MR-LDP: a two-sample Mendelian randomization for GWAS summary statistics accounting for linkage disequilibrium and horizontal pleiotropy

Qing Cheng[1], Yi Yang[1], Xingjie Shi[1,2], Kar-Fu Yeung[1], Can Yang[3], Heng Peng[4], Jin Liu[1*]

[1]Centre for Quantitative Medicine, Health Services & Systems Research, Duke-NUS Medical School

[2]Department of Statistics, Nanjing University of Finance and Economics, Nanjing, China

[3]Department of Mathematics, The Hong Kong University of Science and Technology

[4]Department of Mathematics, Hong Kong Baptist University

[*]To whom the correspondence should be addressed

Abstract

The proliferation of genome-wide association studies (GWAS) has prompted the use of two-sample Mendelian randomization (MR) with genetic variants as instrumental variables (IV) for drawing reliable causal relationships between health risk factors and disease outcomes. However, the unique features of GWAS demand that MR methods account for both linkage disequilibrium (LD) and ubiquitously existing horizontal pleiotropy among complex traits, which is the phenomenon wherein a variant affects the outcome through mechanisms other than exclusively through the exposure. Therefore, statistical methods that fail to consider LD and horizontal pleiotropy can lead to biased estimates and false-positive causal relationships. To overcome these limitations, we propose a probabilistic model for MR analysis to identify the casual effects between risk factors and disease outcomes using GWAS summary statistics in the presence of LD and to properly account for horizontal pleiotropy among genetic variants

[*]Correspondence should be addressed to Jin Liu (jin.liu@duke-nus.edu.sg)

(MR-LDP). MR-LDP utilizes a computationally efficient parameter-expanded variational Bayes expectation-maximization (PX-VBEM) algorithm to estimate the parameter of interest and further calibrates the evidence lower bound (ELBO) for a likelihood ratio test. We then conducted comprehensive simulation studies to demonstrate the advantages of MR-LDP over the existing methods in terms of both type-I error control and point estimates. Moreover, we used two real exposure-outcome pairs (CAD-CAD and Height-Height; CAD for coronary artery disease) to validate the results from MR-LDP compared with alternative methods, showing that our method is more efficient in using all instrumental variants in LD. By further applying MR-LDP to lipid traits and body mass index (BMI) as risk factors for complex diseases, we identified multiple pairs of significant causal relationships, including a protective effect of high-density lipoprotein cholesterol (HDL-C) on peripheral vascular disease (PVD), and a positive causal effect of body mass index (BMI) on hemorrhoids.

1 Introduction

Epidemiological studies have contributed tremendously to improving our understanding of the primary causes of complex diseases. However, numerous cases of significant associations from observational studies have been subsequently contradicted by large clinical trials [1, 2]. Drawing causal inferences from observational studies is particularly challenging because of unmeasured confounding, reverse causation and selection bias [3, 4]. Although the randomized controlled trial (RCT) is considered a gold standard to evaluate causality in an exposure-outcome pair, RCTs have certain limitations including impracticality (no intervention may exist), high expense, and ethical issues [5]. Fortunately, as germline genetic variants (single nucleotide polymorphisms, SNPs) are fixed after random mating and cannot be modified by subsequent factors, e.g., environment factors and living styles, Mendelian randomization (MR) uses genetic variants as instruments to examine the causal effects between health risk factors and disease outcomes, largely excluding the influence from unobserved confounding factors [3]. In the past decade, a large number of genome-wide association studies (GWAS) have been successfully used to identify genetic variants associated with complex traits at the genome-wide significance level, including both health factors and diseases, e.g., lipids, BMI, and type-2 diabetes, and most of completed GWAS are simply observational studies instead of RCTs. The results from completed GWAS are mostly publicly accessible, e.g., GWAS Catalog outlines a list of sources for

summary statistics (`https://www.ebi.ac.uk/gwas/downloads/summary-statistics`). This large amount of publicly available GWAS summary statistics has prompted the widespread use of two-sample MR as an efficient and cost-effective method to interrogate the causal relationships among health risk factors and disease outcomes.

