journal title, 2022, pp. 1–5

doi: DOI HERE
Advance Access Publication Date: Day Month Year
Paper

PAPER

simmr: An open-source tool to perform simulations in Mendelian Randomization

Noah Lorincz-Comi,[1,*] Yihe Yang[1] and Xiaofeng Zhu[1]

[1]Department of Population Quantitative Health Sciences, Case Western Reserve University, Street, Postcode, Cleveland, Ohio, USA
[*]Corresponding author. noahlorinczcomi@gmail.com
FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

Mendelian Randomization (MR) has become a popular tool for inferring causality of risk factors on disease. There are currently over 50 different methods available to perform MR, each supported by their respective simulations demonstrating their advantages over other methods. However, the literature currently lacks an accepted simulation environment in which researchers can fairly evaluate the performance of competing methods. This has sometimes led to conflicting and misleading simulation results from independent researchers intending to simulate the same reality. We present `simmr`, an open-source software for performing simulations to evaluate the performance of different MR methods. Researchers can directly modify the `simmr` source code with the intended purpose of eventually arriving at a universally accepted standard procedure for researchers to evaluate different MR methods.

Key words: Mendelian Randomization, bias, causal inference, instrumental variables, simulation, statistical genetics

Introduction

Mendelian Randomization is a genetic instrumental variable method that has become a popular statistical tool for identifying risk factors that cause disease [35, 34]. The validity of inferences made using MR relies heavily on the satisfaction of three primary assumptions: the genetic instruments are (i) strongly associated with the exposure(s), (i) not associated with the outcome conditional on the exposures, and (iii) not associated with any confounders of the exposure(s)-outcome relationship(s). Violation of any of these assumptions can lead to bias in causal estimation, false positive, and/or false negative findings. Three additional challenges are present in MR analyses: (a) some individuals may have been present in both the exposure(s) and outcome GWAS, (b) applying strict IV selection criteria to satisfy assumption (i) above can introduce a 'winner's curse' bias, and (c) strong LD between IVs, and its imprecise estimation using reference panels, can drastically inflate false positive rates.

Many MR methods have been introduced in the literature to address different subsets of assumptions (i)-(iii) and scenarios (a)-(c) in either the single-exposure (univariable) or multiple-exposure (multivariable) settings [1]. These methods include IVW [4], dIVW [46], pIVW [43], MR-Egger [7], MR-RAPS [49], MRAID [47], MRMix [30], MR-cML [44], MVMR-cML [22], MR-PRESSO [39], IMRP [50], MR-Median [2], MR-MaxLike [5], MR-Corr [9], MR-Robust [33], MR-Lasso [21], MR-Conmix [6], CAUSE [25], MR-CUE [11], MR-Horse [17], MR-BMA [51], MR-Robin [16], EMIC [20], MR-Mode [18], MRBEE [24], MR-Lap [26], the Wald test [28], JAM [27], MR using factor analysis [29], MRMO [13], BMRMO [12], BWMR [48], moPMR-Egger [23], sisVIVE [21], MR-LDP [10], MR-CIP [42], MR-PATH [19], MR-Clust [15], BayesMR [3], BMRE [36], MR-link [38], OMR [41], CoJo [45], MR using PCA [8], and GRAPPLE [40]. Clearly, the literature is saturated with MR estimators. In each of the studies in which each new estimator was introduced, the estimator must have had a demonstrated and meaningful improvement over existing methods as it relates to bias, variance, causal inference, flexibility, or computational efficiency. These improvements are almost always demonstrated using simulations and real data analyses. Despite the numerous simulations that have been performed in the literature of MR method development, still no standard for performing them has been introduced. Therefore, the extent to which the demonstrated improvement of new MR methods is sensitive to the precise simulation settings of independent simulations is unknown. Often independent simulations intended to mimic the same reality can produce conflicting results (e.g., [26] and [46], [22] and [37], [31] and [44]).

