

MRBEE: A bias-corrected multivariable Mendelian randomization method

Noah Lorincz-Comi,^{1,2} Yihe Yang,^{1,2} Gen Li,¹ and Xiaofeng Zhu^{1,3,*}

Summary

Mendelian randomization (MR) is an instrumental variable approach used to infer causal relationships between exposures and outcomes, which is becoming increasingly popular because of its ability to handle summary statistics from genome-wide association studies. However, existing MR approaches often suffer the bias from weak instrumental variables, horizontal pleiotropy and sample overlap. We introduce MRBEE (MR using bias-corrected estimating equation), a multivariable MR method capable of simultaneously removing weak instrument and sample overlap bias and identifying horizontal pleiotropy. Our extensive simulations and real data analyses reveal that MRBEE provides nearly unbiased estimates of causal effects, well-controlled type I error rates and higher power than comparably robust methods and is computationally efficient. Our real data analyses result in consistent causal effect estimates and offer valuable guidance for conducting multivariable MR studies, elucidating the roles of pleiotropy, and identifying total 42 horizontal pleiotropic loci missed previously that are associated with myopia, schizophrenia, and coronary artery disease.

Introduction

Mendelian randomization (MR) is an instrumental variable (IV) approach used to infer causal relationships between exposures and outcomes and can apply to summary statistics from genome-wide association studies (GWASs), providing a cost-effective and generalizable alternative to randomized controlled trials.¹ Inverse-variance weighting (IVW)² is the fundamental approach to perform MR with GWAS summary statistics, and the validity of which relies heavily on three so-called valid IV assumptions: the genetic IVs are (IV1) strongly associated with the exposures; (IV2) not directly associated with the outcome conditional on the exposures; and (IV3) not associated with any confounders of the exposure-outcome relationships. Violations of the (IV1)–(IV3) assumptions will introduce weak instrument,³ unbalanced uncorrelated horizontal pleiotropy (UHP),⁴ and correlated horizontal pleiotropy (CHP)⁵ biases into the causal effect estimation, respectively. As for balanced UHP, which aligns with the instrument strength independent of direct effect (InSIDE) assumption,⁴ the causal effect estimation remains unbiased.

From a statistical perspective, both unbalanced UHP and CHP in an MR model exhibit characteristics similar to outliers in traditional regression analyses. Therefore, these issues can be addressed using robust statistical tools. In the literature, MR-PRESSO⁶ and IMRP⁷ identify and remove horizontal pleiotropic variants through hypothesis tests, while the MR-Lasso⁸ and MRcML⁹ methods detect horizontal pleiotropy through variable selection tools. On the other hand, approaches like MR-Median¹⁰ and MR-Robust¹¹ employ robust loss functions to mitigate the horizontally pleiotropic effects. Furthermore, Gaussian

mixture models are implemented in methods such as MRMix,¹² MR-Conmix,¹³ CAUSE,⁵ MRAID,¹⁴ and MR-CUE.¹⁵ These models offer an advantage over traditional robust tools by utilizing fewer degrees of freedom to describe unbalanced UHP and CHP, thereby increasing efficiency when the mixture models are correctly specified.

While univariable MR (UVMR) methods allow some IVs to have horizontally pleiotropic effects, they generally assume that most IVs influence the outcome solely through the mediation of the exposure. However, this assumption can be problematic when traits share more than 50% causal variants. For instance, both systolic and diastolic blood pressure (SBP and DBP)¹⁶ are revealed to share substantial causal variants. When analyzing the causal effect of SBP on cardiovascular disease, it is often challenging to remove the effect through DBP. A more effective way to address this issue is multivariable MR (MVMR), which accounts for the majority of horizontally pleiotropic variants that are shared by multiple exposures.¹⁷ To date, the multivariable versions of IVW,¹⁸ MR-Egger,¹⁹ MR-Median,¹⁰ and MRcML²⁰ have been developed. As demonstrated by Sanderson et al.,¹⁷ MVMR is reliable in estimating the direct causal effects of one or more exposures.

The issue of weak instrument bias, stemming from the violation of the (IV1) assumption, poses even more challenging to resolve in MVMR than in UVMR. Specifically, it is usually difficult to find a set of IVs that are strongly associated with all exposures under consideration. In contrast, IVs are generally selected if they are associated with at least one exposure.²¹ With the growing identification of causal variants for complex traits, the pool of IVs used in MVMR can easily reach the thousands due to this

¹Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA

²These authors contributed equally

³Lead contact

*Correspondence: xxz10@case.edu

<https://doi.org/10.1016/j.xhgg.2024.100290>.

© 2024 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



IV selection procedure, therefore worsen weak instrument bias. Traditional approaches to mitigate this bias involve discarding weak IVs whose F-statistic or conditional F-statistic is less than 10. This threshold is believed to keep the relative bias in causal effect estimates below 10%.^{3,22} However, the exclusion of IVs can lead to reduced statistical power and introduce a “winner’s curse,” thereby compromising the validity of the causal inference.²³

We propose to resolve the weak instrument bias by using tools in measurement error analysis.²⁴ Specifically, measurement error bias occurs when explanatory variables are measured with random error, leading to biased estimates of model parameters. Since current MR approaches are performed with GWAS summary statistics that always contain estimation errors, the causal effect estimates inevitably suffer from measurement error bias.^{25,26} Therefore, we view a weak instrument as a relatively large measurement error in effect size estimate based on finite sample size and is the primary reason for violating assumption (IV1) in IVW and other MR approaches. Furthermore, unlike traditional measurement error analyses that assume uncorrelated estimation errors in exposures and outcomes, overlapping individuals in exposure and outcome GWAS can result in correlated measurement errors, leading the direction of measurement error bias either toward or away from zero. As we observed in Figure 1, IVW estimates²⁷ exhibit negative bias with small numbers of overlapping samples and positive bias with large numbers of overlapping samples, respectively.

We develop a computationally efficient MVMR method, MR using bias-corrected estimating equations (MRBEE), to eliminate weak instrument bias while simultaneously accounting for horizontal pleiotropy in the presence of weak IVs or sample overlap. In contrast, existing methods only address weak instrument bias in specific cases such as no sample overlap (debiased IVW)²⁶ or no horizontal pleiotropy (MRlap).²⁸ Although the multivariable MRcML methods²⁰ generally provide unbiased causal estimates, they may be vulnerable to horizontal pleiotropy and computationally intensive. To underscore its practical significance, we apply MRBEE to three datasets, each targeting a unique disease, namely, myopia, schizophrenia (SCZ), and coronary artery disease (CAD), with the aim to unravel the distinct causal exposures associated with each. In addition, we extend the pleiotropy test to a genome-wide pleiotropy test (GWPT) for detecting novel loci. These empirical analyses offer valuable guides for conducting MVMR studies, elucidating the roles of pleiotropy and weak instrument bias, and illustrating how to identify novel loci through pleiotropy tests. The study was approved by the institutional review board (IRB number: STUDY20180592) at Case Western Reserve University.

Results

Overview of method

The detailed MRBEE is described in the [material and methods](#) section. Briefly, suppose that there are p expo-

sure having causal effects on an outcome and m genetic variants as IVs. Let $\alpha = (\alpha_1, \dots, \alpha_m)^T$ be a vector of length m , representing the genetic effect sizes of IVs on the outcome, $\mathbf{B} = (\beta_1, \dots, \beta_m)^T$ be an $(m \times p)$ matrix with $\beta_j = (\beta_{j1}, \dots, \beta_{jp})^T$ representing the genetic effect sizes of the j th IV on the p exposures, $\theta = (\theta_1, \dots, \theta_p)^T$ be a vector of length p representing the causal effects of the p exposures on the outcome, and $\gamma = (\gamma_1, \dots, \gamma_m)^T$ be a vector of length m representing horizontal pleiotropy. We model the causal effects of the exposures on the outcome by

$$\alpha_j = \beta_j^T \theta + \gamma_j.$$

The goal in MR analysis is to estimate the causal effects θ unbiasedly. In the above equation, the true genetic effect sizes α and \mathbf{B} are not observed but can be estimated through the GWAS of exposures and outcome and the pleiotropy effect γ is simply unknown. Let $\hat{\alpha}_j$ and $\hat{\beta}_j$ be the effect size estimates of the j th IV from the outcome and exposure GWASs. We have

$$\begin{aligned} \hat{\alpha}_j &= \alpha_j + w_{\alpha_j}, \\ \hat{\beta}_j &= \beta_j + \mathbf{w}_{\beta_j}, \end{aligned}$$

where w_{α_j} and \mathbf{w}_{β_j} represent the measurement errors because of finite sample sizes of the GWASs.

In general, an MVMR analysis is performed by the following linear regression:

$$\hat{\alpha}_j = \hat{\beta}_j^T \theta + \gamma_j + \varepsilon_j,$$

where ε_j represents the residual. When we standardize $\hat{\alpha}_j$ and $\hat{\beta}_j$ by their corresponding standard errors obtained from GWASs, the multivariable IVW (MV-IVW) estimates θ is

$$\hat{\theta}_{\text{IVW}} = \operatorname{argmin}_{\theta} \{ \|\hat{\alpha} - \hat{\mathbf{B}}\theta\|_2^2 \} = (\hat{\mathbf{B}}^T \hat{\mathbf{B}})^{-1} \hat{\mathbf{B}}^T \hat{\alpha},$$

which is equivalent to solve the score equation $\mathbf{S}_{\text{IVW}}(\theta) = \hat{\mathbf{B}}^T (\hat{\mathbf{B}}\theta - \hat{\alpha}) = 0$. However, the MV-IVW fails to consider the weak IVs and the correlations among w_{α_j} and \mathbf{w}_{β_j} induced by sample overlap and assumes pleiotropy $\gamma_j = 0$ for all IVs. Thus, the MV-IVW is biased. To solve this problem, we propose MRBEE by solving the following estimating equation,

$$\mathbf{S}_{\text{BEE}}(\theta) = \mathbf{S}_{\text{IVW}}(\theta) - m(\Sigma_{W_{\beta}W_{\alpha}}\theta - \sigma_{W_{\beta}W_{\alpha}}) = 0,$$

where $\Sigma_{W_{\beta}W_{\alpha}}$ and $\sigma_{W_{\beta}W_{\alpha}}$ represent the covariance matrix among \mathbf{w}_{β_j} and between \mathbf{w}_{β_j} and w_{α_j} ([material and methods](#)) in the set of the m IVs. The score function in $\mathbf{S}_{\text{BEE}}(\theta)$ adds a corrected term, which corrects the bias because of weak IVs and sample overlap, meanwhile assumes there are no pleiotropic IVs ($\gamma_j = 0$). The solution of the equation $\mathbf{S}_{\text{BEE}}(\theta)$ is

$$\hat{\theta}_{\text{BEE}} = (\hat{\mathbf{B}}^T \hat{\mathbf{B}} - m\Sigma_{W_{\beta}W_{\alpha}})^{-1} (\hat{\mathbf{B}}^T \hat{\alpha} - m\sigma_{W_{\beta}W_{\alpha}}).$$

With the presence of pleiotropic IVs, we apply an iterative procedure⁷ with the pleiotropy test S_{pleio} for multiple

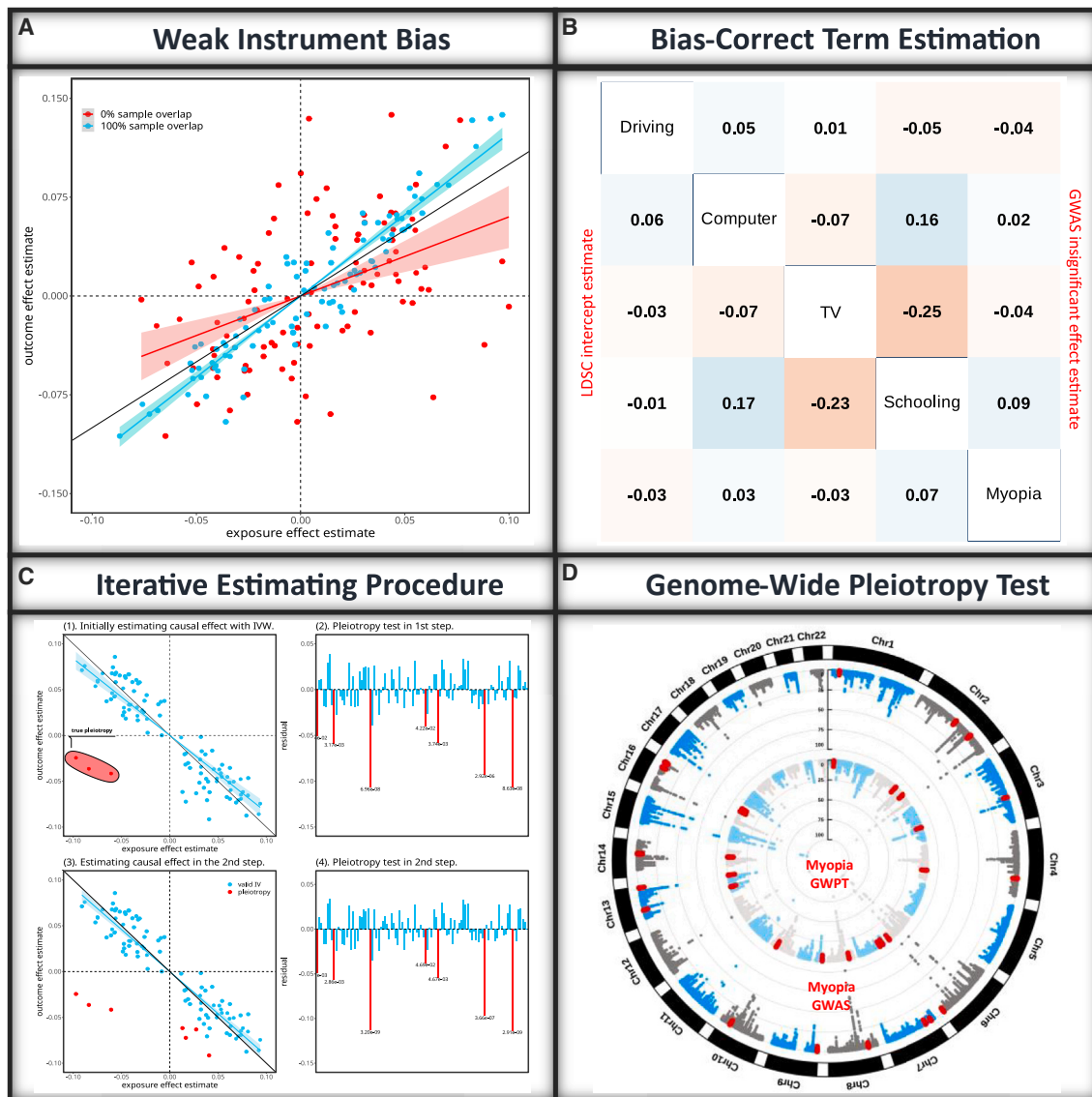


Figure 1. Principle of MRBEE

(A) Traditional MR methods are vulnerable to weak instrument bias arising from the estimation errors in GWAS associations for the exposure(s) and outcome. The direction of the bias is influenced by the degree of sample overlap between the studies where the red and blue points refer to two simulated data with 0% and 100% sample overlap. The shadow regions represent the 95% confidence interval regions.

(B) MRBEE corrects for weak instrument bias using bias-correction terms which are calculated from the matrix of correlations between measurement errors for all exposures and the outcome. In this example with myopia and its four exposures, the numbers in the lower triangle of the table are the correlations estimated using LD score regression and that in the upper triangle of the table are the correlations estimated using non-significant SNPs.