MR is closely related to the instrumental variable (IV) methods, which have a long history of use in econometrics [6]. Classically, an inverse-variance weighted (IVW) and a likelihood-based approach have been used for two-sample MR analysis with summary-level data [7]. These methods must strictly obey assumptions for MR, including two most fundamental ones:

1. IVs affect the outcome exclusively through the risk exposures.

2. IVs are independent from each other, or in a GWAS context, instrumental variants are not in LD.

The first assumption is also referred to as exclusion restriction assumption or no horizontal pleiotropy. The violation of this assumption can distort the statistical inference for MR analysis, leading to biased estimates and false-positive causal relationships. Recent comprehensive surveys reported persuasive pleiotropy among complex traits [8, 9], such as autoimmune diseases [10] and psychiatric disorders [11]. Consequently, methods that do not account for pleiotropy can substantially reduce the power and inflate the false-positive discoveries. To address this issue, sisVIVE was proposed in the presence of individual-level data [12]. To further relax this assumption for two-sample MR analysis using summary-level data, various statistical methods have been proposed and we divide them into two categories. The first group consists of step-wise methods to correct the impact of horizontal pleiotropy. These methods first detect and remove SNPs with horizontal pleiotropy, and MR analysis is performed in the subsequent step, including Q test [13], Cook's distance [14], Studentized residuals [14], GSMR [15], and MR-PRESSO [16]. The drawback of this type of methods is that the number of SNPs after removal is limited given that abundant pleiotropy exists among complex traits, which can substantially reduce the statistical power to detect the causal relationships. In contrast, the second group of methods jointly estimate causal effects by taking into account the horizontal pleiotropy, e.g., MR-Egger [17], MRMix [18] and RAPS [19]. Compared to MR-Egger, RAPS further addressed the measurement error issues, where most of existing methods applicable to GWAS summary statistics assume that sampling error from SNP-exposure is negligible [20].

On the other hand, minimal literature is available regarding the relaxation of the second assumption above. Among the methods mentioned above, only GSMR is capable of accounting weak or moderate linkage disequilibrium (LD) for SNPs, while others demand all instrumental SNPs to be independent, which is typically achieved by conducting SNP pruning and thus reducing the number of instrumental variants for follow-up MR analysis. As SNPs within proximity tend to be highly correlated, MR methods not accounting for LD structure may substantially lose statistical power due to the pruning process to obtain independent SNPs. Moreover, GSMR is a step-wise method to remove instrumental variants with horizontal pleiotropy, making it less powerful due to the removal of invalid variants.

In this paper, we propose a statistically unified and efficient two-sample MR method to utilize all weak instruments within LD (MR-LD), and further consider a MR-LD accounting for horizontal Pleiotropy (MR-LDP). Similar to RAPS, MR-LDP does not require the no measurement error assumption. The key idea is to build a joint probabilistic model for GWAS summary statistics from both exposure and outcome using a reference panel to reconstruct LD among instrumental variants and to conduct a formal hypothesis testing to make inferences about the causal effect that links the exposure and the outcome through a linear relationship. We also develop an efficient variational Bayesian expectation-maximization algorithm accelerated by using the parameter expansion (PX-VBEM) to estimate the causal effect for MR-LD and MR-LDP. Moreover, we calibrate the evidence lower bound (ELBO) to construct the likelihood ratio test for the evaluation of statistical significance of the estimated effect. Simulation studies show that MR-LDP outperforms competing methods in terms of type-I error control and point estimates for making causal inference. Additionally, we used two real exposure-outcome pairs to validate results from MR-LD and MR-LDP compared with alternative methods, particularly showing our methods more efficiently use all SNPs in LD. By further applying MR-LDP to summary statistics from GWAS, we identified multiple pairs of significant causal relationships, including a protective effect of high-density lipoprotein cholesterol (HDL-C) on peripheral vascular disease (PVD), and a positive causal effect of BMI on hemorrhoids.

2 Material and Methods

2.1 Reference panel data

As MR-LD and MR-LDP use the marginal effect sizes and their standard errors from GWAS summary statistics to build a probabilistic model for making causal inference, information regarding correlations among SNPs is missing, i.e., LD denoted as $\mathbf{R}$ is missing. Thus, we need to use a reference panel dataset to assist with reconstructing LD. In this study, given that we primarily focus on European populations, we choose to use samples from the following resource as the external reference panel: UK10K Project (Avon Longitudinal Study of Parents and Children, ALSPAC; TwinsUK) merged with 1000 Genome Project Phase 3 ($N = 4,284$), which is denoted UK10K thereafter. As SNPs from HapMap Project Phase 3 (HapMap3) are more reliable, we choose to limit our analysis using SNPs from HapMap3 ($p = 1,189,556$).