We introduce `simmr`, an open-source and easily accessible tool for generating data to use in MR simulations. `simmr` can be used in both the univariable and multivariable MR settings and users can simultaneously specify different properties of uncorrelated horizontal pleiotropy, correlated horizontal pleitoropy, weak instrument bias, GWAS sample overlap, winner's curse, and linkage disequilibrium between instrumental variables. With no standard set in the literature, the purpose of the publicly modifiable `simmr` software is to begin the construction of a standard simulation environment in which fair comparisons can be made between MR estimators in different settings that are encountered in practice.

Methods

Simulation models

The data-generating simulation models are based partly on the framework of [22]. Simulation data are generated under the directed acyclic graph in Figure 1 and the following models:

$$U = \mathbf{g}^\top \boldsymbol{\gamma}^C + \epsilon_U \tag{1}$$

$$\mathbf{x} = \mathbf{B}^\top \mathbf{g} + \boldsymbol{\Psi}^\top \mathbf{g}_w + \pi_{\mathbf{x}} U + \boldsymbol{\epsilon}_{\mathbf{x}} \tag{2}$$

$$Y = \mathbf{x}^\top \boldsymbol{\theta} + \pi_Y U + \mathbf{g}^\top \boldsymbol{\gamma}^U + \epsilon_Y \tag{3}$$

where U is a confounder of the relationship between exposure(s) $\mathbf{x}$ and outcome Y, $\boldsymbol{\theta}$ represents the corresponding causal effect(s), $\mathbf{B}$ and $\boldsymbol{\Psi}$ respectively are non-weak and weak associations with $\mathbf{g}$ and $\mathbf{g}_w$, and $\boldsymbol{\gamma}^C$ and $\boldsymbol{\gamma}^U$ are respectively CHP and UHP effects. Users of `simmr` can fix the variance explained in each phenotype by the others and partition the variance in $\mathbf{x}$ explained by $\mathbf{g}$ into effects from UHP, CHP, valid, and weak SNPs. This is because `simmr` uses the following distributions to generate the quantites in the above models:

$$\text{Var}(U) = 1, \quad \text{Var}(\mathbf{x}) = \mathbf{S}_P, \quad \text{Var}(x_k) = 1, \quad \text{Var}(Y) = 1, \tag{4}$$

$$\sqrt{2\psi(1-\psi)}g_j + 2\psi \sim \text{Binom}(2, \psi), \qquad j = 1, ..., M \text{ SNPs}, \tag{5}$$

$$\alpha_C^{-1}\gamma_j^C \sim \text{Uniform}(-1, 1)d_j^{CHP}, \tag{6}$$

$$\alpha_U^{-1}\gamma_j^U \sim \text{Uniform}(-1, 1)d_j^{UHP}, \tag{7}$$

$$d_j^{CHP} \sim \text{Binom}(1, p^{CHP}), \quad d_j^{UHP} \sim \text{Binom}(1, p^{UHP}), \tag{8}$$

$$\boldsymbol{\Xi} \sim \text{N}(\mathbf{1}, \mathbf{R}, \alpha_{\beta(S)}^{-1}\mathbf{S}_G)d_j^\beta + \text{N}(\mathbf{1}, \mathbf{R}, \alpha_{\beta(W)}^{-1}\mathbf{S}_G)(1 - d_j^\beta), \tag{9}$$

$$\Big[\mathbf{B}\Big]_{jk} = \Phi\Big(\Big[\boldsymbol{\Xi}\mathbf{S}_G^{-1/2}\Big]_{jk}\Big), \qquad k = 1, ..., p \text{ exposures} \tag{10}$$

