

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 provide data and R code.

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:
- ☐ Data are available as part of the paper's supplementary material.
- ☒ Data are publicly available by request, following the process described here:

Please contact yangfan1987@tsinghua.edu.cn for the data we used in this paper.

- ☐ 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.): pckle
- ☐ Standardized binary format (e.g., netCDF, HDF5, etc.):
- ☐ Other (please specify):

Data dictionary

- ☒ Provided by authors in the following file(s): CSV
- ☐ Data file(s) is(are) self-describing (e.g., netCDF files)
- ☐ Available at the following URL:

Additional Information (optional)

Part 2: Code

Abstract

We provide R code for simulation and data analysis.

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 4.1.3

Libraries and dependencies used by the code mice (3.14.0), dplyr (1.0.8), parallel (4.1.3), xlsx (0.6.5), ggplot2 (3.3.5), nnet (7.3-17), miscF (0.1-5), boot (1.3-28), MASS (7.3-55), scales (1.2.1), table1 (1.4.2)

Supporting system/hardware requirements (optional)

MacOS monterey Version 12.1 Apple M1 Memory 8 GB

x86_64-apple-darwin17.0 (64-bit)

Parallelization used

- ☐ No parallel code used
- ☒ Multi-core parallelization on a single machine/node
 - Number of cores used: 8
- ☐ 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/sushi133/Mediation-analysis-with-the-mediator-and-outcome-missing-not-at-random>
- ☐ 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

Please simply run the R code we provide.

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)

Computation time for each R code is documented in the `Compute_time(hours).xlsx`.

Notes (optional)