As samples from ALSPAC and TwinsUK include populations other than European, we conducted strict quality control for UK10K data using PLINK [21]. First, SNPs were excluded from the analysis if their calling rates were less than 95%, minor allele frequencies were less than 0.01, or p-values were less than 1×10^{-6} in the Hardy-Weinberg equilibrium test. We then removed the individuals with genotype missing rates greater than 5%. To further remove individuals with high relatedness in all samples, we used GCTA [22] to first identify those individual pairs with estimated genetic relatedness greater than 0.05 and then randomly remove one from such a pair. Additionally, we carried out the principal components analysis (PCA) on the individuals to identify the population stratification [23]. In this way , we extracted the clustering subgroup representing the major European ancestry using hierarchical clustering on principal components(HCPC) approach [24]. Finally, there were 3,764 individuals remained with 989,932 SNPs.

2.2 Choice of LD matrix

Since the LD between two SNPs decays exponentially with respect to their distance, we use LDetect [25] to partition the whole genome into L blocks first and then calculate the estimated correlation matrix in each block. For each block, we adopt a shrinkage method to guarantee the sparsity and positive definiteness of the estimated correlation matrix [26]. In particular,

the correlation matrix estimator $\widehat{\mathbf{R}}^{(l)}$ in each block is obtained by optimizing as follows

$$\widehat{\mathbf{R}}^{(l)} = \underset{\mathbf{R}^{(l)} \succ 0}{\arg\min}(\|\mathbf{R}^{(l)} - \widehat{\mathbf{R}}^{(l)}_{\text{emp}}\|_F^2/2 - \tau \log|\mathbf{R}^{(l)}| + \lambda\|\mathbf{R}^{(l)-}\|_1), \tag{2.1}$$

where $\widehat{\mathbf{R}}^{(l)}_{\text{emp}}$ is the empirical correlation matrix in the l-th block, $\lambda \geq 0$ is the shrinkage tuning parameter, and the lasso-type penalty ensures a sparse solution. In addition, $\tau > 0$ is fixed at a small value and the logarithmic barrier term is used to enforce a positive-definite solution. More details can be found in [26]. A corresponding R package named *PDSCE* is available to complete the estimation process. In addition, we fix the shrinkage parameter λ to be 0.055 in simulation studies and vary $\lambda \in \{0.1, 0.15\}$ in real data analysis.

2.3 Likelihood for summary statistics

Before elaborating on our method, we first review the following multiple linear regression model that links a trait to genotype data:

$$\mathbf{y} = \mathbf{G}\boldsymbol{\gamma} + \boldsymbol{\epsilon},$$

where $\mathbf{y}$ is an $n \times 1$ vector for trait among n individuals, $\mathbf{G}$ is an $n \times p$ matrix for genotypes, $\boldsymbol{\gamma}$ is a $p \times 1$ vector for effect sizes, and $\boldsymbol{\epsilon}$ is the vector for random noises. Suppose that the individual-level data $\{\mathbf{G}, \mathbf{y}\}$ are not accessible, but the summary statistics $\{\widehat{\boldsymbol{\gamma}}_k, \widehat{\mathbf{s}}_k^2\}_{k=1,\cdots p}$ from univariate linear regression are available:

$$\widehat{\boldsymbol{\gamma}}_k = (\mathbf{g}_k^{\mathrm{T}}\mathbf{g}_k)^{-1}\mathbf{g}_k^{\mathrm{T}}\mathbf{y}, \quad \widehat{\mathbf{s}}_k^2 = (n\mathbf{g}_k^{\mathrm{T}}\mathbf{g}_k)^{-1}(\mathbf{y} - \mathbf{g}_k\widehat{\boldsymbol{\gamma}}_k)^{\mathrm{T}}(\mathbf{y} - \mathbf{g}_k\widehat{\boldsymbol{\gamma}}_k),$$