$$d_j^\beta \sim \text{Binom}(1, p^\beta), \tag{11}$$

$$\alpha_\pi^{-1}\pi_Y = -1, \tag{12}$$

$$\alpha_\theta^{-1}\boldsymbol{\theta} = (1, ..., 1)^\top \tag{13}$$

where the hyperparameter ψ controls MAF; α_C and α_U respectively control variance in U and Y explained by CHP and UHP; d_j^{CHP} and d_j^{UHP} respectively indicate whether a SNP is CHP or UHP with probabilities p^{CHP} and p^{UHP}; $\alpha_{\beta(S)}$ and $\alpha_{\beta(W)}$ respectively control the variance in $\mathbf{x}$ explained by non-weak and weakly associated SNPs; d_j^β is a binary indicator used to distinguish SNPs weakly and non-weakly associated with $\mathbf{x}$ at probability p^β; α_π controls the variance in Y explained by the confounder U; $\mathbf{S}_P$ and $\mathbf{S}_G$ respectively are the phenotypic and genetic correlation matrices between the exposures; Φ is the cumulative density function of the standard normal distribution; and α_θ^{-1} controls the variance in Y explained by $\mathbf{x}$. Use of the cdf Φ is to convert the normally-distributed values in $\boldsymbol{\Xi}$ to the uniformly-distributed values in $\mathbf{B}$ while retaining the column-wise covariance structure $\mathbf{S}_G$ and row-wise covariance structure $\mathbf{R}$. `simmr` will automatically partition the SNPs causing $\mathbf{x}$ into mutually exclusive valid, UHP, and CHP groups (see Figure 1). Figure 2 shows an example of data generated under the above models alongside computation times for different combinations of GWAS sample overlap, GWAS sample sizes, and nubmers of causal exposure SNPs.

`simmr` generates individual-level genotypes and phenotypes for the exposures, outcome, and their confounder then efficiently performs GWAS for each SNP to produce GWAS summary statistics. This set of SNPs is then reduced by the user-specified significance and LD pruning thresholds [14], both of which are implemented in the `simmr` software. Some MR methods use the matrix of correlations between GWAS estimation errors $\boldsymbol{\Omega}$ across cohorts to correct for weak instrument bias (e.g., [22], [24], [10], [26]). These methods calculate $\boldsymbol{\Omega}$ using non-significant GWAS summary statistics [24, 26] or LD score regression [22, 10]. `simmr` will directly calculate $\boldsymbol{\Omega}$ from realized values of $(\mathbf{x}, Y)$ without any estimation error and return it to the user.

Software

`simmr` uses the popular R software [32] and only requires downloading 3 files: `basicfunctions.R`, `set_params.R`, and `generate_data.R`. Users of `simmr` change the parameters in the `set_params.R` file to their desired quantities. Then, another file named `generate_data.R` is sourced and the simulation data is automatically loaded into the user's R environment. `simmr` is hosted at `https://github.com/noahlorinczcomi/simmr`, where anyone can directly change any of the simulation code to improve it. Changes are tracked automatically and earlier versions can be restored at any time.

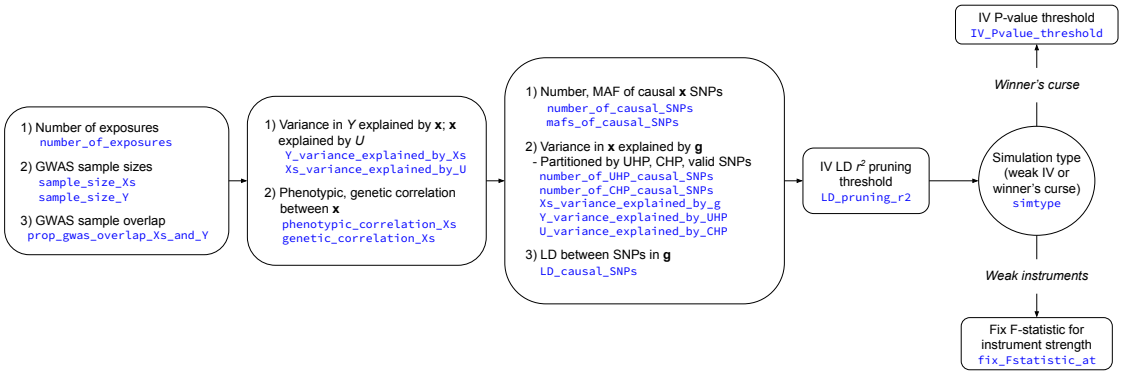
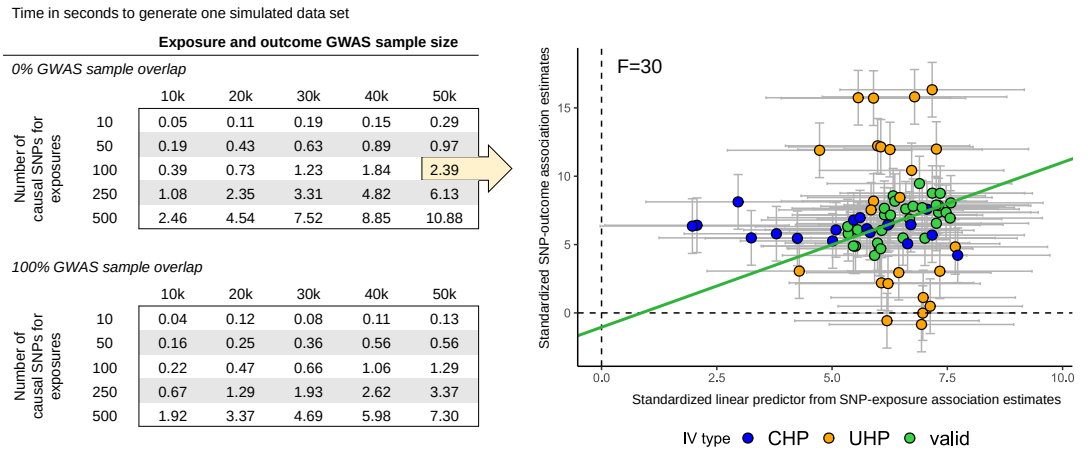