(C) MRBEE uses an iterative estimation procedure, where horizontally pleiotropic IVs are removed at each iteration until convergence. The y axis in panels (2) and (4) reflect the SNP association with the outcome not mediated by the exposures. The numbers under the red vertical lines represent p values.

(D) After estimating causal effects, MRBEE performs genome-wide horizontal pleiotropy testing to find loci associated with the outcome (e.g., myopia) that were not detected in the original GWAS.

exposures and an outcome, which uses the following statistic S_{pleio} for the j th IV,

$$S_{\text{pleio}}(\hat{\theta}) = \frac{(\hat{\alpha}_j - \hat{\beta}_j^T \hat{\theta})^2}{\text{var}(\hat{\alpha}_j - \hat{\beta}_j^T \hat{\theta})}$$

Thus, MRBEE estimates the causal effect θ , and identifies pleiotropic IVs with the current estimated causal effect θ iteratively. The entire pipeline from inputting summary

statistics, estimating causal effects, and identifying pleiotropic IVs is illustrated in Figure 1. Note that after estimating the causal effect θ , we can further search pleiotropic variants across the entire genome.

Simulation

We compared MRBEE with the multivariable MR of IVW, MR-Egger, MR-Median, MR-Lasso, MRcML-DP, and MRcML-BIC. MRBEE is implemented with the R package

MRBEE and the other methods are implemented through the R package MendelianRandomization.²⁹ We call IVW, MR-Egger, MR-Median, and MR-Lasso the traditional MVMR methods, as they either do not account for estimation error of effect size or the sample overlap. Our simulation setting was adapted from the ones considered by Lin et al.,²⁰ but with specific adjustments to better reflect real-world situations. Specifically, we set the heritability of both exposures and confounders at 0.1, introduced moderate genetic correlations among the exposures, and added correlations among random errors of exposure and outcome GWAS cohorts. In our analysis, we consider three scenarios: no pleiotropy, 30% unbalanced UHP, and 30% CHP. All exposures are assumed to come from the same GWAS sample, while the outcome may overlap completely (100% sample overlap), partially (50% and 77% overlap), or be entirely independent (0% sample overlap). In addition, the sample size was set at 50,000, the number of IVs was set at 50, 100, and 200, representing the increasing of weak IVs, the number of exposures was 4, and the causal effect was $\theta = (0, 0.2, -0.2, 0.4)^T$, which represents no causal effect, and positive, negative, and large causal effects, respectively. Simulation settings are fully presented in [supplement 1](#) and the R code used to generate simulated data is available at the GitHub repository of this paper. The number of simulation replicates was 500, and additional simulations can be found in [supplement 1](#).

Bias of causal effect estimates

[Figures 2A, 2D, 2G, and 2J](#) demonstrate that the bias in traditional MVMR methods (IVW, MR-Egger, MR-Median, MR-Lasso) is proportional to the number of IVs used, especially in the absence of horizontal pleiotropy. The direction of this bias is influenced by sample overlap: no overlap results in bias toward the null, while sample overlap leads to bias away from the null. On the contrary, MRBEE, MRcML-BIC, and MRcML-DP are unbiased under these conditions. The unbiasedness for MRcML methods is likely attributed to the fact that the objective function of MRcML methods²⁰ accounts for the covariance of estimation errors. Our results suggest that incorporating the estimation error covariance matrix mitigates measurement error bias.

[Figures 2B, 2E, 2H, and 2K](#) demonstrate that when there was 30% unbalanced UHP, IVW, and MR-Egger generally incurred substantial bias. Moreover, there were inflated standard errors in the causal estimates due to the horizontal pleiotropy. MR-Median and MR-Lasso also incurred substantial bias, but the standard errors of their causal estimates were smaller than that from IVW and MR-Egger. These methods apply robust tools to estimate the causal effects in the presence of horizontal pleiotropy but are not able to remove the bias by weak instrument or sample overlap. MRcML-BIC and MRcML-DP generally provided unbiased causal estimates when there was no sample overlap. When the sample overlap percentage was 100%, both MRcML-BIC and MRcML-DP incurred biases in different directions. The magnitude of this bias was proportional to the number of IVs used. For example, for exposure 1

with true $\theta_1 = 0$, MRcML-BIC and MRcML-DP had bias away from the null; for exposure 3 where $\theta_3 = -0.2$, the two methods had bias toward, and even past, the null. In comparison, MRBEE was unbiased in all scenarios except when there were 200 IVs and 100% overlap. In this case, MRBEE still had a smaller upward bias for exposure 3 with $\theta_3 = -0.2$ than other methods.

[Figures 2C, 2F, 2I, and 2L](#) demonstrate that when there was 30% CHP, IVW, and MR-Egger had larger bias and standard errors in their causal estimates than the rest of methods. Both had bias away from the null for exposure 1, and the magnitude of which depended on the number of IVs used. MR-Median and MR-Lasso generally were less biased than IVW and MR-Egger, as they are more robust in handling of CHP IVs. The weak instrument bias of MR-Median and MR-Lasso followed the same bias patterns as no pleiotropy. MRcML-BIC and MRcML-DP were both biased when the sample overlapping percentage was 100% or 0%, potentially due to the instability of algorithm when horizontal pleiotropy is present. MRBEE was unbiased in all cases and generally had standard errors comparable to other methods excluding IVW and MR-Egger. Finally, when the number of IVs increased from 50 to 200, representing the increasing of weak IVs, MRBEE was always performing better than the comparing methods ([Figure 2](#)).

Type I error and power

[Figures 3A–3C](#) present the type I error rates for all the methods when the true causal effect $\theta_1 = 0$, which corresponds to the first exposure in our simulations. When there was no sample overlap between exposures and the outcome, the type I error was well controlled for MRBEE in all three scenarios, i.e., no pleiotropy, 30% unbalanced UHP, and 30% unbalanced CHP. In comparison, MRcML-DP, MR-Median, MR-Egger, and IVW was generally conservative, while MRcML-BIC and MR-LASSO usually had inflated type I error rates. When 100% overlap between exposures and the outcome, MRBEE still controlled type I error rate well. The rest of the methods either had inflated or extremely conservative type I error rate.

[Figures 3D–3L](#) present power for different causal effects in the three scenarios. Overall, MRBEE has comparable power with the best of the other methods but maintains a type I error rate. We specifically compared MRBEE and MRcML-DP, where the latter controlled type I error rate well under all the simulation scenarios. We observed that MRBEE either had similar or better power than MRcML-DP. The power pattern across the seven methods does not align well with the type I error pattern, that is, high type I error rate corresponds to high power and vice versa. We observed that the reason is the bias direction in causal effect estimates as illuminated in [Figure 2](#), i.e., the bias direction could be in opposite to the true causal effect.

Again, when the number of IVs increased from 50 to 200, the performance of type I error and power of MRBEE was either equal well or better than the comparing methods. We further evaluated these approaches in terms

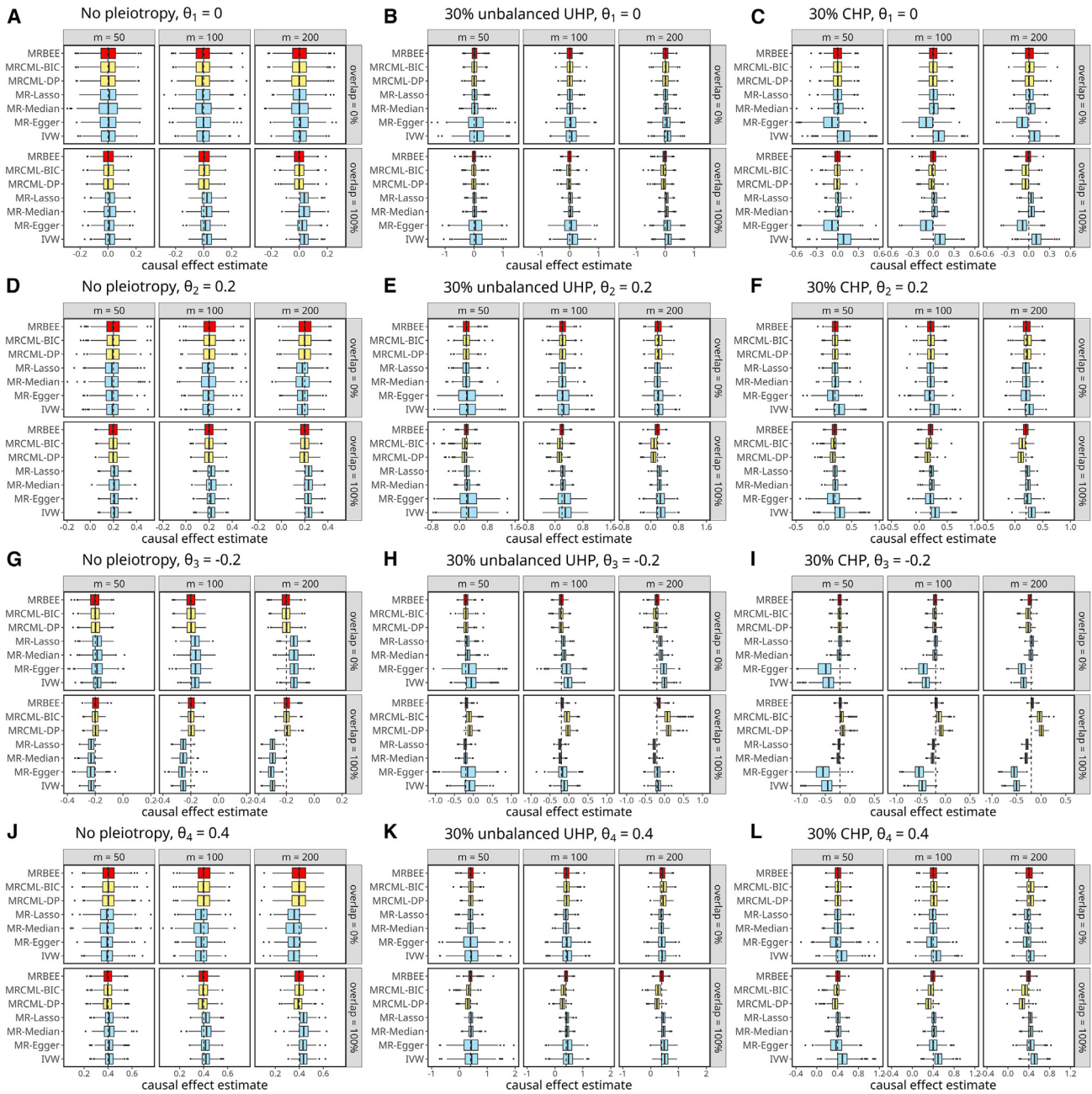


Figure 2. Comparison of the causal effect estimates by the 7 MVMR methods

(A–L) Boxplots display the causal effect estimates from seven methods in the MVMR simulation. The four rows represent the four causal effects θ_j , $j = 1, 2, 3, 4$. Each column corresponds to one of the three pleiotropy scenarios for IVs (i.e., No pleiotropy; 30% unbalanced UHP IVs; 30% CHP IVs). The x axis indicates the value of the causal effect estimate, while the y axis lists the seven methods. The true values of causal effects are denoted by dashed lines. Plots in (A), (D), (G), and (J) when the sample overlap proportion is 0% can be used to infer the magnitude of weak instrument bias since differences between MRBEE and IWV causal estimates in these scenarios are proportional to the degree of weak IV bias. The left and right vertical edges of each box plot represent the 25th and 75th percentiles of causal effect estimate, and the vertical middle line represent the 50th percentile.

of the root-mean-square error (RMSE), standard error (SE) estimation, and coverage frequency of causal effects. MRBEE was again the best among the methods evaluated (Figures S1–S3, and Tables S1–S24 in supplement 1).

We have performed additional simulations in which the overlapping proportion takes values 0, 0.5, 0.77, and 1. In these scenarios MRBEE still performs well. The results are presented in supplement 1.

Computational efficiency

Figure 4 illuminates the computation efficient across seven methods. We observed that MRBEE is computationally as efficient as MR-Median, MR-Lasso, MR-Egger, and MR-IWV. We attribute the computational requirements of MRCML to two potential factors. First, MRCML methods utilize an algorithm similar to the best subset selection to identify the optimal subset of pleiotropic variants. This involves

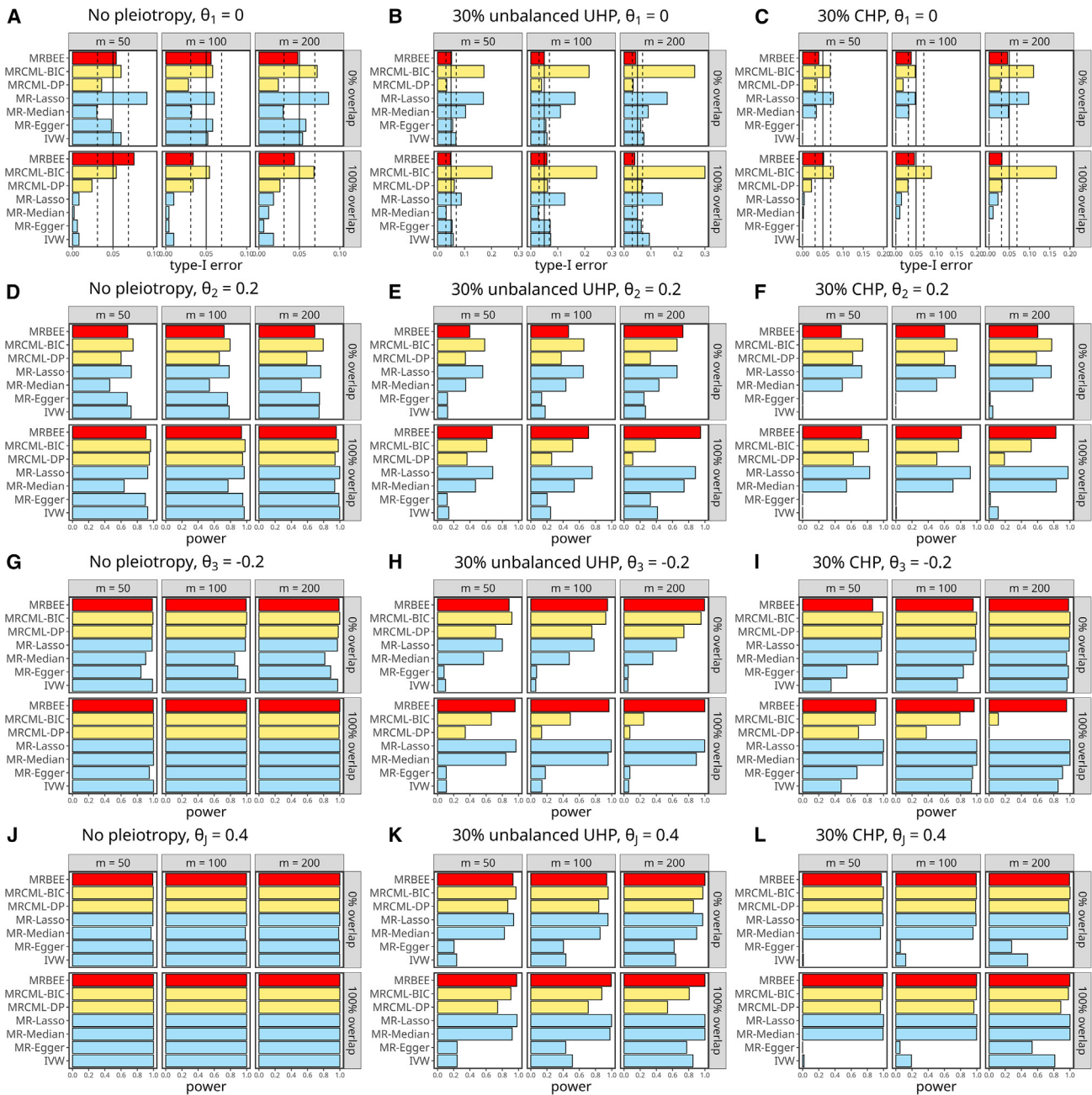


Figure 3. Comparisons of type I error and power of the seven MVMR methods

(A–C) Type I error of the seven methods (MRBEE, MRcML-BIC, MRcML-DP, MR-LASSO, MR-Median, MR-Egger and IVW). (D–L) Power of the seven methods. The three columns correspond to no pleiotropy, 30% unbalanced UHP IVs and 30% CHP IVs, respectively. In each figure, the top and bottom panels represent 0% and 100% sample overlap between exposures and outcome, respectively. Each row represents different causal effects. Simulation settings are described in the [simulation](#) section in the main text, [supplement 1](#), and at our GitHub repository. Displayed are bar plots of rejection frequency estimations across 500 simulations for each scenario, which represents the type I error or power depending on the true causal effect is zero or not. The two dotted vertical lines in (A)–(C) represent the 95% confidence interval.

performing MVMR iteratively by considering numbers of pleiotropic variants ranging from 1 to K (defaulting to $m/2$), and determining the optimal number based on the BIC criteria. In contrast, MRBEE automatically detects pleiotropic variants using a hypothesis test, and MR-Lasso utilizes lasso for pleiotropic variant selection, both of which are computationally efficient. Second, MRcML-DP relies on permutations to derive the SE, which further increases computa-

tional burden. Conversely, MRBEE uses the sandwich formula to estimate its SE, which appears to be accurate in our simulations ([Equation 19](#) in [material and methods](#)).