where $\mathbf{g}_k$ is the k-th column of $\mathbf{G}$, $\widehat{\boldsymbol{\gamma}}_k$ and $\widehat{\mathbf{s}}_k^2$ are estimated effect sizes and its variance for SNP k, respectively. $\widehat{\mathbf{R}}$ denotes the correlation among all genotyped SNPs and $\widehat{\mathbf{S}} = \text{diag}([\widehat{\mathbf{s}}_1, \cdots, \widehat{\mathbf{s}}_p])$, which is a diagonal matrix for corresponding standard errors. Provided that sample size n is large enough and the trait is highly polygenic (i.e., the squared correlation coefficient between the trait and each genetic variant is close to zero), we can use the following formula to approximate the distribution of $\boldsymbol{\gamma}$ based on the summary statistics in a similar fashion as [27, 28, 29, 30]:

$$\widehat{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \widehat{\mathbf{R}}, \widehat{\mathbf{S}} \sim \mathcal{N}(\widehat{\mathbf{S}}\widehat{\mathbf{R}}\widehat{\mathbf{S}}^{-1}\boldsymbol{\gamma}, \widehat{\mathbf{S}}\widehat{\mathbf{R}}\widehat{\mathbf{S}}). \tag{2.2}$$

Analogously, we apply this distribution to the two-sample MR analysis. The summary statistics for SNP-exposure and SNP-outcome are denoted by $\{\widehat{\gamma}_k, \widehat{\mathbf{s}}^2_{\gamma_k}\}_{k=1,\cdots,p}$ and $\{\widehat{\Gamma}_k, \widehat{\mathbf{s}}^2_{\Gamma_k}\}_{k=1,\cdots,p}$,

respectively. Therefore, the likelihood for two-sample summary statistics can be written as:

$$\begin{aligned} \widehat{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \widehat{\mathbf{R}}, \widehat{\mathbf{S}}_{\gamma} &\sim \mathcal{N}(\widehat{\mathbf{S}}_{\gamma}\widehat{\mathbf{R}}\widehat{\mathbf{S}}_{\gamma}^{-1}\boldsymbol{\gamma}, \widehat{\mathbf{S}}_{\gamma}\widehat{\mathbf{R}}\widehat{\mathbf{S}}_{\gamma}), \\ \widehat{\boldsymbol{\Gamma}}|\boldsymbol{\Gamma}, \widehat{\mathbf{R}}, \widehat{\mathbf{S}}_{\Gamma} &\sim \mathcal{N}(\widehat{\mathbf{S}}_{\Gamma}\widehat{\mathbf{R}}\widehat{\mathbf{S}}_{\Gamma}^{-1}\boldsymbol{\Gamma}, \widehat{\mathbf{S}}_{\Gamma}\widehat{\mathbf{R}}\widehat{\mathbf{S}}_{\Gamma}), \end{aligned} \tag{2.3}$$

where $\widehat{\mathbf{S}}_{\gamma} = \text{diag}([\widehat{\mathbf{s}}_{\gamma_1}, \cdots, \widehat{\mathbf{s}}_{\gamma_p}])$ and $\widehat{\mathbf{S}}_{\Gamma} = \text{diag}([\widehat{\mathbf{s}}_{\Gamma_1}, \cdots, \widehat{\mathbf{s}}_{\Gamma_p}])$ are both diagonal matrices. In this formulation, the correlations among all p SNPs, $\widehat{\mathbf{R}}$, are not estimable from summary statistics itself. Zhu and Stephens [29] showed that $\widehat{\mathbf{R}}$ could be replaced with $\widehat{\mathbf{R}}^{\text{ref}}$ that is estimated from independent samples, where the difference in log-likelihood between individual-level data and summary statistics is a constant that does not depend on the effect size assuming that polygenicity holds and the sample size of individual-level data is large. Thus, distributions for summary statistics (2.3) will produce approximately the same inferential results as its counterpart for individual-level data. Hereafter, we use $\widehat{\mathbf{R}}$ implicitly for $\widehat{\mathbf{R}}^{\text{ref}}$ and details on estimating $\widehat{\mathbf{R}}$ can be found in Section 2.2.

