A new Mendelian Randomization method to estimate causal effects of multivariable brain imaging exposures

Chen Mo, Zhenyao Ye, Hongjie Ke, Tong Lu, Travis Canida, Song Liu, Qiong Wu, Zhiwei Zhao, Yizhou Ma, L. Elliot Hong, Peter Kochunov, Tianzhou Ma, Shuo Chen

Division of Biostatistics and Bioinformatics and Maryland Psychiatric Research Center, School of Medicine, University of Maryland

INTRODUCTION

The advent of simultaneously collected imaging-genetics data in large study cohorts provides an unprecedented opportunity to assess the causal effect of brain imaging traits on externally measured experimental results (e.g., cognitive tests) by treating genetic variants as instrumental variables (IVs).

However, classic Mendelian Randomization (MR) methods are limited when handling high-throughput imaging traits as exposures to identify causal effects due to the complex covariance structure of exposures and horizontal pleiotropy. We propose a new MR framework to jointly select IVs and imaging exposure variables, and then estimate the causal effect of multivariable imaging data on the outcome variable.

METHODS

Our MR analysis method consists of three main steps. (see Figure 1 for an overview of our MR pipeline).

Step 1: We screen the causal effect of each imaging exposure on the outcome via MR analysis using one IV at a time and assess the validity of IVs, following the guideline for MR investigations proposed by Burgess et al. (2019) [1]. We record the significance of MR analyses (e.g., $-logP$-value) in a matrix (W_{SM}, where S and M are the number of SNPs and imaging exposures, respectively).

Step 2: We detect a submatrix (W) from W_{SM}, which contains the most informative features, by employing greedy algorithms[2] or exhausting search algorithms [3] to maximize the information included based on $\widetilde{W}$.

Step 3: Given $\widetilde{W}$ from step 2, we adopt statistical techniques in imaging causal mediation analysis to transform the imaging exposures into a set of orthogonal variables using algorithms such as the procedures in Chén et al. (2018)[4]. Next, we can estimate the causal effect of the orthogonal imaging factors through MR analysis followed existing MR methods [1].

Evaluate the causal effect of white matter microstructure integrity on cognitive function

We apply our method to evaluate the causal effect of white matter microstructure integrity (WM) on cognitive function using data from UK Biobank.

Exposures: WM was quantified by fractional anisotropy (FA) measures derived from diffusion MRI data in 40 tracts. 31 out of 40 FAs had significant association with the outcome variable based on p-value < 0.05 adjusted by FDR) in data prepossessing and therefore, used for subsequent analyses.

Outcome: An **intelligence** g factor of cognitive function derived from 5 cognitive tests.

IVs: Genetic variants highly associated (p-value $< 5E^{-7}$) with FAs in gene VCAN, based on GWAS results after genetic data preprocessing (e.g., QC, LD clumping) [1].

We selected a submatrix from 31 FAs and 27 variants (see Figure 2). We estimated a general factor (gFA) based on 22 selected FAs. Next, we assessed the causal effect of gFA on intelligence g using the 3 selected variants as IVs in MR analysis and found significant causal effect (see Figure 2).