Real data analysis

Data sources

To demonstrate the MRBEE performance in real data analysis, we analyzed three outcomes, including myopia, SCZ, and

Computation Time of MVMR Approaches in Simulation

	no pleiotropy			30% unbalanced UHP			30% CHP			
MRBEE	6 ms	4 ms	6 ms	6 ms	7 ms	8 ms	4 ms	4 ms	7 ms	overlap = 0%
MRCML-BIC	0.3 s	0.2 s	0.4 s	0.3 s	0.4 s	0.5 s	0.3 s	0.3 s	0.4 s	
MRCML-DP	24.1 s	45.3 s	3 min	16.5 s	56.6 s	3 min	15.5 s	43.1 s	3 min	
MR-Lasso	53 ms	40 ms	96 ms	47 ms	62 ms	0.1 s	38 ms	43 ms	84 ms	
MR-Median	0.1 s	68 ms	0.1 s	84 ms	0.1 s	92 ms	66 ms	67 ms	0.1 s	
MR-Egger	1 ms	1 ms	2 ms	1 ms	2 ms	1 ms	1 ms	1 ms	2 ms	
IVW	2 ms	1 ms	2 ms	1 ms	2 ms	1 ms	1 ms	1 ms	2 ms	
MRBEE	5 ms	4 ms	6 ms	6 ms	7 ms	7 ms	4 ms	5 ms	7 ms	overlap = 100%
MRCML-BIC	0.3 s	0.2 s	0.4 s	0.3 s	0.4 s	0.4 s	0.3 s	0.3 s	0.4 s	
MRCML-DP	49.5 s	2 min	9 min	37.1 s	2 min	8 min	33.0 s	2 min	8 min	
MR-Lasso	51 ms	41 ms	74 ms	48 ms	75 ms	97 ms	39 ms	95 ms	81 ms	
MR-Median	99 ms	70 ms	94 ms	83 ms	93 ms	89 ms	67 ms	75 ms	89 ms	
MR-Egger	1 ms	1 ms	1 ms	1 ms	1 ms	1 ms	1 ms	1 ms	1 ms	
IVW	1 ms	1 ms	1 ms	1 ms	1 ms	1 ms	1 ms	1 ms	1 ms	
	m = 50	m = 100	m = 200	m = 50	m = 100	m = 200	m = 50	m = 100	m = 200	

Figure 4. Comparison of computation efficiency of the seven MVMR methods

This figure depicts the average computation time of the seven methods over 500 simulations.

Data preprocessing

For UVMR, we applied the C + T procedure with the linkage disequilibrium (LD) parameter $r^2 < 0.01$ in ± 500 Kb window to SNPs with association p value at least as small as $5E-5$. We performed this operation for each exposure separately and used the 1000 Genomes project (phase 3) data as the LD reference panel.³⁴ For MVMR, we initially considered the union set of all exposure-specific IV sets from UVMR, then restricted this set to only include SNPs with a joint χ^2 -test p value reaching genome-wide significance and again passing C + T using the same parameters as before. The

CAD. Myopia is known to be influenced by a combination of genetic and environmental factors, including educational attainment (EDU), near-work activity, and outdoor activities³⁰ but their direct causality to myopia is not clear. In this MVMR analysis, we considered refractive error, the measure of myopia degree, as the outcome. The exposures include EDU, near-work activity measured by time spent watching TV and playing on the computer (TV and Computer), and outdoor activity measured by time spent driving (Driving).

Attention-deficit/hyperactivity disorder (ADHD), cannabis use disorder (CAN), EDU, intelligence (INT), left-handedness (LH), intelligence (INT), neuroticism (SESA), and sleep duration (SLP) have been reported as risk factors for SCZ. Of these risk factors, CAN arguably has the strongest evidence of causality with respect to SCZ, with studies reporting dose-response³¹ and strong temporal³² relationships for at-risk individuals. The direct causality of the risk factors on SCZ is also not clear.

Many studies have been published to understand the causal effects of risk factors on CAD. However, the findings in these studies have been inconsistent. For instance, Holmes et al. and Lin et al.^{20,33} found that HDL-C is not significant, while Zhu et al.,⁷ using the GWAS summary data with a much larger sample size (1.3M vs. 90K), found it to be significant. Besides, Wang et al.²¹ found that low-density lipoprotein cholesterol (LDL-C) is not significant in European populations, which seems unreasonable. In this data analysis, we investigated the causal relationships of these risk factors on CAD using the GWAS summary data with the largest sample sizes to date. We focus on the same eight factors studied in Lin et al.,²⁰ i.e., body mass index (BMI), DBP, fasting plasma glucose (FPG), height, HDL-C, LDL-C, triglycerides (TG), and SBP.

joint χ^2 -test for the exposures is presented in Equation 22 in material and methods and is used to assess the null hypothesis that an SNP is not associated with any exposure. We additionally standardized the GWAS effect size estimates so that their SEs were the inverse of the sample sizes. This procedure leads to comparable causal effect estimates across different exposures. We used false discovery rate (FDR) correction in MRBEE to identify and remove SNPs with evidence of horizontal pleiotropy (see Algorithm 1).

Table 1 summarizes the information of GWAS data in this study. In Table 1, the last three columns present the SNP heritability estimated by the LD score regression (LDSC),³⁵ the variances explained by the IVs in UVMR, and the variances explained by the IVs in MVMR. It is evident that for the trait with lower heritability and small sample sizes, the UVMR IVs account for about 1% of its SNP heritability, which may reduce power to detect causal effects using UVMR. However, for most traits, IVs in MVMR analyses explain a substantial portion of the variance, which will provide good power to detect causal effects. This is because the standard error of causal estimate(s) is inversely proportional to the variance(s) explained by the IVs. The last column of Table 1 shows the reliability ratios, a measure of IV strength, for exposures used in real data analysis. The estimation errors averagely account for $\sim 20\%$ variance of the GWAS effect estimates.

Myopia

All the MVMR methods consistently showed that EDU (MRBEE $p = 9.3E-21$) and Driving ($p = 3.8E-11$) are directly causal on myopia, but not TV ($p = 0.136$) or Computer ($p = 0.972$) (Figure 5A). The no direct causal effect of TV or Computer on myopia risk although all exposures were observed to have significant causal effects on myopia in the UVMR

Table 1. Summary of GWAS data used in real data analyses

	Trait	Source	Sample size	Significant IVs	LDSC heritability	UVMR variance ^a	MVMR variance ^b	Reliability ratio ^c	
Myopia	Driving	van De Vegte et al. ⁵⁴	422K	4	0.0365	0.00034	0.00400	0.705	
	Playing computer	Arns et al. ⁵⁵	422K	46	0.0719	0.00408	0.01154	0.873	
	Watching TV	Rustad et al. ⁵⁶	422K	189	0.1321	0.01788	0.02775	0.943	
	EDU	Okbay et al. ⁵⁷	765K	656	0.1352	0.03954	0.03683	0.976	
	Joint 4 test			707					
	Refractive error	Hysi et al. ⁵⁸	246K	420	0.2702	0.11079	0.01433	0.838	
Schizophrenia	ADHD	Demontis et al. ⁵⁹	55K	12	0.0956	0.00279	0.01871	0.832	
	CAN	Johnson et al. ⁶⁰	384K	5	0.0174	0.00033	0.00272	0.552	
	EDU	Okbay et al. ⁵⁷	765K	656	0.1222	0.03954	0.03728	0.973	
	Intelligence	Neale's Lab	430K	48	0.2326	0.01527	0.06023	0.900	
	Left handedness	Cuellar-Partida et al. ⁶¹	205K	4	0.0338	0.00086	0.00533	0.576	
	Neuroticism (SESA)	Nagel et al. ⁶²	450K	42	0.0800	0.00476	0.01056	0.825	
	Sleep duration	Dashti et al. ⁶³	493K	66	0.0649	0.00589	0.00998	0.850	
	Joint 7 test			1,227					
		SCZ	Trubetskoy et al. ⁶⁴	320K	287	0.3380	0.06378	0.03570	0.855
	Coronary artery disease	BMI	Loh et al. ⁶⁵	458K	882	0.2076	0.09494	0.10612	0.918
DBP		Evangelou et al. ⁶⁶ +MVP	1.00M	942	0.1095	0.06022	0.04525	0.823	
FPG		Neale's Lab	361K	115	0.0848	0.03729	0.05156	0.789	
Height		Loh et al. ⁶⁵	458K	2,728	0.6023	0.48986	0.49156	0.981	
HDL-C		Graham et al. ⁶⁷	1.32M	1,031	0.1779	0.09207	0.09745	0.965	
LDL-C		Graham et al. ⁶⁷	1.32M	754	0.1293	0.08435	0.08713	0.961	
TG		Graham et al. ⁶⁷	1.32M	900	0.1251	0.07298	0.08105	0.959	
SBP		Evangelou et al. ⁶⁶ +MVP	1.00M	895	0.1152	0.05626	0.04550	0.829	
Joint 8 test				4,336					
		CAD	Aragam et al. ⁶⁸ +MVP	1.45M	343	0.0500	0.01610	0.01712	0.850

^aVariance explained by the IVs in UVMR analysis.

^bVariance explained by the IVs in MVMR analysis.

^cReliability ratios of exposures in MVMR analysis.

analysis (Figure 5B). The insignificance of both TV and Computer in MVMR analysis suggests their correlations with myopia could be attributed to the confounding with EDU and Driving time. MRBEE provided larger protective causal estimate of driving time than that by IVW (i.e., MRBEE odds ratio [OR] = 0.71 vs. IVW OR = 0.84), likely due to a correction for weak instrument bias given that the driving time variance explained by the IVs was less than 1%. The causal effect of driving time estimated by MRcML-BIC and MRcML-DP was 3–5 times larger in magnitude than those from other methods. (MRcML-BIC OR = 0.38 and MRcML-DP OR = 0.58, respectively). In the iterative pleiotropy test, we detected 31 IVs demonstrating pleiotropy of the exposures and myopia (Figure 5C). Figure 5D compares the computational efficiency of the MR

methods. We observed that IVW was the fastest (<0.1 s), followed by MRBEE (0.1 s), MR-Lasso (0.9 s), MR-Median (1.4 s), MRcML-BIC (1 min), and MRcML-DP (107 min), which were consistent with the simulations.

Schizophrenia

All MVMR methods consistently estimated that CAN (MRBEE $p = 3.7E-8$), EDU ($p = 3.6E-15$), INT ($p = 7.7E-12$), and SESA ($p = 1.8E-7$) have direct causal contributions on schizophrenia (Figure 6A). MRBEE, MR-Lasso, and MR-Median suggested that SLP ($p = 3.4E-4$) has direct causal contribution on schizophrenia but not MRcML-BIC or MRcML-DP. It is not clear why both MRcML-DP and MRcML-BIC failed to detect this causal contribution given our simulations suggest MRcML-BIC could be more powerful although with inflated type I error. A potential reason could

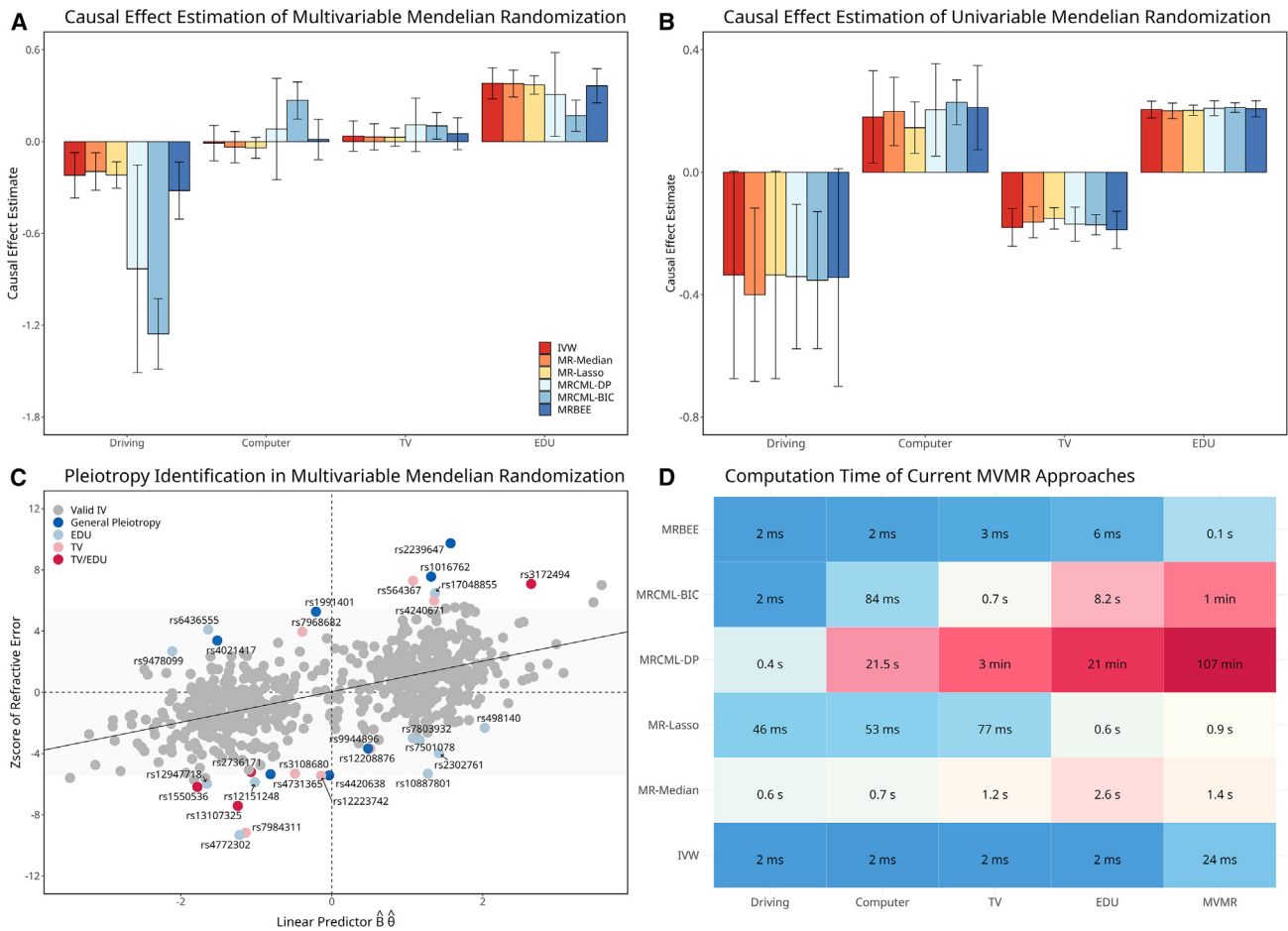


Figure 5. Myopia data analysis

(A) Causal effect estimates by the six MVMR methods. The corresponding 95% confidence interval shown as vertical error bars. (B) Corresponding causal effect estimates by UVMR approaches. The corresponding 95% confidence interval shown as vertical error bars. The radius of the confidence interval equals $\sqrt{a} \times \text{SE}$, where $a = F_{\chi^2}^{-1}(1 - 0.05/4, 1)$, $F_{\chi^2}^{-1}(x, \text{df})$ is the inverse cumulative distribution function of a χ^2 distribution. (C) Pleiotropy test where the x axis represents the linear predictor $\widehat{B}\widehat{\theta}_{\text{BEE}}$ and the y axis represents the corresponding standardized association with myopia from GWAS. The annotation of a pleiotropic variant is made if its pleiotropic test p value is $< 5\text{E}-8$. Only IVs are present in the figure. (D) Computation times of the six comparing methods. Columns 1–4 represent computation time for UVMR of four exposures and the last column represents computation time for MVMR.