2.4 MR-LDP model overview

The fundamental assumptions for two-sample MR analysis include the independence among instrumental variables, and three IV assumptions for a genetic instrument: (1) associated with health risk factors ($\gamma \neq 0$); (2) independent of unobserved confounding factors between the risk factors and the disease outcomes; (3) independent of $\mathbf{Y}$ given risk factors and confounders. Given strong LD structure among SNPs and abundant horizontal pleiotropy in GWAS, these unique features invalidate the independence assumption for genetic variants and two IV assumptions (2) and (3). Our proposed MR-LDP aims to make causal inference of the risk factors on a disease outcome using a probabilistic model by accounting for both the LD structure and the influence of horizontal pleiotropy as depicted in Figure 1. We first utilize an approximated likelihood to depict the distribution of correlated SNPs from GWAS summary statistics for the risk exposure and the disease outcome, respectively, as shown Equation (2.3). Given p instrumental variants, the inputs for MR-LDP are GWAS summary statistics for SNP-exposure and SNP-outcome, respectively, and a genotype reference panel (Figure 1A). By introducing an additional random effect $\boldsymbol{\alpha}$, we would further eliminate the variance in the disease outcome due to pervasive horizontal pleiotropy. Since MR-LDP uses an approximated likelihood to jointly delineate the distribution for summary statistics (i.e., estimated effect sizes and their standard errors) from GWAS, it is free of the assumption for no measurement errors, requiring that

sample sizes used to generate GWAS summary statistics are large [31, 20]. Figure 1B depicts MR-LDP as a probabilistic graphical model, where the observed variables of our model include GWAS summary statistics from both the SNP-exposure and the SNP-outcome, and an external reference panel for genotype data. We assume that α_k and γ_k follow two independent Gaussian distributions. The latent variable γ_k and parameter β_0 jointly assist with formulating the distribution for SNP-outcome. Then, we can formalize the hypothesis testing for β_0 as shown in Figure 1B. The scatter plots of estimated effect sizes for SNP-exposure against SNP-outcome together with the MR-LDP analysis results ($\hat{\beta}_0$ and p-value) are shown in Figure 1C. In both BMI-T2D and BMI-VV, there is a dominant proportion of instrumental variants in the center that is largely due to LD, and methods that do not account for LD tend to inflate findings.

2.5 Details of MR-LDP

Parameterization for causal relationship The relationship between $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$ can be constructed using linear structural models as follows:

$$\Gamma_j = \beta_0\gamma_j, \quad \text{or} \quad \Gamma_j = \alpha_j + \beta_0\gamma_j, \tag{2.4}$$

where $j = 1, \ldots, p$, considering without/with horizontal pleiotropy, respectively [12, 32]. Note that β_0 is the effect size of the exposure on the outcome and $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]^T$ is the vector of effects of genetic variants on the outcome due to horizontal pleiotropy. Importantly, β_0 can be interpreted as the causal effect between exposure and outcome in the study [32]. More details regarding linear structural models incorporating the relationship (2.4) are available in the supplementary document. As MR-LD can be taken as a special case of MR-LDP by taking all $\boldsymbol{\alpha}$ to be zero, we focus on deriving MR-LDP in the main text and provide the supplementary document for details on MR-LD.

Empirical Bayes model By assuming that $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$ are two latent variables coming from two independent Gaussian distributions, the complete-data likelihood can be written as follows:

$$\Pr(\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}, \widehat{\mathbf{R}}; \boldsymbol{\theta}) = \Pr(\widehat{\boldsymbol{\Gamma}} | \boldsymbol{\gamma}, \boldsymbol{\alpha}, \widehat{\mathbf{R}}, \widehat{\mathbf{S}}_{\Gamma}; \beta_0) \Pr(\widehat{\boldsymbol{\gamma}} | \boldsymbol{\gamma}, \widehat{\mathbf{R}}, \widehat{\mathbf{S}}_{\gamma}) \Pr(\boldsymbol{\alpha} | \sigma^2_{\boldsymbol{\alpha}}) \Pr(\boldsymbol{\gamma} | \sigma^2_{\boldsymbol{\gamma}}), \tag{2.5}$$

where $\boldsymbol{\theta} \overset{\text{def}}{=} \{\beta_0, \sigma^2_{\boldsymbol{\gamma}}, \sigma^2_{\boldsymbol{\alpha}}\}$ denotes the collection of model parameters. Integrating out the latent variables $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$, the marginal likelihood can be written as:

$$\Pr(\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\gamma}} | \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}, \widehat{\mathbf{R}}; \boldsymbol{\theta}) = \int \int \Pr(\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\gamma}}, \boldsymbol{\alpha}, \boldsymbol{\gamma} | \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}, \widehat{\mathbf{R}}; \boldsymbol{\theta}) d\boldsymbol{\gamma}\, d\boldsymbol{\alpha}.$$