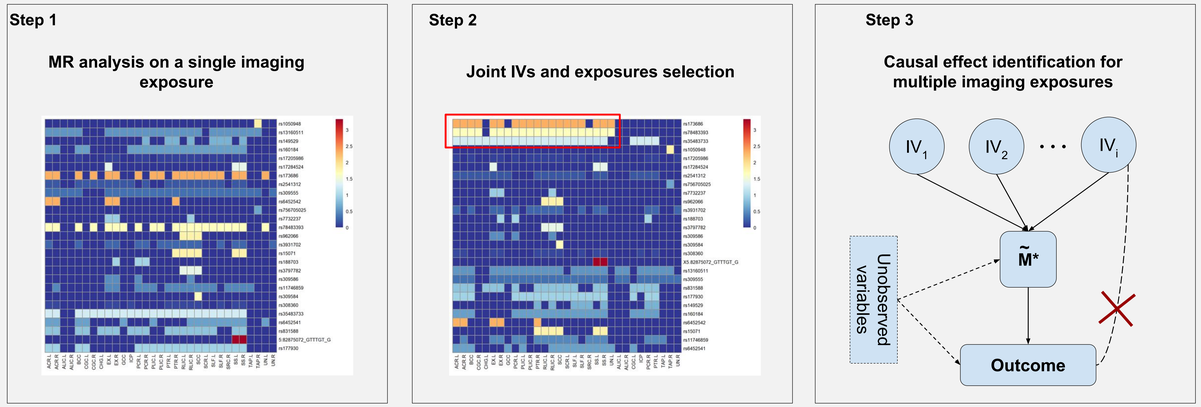

Figure 1:**Overview of DenseLD framework.** Our MR analysis method consists of three main steps. The heatmap (left) shows the raw unorganized matrix of $-logP$-value in the first analysis step, where the $-logP$-value represents the significance of "single exposure → outcome" using one single IV at a time in each MR analysis. The heatmap (middle) shows the matrix after submatrix identification in the second step. This submatrix contains the most informative features (exposure and genetic variants). The diagram (right) shows the MR analysis on the identified features with selected variants (i.e., IVs) in the last step.

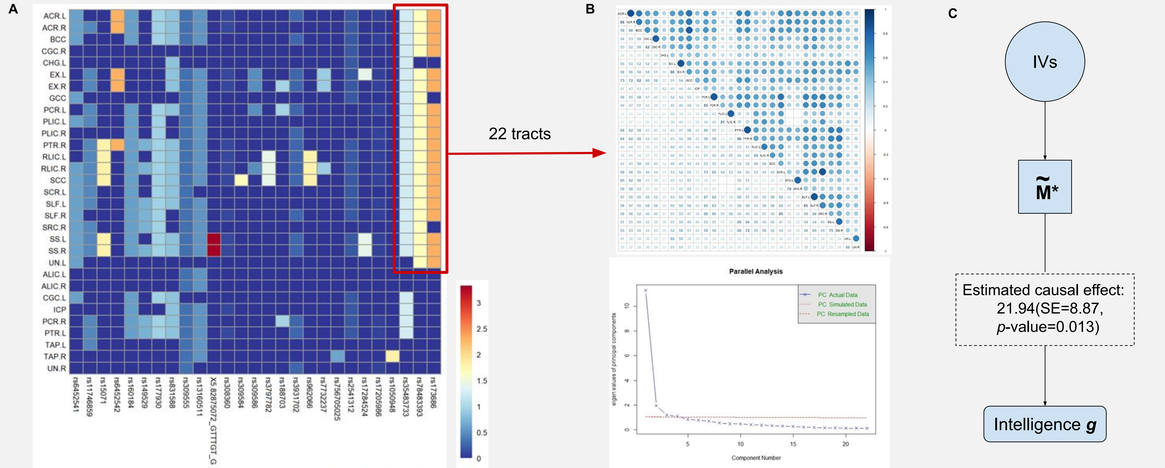

Figure 2:**Mendelian randomization analysis results of imaging exposures and cognitive function.** A shows the submatrix consisting of 22 FA tracts and 3 genetic variants (i.e., IVs) selected from 31 FAs and 27 variants, using LAS [3]. Given $-logP$-value revealed the significance of causal effect, the lowest significance was shown in dark blue whereas red indicated the highest significance. B shows the matrix of pair-wise correlation matrix of the 22 tracts along with their parallel analysis based on PCA for estimating orthogonal factors. The gFA explained 59% common variations of 22 tracts. C shows the final MR analysis results of the causal effect from gFA on intelligence g. The results indicated that WM measures across different tracts have a joint causal effect significantly impact the cognitive function among the UK Biobank participants.

More information about this study can be found at https://github.com/kehongjie/ImagingMR

Correspond to: *(Shuo Chen) shuochen@som.umaryland.edu; (Tianzhou Ma) tma0929@umd.edu*

SIMULATION STUDY

We compared the performance of our framework to two other types of MR analyses in 500 repeated simulations using $n = 500$ individuals: **a)** used a submatrix of imaging exposures (M) and IVs selected by LAS [3] in the MR model (our method); **b)** included all 40 imaging exposures in the MR model; and, **c)** simply ran 40 MR models with single exposure separately.