be the instability of MRcML, which may converge to local maximum. However, this requires additional investigation. MRBEE suggested no direct causal effect of ADHD ($p = 0.510$) or LH ($p = 0.096$), possibly due to the relatively low exposure variance explained by the IV set (i.e., 0.018 for ADHD and 0.005 for LH). We observed relatively larger odds ratios of EDU and CAN for MRBEE than MR-Median, MR-Lasso, and IVW, but less than MRcML-DP and MRcML-BIC. In comparison, UVMR analyses by all methods suggested evidence of total causal effects of CAN (MRBEE $p = 1.6\text{E}-4$), INT ($p = 3.2\text{E}-7$), SESA ($p = 2.0\text{E}-10$), ADHD ($p = 0.017$), and SLP ($p = 1.4\text{E}-4$), but not EDU ($p = 0.542$) or LH ($p = 0.716$) (Figure 6B). We did not observe any IVs with evidence of horizontal pleiotropy at the Bonferroni-corrected $p = 0.05$ level (Figure 6C), suggesting that the genetic association of the IVs with SCZ are strictly mediated by the five significant exposures. Again, we observed

similar computational efficiency for these methods as before (Figure 6D).

Coronary artery disease

Figure 7A presents MVMR causal estimates for the effects of BMI, DBP, SBP, FPG, height, HDL-C, LDL-C, TG, and SBP on CAD. Using MRBEE, we identified the following significant direct causal effects on CAD, including BMI (MRBEE $p = 3.8\text{E}-39$), FPG ($p = 6.7\text{E}-10$), HDL-C ($p = 8.4\text{E}-21$), LDL-C ($p = 1.5\text{E}-87$), TG ($p = 1.9\text{E}-7$), and SBP ($p = 1.1\text{E}-25$). We observed that MRBEE estimates were generally consistent with estimates from IVW, MR-Median, and MR-Lasso. Conversely, MRcML-BIC and MRcML-DP estimates diverged from all other methods for DBP and FPG. For example, MRcML-DP/BIC were the only methods that simultaneously produced significant causal effects for SBP ($p = 1.6\text{E}-39$) and DBP ($p = 1.8\text{E}-45$) on CAD, two traits that are highly genetically correlated. In

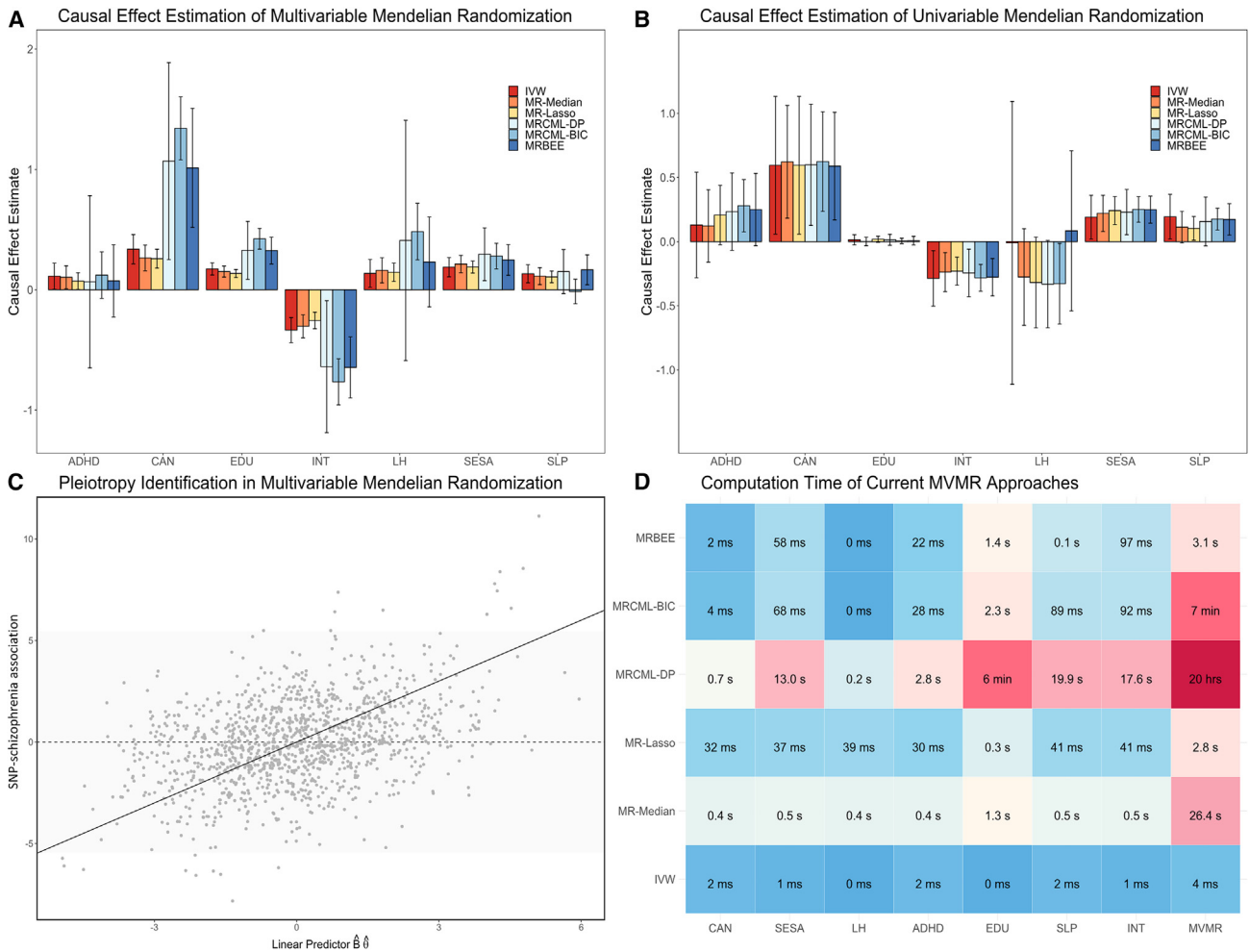


Figure 6. Data analysis of schizophrenia

(A) Causal effect estimates by the six MVMR methods. The corresponding 95% confidence interval shown as vertical error bars. (B) Causal effect estimates by UVMR approaches. The corresponding 95% confidence interval shown as vertical error bars. The radius of the confidence interval equals $\sqrt{a} \times SE$ where $a = F_{\chi^2}^{-1}(1 - 0.05/7, 1)$. (C) Pleiotropy test where the x axis represents the linear predictor $\hat{B}\hat{\theta}_{BEE}$ and the y axis represents the standardized association of the IV with SCZ from GWAS. (D) Computation times of the comparing methods. Columns 1–7 represent computation time for UVMR of seven exposures and the last column represents computation time for MVMR.

comparison, UVMR analyses by all methods consistently suggested that all the exposures have causal contributions on CAD (Figure 7B). We also observed 173 IVs demonstrating horizontal pleiotropy by the horizontal pleiotropy test in MVMR (Figure 7C). Again, our proposed MRBEE is computationally efficient (Figure 7D).

Pleiotropic variants detected by GWPT

GWPT uses the S_{pleio} statistic (Equation 21 in materials and methods) to test whether a genetic variant is associated with the outcome phenotype strictly through the mediation of a select group of exposures. In our GWPT analyses, these groups of exposures are those that were used in each MVMR. This test can be used to find these outcome-associated loci¹⁶ that do not reach the level of genome-wide significance in the original outcome phenotype GWAS but are genome-wide significant in GWPT. In these regions, it is possible that the local genetic correlations between the

exposures and outcome are of different sign or magnitude than the genome-wide genetic correlations.³⁶ To ensure that the loci identified in GWPT were not primarily influenced by other exposures, we excluded any loci that showed even a marginal association with any of the exposures at a genome-wide significance level (i.e., $p < 5E-8$). We also compared GWPT with cross-phenotype association analysis (CPASSOC),³⁷ multi-trait analysis of GWAS (MTAG),³⁸ which are joint tests of association between all exposures in the outcome.

Table 2 lists the variants detected by GWPT for myopia, SCZ, and CAD but missed in the original GWAS. For comparison, we listed the p values for association from the original outcome GWAS, cross-phenotype tests by CPASSOC and MTAG, and by GWPT, respectively. The GWPT identified 18 genome-wide significant loci for myopia, four for SCZ, and 20 for CAD, respectively. All these loci did not

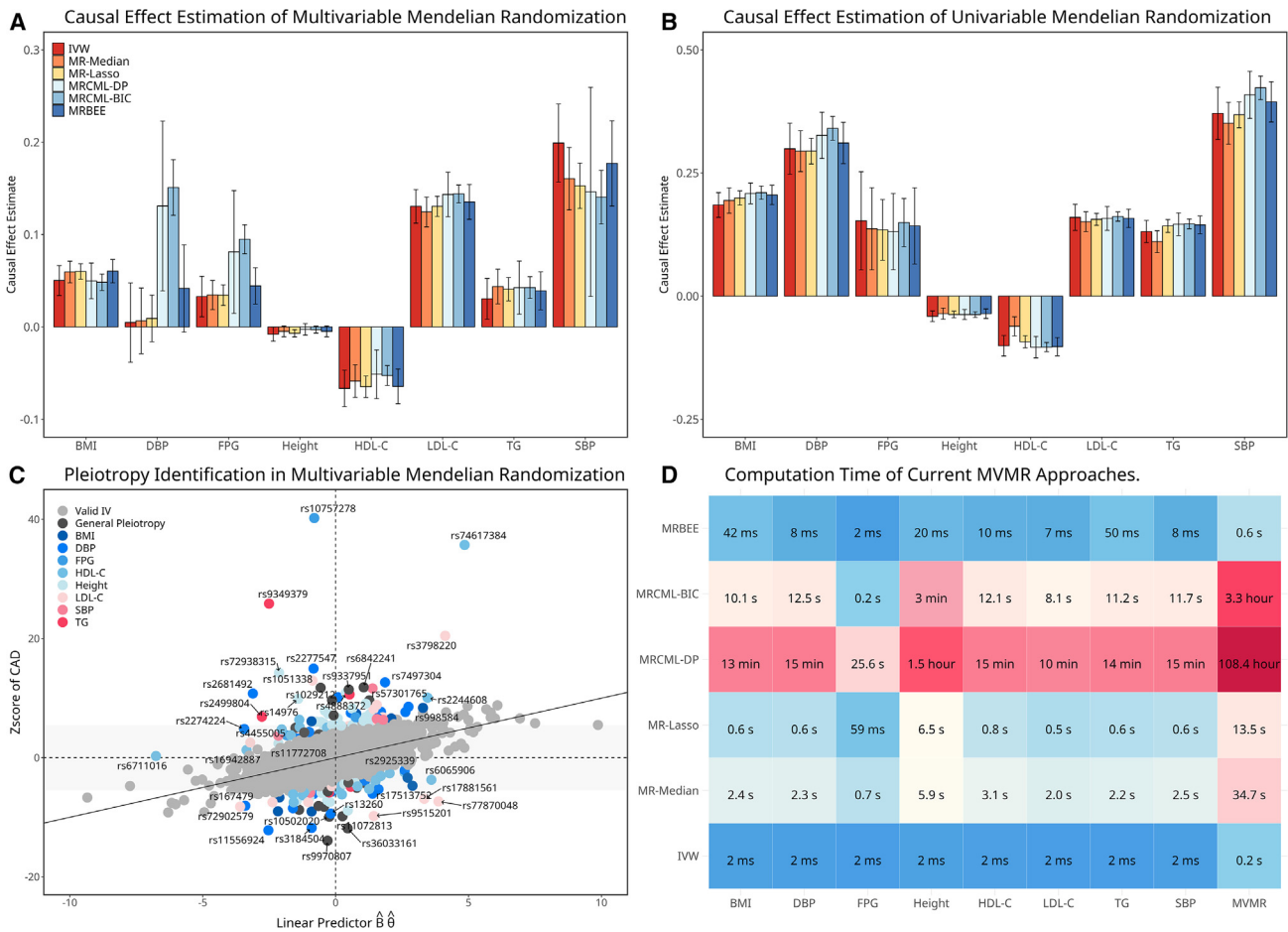


Figure 7. Data analysis of coronary artery disease

(A) Causal effect estimates by the six MVMR methods. The corresponding 95% confidence interval shown as vertical error bars.

(B) Causal effect estimates by UVMR approaches. The corresponding 95% confidence interval shown as vertical error bars. The radius of the confidence interval equals $\sqrt{a} \times \text{SE}$ where $a = F_{\alpha}^{-1}(1 - 0.05/8, 1)$.

(C) Pleiotropy test where the x axis represents the linear predictor $\hat{B}\hat{\theta}_{\text{BEE}}$ and the y axis represents the standardized association between the IV and CAD from GWAS. The annotation of this pleiotropic variant is made if it is associated with the most significant exposures with $p < 5E-8$.

(D) Computation times of the comparing methods. Columns 1–8 represent computation time for UVMR of eight exposures and the last column represents computation time for MVMR.

reach genome-wide significance level in the outcome GWASs, suggesting GWPT captures pleiotropic evidence and the standard GWAS does not. We also performed expression quantitative trait loci (eQTL) mapping for the identified loci using functional mapping of GWAS (FUMA GWAS).³⁹ Each SNP that tagged a locus had marginal evidence of association with the expression of a gene in that locus in at least one tissue, where association p values ranged from $3.3E-310$ to $6.7E-5$. This suggests that these loci may have functionally relevant consequences in their conferred risk for myopia, SCZ, or CAD.

