

## Data

### Abstract (Mandatory)

We use the following three datasets in our analysis:

- 1) The GWAS summary statistics of reaction time (Davies et al., 2018, <https://www.ncbi.nlm.nih.gov/pubmed/29844566>);
- 2) The GWAS summary statistics of total brain volume (TBV) (Zhao et al., 2019, <https://pubmed.ncbi.nlm.nih.gov/31676860/>); and
- 3) The individual-level single nucleotide polymorphism (SNP), brain imaging, and covariates (age, gender) data in the Pediatric Imaging, Neurocognition, and Genetics (PING) study (Jernigan et al., 2016, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4628902/>).

### Availability (Mandatory)

1) The GWAS summary statistics of reaction time (Davies et al., 2018) are publicly available at <https://www.ccace.ed.ac.uk/node/335>.

2) The GWAS summary statistics of TBV (Zhao et al., 2019) are publicly available at <https://bigkp.web.unc.edu/>

3) The individual-level SNP, brain imaging, and covariates (age, gender) data in the PING study have been made publicly available at <http://pingstudy.ucsd.edu>. However, we do not have the permission to further release their individual-level data to the public. Here is more information about how to access the PING data (Jernigan et al., 2016):

*“Access to the PING Data Resource is available through an online web interface at <http://pingstudy.ucsd.edu>. Here, information about the study, the consortium, and the methods are available for browsing, and instructions are given for applying for approval to explore, download, or request bulk shipping of data (for a fee covering media and shipping charges). Full sharing of all data is not permitted by the IRB for some PING sites. This includes restrictions on the sharing of some raw image data on the NITRC site and restrictions on sharing of some genetics data. All sharable data are available through the PING Data Portal to any researcher who holds a position in a research institution and is at least at the postdoctoral level (upon assent to the PING Data Use Agreement and approval of a brief data use description). Students can gain access to data if sponsored by eligible supervising researchers who agree to supervise the students’ compliance with the data use agreement. Raw image data for a subset of the participants is available through NITRC after an account is approved through the PING Portal.”*

**Description (Mandatory if data available)**

- Author permissions:  
The GWAS summary statistics (Davies et al., 2018 and Zhao et al., 2019) are publicly available. No additional permissions from the present authors are required to access the data.
- Licensing information or terms of use:  
1) The terms of use for GWAS summary statistics of reaction time (Davies et al., 2018) are  
  
*“When reporting results of research that utilizes CCACE data we request that you cite the original publication. By downloading these data, you acknowledge that they will be used for research purposes and that you are in compliance with applicable rules, policies and regulations. To protect subject confidentiality, we are not releasing sample allele frequencies.”*  
  
2) The GWAS summary statistics of TBV (Zhao et al., 2019) are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).
- Link to data:  
1) The GWAS summary statistics of reaction time (Davies et al., 2018) are publicly available at <https://www.ccace.ed.ac.uk/node/335>.  
The file name is “Davies\_NC\_2018\_OPEN\_DATASET.zip” on that website.  
  
2) The GWAS summary statistics of TBV (Zhao et al., 2019) are publicly available at <https://www.med.unc.edu/bigs2/data/gwas-summary-statistics/>  
The download link can be found in “gwas\_ROI\_v20190524.xlsx” of that website, and the ID is “pheno101”.
- Data provenance, including identifier or link to original data if different than above: N/A.
- File format: zipped txt files.
- Metadata (including data dictionary): N/A.
- Version information: N/A for GWAS summary statistics of reaction time (Davies et al., 2018); v20190524 for GWAS summary statistics of TBV (Zhao et al., 2019).

## Code

### Abstract (Mandatory)

We provide the codes for the simulation studies (Figures 1 and 2), and the Linux commands and codes to run the real data analysis (Figure 3). We also implement the bias-corrected estimators in R package bcPRS. The package has been uploaded with submission and can be accessed at <https://github.com/xm1701/bcPRS>.

### Description (Mandatory)

- How delivered

The R codes for the simulation studies are uploaded with the submission. They can be used to reproduce our simulation results in Figures 1 and 2.

For real data analysis in Figure 3, we use the publicly available toolset Plink (Purcell et al., 2007, doi: 10.1086/519795) to generate the polygenic risk scores (PRS), and then perform follow-up analysis in R. We provide the Linux commands to reproduce the PRS with Plink, and the R codes for the follow-up analysis.

In addition, all these codes have been made publicly available at [https://github.com/xm1701/paper\\_figures](https://github.com/xm1701/paper_figures).

- Licensing information

The Plink tool set is licensed under the GNU General Public License (<https://www.gnu.org/licenses/gpl-3.0.en.html>).

R is a free software and the licenses can be found at <https://www.r-project.org/Licenses/>.

- Link to code/repository

The codes and commands have been uploaded with submission, and they are also publicly available at [https://github.com/xm1701/paper\\_figures](https://github.com/xm1701/paper_figures) and <https://github.com/xm1701/bcPRS>.

The Plink toolset can be downloaded at <https://www.cog-genomics.org/plink/1.9/>.

R can be downloaded at <https://www.r-project.org/>.

- Version information: Plink v1.90 beta; R version 3.5.0.

- Supporting software requirements:

Library (MASS) and Library (MatrixEQTL) (version 2.2) for the R codes in simulation studies ([http://www.bios.unc.edu/research/genomic\\_software/Matrix\\_eQTL/](http://www.bios.unc.edu/research/genomic_software/Matrix_eQTL/)).

# Instructions for Use

## Reproducibility (Mandatory)

- What is to be reproduced: Figures 1-3 in the main text of the paper.
- How to reproduce analyses:  
Figure 1 can be reproduced with the file “Figure1\_v2.R”;  
Figure 2 can be reproduced with the file “Figure2\_v2.R”; and  
Figure 3 can be reproduced with files “Figure3\_1\_v2.txt” and “Figure3\_2\_v2.R”;
- Expected run-time of the workflow  
  
Figure 1: one replicate takes about three minutes on a MacBook Pro with 2.9 GHz Intel Core i7 processor, and 16 GB memory. We run 200 replicates in our simulation.  
  
Figure 2: one replicate takes about three minutes on a MacBook Pro with 2.9 GHz Intel Core i7 processor, and 16 GB memory. We run 200 replicates in our simulation.  
  
Figure 3: We run the real data analysis on Linux computing cluster, the total time of analysis is about one hour.

## Notes

The individual-level SNP, brain imaging, and covariates (age, gender) data in the PING study have been made publicly available at <http://pingstudy.ucsd.edu>. However, we do not have the permission to further release their data to the public. Researchers can directly apply data from the PING study research team, see Availability section above for details.

The GWAS summary statistics of reaction time and TBV are publicly available and can be freely downloaded from their resources.