

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

The Mentoring dataset is a randomized controlled trial conducted on German adolescents. It contains baseline covariates of the following categories: demographics (age, gender, migrant status), home environment (books at home, parental support), academic grades (math, German, English), personality (Big Five scales), and socioeconomic status. Treatment is a binary indicator for assignment to a mentoring intervention. A standardized labor market outcome was measured for each child.

## 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: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/IP98QW>
- ☒ 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.): pkle
- ☐ 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: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/IP98QW>

## Additional Information (optional)

Data can be accessed through the Harvard Dataverse, but for convenience is also attached in the data subfolder.

## Part 2: Code

### Abstract

The code is organized into three parts: `R/replication_functions.R` defines all relevant functions for analysis, then, two analysis scripts (`scripts/generate_figures.R` and `scripts/generate_tables.R`) call those functions to produce all figures and tables. The master script (`scripts/run_analysis.R`) activates the `renv` environment, creates the necessary output directories, and runs the figure and table scripts.

### 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 version 4.4.2

## Libraries and dependencies used by the code

Package	Version
KernSmooth	2.23-24
MASS	7.3-61
Matrix	1.7-1
R6	2.6.1
RColorBrewer	1.1-3
ROCR	1.0-11
Rcpp	1.0.14
RcppEigen	0.3.4.0.2
SuperLearner	2.0-29
abind	1.4-8
bit	4.6.0
bit64	4.6.0-1
bitops	1.0-9
caTools	1.18.3
cli	3.6.5
clipr	0.8.0
codetools	0.2-20
cpp11	0.5.2
crayon	1.5.3
cvAUC	1.1.4
data.table	1.17.4
dirmult	0.1.3-5
doParallel	1.0.17
dplyr	1.1.4
farver	2.1.2
fastDummies	1.7.5
forcats	1.0.0
foreach	1.5.2
gam	1.22-5
gbm	2.2.2
generics	0.1.4
ggplot2	3.5.2
glmnet	4.1-9
glue	1.8.0
gplots	3.2.0
gtable	0.3.6
gtools	3.9.5
haven	2.5.5
here	1.0.1
hms	1.1.3
isoband	0.2.7
iterators	1.0.14
labeling	0.4.3
lattice	0.22-6
lifecycle	1.0.4
magrittr	2.0.3
mgcv	1.9-1

nlme	3.1-166
nnls	1.6
pillar	1.10.2
pkgconfig	2.0.3
plyr	1.8.9
pracma	2.4.4
prettyunits	1.2.0
progress	1.2.3
ranger	0.17.0
readr	2.1.5
renv	1.1.4
reshape2	1.4.4
rlang	1.1.6
rprojroot	2.0.4
scales	1.4.0
shape	1.4.6.1
stringi	1.8.7
stringr	1.5.1
survival	3.7-0
tibble	3.3.0
tidyselect	1.2.1
tzdb	0.5.0
utf8	1.2.6
vetrs	0.6.5
viridisLite	0.4.2
vroom	1.6.5
withr	3.0.2

---

## Supporting system/hardware requirements (optional)

### Parallelization used

- ☐ No parallel code used
- ☒ Multi-core parallelization on a single machine/node
  - Number of cores used: 15
- ☐ 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://anonymous.4open.science/r/rerandomization-quadratic-forms-3DAC>
- ☐ 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

To reproduce the full analysis, including Figures 1–3 and Table 2:

#### 1. Download and unzip the anonymized repository

```
https://anonymous.4open.science/r/rerandomization-quadratic-forms-3DAC
```

#### 2. Restore R environment

```
Rscript -e "install.packages('renv', repos='https://cloud.r-project.org')"  
Rscript -e "renv::restore(repos='https://cloud.r-project.org')"
```

#### 3. Run master script

```
Rscript scripts/run_analysis.R
```

### 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: It is possible to run on a desktop machine, but it will take several days. Analysis from the paper was run in batches.

### Additional information (optional)

None.

### Notes (optional)

None.