Discussion

We proposed MRBEE to overcome the weak instrument, pleiotropy and sample overlap bias in MVMR analysis. We pointed out that weak instrument bias is essentially

driven by measurement error of GWAS effect estimates, whose scale and bias direction are influenced by the degree of weakness of IVs and the GWAS sample overlap, respectively. An IV is not considered weak when the estimation error is negligible, which can be achieved with a sufficiently large GWAS sample size, no matter how large or small the effect size is. In genetics, Burgess et al.³ suggested using the F-statistics to define the strength of an IV, whereas we recommend the reliability ratio ([material and methods, Equation 11](#)), a commonly used statistic in measurement error analysis. Both metrics are equivalent and will be influenced by the GWAS sample size and the number of IVs, highlighting that the definition of a weak instrument is dynamic. MRBEE removes the measurement error bias by using an unbiased estimating function. Although this estimating function has a long history in the literature of measurement error analysis,⁴⁰ it has not been utilized to modify the current MVMR approaches.

Table 2. Loci detection of GWPT and eQTL mapping of leading variant

	SNP information		Association test				eQTL mapping			
	SNP	CHR:BP	GWAS	MTAG	CPASSOC	GWPT	Symbol	Tissue	Database	<i>p</i>
Myopia	rs55761633	1:20757820	9.9e-07	3.1e-05	5.6e-07	9.6e-09	CAMK2N1	Muscle Skeletal	GTEEx/v8	3.9e-06
	rs2419964	2:124252256	1.3e-07	1.2e-05	2.6e-10	6.0e-10	NA	NA	NA	NA
	rs7602460	2:182261869	4.9e-07	9.2e-06	6.6e-07	3.2e-08	ITGA4	Blood	eQTLGen	2.0e-19
	rs61548163	2:184349492	9.2e-07	1.3e-05	3.3e-06	2.7e-08	NA	NA	NA	NA
	rs6764842	3:123106287	1.6e-06	NA	2.0e-07	3.1e-08	ADCY5	Artery Tibial	GTEEx/v8	3.0e-17
	rs9761983	4:138482973	1.5e-07	NA	9.0e-06	4.4e-08	RP11-714L20.1	Cortex	GTEEx/v8	2.2e-06
	rs2461726	6:166316838	5.1e-07	1.7e-05	1.2e-06	1.1e-08	SDIM1	Pituitary	GTEEx/v7	2.8e-07
	rs12699288	7:11975557	3.8e-06	1.4e-04	3.4e-08	1.4e-08	THSD7A	Nerve Tibial	GTEEx/v8	6.7e-05
	rs2970498	7:30478056	1.1e-07	1.8e-06	7.0e-06	3.1e-08	NOD1	Blood	eQTLGen	1.0e-07
	rs1532278	8:27466315	3.4e-07	1.1e-05	9.6e-07	6.7e-09	CLU	Eye	EyeGEx	1.1e-26
	rs7048915	9:4206388	1.0e-07	2.0e-06	6.2e-06	1.3e-08	NA	NA	NA	NA
	rs902997	10:105384262	1.9e-07	9.2e-07	1.2e-08	3.2e-09	USMG5	Blood	eQTLGen	3.0e-37
	rs17065719	13:44925021	4.3e-07	1.1e-05	2.6e-05	3.7e-08	SERP2	Blood	eQTLGen	7.7e-48
	rs1926715	13:111538590	8.6e-08	1.4e-05	2.4e-12	2.0e-09	ANKRD10	Eye	EyeGEx	3.0e-48
	rs7141076	14:67922172	9.2e-08	1.8e-05	5.7e-06	1.7e-08	TMEM229B	Pituitary	GTEEx/v8	1.1e-08
rs12889206	14:68769182	8.8e-08	1.5e-06	1.2e-06	3.9e-08	NA	NA	NA	NA	
rs7198357	16:67884619	2.5e-07	4.2e-06	2.1e-09	4.3e-09	DUS2	Blood	eQTLGen	3.3e-310	
rs35594082	16:84796864	8.5e-07	1.8e-05	1.8e-06	3.1e-08	USP10	Eye	EyeGEx	2.9e-09	
Schizophrenia	rs17672204	5:74946518	1.1e-06	8.2e-06	1.9e-06	2.4e-08	COL4A3BP	Muscle Skeletal	GTEEx/v8	5.8e-15
	rs79650876	3:187997616	1.7e-07	4.3e-08	4.5e-07	3.2e-08	AC022498.1	Blood	eQTLGen	5.7e-06
	rs2300921	3:185651001	8.0e-06	8.9e-06	4.9e-06	3.2e-08	TRA2B	Breast	GTEEx/v8	1.8e-06
	rs7225476	17:78561603	8.2e-07	2.1e-06	7.0e-05	3.3e-08	RPTOR	Blood	eQTLGen	4.0e-89
Coronary artery disease	rs2045886	2:29010517	3.6e-07	4.6e-05	2.7e-21	7.7e-11	PPP1CB	Blood	eQTLGen	3.3e-310
	rs6727524	2:238570309	8.9e-07	6.3e-05	4.4e-09	2.8e-08	LRRFIP1	Blood	eQTLGen	3.4e-76
	rs1868217	3:98445534	2.1e-05	4.5e-04	1.3e-10	3.6e-08	ST3GAL6	Blood	eQTLGen	2.0e-30
	rs73070809	3:186885760	1.1e-07	NA	3.7e-07	1.0e-08	RPL39L	Adipose	GTEEx/v8	4.0e-05
	rs12523133	5:86297919	8.5e-08	2.3e-05	3.2e-19	2.5e-10	RP11-72L22.1	Spinal Cord	GTEEx/v8	6.6e-08
	rs6899197	5:111250597	8.8e-06	NA	1.2e-20	2.2e-09	EPB41L4A	Esophagus	GTEEx/v8	5.0e-05
	rs13202921	6:41687366	3.3e-07	2.9e-06	1.8e-08	3.8e-08	CCDC77	Artery Coronary	GTEEx/v7	4.9e-11

(Continued on next page)

Table 2. Continued

SNP information		Association test				eQTL mapping			
SNP	CHR:BP	GWAS	MTAG	CPASSOC	GWPT	Symbol	Tissue	Database	<i>p</i>
rs2073533	7:14029739	1.1e-07	2.6e-05	5.2e-14	3.2e-11	NA	NA	NA	NA
rs7822979	8:106468592	6.0e-08	NA	1.5e-08	1.6e-10	NA	NA	NA	NA
rs4734881	8:106587829	7.8e-08	3.1e-05	5.3e-06	1.3e-10	ZPPM2	Esophagus	GTEX/v8	3.8e-08
rs12375254	8:125054365	3.2e-07	4.0e-06	2.5e-04	1.0e-08	TRMT12	Blood	eQTLGen	4.3e-15
rs12412313	10:134456762	2.4e-06	6.8e-03	2.7e-19	2.3e-10	INPP5A	Blood	eQTLGen	6.6e-13
rs7113595	11:70236819	7.8e-07	2.2e-05	2.7e-08	2.4e-08	PPFIA1	Blood	eQTLGen	5.3e-305
rs7315852	12:417633	3.0e-07	NA	7.7e-19	2.1e-11	CCDC77	Blood	eQTLGen	3.3e-310
rs5893521	15:8395536	2.2e-07	1.7e-04	3.4e-20	2.9e-09	BTBD1	Brain	xQTLServer	3.4e-75
rs12918327	16:30626616	6.8e-06	9.7e-04	2.8e-32	5.2e-10	STX4	Blood	eQTLgen	1.1e-85
rs9958798	18:52769637	4.0e-06	NA	9.9e-10	4.2e-08	RP11-99A1.2	Testis	GTEX/v8	6.3e-08
rs35496634	22:39147235	8.5e-06	1.2e-03	4.6e-17	3.4e-08	SUN2	Blood	eQTLGen	7.0e-79
rs757949	22:40820151	2.9e-06	9.3e-06	5.2e-15	3.8e-08	MKL1	Eye	EyeGEx	1.6e-07

Our simulations suggested that MRBEE in general leads to equal or less bias of causal effect estimate than the comparing methods when weak IVs, pleiotropy, and sample overlap are present (Figure 2). Similarly, MRBEE also has equal or better type I error control and statistical power than robust comparing methods (Figure 3). MRcML-DP and MRcML-BIC were robust to weak IVs and consistently yield unbiased causal effect estimates under the “no pleiotropy” case. However, in the presence of horizontal pleiotropy, our simulations suggested that MRcML methods may produce local minimizers in some specific scenarios in which horizontal pleiotropy was not completely removed. MRcML employs the best subset selection for detecting pleiotropy and the algorithm’s stability and time consumption could be a challenge,⁴¹ as we observed in our simulations. MRBEE uses an iterative pleiotropy test, whose reliability has been validated in MR-PRESSO and IMRP.

In the myopia analysis, our detected causal effects for outdoor activities are consistent with the literature. For example, spending more time outdoors reducing incident myopia was confirmed by a randomized clinical trial.⁴² On the other hand, near-work activities such as time spent watching TV or using the computer have not been found to be associated with myopia risk.⁴³ The potential biological mechanism is that outdoor activities increase the exposure time to natural light, which induces the release of dopamine and thereby inhibits axial elongation, thus suppressing the development of myopia.³⁰ Moreover, MRBEE yields a relatively large causal estimate for time spent driving, likely correcting for weak instrument bias given the small variation of driving time explained by the IVs. Although MRcML-DP and MRcML-BIC can effectively reduce weak instrument bias in simulations, their estimates for the effects of driving time were 3–5 times larger than those from other methods.

We observed that cannabis use disorder and education have substantially larger causal effects on SCZ than other exposures we examined. For LH, the current GWAS has identified four genome-wide significant IVs together explaining its 0.086% variation. As a result, we did not have sufficient power to confirm their causal effect due to its relatively smaller variance of LH explained by IVs. MRBEE did not identify pleiotropic variants in these data, suggesting that our study may already include most of the direct causal risk factors for SCZ.

Our MVMR analysis seems to suggest that HDL-C is likely a protective factor against CAD but with a weaker effect size than that from UVMR analysis, aligning with recent pharmaceutical trial outcomes.⁴⁴ The previously observed negative results^{20,33} are likely because they did not utilize the lipid GWAS summary data with the largest available sample sizes. When using the largest GWAS summary statistics of CAD as in this study, all methods including IVW, MRcML-DP, and MRBEE resulted in significant protective causal effect of HDL-C on CAD (Figure 7A). We noted that the estimated equal contributions of DBP and SBP on CAD risk by MRcML-BIC and MRcML-DP,

which is in direct conflict with all other MR methods we tested and the literature.⁴⁵ In addition, MRBEE identified 173 pleiotropic IVs, one of which (rs10757278) is strongly associated with CAD ($p < 5E-300$) but whose biological mechanism warrants for further investigation.

We introduced the GWPT using the statistic S_{pleio} , which can be applied in UVMR or MVMR to identify specific IVs with evidence of horizontal pleiotropy. When S_{pleio} was applied to the whole genome, we identified genetic loci associated with myopia, SCZ, and CAD that were missed in their original GWAS. These loci also reflect their direct association with the outcomes or through exposures not included in this study. Genes in these loci had genome-wide significant eQTLs across a range of tissues, suggesting that these genes might be functionally relevant in modifying disease risk. For example, we identified the *RPTOR* gene for SCZ, which has previously been found to be associated with BMI⁴⁶ and blood pressure.⁴⁷ This gene also has significant eQTLs (smallest $p = 4E-89$) in blood tissue. This and other examples highlight the potential utility of S_{pleio} in identifying trait-associated loci and functionally relevant genes.

In our theoretical study (supplement 2), we consider the effect size of IVs to follow a normal distribution, representing a genomic random effect model.⁴⁸ We observed that increasing the sample size of GWAS often yields more novel loci, hence more IVs with non-zero effects can be used in a corresponding MR analysis. Therefore, in our theoretical investigation, we allow the number of IVs m to increase with the sample size n and examine the outcomes of MRBEE and MV-IVW under different rates of m and n . Our conclusion can be summarized as follows: for scenarios like those in our myopia and CAD data, where GWAS sample sizes for exposures are approximately half a million or more, MRBEE and MV-IVW are equally efficient (supplement 2, Theorem 1.3 (i)). In this case, MRBEE's inference is asymptotically valid, whereas MV-IVW may lead to incorrect inferences. For the SCZ data involving CUD, with GWAS sample sizes in the tens of thousands, MRBEE is less efficient than MV-IVW, but the inference made by MRBEE remains valid (supplement 2, Theorem 1.3 (ii)-(iii)). In these cases, the confidence intervals of MRBEE will be wider than MV-IVW but ensure the 95% coverage frequency. Although MRBEE can remove the weak instrument bias in general, we still recommend including the IVs with the association p values below a significance threshold. The reason is that weak IVs still require to be truly associated with an exposure although their effect sizes can be extremely small. Variants with the association p values above the threshold are likely to be false positive and including false positive IVs will lead to bias for MRBEE because of the violation of assumption (IV1). The purpose of developing MRBEE is to enhance existing methods, making causal effect estimation and inference more robust to weak IVs.

The comparison between MRBEE and MRcML in terms of statistical principle is as follows. MRBEE employs the unbi-

ased estimating function method which constructs its unbiased score function from the score function of the MV-IVW method. In contrast, the MRcML method is a conditional score function method, characterized by first estimating the sufficient statistic containing parameters to be estimated and then estimating the parameter based on this sufficient statistic through an iterative method.⁴⁰ Although Stefanski and Carroll²⁴ demonstrated that the conditional score function possesses statistical efficiency, whether this conclusion can be directly applied to the MRcML method requires further investigation. In contrast, our investigation shows that MRBEE reaches statistical efficiency if $m/n_{\text{min}} \rightarrow 0$ where n_{min} is the minimum GWAS sample size (supplement 2, Theorem 1.3 (i)). Furthermore, our simulations in Section S1.3 of supplement 1 suggest that MRcML-DP tends to overestimate its SD (i.e., $SE > SD$), MRcML-BIC underestimates its SD (i.e., $SE < SD$), and MRBEE estimates its SD well in most cases, suggesting MRBEE can achieve more efficiency than MRcML. The exact reason that MRcML does not estimate SD well warrants further investigation.

MRBEE also has some limitations. MRBEE is ineffective in handling exposures associated with significantly weaker IVs, such as CUD and LH in the SCZ data. This is also a challenge inherited from the field of measurement error analysis. In this case, MRBEE and analogous methods such as MRcML tend to produce causal effect estimates with relatively large SE. MRBEE is effective when the proportion of pleiotropic variants is relatively low (e.g., below 30%). Incorporating a Gaussian mixture model with MRBEE might improve the robustness for scenarios with a high proportion of pleiotropic variants. Finally, MRBEE is designed to handle a fixed number of exposures. Expanding its capability to a high-dimensional MR model is warranted in future research.⁴⁹

Last, it is worth offering guidance on how to perform MVMR analysis from our perspective. First, rather than selecting the optimal number of IVs such that the F-statistics and conditional F-statistics are larger than 10,^{3,22} we suggest including all independent IVs that are genome-wide significantly associated with at least one exposure. The main purpose of doing this is to reduce the winner's curse. Our simulations found that all methods, including MRBEE, were affected by the winner's curse, and the only way to alleviate the winner's curse was to include as many causal variants as possible (supplement 1, and Figure S10). Besides, our theory (supplement 2, Theorem 1.2 and 1.3) illustrates that the asymptotic variance of a causal effect estimate is related to the cumulative variance explained by all specified IVs instead of the average variance explained by each IV. Hence, including more IVs in the MR model can reduce the variance of the related causal effect estimate. Second, when performing MVMR analysis, it is not necessary to remove variants that are pleiotropic between the exposures. The reason why Wang et al.²¹ found that LDL-C was not significant in European populations is likely caused by this procedure. In contrast, simultaneously including all the relevant exposures and their

IVs is recommended because the multivariable regression can automatically account for the pleiotropic variants shared by the specified exposures. Third, we suggest conducting a GWPT after performing the MR analysis, which represents an effective multi-trait approach for discovering loci with pleiotropy effect, beyond current methods such as CPASSOC and MTAG. In statistical principle, GWPT is likely to identify new loci associated with the outcome if the effect directions of pleiotropy and exposure mediation are opposite in these genome regions.