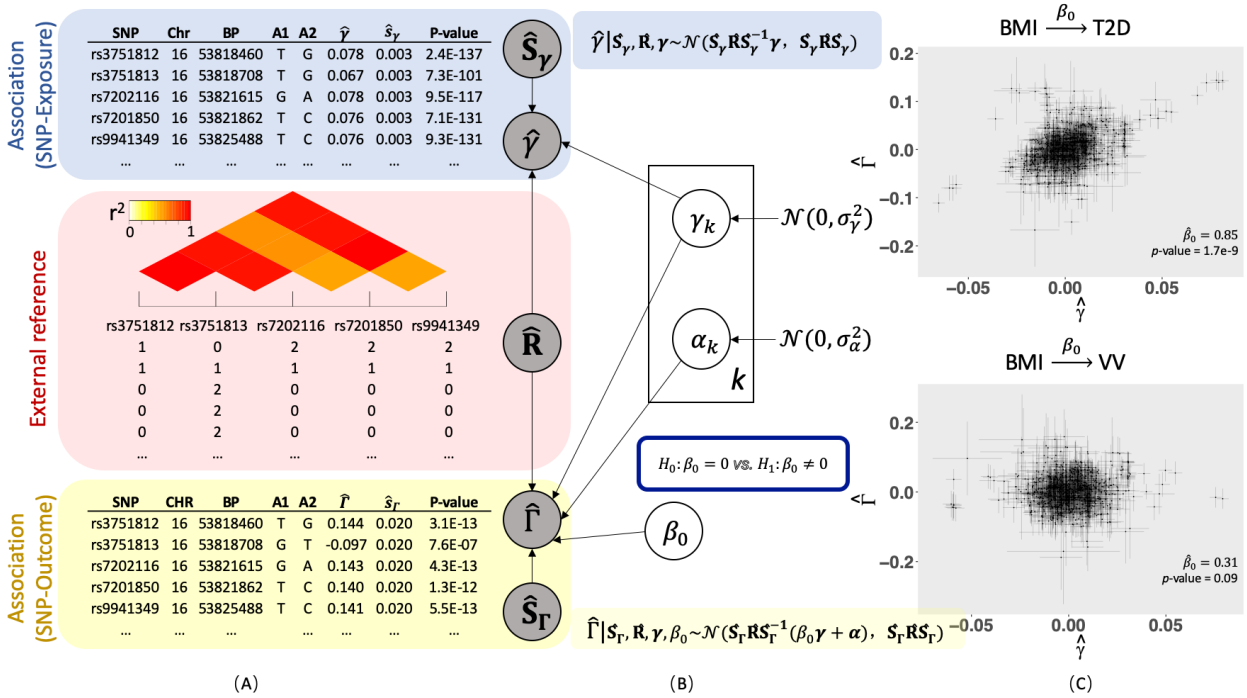


Figure 1: MR-LDP model overview: (A) Inputs for MR-LDP include GWAS summary statistics from both the risk factor (Blue) and the disease outcome (Yellow), and an external reference panel data (Red). (B) A probabilistic graphical model representation of MR-LDP. The box is the "plate" representing SNPs, $k = 1, \ldots, p$. The circles are either variables or parameters. The circles at the root are parameters. The variables in shaded circles are observed (i.e., GWAS summary statistics $\{\widehat{\gamma}_k, \widehat{\mathbf{s}}_{\gamma_k}\}_{k=1,\cdots,p}$ and $\{\widehat{\Gamma}_k, \widehat{\mathbf{s}}_{\Gamma_k}\}_{k=1,\cdots,p}$, and the estimated $\widehat{\mathbf{R}}$ for these p SNPs from a reference panel) and variables in unshaded circles are latent variables (i.e., γ_k and α_k , $k = 1, \ldots, p$). The primary goal is to conduct a formal hypothesis testing for $\mathcal{H}_0 : \beta_0 = 0 \quad \text{vs} \quad \mathcal{H}_1 : \beta_0 \neq 0$. (C) Scatter plots of effect sizes with their standard errors for two exposure-outcome pairs: BMI-T2D and BMI-VV; T2D for type-2 diabetes and VV for varicose veins. Dots represent the effect sizes from SNP-exposure against these from SNP-outcome,and horizontal and vertical bars represent the standard errors from SNP-exposure and SNP-outcome, respectively. The estimated β_0 and its p-value from MR-LDP are not shown in each subfigure.

Algorithm The standard expectation-maximization (EM) algorithm is a common choice to find the maximum likelihood for probabilistic models in the presence of latent variables [33]. However, it may cause instability or numerical failure as $\widehat{\mathbf{R}}$ can be non-positive definite due to the relative small sample size in the reference panel. To address these issues, we develop an accelerated variational Bayes (VB) EM algorithm in light of [34], namely, PX-VBEM. Starting with the algorithm, we expand the original MR-LD/MR-LDP model (2.5) as follows:

$$\widehat{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \widehat{\mathbf{R}}, \widehat{\mathbf{S}}_\gamma \quad \sim \quad \mathcal{N}(\xi\widehat{\mathbf{S}}_\gamma\widehat{\mathbf{R}}\widehat{\mathbf{S}}_\gamma^{-1}\boldsymbol{\gamma}, \widehat{\mathbf{S}}_\gamma\widehat{\mathbf{R}}\widehat{\mathbf{S}}_\gamma). \tag{2.6}$$