Fig. 1. This figure provides a conceptual overview of the `simmr` data-generating process. First, users specify the sizes of exposure and outcome GWAS sample sizes and their degree of overlap. Next, users specify the true causal model by changing variances in each phenotype explained by each other and by the SNPs that are causally related to the exposures. Next, the user specifies the particular characteristics of the causal SNPs, inputting the degree of UHP, CHP, and LD between them. Finally, the user can perform IV selection to evaluate winner's curse or weak instrument bias. `simmr` commands are presented in blue text underneath the description for each parameter.



Time in seconds to generate one simulated data set

Exposure and outcome GWAS sample size

0% GWAS sample overlap

Number of causal SNPs for exposures	10k	20k	30k	40k	50k
10	0.05	0.11	0.19	0.15	0.29
50	0.19	0.43	0.63	0.89	0.97
100	0.39	0.73	1.23	1.84	2.39
250	1.08	2.35	3.31	4.82	6.13
500	2.46	4.54	7.52	8.85	10.88

100% GWAS sample overlap

Number of causal SNPs for exposures	10k	20k	30k	40k	50k
10	0.04	0.12	0.08	0.11	0.13
50	0.16	0.25	0.36	0.56	0.56
100	0.22	0.47	0.66	1.06	1.29
250	0.67	1.29	1.93	2.62	3.37
500	1.92	3.37	4.69	5.98	7.30

Fig. 2. This figure shows computation times to generate one simulated data set using `simmr` (left) and an example of the generated summary data for evaluation of MR methods (right). The data in the right panel is only of the final selected and pruned set of instruments. The original number of causal SNPs was 100, but the selected number of IVs to achieve an F-statistic of 30 was 91. The full code used to produce this figure is available at `https://github.com/noahlorinczcomi/simmr`. The x-axis in the right panel is the standardized linear predictor used in multivariable MR $\hat{\mathbf{B}}\boldsymbol{\theta}$ and the y-axis is the estimated association of the IVs with the outcome in standardized scale. The green line is the linear fit to the green points representing valid instrumental variables. The yellow arrow in the left panel indicates that the time it took to generate the data in the right panel was 2.4 seconds when there is no GWAS sample overlap. When the exposure and outcome GWAS sample overlap is 100%, computation time was reduced to 1.39 seconds.

Conclusion

There is currently no standard simulation framework for performing simulations in Mendelian Randomization research. Different researchers have independently performed simulations designed to reflect the same real world conditions, but the performance of the same methods in different simulations may vary. The MR literature is replete with MR estimators, with each at some point having a demonstrated advantage over others in simulation. The transferability of their performance to real world settings may be in question if their performance in independent simulations cannot be replicated. We present `simmr`, an open-source software for performing simulations to evaluate the performance of Mendelian Randomization methods. Researchers can directly modify the `simmr` software. It is our intention that the community will use this opportunity to establish an accepted procedure for perfoming simulations using MR. As `simmr` is refined and expanded, the performance of new and existing MR methods should become more transferable to real-world conditions.