During the revision of this manuscript, we noted that a recent preprint⁵³ claimed that MRBEE was biased with extremely large SD and SE for some of simulation scenarios. We observed that the reason was that the authors of the preprint did not perform the standardization for the instrument effects of both exposures and outcome, which was documented in the MRBEE software. We present the reproduction of Table 1 in the preprint before and after the standardization in Table S25 in supplement 1. The result indicates MRBEE has reasonable SD and SE. We have updated MRBEE software on GitHub and it does not need the standardization now.

Material and methods

MR model

We describe MRBEE with details here. As in the main text, let $\mathbf{g}_i = (g_{i1}, \dots, g_{im})^\top$ be a vector of m independent genetic variants where each variant is standardized with mean zero and variance one, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ be a vector of p exposures, and y_i be an outcome. Consider the following linear structural model:

$$\mathbf{x}_i = \mathbf{B}^\top \mathbf{g}_i + \mathbf{u}_i, \quad (\text{Equation 1})$$

$$y_i = \theta^\top \mathbf{x}_i + \gamma^\top \mathbf{g}_i + v_i, \quad (\text{Equation 2})$$

where $\mathbf{B} = (\beta_1, \dots, \beta_m)^\top$ is an $(m \times p)$ matrix of genetic effects on exposures with $\beta_j = (\beta_{j1}, \dots, \beta_{jp})^\top$ being a vector of length p , $\theta = (\theta_1, \dots, \theta_p)^\top$ is a vector of length p representing the causal effects of the p exposures on the outcome, $\gamma = (\gamma_1, \dots, \gamma_m)^\top$ is a vector of length m representing horizontal pleiotropy, which may violate the (IV2) or (IV3) conditions, and \mathbf{u}_i and v_i are noise terms. Substituting for \mathbf{x}_i in (2), we obtain the reduced-form model:

$$y_i = \mathbf{g}_i^\top \alpha + \mathbf{u}_i^\top \theta + v_i, \quad (\text{Equation 3})$$

where

$$\alpha = \mathbf{B}\theta + \gamma. \quad (\text{Equation 4})$$

In practice, \mathbf{u}_i and v_i are usually correlated, and hence traditional linear regression between \mathbf{x}_i and y_i cannot obtain a consistent estimate of θ . In contrast, the genetic variant vector \mathbf{g}_i is assumed to be independent of the noise terms \mathbf{u}_i and v_i because the genotypes of individuals are randomly inherited from their parents and do not change during their lifetime.⁵⁰ Hence, \mathbf{g}_i can be used as IVs to remove the confounding effect of \mathbf{u}_i and v_i .

We assume that the genetic effect β_j ($j = 1, \dots, m$) is a p -dimensional random vector with zero-mean, covariance matrix $\Sigma_{\beta\beta}$, and cumulative covariance matrix $\Psi_{\beta\beta}$:

$$\Sigma_{\beta\beta} = E(\beta_j \beta_j^\top), \Psi_{\beta\beta} = m \Sigma_{\beta\beta}.$$

The covariance matrix $\Sigma_{\beta\beta}$ will vanish as $m \rightarrow \infty$, but the cumulative covariance matrix $\Psi_{\beta\beta}$ is still a constant matrix, representing the total genetic covariance contributed from the m IVs. The genetic variant g_{ij} ($i = 1, \dots, n$, $j = 1, \dots, m$) is standardized so that $E(g_{ij}) = 0$ and $\text{var}(g_{ij}) = 1$, and all IVs are assumed to be in linkage equilibrium (LE), i.e., $\text{cov}(g_{ij}, g_{ik}) = 0$ for $j \neq k$. Next, the noise terms \mathbf{u}_i and v_i have zero-means and joint covariance matrix:

$$\Sigma_{\mathbf{u} \times v} = \text{cov}((\mathbf{u}_i^\top, v_i)^\top) = \begin{pmatrix} \Sigma_{uu} & \sigma_{uv} \\ \sigma_{uv}^\top & \sigma_{vv} \end{pmatrix}.$$

Thus, the exposure \mathbf{x}_i and outcome y_i have zero-means and joint covariance matrix:

$$\Sigma_{\mathbf{x} \times y} = \text{cov}((\mathbf{x}_i^\top, y_i)^\top) = \begin{pmatrix} \Sigma_{xx} & \sigma_{xy} \\ \sigma_{xy}^\top & \sigma_{yy} \end{pmatrix},$$

$\Sigma_{xx} = \Psi_{\beta\beta} + \Sigma_{uu}$, $\sigma_{xy} = \Psi_{\beta\beta}\theta + \Sigma_{uu}\theta + \sigma_{uv}$, and $\sigma_{yy} = \theta^\top \Psi_{\beta\beta}\theta + \theta^\top \Sigma_{uu}\theta + 2\theta^\top \sigma_{uv} + \sigma_{vv}$. Note that $\sigma_{uv} \neq \mathbf{0}$ means the confounders affect both \mathbf{x}_i and y_i .

Bias of multivariable IVW estimate

Since large individual-level data from GWAS are less publicly available, most of the current MR analyses are performed with summary statistics of IVs through the following linear regression:

$$\hat{\alpha}_j = \hat{\beta}_j^\top \theta + \gamma_j + \varepsilon_j, \quad (\text{Equation 5})$$

where $\hat{\alpha}_j$ and $\hat{\beta}_j$ are respectively estimated from the outcome and exposure GWASs, γ_j is the horizontal pleiotropy, ε_j represents the residual of this regression model, and $j = 1, \dots, m$ referring to the m IVs. MV-IVW, which is the foundation of most existing MR methods, estimates θ by

$$\begin{aligned} \hat{\theta}_{\text{IVW}} &= \underset{\theta}{\text{argmin}} \{ (\hat{\alpha} - \hat{\mathbf{B}}\theta)^\top \mathbf{V}^{-1} (\hat{\alpha} - \hat{\mathbf{B}}\theta) \} \\ &= (\hat{\mathbf{B}}^\top \mathbf{V}^{-1} \hat{\mathbf{B}})^{-1} \hat{\mathbf{B}}^\top \mathbf{V}^{-1} \hat{\alpha} \end{aligned}, \quad (\text{Equation 6})$$

where \mathbf{V} is a diagonal matrix consisting of the variance of estimation errors of $\hat{\alpha}$. In practice, it is routine to standardize $\hat{\alpha}_j$ and $\hat{\beta}_{jk}$ by $\hat{\alpha}_j / \text{se}(\hat{\alpha}_j)$ and $\hat{\beta}_{jk} / \text{se}(\hat{\beta}_{jk})$ to remove the minor allele frequency effect.¹⁶ With this standardization, the MV-IVW estimates θ by

$$\hat{\theta}_{\text{IVW}} = \underset{\theta}{\text{argmin}} \{ \|\hat{\alpha} - \hat{\mathbf{B}}\theta\|_2^2 \} = (\hat{\mathbf{B}}^\top \hat{\mathbf{B}})^{-1} \hat{\mathbf{B}}^\top \hat{\alpha}. \quad (\text{Equation 7})$$

However, the MV-IVW estimate $\hat{\theta}_{\text{IVW}}$ is biased due to the estimation errors of $\hat{\alpha}_j$ and $\hat{\beta}_j$:

$$\hat{\alpha}_j = \alpha_j + w_{\alpha_j}, \quad (\text{Equation 8})$$

$$\hat{\beta}_j = \beta_j + \mathbf{w}_{\beta_j}. \quad (\text{Equation 9})$$

To see this, observe the estimating equation and Hessian matrix of $\hat{\theta}_{\text{IVW}}$:

$$\mathbf{S}_{\text{IVW}}(\theta) = \hat{\mathbf{B}}^\top (\hat{\mathbf{B}}\theta - \hat{\alpha}), \mathbf{H}_{\text{IVW}} = \hat{\mathbf{B}}^\top \hat{\mathbf{B}}.$$

That is, $\mathbf{S}_{\text{IVW}}(\theta)$ is the score function of Equation 7 and $\hat{\theta}_{\text{IVW}}$ is estimated by solving $\mathbf{S}_{\text{IVW}}(\theta_{\text{IVW}}) = \mathbf{0}$, and \mathbf{H}_{IVW} is the 2nd derivative matrix of Equation 7. As shown in supplement 2, since $\hat{\theta}_{\text{IVW}} - \theta = -\mathbf{H}_{\text{IVW}}^{-1} \mathbf{S}_{\text{IVW}}(\theta)$, the bias of $\hat{\theta}_{\text{IVW}}$ is approximately:

$$\begin{aligned} E(\widehat{\theta}_{IVW} - \theta) &\approx -E(\mathbf{H}_{IVW})^{-1}E(\mathbf{S}_{IVW}(\theta)) \\ &= -\{\Sigma_{\beta\beta} + \Sigma_{W_\beta W_\beta}\}^{-1}\{\Sigma_{W_\beta W_\beta}\theta - \sigma_{W_\beta w_\alpha} + \sigma_{\beta\gamma}\}, \end{aligned} \quad (\text{Equation 10})$$

where

$$\begin{aligned} \text{cov}\left(\left(\mathbf{w}_{\beta_j}^\top, w_{\alpha_j}\right)^\top\right) &= \Sigma_{W_\beta \times w_\alpha} = \begin{pmatrix} \Sigma_{W_\beta W_\beta} & \sigma_{W_\beta w_\alpha} \\ \sigma_{W_\beta w_\alpha}^\top & \sigma_{w_\alpha w_\alpha} \end{pmatrix}, \text{cov}(\beta_j, \gamma_j) \\ &= \sigma_{\beta\gamma}. \end{aligned}$$

Interpretation of weak instrument bias

Here, $\Sigma_{\beta\beta}$ can be regarded as the average information carried by each IV, while $\Sigma_{W_\beta W_\beta}$ can be regarded as the information carried by each estimation error. If $\Sigma_{\beta\beta}$ is not substantially larger than $\Sigma_{W_\beta W_\beta}$, then the weak instrument will inflate the measurement error bias by the multiplier $(\Sigma_{\beta\beta} + \Sigma_{W_\beta W_\beta})^{-1}$. This is the primary reason why violating assumption (IV1) introduces bias into causal effect estimates in IVW and other MR approaches.²⁶

The covariance between the estimation errors of SNP-exposure and SNP-outcome associations $\sigma_{W_\beta w_\alpha}$ can be affected by the fraction of overlapping samples of the exposures and outcome GWAS. If the exposures and outcome GWAS are independent of each other, then $\sigma_{W_\beta w_\alpha} = \mathbf{0}$ and hence the measurement error bias always shrinks $\widehat{\theta}_{IVW}$ toward the null. In contrast, if the exposures GWAS and outcome GWAS are estimated from the same cohorts, $\sigma_{W_\beta w_\alpha}$ usually introduces bias toward the direction of σ_{IV} , reflecting the degree of sample overlap between exposures and outcome. This is the reason why in some empirical studies,^{23,27} IVW cannot completely remove the confounding bias if the overlapping sample fraction is large.

If $\sigma_{\beta\gamma} \neq \mathbf{0}$, there is additional pleiotropy bias due to the horizontal pleiotropy that violates the InSIDE assumption. In UVMR, it is challenging to guarantee $\gamma_j = 0$ or $\text{cov}(\gamma_j, \beta_j) = \mathbf{0}$ for all $1 \leq j \leq m$, resulting in a potentially biased IVW estimate. Traditional solutions to horizontal pleiotropy bias require that only a small proportion of IVs exhibit horizontally pleiotropic effects.^{5,7,12} However, for complex traits, it is plausible that a large portion of IVs (even possibly > 50%) possess horizontally pleiotropic effects, leading to the failure of UVMR methods. MVMR can balance these pleiotropic effects shared by multiple exposures, significantly reducing the number of IVs with horizontal pleiotropy evidence when conditioned on specified exposures. In other words, it is more likely to guarantee that only few IVs violate the InSIDE assumption $\sigma_{\beta\gamma} = \mathbf{0}$ after accounting for multiple exposures, which can be then detected and removed using the robust tools such as a pleiotropy hypothesis test.

Reliability ratio

In practice, we suggest using the reliability ratio⁴⁰:

$$\omega_k = \frac{\text{var}(\beta_{jk})}{\text{var}(\widehat{\beta}_{jk})} \quad (\text{Equation 11})$$

to measure the degree of bias in $\widehat{\theta}_{k,IVW}$, which can be empirically estimated by

$$\widehat{\omega}_k = \frac{\sum_{j=1}^m (\widehat{\beta}_{jk}^2 - \text{var}(w_{\beta_{jk}}))}{\sum_{j=1}^m \widehat{\beta}_{jk}^2}. \quad (\text{Equation 12})$$

$\widehat{\omega}_k$ reflects the proportion of variability in the estimated effects attributable to the underlying true genetic effects. For example, a reliability ratio of 0.6 indicates that 60% of the variance of the estimated effects is attributable to the true effects and the rest is attributable to their estimation errors. From the perspective of measurement error theory, the IVW estimate $\widehat{\theta}_{IVW}$ converges to $\omega\theta$ in a univariable MR analysis when there is no sample overlap, where ω is equal to $\text{var}(\beta_j)/\text{var}(\widehat{\beta}_j)$.⁴⁰ Here ω is less than 1 and is viewed as a shrinkage coefficient for $\widehat{\theta}_{IVW}$ relative to the true effect θ . We adopt this reliability ratio to much broader contexts, such as multivariable MR and sample overlap. In our real data analysis, we found this reliability ratio works reasonably well although additional investigation is warranted. While the reliability ratio and the F-statistics³ are similar, the former has a simpler calculation and can more clearly reflect the proportion of weak IV bias than the latter.

MR using bias-corrected estimating equation

We propose MRBEE, which estimates causal effects by solving a new unbiased estimating equation of causal effects. The unbiased estimating equation of θ is

$$\mathbf{S}_{BEE}(\theta) = \mathbf{S}_{IVW}(\theta) - m(\Sigma_{W_\beta W_\beta}\theta - \sigma_{W_\beta w_\alpha}), \quad (\text{Equation 13})$$

where $\mathbf{S}_{IVW}(\theta) = -\widehat{\mathbf{B}}^\top(\widehat{\alpha} - \widehat{\mathbf{B}}\theta)$. Equation 13 states that the MRBEE estimating function is equal to the IVW estimating equation minus its bias. Unbiasedness of the MRBEE estimating equation implies unbiasedness of the MRBEE estimator. The solution $\widehat{\theta}_{BEE}$ such that $\mathbf{S}_{BEE}(\widehat{\theta}_{BEE}) = \mathbf{0}$ is

$$\widehat{\theta}_{BEE} = (\widehat{\mathbf{B}}^\top \widehat{\mathbf{B}} - m\Sigma_{W_\beta W_\beta})^{-1}(\widehat{\mathbf{B}}^\top \widehat{\alpha} - m\sigma_{W_\beta w_\alpha}). \quad (\text{Equation 14})$$

Note that unlike other optimizations such as generalized linear model in measurement error,⁴⁰ the Hessian matrix $\mathbf{H}_{IVW} = \widehat{\mathbf{B}}^\top \widehat{\mathbf{B}}$ does not involve θ and hence $\mathbf{S}_{BEE}(\theta)$ can be directly obtained from $\mathbf{S}_{IVW}(\theta)$ without any iterative approximation.

