurements per individual, resulting in a highly unbalanced dataset. **Serum Cholesterol Data** from the National Cooperative Gallstone Study includes patients with floating gallstones in high-dose and placebo groups, with cholesterol measured at baseline and follow-ups. Many measurements are missing due to missed visits, inadequate lab specimens, or early termination.

Availability

☒ Data **are** publicly available.
☐ Data **cannot be made** publicly available.

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available data* section, below.

Publicly available data

☒ Data are available online at:

- **CD4 Data**: https://search.r-project.org/CRAN/refmans/lqmix/html/cd4.html
- **Serum cholesterol data**: https://content.sph.harvard.edu/fitzmaur/ala/cholesterol.txt

☐ Data are available as part of the paper's supplementary material.

☐ Data are publicly available by request, following the process described here:

☐ Data are or will be made available through some other mechanism, described here:

Non-publicly available data

Description

File format(s)

☒ CSV or other plain text.
☒ Software-specific binary format (.Rda, Python pickle, etc.): Rda
☐ Standardized binary format (e.g., netCDF, HDF5, etc.):
☐ Other (please specify):

Data dictionary

- ☐ Provided by authors in the following file(s):
- ☐ Data file(s) is(are) self-describing (e.g., netCDF files)
- ☒ Available at the following URL:
  - **CD4 Data**: https://search.r-project.org/CRAN/refmans/lqmix/html/cd4.html
  - **Serum cholesterol data**: https://content.sph.harvard.edu/fitzmaur/ala/cholesterol.txt

Additional Information (optional)

Part 2: Code

Abstract

Our code investigates the performance of the HCP method across various aspects, including **marginal coverage** (`HCP_marginal.R`), **conditional coverage** (`HCP_conditional.R`), **local coverage** (`HCP_local.R`.), and **simultaneous prediction** (`HCP_simul.R`). These codes are located in different folders for tables and figures. We provide the code for all tables and figures in the paper, each with a corresponding folder containing a `main.R` file that calls all necessary scripts and can be run directly to reproduce the results.

Description

Code format(s)

- ☒ Script files
  - ☒ R
  - ☐ Python
  - ☐ Matlab
  - ☐ Other:
- ☒ Package
  - ☒ R
  - ☐ Python
  - ☐ MATLAB toolbox
  - ☐ Other:
- ☒ Reproducible report
  - ☒ R Markdown
  - ☐ Jupyter notebook
  - ☐ Other:
- ☐ Shell script
- ☐ Other (please specify):

Supporting software requirements

Version of primary software used R 4.4.2 (Big Sur ARM build, GUI 1.81, Build 8462)

Libraries and dependencies used by the code

- MASS version 7.3.61
- stats version 4.4.2
- grf version 2.4.0
- quantreg version 5.99.1
- doParallel version 1.0.17
- doRNG version 1.8.6.1
- lme4 version 1.1.36
- merTools version 0.6.2
- randomForest version 4.7.1.2
- rstudioapi version 0.17.1

Supporting system/hardware requirements (optional)

Parallelization used

☐ No parallel code used
☒ Multi-core parallelization on a single machine/node
  – Number of cores used: 10
☐ Multi-machine/multi-node parallelization
  – Number of nodes and cores used:

License

☒ MIT License (default)
☐ BSD
☐ GPL v3.0
☐ Creative Commons
☐ Other: (please specify)

Additional information (optional)

Part 3: Reproducibility workflow

Scope

The provided workflow reproduces:

☐ Any numbers provided in text in the paper
☒ The computational method(s) presented in the paper (i.e., code is provided that implements the method(s))
☒ All tables and figures in the paper
☐ Selected tables and figures in the paper, as explained and justified below:

Workflow

Location

The workflow is available:

- ☐ As part of the paper's supplementary material.
- ☒ In this Git repository: https://github.com/judywangstat/HCP
- ☐ Other (please specify):

Format(s)

- ☐ Single master code file
- ☒ Wrapper (shell) script(s)
- ☐ Self-contained R Markdown file, Jupyter notebook, or other literate programming approach
- ☐ Text file (e.g., a readme-style file) that documents workflow
- ☐ Makefile
- ☐ Other (more detail in *Instructions* below)

Instructions

Directly access the Git webpage: https://github.com/judywangstat/HCP, which contains all the folders corresponding to the tables and figures in the paper. Each table and figure has its own folder, which includes all the necessary scripts and a `main.R` file to call them. Download the files and install the required R packages. Our program can **automatically set the paths** in the RStudio environment, so you do not need to manually configure them. However, the original folder structure must remain unchanged. Once the **paths are correctly set**, you can simply click on `main.R` to run the program directly. For a single master code file implementing the proposed method, simply run `CD4_HCP.R` in Table 3 folder.

Expected run-time

Approximate time needed to reproduce the analyses on a standard desktop machine:

- ☐ $<$ 1 minute
- ☐ 1-10 minutes
- ☒ 10-60 minutes
- ☐ 1-8 hours
- ☐ $>$ 8 hours
- ☐ Not feasible to run on a desktop machine, as described here:

Additional information (optional)

Notes (optional)