Competing interests

No competing interest is declared.

Author contributions statement

Acknowledgments

References

1. Frederick J Boehm and Xiang Zhou. Statistical methods for mendelian randomization in genome-wide association studies: A review. *Computational and Structural Biotechnology Journal*, 20:2338–2351, 2022.
2. Jack Bowden, George Davey Smith, Philip C Haycock, and Stephen Burgess. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.*, 40(4):304–314, 2016.
3. Ioan Gabriel Bucur, Tom Claassen, and Tom Heskes. Inferring the direction of a causal link and estimating its effect via a bayesian mendelian randomization approach. *Statistical Methods in Medical Research*, 29(4):1081–1111, 2020.
4. Stephen Burgess, Adam Butterworth, and Simon G Thompson. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic epidemiology*, 37(7):658–665, 2013.
5. Stephen Burgess, Frank Dudbridge, and Simon G Thompson. Combining information on multiple instrumental variables in mendelian randomization: comparison of allele score and summarized data methods. *Statistics in medicine*, 35(11):1880–1906, 2016.
6. Stephen Burgess, Christopher N Foley, Elias Allara, James R Staley, and Joanna MM Howson. A robust and efficient method for mendelian randomization with hundreds of genetic variants. *Nat. Commun.*, 11(1):1–11, 2020.
7. Stephen Burgess and Simon G Thompson. Interpreting findings from mendelian randomization using the mr-egger method. *European journal of epidemiology*, 32:377–389, 2017.
8. Stephen Burgess, Verena Zuber, Elsa Valdes-Marquez, Benjamin B Sun, and Jemma C Hopewell. Mendelian randomization with fine-mapped genetic data: choosing from large numbers of correlated instrumental variables. *Genetic epidemiology*, 41(8):714–725, 2017.
9. Qing Cheng, Tingting Qiu, Xiaoran Chai, Baoluo Sun, Yingcun Xia, Xingjie Shi, and Jin Liu. Mr-corr2: a two-sample mendelian randomization method that accounts for correlated horizontal pleiotropy using correlated instrumental variants. *Bioinformatics*, 38(2):303–310, 2022.
10. Qing Cheng, Yi Yang, Xingjie Shi, Kar-Fu Yeung, Can Yang, Heng Peng, and Jin Liu. Mr-ldp: a two-sample mendelian randomization for gwas summary statistics accounting for linkage disequilibrium and horizontal pleiotropy. *NAR genomics and bioinformatics*, 2(2):lqaa028, 2020.
11. Qing Cheng, Xiao Zhang, Lin S Chen, and Jin Liu. Mendelian randomization accounting for complex correlated horizontal pleiotropy while elucidating shared genetic etiology. *Nat. Commun.*, 13(1):1–13, 2022.
12. Yangqing Deng, Dongsheng Tu, Chris J O'Callaghan, Derek J Jonker, Christos S Karapetis, Jeremy Shapiro, Geoffrey Liu, and Wei Xu. A bayesian approach for two-stage multivariate mendelian randomization with mixed outcomes. *Statistics in Medicine*, 2023.
13. Yangqing Deng, Dongsheng Tu, Chris J O'Callaghan, Geoffrey Liu, and Wei Xu. Two-stage multivariate mendelian randomization on multiple outcomes with mixed distributions. *Statistical Methods in Medical Research*, page 09622802231181220, 2022.
14. Frank Dudbridge and Paul J Newcombe. Accuracy of gene scores when pruning markers by linkage disequilibrium. *Human heredity*, 80(4):178–186, 2016.
15. Christopher N Foley, Amy M Mason, Paul DW Kirk, and Stephen Burgess. Mr-clust: clustering of genetic variants in mendelian randomization with similar causal estimates. *Bioinformatics*, 37(4):531–541, 2021.
16. Kevin J Gleason, Fan Yang, and Lin S Chen. A robust two-sample mendelian randomization method integrating gwas with multi-tissue eqtl summary statistics. *bioRxiv*, pages 2020–06, 2020.
17. Andrew J Grant and Stephen Burgess. A bayesian approach to mendelian randomization using summary statistics in the univariable and multivariable settings with correlated pleiotropy. *bioRxiv*, pages 2023–05, 2023.
18. Fernando Pires Hartwig, George Davey Smith, and Jack Bowden. Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption. *International journal of epidemiology*, 46(6):1985–1998, 2017.
19. Daniel Iong, Qingyuan Zhao, and Yang Chen. A latent mixture model for heterogeneous causal mechanisms in mendelian randomization. *arXiv preprint arXiv:2007.06476*, 2020.
20. Lin Jiang, Lin Miao, Guorong Yi, Xiangyi Li, Chao Xue, Mulin Jun Li, Hailiang Huang, and Miaoxin Li. Powerful and robust inference of complex phenotypes' causal genes with dependent expression quantitative loci by a median-based mendelian randomization. *The American Journal of Human Genetics*, 109(5):838–856, 2022.
21. Hyunseung Kang, Anru Zhang, T Tony Cai, and Dylan S Small. Instrumental variables estimation with some invalid instruments and its application to mendelian randomization. *J. Am. Stat. Assoc.*, 111(513):132–144, 2016.
22. Zhaotong Lin, Haoran Xue, and Wei Pan. Robust multivariable mendelian randomization based on constrained maximum likelihood. *The American Journal of Human Genetics*, 110(4):592–605, 2023.
23. Lu Liu, Ping Zeng, Fuzhong Xue, Zhongshang Yuan, and Xiang Zhou. Multi-trait transcriptome-wide association studies with probabilistic mendelian randomization. *The American Journal of Human Genetics*, 108(2):240–256, 2021.