Bias-correction terms estimation

We estimate the bias-correction terms $\Sigma_{W_\beta W_\beta}$ and $\sigma_{W_\beta w_\alpha}$ from the insignificant and independent GWAS summary statistics.³⁷ Let $\widehat{\alpha}_j^*, \widehat{\beta}_{j1}^*, \dots, \widehat{\beta}_{jp}^*$ ($j = 1, \dots, M$) be M insignificant GWAS effect size estimates of outcome and exposures, where the insignificance means that the p value of the genetic variants are larger than 0.05 for all exposures and outcome, and the independence means that they are not in LD. Because $\widehat{\alpha}_j^*$ and $\widehat{\beta}_{jk}^*$ follow the same distributions of w_{α_j} and $w_{\beta_{jk}}$, $\Sigma_{W_\beta \times w_\alpha}$ can be estimated by

$$\widehat{\Sigma}_{W_\beta \times w_\alpha} = \frac{1}{M} \sum_{j=1}^M \left(\widehat{\beta}_{j1}^*, \dots, \widehat{\beta}_{jp}^*, \widehat{\alpha}_j^* \right)^\top \left(\widehat{\beta}_{j1}^*, \dots, \widehat{\beta}_{jp}^*, \widehat{\alpha}_j^* \right). \quad (\text{Equation 15})$$

Here, $\widehat{\Sigma}_{W_\beta W_\beta}$ is the first $(p \times p)$ sub-matrix of $\widehat{\Sigma}_{W_\beta \times w_\alpha}$ and $\widehat{\sigma}_{W_\beta w_\alpha}$ consists of the first p elements of the last column of $\widehat{\Sigma}_{W_\beta \times w_\alpha}$. The intercept provided by LDSC³⁵ is also a consistent estimate of $\text{cov}(w_{\alpha_j}, w_{\beta_{jk}})$. Each of these two estimators may be used by MRBEE and experience with real data suggests that they generally produce similar results. LDSC requires specification of an LD reference panel that is from an ancestrally similar population to that under study in MR. Differences in genetic architecture between the LD reference panel and the MR study population could introduce bias. Use of Equation 15 does not require an LD reference panel and so will not be biased for this reason. Additionally, use of Equation 15 is computationally simpler.

Algorithm 1. Pseudo-code of MRBEE + pleiotropy test

1. Input: Initial estimates $\hat{\theta}^{(0)}$, Bias-correction terms $\hat{\Sigma}_{W_\beta W_\beta}$ and $\hat{\sigma}_{W_\beta W_\alpha}$, S_{pleio} , FDR q-value threshold κ , Tolerance ϵ , Full set of m^* IVs
2. Output: Causal effect estimates $\hat{\theta}_{\text{BEE}}$, Set of m non-UHP/CHP IVs \mathcal{F}_θ
3. Pseudo-code:
Initialize $\mathcal{F}_\theta^{(0)} = \{j : j = 1, \dots, m^*\}$
While $\|\hat{\theta}^{(t+1)} - \hat{\theta}^{(t)}\|_2 > \epsilon$
1. Calculate $S_{\text{pleio}}^{(t)}(\hat{\theta}^{(t)})$ for all $j = 1, \dots, m^*$,
2. Update $\mathcal{F}_\theta^{(t+1)} = \{j : S_{\text{pleio}}^{(t)}(\hat{\theta}^{(t)}) < F_{\chi^2}^{-1}(1 - \kappa)\}$,
3. Update $\hat{\theta}^{(t+1)}$ using Equation 14 and IVs in $\mathcal{F}_\theta^{(t+1)}$
End While

SE estimation

The covariance matrix of $\hat{\theta}_{\text{BEE}}$ is yielded through the sandwich formula:

$$\text{cov}(\hat{\theta}_{\text{BEE}}) := \Sigma_{\text{BEE}}(\theta) = F_{\text{BEE}}^{-1} V_{\text{BEE}}(\theta) F_{\text{BEE}}^{-1}, \quad (\text{Equation 16})$$

where the outer matrix F_{BEE} is the Fisher information matrix, i.e., the expectation of the Hessian matrix of $\mathbf{S}_{\text{BEE}}(\theta)$, and the inner matrix $V_{\text{BEE}}(\theta)$ is the covariance matrix of $\mathbf{S}_{\text{BEE}}(\theta)$. A consistent estimate of $\Sigma_{\text{BEE}}(\theta)$ is

$$\hat{\Sigma}_{\text{BEE}}(\theta) = \hat{F}_{\text{BEE}}^{-1} \hat{V}_{\text{BEE}}(\hat{\theta}_{\text{BEE}}) \hat{F}_{\text{BEE}}^{-1}, \quad (\text{Equation 17})$$

where $\hat{F}_{\text{BEE}} = \mathbf{B}^\top \mathbf{B} - \hat{\Sigma}_{W_\beta W_\beta}$, $\hat{V}_{\text{BEE}}(\hat{\theta}_{\text{BEE}}) = \sum_{j=1}^m \hat{\mathbf{S}}_j(\hat{\theta}_{\text{BEE}}) \hat{\mathbf{S}}_j(\hat{\theta}_{\text{BEE}})^\top$, and $\hat{\mathbf{S}}_j(\hat{\theta}_{\text{BEE}}) = -(\hat{\alpha}_j - \hat{\theta}_{\text{BEE}}^\top \hat{\beta}_j) \hat{\beta}_j - \hat{\Sigma}_{W_\beta W_\alpha} \hat{\theta}_{\text{BEE}} + \hat{\sigma}_{W_\beta W_\alpha}$.

When the number of IVs m is small, the standard sandwich formula has been observed to underestimate the SE.⁵¹ We apply the MD correction⁵² to solve this problem. Consider the so-called hat matrix:

$$\mathbf{H} = \hat{\mathbf{B}} (\hat{\mathbf{B}}^\top \hat{\mathbf{B}} - m \hat{\Sigma}_{W_\beta W_\beta})^{-1} \hat{\mathbf{B}}^\top$$

and H_{jj} is its j th diagonal entries. The MD correction adjusts the inner matrix as

$$\hat{V}_{\text{BEE}}^{\text{MD}}(\hat{\theta}_{\text{BEE}}) = \sum_{j=1}^m (1 - H_{jj})^{-2} \hat{\mathbf{S}}_j(\hat{\theta}_{\text{BEE}}) \hat{\mathbf{S}}_j(\hat{\theta}_{\text{BEE}})^\top. \quad (\text{Equation 18})$$

Their theory shows that

$$E(\hat{\mathbf{S}}_j(\hat{\theta}_{\text{BEE}}) \hat{\mathbf{S}}_j(\hat{\theta}_{\text{BEE}})^\top) \approx (1 - H_{jj})^2 V_{\text{BEE}}(\theta),$$

and hence it can obtain a more reliable covariance matrix by adjusting $(1 - H_{jj})^{-2}$ when estimating $V_{\text{BEE}}(\theta)$ with the moment method. When there is horizontal pleiotropy, we adjust Equation 18 as

$$\hat{V}_{\text{BEE}}^{\text{MD}}(\hat{\theta}_{\text{BEE}}) = \frac{m + m_{\text{pleiotropy}}}{m} \sum_{j=1}^m (1 - H_{jj})^{-2} \hat{\mathbf{S}}_j(\hat{\theta}_{\text{BEE}}) \hat{\mathbf{S}}_j(\hat{\theta}_{\text{BEE}})^\top, \quad (\text{Equation 19})$$

where m is number of valid IVs and $m_{\text{pleiotropy}}$ is the number of detected pleiotropies. Section S1.3 of supplement 1 compares the estimated and true standard errors of causal effect estimates for MRBEE and other MVMR estimators. These results demonstrate that the MD correction described above controls the Type I error rate well. It is also worth noting that the standard errors of the MRBEE causal estimates will generally become smaller as the degree of weak instrument bias becomes smaller.

Horizontal pleiotropy detection

We illustrate how to remove specific IVs with evidence of UHP or CHP effects with the pleiotropy test S_{pleio} which tests the same null hypothesis for each SNP as MR-PRESSO and IMRP. The null hypothesis for the j th IV not having any horizontally pleiotropic effects on the outcome is

$$\mathbf{H}_{0j} : \gamma_j = 0 \text{ v.s. } \mathbf{H}_{1j} : \gamma_j \neq 0. \quad (\text{Equation 20})$$

The statistic S_{pleio} for the j th IV is defined as

$$S_{\text{pleio}_j}(\hat{\theta}) = \frac{(\hat{\alpha}_j - \hat{\beta}_j^\top \hat{\theta})^2}{\text{var}(\hat{\alpha}_j - \hat{\beta}_j^\top \hat{\theta})}, \quad (\text{Equation 21})$$

which follows a χ_1^2 distribution under \mathbf{H}_{0j} . The only assumption here is that $\hat{\alpha}_j - \hat{\beta}_j^\top \hat{\theta}$ is asymptotically normal distributed. In fact, this test examines whether the outcome effect can be explained by the mediation effects through all exposures. In practice, we estimate $\text{var}(\hat{\alpha}_j - \hat{\beta}_j^\top \hat{\theta})$ using the delta method:

$$\widehat{\text{var}}(\hat{\alpha}_j - \hat{\beta}_j^\top \hat{\theta}) = \sigma_{w_\alpha}^2 + \hat{\theta}^\top \hat{\Sigma}_{W_\beta W_\beta} \hat{\theta} + \hat{\beta}_j^\top \hat{\Sigma}_{\text{BEE}} \hat{\beta}_j - 2 \hat{\theta}^\top \hat{\sigma}_{W_\beta W_\alpha}.$$

Other methods such as empirical variance and robust variance estimates of the residual can also be used here. We calculate S_{pleio} for all candidate IVs and remove IVs with large S_{pleio} values in an iterative manner. Algorithm 1 uses an FDR Q-value threshold to exclude IVs showing potential pleiotropy evidence. We suggest a threshold Q-value < 0.05 in general. Additional simulation results presented in Section S2.5 of supplement 1 show that FDR correction generally performs well.

GWPT

Since S_{pleio} tests a very general null hypothesis, we can also calculate S_{pleio} for all SNPs across the genome after obtaining the causal effect estimates of p exposures on the outcome. Results from these tests can be used to (1) find novel loci associated with the MR outcome and (2) draw inferences about pathways of genetic association with the MR outcome. Specifically, when an SNP has a negative effect on the exposure β_j and a positive pleiotropic effect on the outcome γ_j , and simultaneously the causal effect θ is positive, then the total effect of this variant on the outcome α_j is canceled and hence cannot be detected in the outcome GWAS. In contrast, the pleiotropy test directly tests the effect γ_j and therefore can detect novel loci. For example, Zhu et al.¹⁶ successfully detected many blood pressure loci missed previously by using this

GWPT with IMRP as the estimator of the causal effect. The results indicated that most detected pleiotropic variants influenced SBP and DBP in opposite directions, providing support for the principle of the GWPT.

Joint χ^2 -test for IVs selection

We applied the joint χ^2 -test to select a set of IVs that are strongly associated with multiple exposures. Let $\beta_j = (\beta_{j1}, \dots, \beta_{jp})^\top$ be the p -length vector of standardized associations between the j th SNP and the p exposures. We performed the following hypothesis test:

$$\mathbf{H}_{0j} : \beta_{j1} = \dots = \beta_{jp} = 0, \text{ v.s. } \mathbf{H}_{1j} : \beta_{j1} \neq 0, \text{ or } \dots \text{ or } \beta_{jp} \neq 0. \quad (\text{Equation 22})$$

The test statistic is

$$t_j = \widehat{\beta}_j^\top \widehat{\Sigma}_{W_\theta W_\theta}^{-1} \widehat{\beta}_j, \quad (\text{Equation 23})$$

which follows a χ_p^2 distribution when the null hypothesis holds, where $\widehat{\Sigma}_{W_\theta W_\theta}$ is the estimated matrix of covariances between estimation errors. We only considered variants as IVs if they are genome-wide significant in the joint χ^2 -test.

Estimation of variance explained by instrument variables

Assume that we intend to estimate the SNP heritability of a trait Y using a set of m IVs in the m -length vector $\mathbf{g} = (G_1, \dots, G_m)^\top$ with corresponding associations with Y in the vector $\beta = (\beta_1, \dots, \beta_m)^\top$. If the variance of Y is 1 and $E(G_j) = 0$, we can estimate the variance in Y explained by the m IVs using the following equation:

$$R^2 = \sum_{j=1}^m 2\widehat{\beta}_j^2 p_j (1 - p_j) \quad (\text{Equation 24})$$

where p_j is the minor allele frequency of G_j . We used Equation 24 to produce the heritability estimates in Table 1.

Asymptotic results

We assume that both total number of IVs m and the minimum sample size among the exposure and outcome GWAS n_{\min} can approach infinity, while the number of exposures p and the p -dimensional causal effect vector θ are fixed and bounded. Our goal is to identify the scenarios when MV-IVW outperforms MRBEE, when they perform equally well, and when MRBEE outperforms MV-IVW in terms of unbiased estimation of causal effects and the asymptotic validity of causal inference. We demonstrate the related theorems and the related regularity conditions and lemmas in supplement 2.

Data and code availability

The data referenced in this study can be accessed through the GWAS Catalog (<https://www.ebi.ac.uk/gwas/home>), with the corresponding GWAS summary data available for download in the "data availability" section of the respective papers. Some of the GWAS summary data are exclusive of the Million Veteran Program (MVP) summary results, which are available through dbGAP under the accession number phs001672.v3.p1.

The MRBEE R package generated during this study is available at <https://github.com/noahlorinczcomi/MRBEE>.

Simulation codes generated during this study are available at <https://github.com/harryiheyang/MRBEE.Simulation>.

Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.xhgg.2024.100290>.

Acknowledgments

This work was supported by grant HG011052 and HG011052-03S1 (to X.Z.) from the National Human Genome Research Institute (NHGRI), USA. N.L.-C. was partially supported by grant T32 HL007567 from the National Heart, Lung, and Blood Institute (NHLBI).

Declaration of interests

The authors declare no competing interests.