Simulation setting:

1). Randomly simulated genotypes $\mathbf{X}_{500\times 20}$ for 20 uncorrelated IVs, while assumed an underlying true factor of $Mf_{500\times 1} = \mathbf{X}\alpha_{20\times 1}$, where $\alpha^T = (2, \ldots, 2, 0, \ldots, 0)$ measured the effect that IVs had on exposures. Among these IVs, assumed 10 had true effect on $Mf_{500\times 1}$ whereas the rest 10 had no true effect.

2). Generated 20 imaging exposures with true casual effects on the outcome by $M_i = \mathbf{X}\alpha_i^* + \varepsilon_i^*$, $\alpha_i^* = \alpha + (\delta_{i,1}^*, \ldots, \delta_{i,20}^*)^T$ and $\varepsilon_i^* = (\varepsilon_{i,1}^*, \ldots \varepsilon_{i,500}^*)^T$. In addition, simulated another 20 imaging exposures without true casual effects on the outcome by $M_i' = \varepsilon_i'$, $\varepsilon_i' = (\varepsilon_{i,1}', \ldots \varepsilon_{i,500}')^T$. Here $\varepsilon_{i,k}^*$, $\varepsilon_{i,k}'$ and $\delta_{i,j}^*$ are all $i.i.d$ random noise with standard normal distribution, where $i, j \in \{1, \ldots, 20\}$ and $k \in \{1, \ldots, 500\}$.

3). simulated the outcome data using the true exposure factor, i.e. $Y_{500\times 1} = \beta * Mf + \varepsilon_{500\times 1}$, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_{500})^T$ is another set of standard normal random noises.

Results: Our method achieved smaller bias in estimating the causal effect compared to the method using all 40 imaging exposures (see Table 1). In terms of the selection of causal imaging exposures, our method had substantially decreased FDR while still maintaining a sensitivity closed to 1.

Table 1:**Simulation results:** mean (standard error) of 100 simulation.

Method	Bias of $\hat{\beta}$	Sensitivity	FDR
Simulation results with $\beta = 1$ (large effect size)			
Exposures selected (our method)	0.108 (0.084)	0.947 (0.075)	0.15 (0.157)
All exposures	0.924 (0.213)	1 (0)	0.5 (0)
A single exposure	-	1 (0)	0.5 (0)
Simulation results with $\beta = 0.5$ (small effect size)			
Exposures selected (our method)	0.05 (0.045)	0.945 (0.077)	0.148 (0.155)
All exposures	0.473 (0.107)	1 (0)	0.5 (0)
A single exposure	-	1 (0)	0.5 (0)

FINAL REMARKS

Compared to previous studies that only repeatedly tested the associations between WM and cognitive function, our analysis revealed a comprehensive *causal* relationship between them and opens a new avenue for imaging-genetics data analysis in causal inference. The selected imaging variables provide spatially-specific causes for the externally measured test results. The shared IVs also becomes the foundation to transform the imaging exposures to causal independent factors as the IVs are valid for any of the imaging variables.

Selected References

[1] Stephen Burgess, George Davey Smith, Neil M Davies, Frank Dudbridge, Dipender Gill, M Maria Glymour, Fernando P Hartwig, Michael V Holmes, Cosetta Minelli, Caroline L Relton, et al.
Guidelines for performing mendelian randomization investigations.
Wellcome Open Research, 4, 2019.

[2] Qiong Wu, Xiaoqi Huang, Adam Culbreth, James Waltz, L Elliot Hong, and Shuo Chen.
Extracting brain disease-related connectome subgraphs by adaptive dense subgraph discovery.
Biometrics, 2021.

[3] Andrey A Shabalin, Victor J Weigman, Charles M Perou, and Andrew B Nobel.
Finding large average submatrices in high dimensional data.
The Annals of Applied Statistics, pages 985–1012, 2009.

[4] Oliver Y Chén, Ciprian Crainiceanu, Elizabeth L Ogburn, Brian S Caffo, Tor D Wager, and Martin A Lindquist.
High-dimensional multivariate mediation with application to neuroimaging data.
Biostatistics, 19(2):121–136, 2018.