24. Noah Lorincz-Comi, Yihe Yang, Gen Li, and Xiaofeng Zhu. Mrbee: A novel bias-corrected multivariable mendelian randomization method. *bioRxiv*, pages 2023–01, 2023.
25. Jean Morrison, Nicholas Knoblauch, Joseph H Marcus, Matthew Stephens, and Xin He. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat. Genet.*, 52(7):740–747, 2020.
26. Ninon Mounier and Zoltán Kutalik. Bias correction for inverse variance weighting mendelian randomization. *Genetic Epidemiology*, 2023.
27. Paul J Newcombe, David V Conti, and Sylvia Richardson. Jam: a scalable bayesian framework for joint analysis of marginal snp effects. *Genetic epidemiology*, 40(3):188–201, 2016.
28. Tom M Palmer, John R Thompson, Martin D Tobin, Nuala A Sheehan, and Paul R Burton. Adjusting for bias and unmeasured confounding in mendelian randomization studies with binary responses. *International journal of epidemiology*, 37(5):1161–1168, 2008.
29. A Patel, D Gill, P Newcombe, and Burgess S. Conditional inference in cis-mendelian randomization using weak genetic factors. *Biometrics*, 2023.
30. Guanghao Qi and Nilanjan Chatterjee. Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. *Nature communications*, 10(1):1941, 2019.
31. Guanghao Qi and Nilanjan Chatterjee. A comprehensive evaluation of methods for mendelian randomization using realistic simulations and an analysis of 38 biomarkers for risk of type 2 diabetes. *International journal of epidemiology*, 50(4):1335–1349, 2021.
32. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.
33. Jessica MB Rees, Angela M Wood, Frank Dudbridge, and Stephen Burgess. Robust methods in mendelian randomization via penalization of heterogeneous causal estimates. *PloS One*, 14(9):e0222362, 2019.
34. Rebecca C Richmond and George Davey Smith. Mendelian randomization: concepts and scope. *Cold Spring Harbor perspectives in medicine*, 12(1):a040501, 2022.
35. Eleanor Sanderson, M Maria Glymour, Michael V Holmes, Hyunseung Kang, Jean Morrison, Marcus R Munafò, Tom Palmer, C Mary Schooling, Chris Wallace, Qingyuan Zhao, et al. Mendelian randomization. *Nature Reviews Methods Primers*, 2(1):6, 2022.
36. AF Schmidt and F Dudbridge. Mendelian randomization with egger pleiotropy correction and weakly informative bayesian priors. *International journal of epidemiology*, 47(4):1217–1228, 2018.
37. Eric AW Slob and Stephen Burgess. A comparison of robust mendelian randomization methods using summary data. *Genetic epidemiology*, 44(4):313–329, 2020.
38. Adriaan van Der Graaf, Annique Claringbould, Antoine Rimbert, BIOS Consortium Heijmans Bastiaan T. 8 Hoen Peter AC't 9 van Meurs Joyce BJ 10 Jansen Rick 11 Franke Lude 1 2, Harm-Jan Westra, Yang Li, Cisca Wijmenga, and Serena Sanna. Mendelian randomization while jointly modeling cis genetics identifies causal relationships between gene expression and lipids. *Nature communications*, 11(1):4930, 2020.