Received: November 17, 2023

Accepted: April 1, 2024

References

- Sanderson, E., Glymour, M.M., Holmes, M.V., Kang, H., Morrison, J., Munafò, M.R., Palmer, T., Schooling, C.M., Wallace, C., Zhao, Q., and Davey Smith, G. (2022). Mendelian Randomization. *Nat. Rev. Methods Primers* 2, 6–21.
- Burgess, S., Butterworth, A., and Thompson, S.G. (2013). Mendelian Randomization Analysis with Multiple Genetic Variants Using Summarized Data. *Genet. Epidemiol.* 37, 658–665.
- Burgess, S., Thompson, S.G.; and CRP CHD Genetics Collaboration (2011). Avoiding Bias from Weak Instruments in Mendelian Randomization Studies. *Int. J. Epidemiol.* 40, 755–764.
- Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian Randomization with Invalid Instruments: Effect Estimation and Bias Detection Through Egger Regression. *Int. J. Epidemiol.* 44, 512–525.
- Morrison, J., Knoblauch, N., Marcus, J.H., Stephens, M., and He, X. (2020). Mendelian Randomization Accounting for Correlated and Uncorrelated Pleiotropic Effects Using Genome-Wide Summary Statistics. *Nat. Genet.* 52, 740–747.
- Verbanck, M., Chen, C.-Y., Neale, B., and Do, R. (2018). Detection of Widespread Horizontal Pleiotropy in Causal Relationships Inferred from Mendelian Randomization Between Complex Traits and Diseases. *Nat. Genet.* 50, 693–698.
- Zhu, X., Li, X., Xu, R., and Wang, T. (2021). An Iterative Approach to Detect Pleiotropy and Perform Mendelian Randomization Analysis Using GWAS Summary Statistics. *Bioinformatics* 37, 1390–1400.
- Kang, H., Zhang, A., Cai, T.T., and Small, D.S. (2016). Instrumental Variables Estimation with Some Invalid Instruments and Its Application to Mendelian Randomization. *J. Am. Stat. Assoc.* 111, 132–144.
- Xue, H., Shen, X., and Pan, W. (2021). Constrained Maximum Likelihood-Based Mendelian Randomization Robust to Both Correlated and Uncorrelated Pleiotropic Effects. *Am. J. Hum. Genet.* 108, 1251–1269.
- Bowden, J., Davey Smith, G., Haycock, P.C., and Burgess, S. (2016). Consistent Estimation in Mendelian Randomization

- with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* *40*, 304–314.
11. Rees, J.M.B., Wood, A.M., Dudbridge, F., and Burgess, S. (2019). Robust Methods in Mendelian Randomization via Penalization of Heterogeneous Causal Estimates. *PLoS One* *14*, e0222362.
 12. Qi, G., and Chatterjee, N. (2019). Mendelian Randomization Analysis Using Mixture Models for Robust and Efficient Estimation of Causal Effects. *Nat. Commun.* *10*, 1941.
 13. Burgess, S., Foley, C.N., Allara, E., Staley, J.R., and Howson, J.M.M. (2020). A Robust and Efficient Method for Mendelian Randomization with Hundreds of Genetic Variants. *Nat. Commun.* *11*, 376.
 14. Yuan, Z., Liu, L., Guo, P., Yan, R., Xue, F., and Zhou, X. (2022). Likelihood-Based Mendelian Randomization Analysis with Automated Instrument Selection and Horizontal Pleiotropic Modeling. *Sci. Adv.* *8*, eabl5744.
 15. Cheng, Q., Zhang, X., Chen, L.S., and Liu, J. (2022). Mendelian Randomization Accounting for Complex Correlated Horizontal Pleiotropy While Elucidating Shared Genetic Etiology. *Nat. Commun.* *13*, 6490.
 16. Zhu, X., Zhu, L., Wang, H., Cooper, R.S., and Chakravarti, A. (2022). Genome-Wide Pleiotropy Analysis Identifies Novel Blood Pressure Variants and Improves Its Polygenic Risk Scores. *Genet. Epidemiol.* *46*, 105–121.
 17. Sanderson, E., Davey Smith, G., Windmeijer, F., and Bowden, J. (2019). An Examination of Multivariable Mendelian Randomization in the Single-Sample and Two-Sample Summary Data Settings. *Int. J. Epidemiol.* *48*, 713–727.
 18. Burgess, S., and Thompson, S.G. (2015). Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. *Am. J. Epidemiol.* *181*, 251–260.
 19. Rees, J.M.B., Wood, A.M., and Burgess, S. (2017). Extending the MR-Egger Method for Multivariable Mendelian Randomization to Correct for Both Measured and Unmeasured Pleiotropy. *Stat. Med.* *36*, 4705–4718.
 20. Lin, Z., Xue, H., and Pan, W. (2023). Robust Multivariable Mendelian Randomization Based on Constrained Maximum Likelihood. *Am. J. Hum. Genet.* *110*, 592–605.
 21. Wang, K., Shi, X., Zhu, Z., Hao, X., Chen, L., Cheng, S., Foo, R.S.Y., and Wang, C. (2022). Mendelian Randomization Analysis of 37 Clinical Factors and Coronary Artery Disease in East Asian and European Populations. *Genome Med.* *14*, 63.
 22. Sanderson, E., Spiller, W., and Bowden, J. (2021). Testing and Correcting for Weak and Pleiotropic Instruments in Two-Sample Multivariable Mendelian Randomization. *Stat. Med.* *40*, 5434–5452.
 23. Sadreev, I.I., Elsworth, B.L., Mitchell, R.E., Paternoster, L., Sanderson, E., Davies, N.M., Millard, L.A.C., Davey Smith, G., Haycock, P.C., Bowden, J., et al. (2021). Navigating sample overlap, winner's curse and weak instrument bias in mendelian randomization studies using the UK Biobank. Preprint at medRxiv. <https://doi.org/10.1101/2021.06.28.21259622>.
 24. Carroll, R.J., Ruppert, D., Stefanski, L.A., and Crainiceanu, C.M. (2006). Measurement Error in Nonlinear Models: A Modern Perspective. Chapman (Hall/CRC).
 25. VanderWeele, T.J., Tchetgen Tchetgen, E.J., Cornelis, M., and Kraft, P. (2014). Methodological Challenges in Mendelian Randomization. *Epidemiology* *25*, 427–435.
 26. Ye, T., Shao, J., and Kang, H. (2021). Debiased Inverse-Variance Weighted Estimator in Two-Sample Summary-Data Mendelian Randomization. *Ann. Stat.* *49*, 2079–2100.
 27. Burgess, S., Davies, N.M., and Thompson, S.G. (2016). Bias Due to Participant Overlap in Two-Sample Mendelian Randomization. *Genet. Epidemiol.* *40*, 597–608.
 28. Mounier, N., and Kutalik, Z. (2023). Bias Correction for Inverse Variance Weighting Mendelian Randomization. *Genet. Epidemiol.* *47*, 314–333.
 29. Yavorska, O.O., and Burgess, S. (2017). MendelianRandomization: An R Package for Performing Mendelian Randomization Analyses Using Summarized Data. *Int. J. Epidemiol.* *46*, 1734–1739.
 30. Morgan, I.G., French, A.N., Ashby, R.S., Guo, X., Ding, X., He, M., and Rose, K.A. (2018). The Epidemics of Myopia: Aetiology and Prevention. *Prog. Retin. Eye Res.* *62*, 134–149.
 31. Marconi, A., Di Forti, M., Lewis, C.M., Murray, R.M., and Vassos, E. (2016). Meta-Analysis of the Association Between the Level of Cannabis Use and Risk of Psychosis. *Schizophr. Bull.* *42*, 1262–1269.
 32. Corcoran, C.M., Kimhy, D., Stanford, A., Khan, S., Walsh, J., Thompson, J., Schobel, S., Harkavy-Friedman, J., Goetz, R., Colibazzi, T., et al. (2008). Temporal Association of Cannabis Use with Symptoms in Individuals at Clinical High Risk for Psychosis. *Schizophr. Res.* *106*, 286–293.
 33. Holmes, M.V., Asselbergs, F.W., Palmer, T.M., Drenos, F., Lanktree, M.B., Nelson, C.P., Dale, C.E., Padmanabhan, S., Finan, C., Swerdlow, D.I., et al. (2015). Mendelian Randomization of Blood Lipids for Coronary Heart Disease. *Eur. Heart J.* *36*, 539–550.
 34. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A Global Reference for Human Genetic Variation. *Nature* *526*, 68–74.
 35. Bulik-Sullivan, B.K., Neale, B.M., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., and Price, A.L. (2015). LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies. *Nat. Genet.* *47*, 291–295.
 36. Smith, C.J., Sinnott-Armstrong, N., Cichońska, A., Julkunen, H., Fauman, E.B., Würtz, P., and Pritchard, J.K. (2022). Integrative Analysis of Metabolite GWAS Illuminates the Molecular Basis of Pleiotropy and Genetic Correlation. *Elife* *11*, e79348.
 37. Zhu, X., Feng, T., Tayo, B.O., Liang, J., Young, J.H., Franceschini, N., Smith, J.A., Yanek, L.R., Sun, Y.V., Edwards, T.L., et al. (2015). Meta-Analysis of Correlated Traits via Summary Statistics from GWASs with an Application in Hypertension. *Am. J. Hum. Genet.* *96*, 21–36.
 38. Turley, P., Walters, R.K., Maghzian, O., Okbay, A., Lee, J.J., Fontana, M.A., Nguyen-Viet, T.A., Wedow, R., Zacher, M., Furlotte, N.A., et al. (2018). Multi-Trait Analysis of Genome-Wide Association Summary Statistics Using MTAG. *Nat. Genet.* *50*, 229–237.
 39. Watanabe, K., Taskesen, E., Van Bochoven, A., and Posthuma, D. (2017). Functional Mapping and Annotation of Genetic Associations with FUMA. *Nat. Commun.* *8*, 1826.
 40. Yi, G.Y. (2017). Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application (Springer).

41. Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *Ann. Statist.* *24*, 2350–2383.
42. He, M., Xiang, F., Zeng, Y., Mai, J., Chen, Q., Zhang, J., Smith, W., Rose, K., and Morgan, I.G. (2015). Effect of Time Spent Outdoors at School on the Development of Myopia Among Children in China: A Randomized Clinical Trial. *JAMA* *314*, 1142–1148.
43. Lin, Z., Vasudevan, B., Jhanji, V., Mao, G.Y., Gao, T.Y., Wang, F.H., Rong, S.S., Ciuffreda, K.J., and Liang, Y.B. (2014). Near Work, Outdoor Activity, and Their Association with Refractive Error. *Optom. Vis. Sci.* *91* (4), 376–382.
44. The HPS3/TIMI55–REVEAL Collaborative Group (2017). Effects of Anacetrapib in Patients with Atherosclerotic Vascular Disease. *N. Engl. J. Med.* *377*, 1217–1227.
45. Chobanian, A.V., Bakris, G.L., Black, H.R., Cushman, W.C., Green, L.A., Izzo, J.L., Jr, Jones, D., Materson, B.J., Oparil, S., Wright, J.T., Jr, et al. (2004). The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. *Hypertension* *42*, 1206–1252.
46. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* *518*, 197–206.
47. Takeuchi, F., Akiyama, M., Matoba, N., Katsuya, T., Nakatochi, M., Tabara, Y., Narita, A., Saw, W.Y., Moon, S., Spracklen, C.N., et al. (2018). Interethnic analyses of blood pressure loci in populations of East Asian and European descent. *Nat. Commun.* *9*, 5052.
48. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs Explain a Large Proportion of the Heritability for Human Height. *Nat. Genet.* *42*, 565–569.
49. Zuber, V., Colijn, J.M., Klaver, C., and Burgess, S. (2020). Selecting likely causal risk factors from high-throughput experiments using multivariable Mendelian randomization. *Nat. Commun.* *11*, 29.
50. Lawlor, D.A., Harbord, R.M., Sterne, J.A.C., Timpson, N., and Davey Smith, G. (2008). Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology. *Stat. Med.* *27*, 1133–1163.
51. Wang, M., Kong, L., Li, Z., and Zhang, L. (2016). Covariance Estimators for Generalized Estimating Equations (GEE) in Longitudinal Analysis with Small Samples. *Stat. Med.* *35*, 5318–5319.
52. Mancl, L.A., and DeRouen, T.A. (2001). A Covariance Estimator for GEE with Improved Small-Sample Properties. *Biometrics* *57*, 126–134.
53. Wu, Y., Kang, H., and Ye, T. (2024). Debiased Multivariable Mendelian Randomization. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2402.00307>.
54. van De Vegte, Y.J., Said, M.A., Rienstra, M., van Der Harst, P., and Verweij, N. (2020). Genome-wide association studies and Mendelian randomization analyses for leisure sedentary behaviours. *Nature Commun.* *11*, 1770. <https://doi.org/10.1038/s41467-020-15553-w>.
55. Arns, A., Wahl, T., Wolff, C., Vafeidis, A.T., Haigh, I.D., Woodworth, P., Niehüser, S., and Jensen, J. (2020). Non-linear interaction modulates global extreme sea levels, coastal flood exposure, and impacts. *Nature Commun.* *11*, 1918. <https://doi.org/10.1038/s41467-020-15752-5>.
56. Rustad, E.H., Yellapantula, V., Leongamornlert, D., Bolli, N., Ledergor, G., Nadeu, F., Angelopoulos, N., Dawson, K.J., Mitchell, T.J., Osborne, R.J., et al. (2020). Timing the initiation of multiple myeloma. *Nature Commun.* *11*, 1917. <https://doi.org/10.1038/s41467-020-15740-9>.
57. Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S.M., Sidorenko, J., Kweon, H., Goldman, G., Gjorgjieva, T., et al. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genet.* *54*, 437–449. <https://doi.org/10.1038/s41588-022-01016-z>.
58. Hysi, P.G., Choquet, H., Khawaja, A.P., Wojciechowski, R., Tedja, M.S., Yin, J., Simcoe, M.J., Patasova, K., Mahroo, O.A., Thai, K.K., et al. (2020). Meta-analysis of 542,934 subjects of European ancestry identifies new genes and mechanisms predisposing to refractive error and myopia. *Nature genetics* *52*, 401–407. <https://doi.org/10.1038/s41588-020-0599-0>.
59. Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genet.* *51*, 63–75. <https://doi.org/10.1038/s41588-018-0269-7>.
60. Johnson, E.C., Demontis, D., Thorgeirsson, T.E., Walters, R.K., Polimanti, R., Hatoum, A.S., Sanchez-Roige, S., Paul, S.E., Wendt, F.R., Clarke, T.K., et al. (2020). A large-scale genome-wide association study meta-analysis of cannabis use disorder. *The Lancet Psychiatry* *7*, 1032–1045.
61. Cuellar-Partida, G., Tung, J.Y., Eriksson, N., Albrecht, E., Aliev, F., Andreassen, O.A., Barroso, I., Beckmann, J.S., Boks, M.P., Boomsma, D.I., et al. (2021). Genome-wide association study identifies 48 common genetic variants associated with handedness. *Nature Human Behav.* *5*, 59–70. <https://doi.org/10.1038/s41562-020-00956-y>.
62. Nagel, M., Jansen, P.R., Stringer, S., Watanabe, K., De Leeuw, C.A., Bryois, J., Savage, J.E., Hammerschlag, A.R., Skene, N.G., and Muñoz-Manchado, A.B. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genet.* *50*, 920–927. <https://doi.org/10.1038/s41588-018-0151-7>.
63. Dashti, H.S., Jones, S.E., Wood, A.R., Lane, J.M., Van Hees, V.T., Wang, H., Rhodes, J.A., Song, Y., Patel, K., Anderson, S.G., et al. (2019). Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *Nature Commun.* *10*, 1100.
64. Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T.B., Bryois, J., Chen, C.Y., Dennison, C.A., Hall, L.S., et al. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* *604*, 502–508.
65. Loh, P.R., Kichaev, G., Gazal, S., Schoech, A.P., and Price, A.L. (2018). Mixed-model association for biobank-scale datasets. *Nature Genet.* *50*, 906–908. <https://doi.org/10.1038/s41588-018-0144-6>.
66. Evangelou, E., Warren, H.R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H., Ntritsos, G., Dimou, N., Cabrera, C.P., Karaman, I., et al. (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nature Genet.* *50*, 1412–1425. <https://doi.org/10.1038/s41588-018-0205-x>.

67. Graham, S.E., Clarke, S.L., Wu, K.H.H., Kanoni, S., Zajac, G.J., Ramdas, S., Surakka, I., Ntalla, I., Vedantam, S., Winkler, T.W., et al. (2021). The power of genetic diversity in genome-wide association studies of lipids. *Nature* 600, 675–679. <https://doi.org/10.1038/s41586-021-04064-3>.
68. Aragam, K.G., Jiang, T., Goel, A., Kanoni, S., Wolford, B.N., Atri, D.S., Weeks, E.M., Wang, M., Hindy, G., Zhou, W., et al. (2022). Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nature Genet.* 54, 1803–1815. <https://doi.org/10.1038/s41588-022-01233-6>.