39. Marie Verbanck, Chia-Yen Chen, Benjamin Neale, and Ron Do. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat. Genet.*, 50(5):693–698, 2018.
40. Jingshu Wang, Qingyuan Zhao, Jack Bowden, Gibran Hemani, George Davey Smith, Dylan S Small, and Nancy R Zhang. Causal inference for heritable phenotypic risk factors using heterogeneous genetic instruments. *PLoS genetics*, 17(6):e1009575, 2021.
41. Lu Wang, Boran Gao, Yue Fan, Fuzhong Xue, and Xiang Zhou. Mendelian randomization under the omnigenic architecture. *Briefings in Bioinformatics*, 22(6):bbab322, 2021.
42. Siqi Xu, Wing Kam Fung, and Zhonghua Liu. Mrcip: a robust mendelian randomization method accounting for correlated and idiosyncratic pleiotropy. *Briefings in Bioinformatics*, 22(5):bbab019, 2021.
43. Siqi Xu, Peng Wang, Wing Kam Fung, and Zhonghua Liu. A novel penalized inverse-variance weighted estimator for mendelian randomization with applications to covid-19 outcomes. *Biometrics*, 2022.
44. Haoran Xue, Xiaotong Shen, and Wei Pan. Constrained maximum likelihood-based mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. *The American Journal of Human Genetics*, 108(7):1251–1269, 2021.
45. Jian Yang, Teresa Ferreira, Andrew P Morris, Sarah E Medland, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication, Meta analysis (DIAGRAM) Consortium, Pamela AF Madden, Andrew C Heath, Nicholas G Martin, Grant W Montgomery, et al. Conditional and joint multiple-snp analysis of gwas summary statistics identifies additional variants influencing complex traits. *Nature genetics*, 44(4):369–375, 2012.
46. Ting Ye, Jun Shao, and Hyunseung Kang. Debiased inverse-variance weighted estimator in two-sample summary-data mendelian randomization. *The Annals of statistics*, 49(4):2079–2100, 2021.
47. Zhongshang Yuan, Lu Liu, Ping Guo, Ran Yan, Fuzhong Xue, and Xiang Zhou. Likelihood-based mendelian randomization analysis with automated instrument selection and horizontal pleiotropic modeling. *Science Advances*, 8(9):eabl5744, 2022.
48. Jia Zhao, Jingsi Ming, Xianghong Hu, Gang Chen, Jin Liu, and Can Yang. Bayesian weighted mendelian randomization for causal inference based on summary statistics. *Bioinformatics*, 36(5):1501–1508, 2020.
49. Qingyuan Zhao, Jingshu Wang, Gibran Hemani, Jack Bowden, and Dylan S Small. Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score. 2020.
50. Xiaofeng Zhu, Xiaoyin Li, Rong Xu, and Tao Wang. An iterative approach to detect pleiotropy and perform mendelian randomization analysis using gwas summary statistics. *Bioinformatics*, 37(10):1390–1400, 2021.
51. Verena Zuber, Johanna Maria Colijn, Caroline Klaver, and Stephen Burgess. Selecting likely causal risk factors from high-throughput experiments using multivariable mendelian randomization. *Nature communications*, 11(1):29, 2020.