Next, we sketch the VBEM algorithm using the parameter expanded in Equation (2.6) for

MR-LDP and algorithmic details for MR-LD can be found in the supplementary document. The model parameters for MR-LDP after parameter expansion become $\boldsymbol{\theta} = \{\beta_0, \sigma^2_{\boldsymbol{\gamma}}, \sigma^2_{\boldsymbol{\alpha}}, \xi\}$. Given variational posterior distribution $q(\boldsymbol{\gamma}, \boldsymbol{\alpha})$, it is straightforward to evaluate the marginal log-likelihood by decomposing it into two parts, the evidence lower bound (ELBO) and the Kullback-Leibler (KL) divergence, which is denoted as follows:

$$\log \Pr(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\Gamma}} | \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}, \widehat{\mathbf{R}}; \boldsymbol{\theta}) = \mathcal{L}(q) + \mathbb{KL}(q \| p), \tag{2.7}$$

where

$$\begin{aligned} \mathcal{L}(q) &= \iint_{\boldsymbol{\gamma}, \boldsymbol{\alpha}} q(\boldsymbol{\gamma}, \boldsymbol{\alpha}) \log \frac{\Pr(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\Gamma}}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}, \widehat{\mathbf{R}}; \boldsymbol{\theta})}{q(\boldsymbol{\gamma}, \boldsymbol{\alpha})} d\boldsymbol{\gamma}\, d\boldsymbol{\alpha}, \\ \mathbb{KL}(q \| p) &= \iint_{\boldsymbol{\gamma}, \boldsymbol{\alpha}} q(\boldsymbol{\gamma}, \boldsymbol{\alpha}) \log \frac{q(\boldsymbol{\gamma}, \boldsymbol{\alpha})}{p(\boldsymbol{\gamma}, \boldsymbol{\alpha} | \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\Gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}, \widehat{\mathbf{R}}; \boldsymbol{\theta})} d\boldsymbol{\gamma}\, d\boldsymbol{\alpha}, \end{aligned} \tag{2.8}$$

where $\mathcal{L}(q)$ is the ELBO of the marginal log-likelihood, and $\mathbb{KL}(q \| p)$ is the KL divergence between two distributions. Moreover, $\mathbb{KL}(q \| p) \geq 0$ with equality holding if and only if the variational posterior distribution (q) is equal to the true posterior distribution (p). As a consequence, minimizing the KL divergence is equivalent to maximizing ELBO. Compared with the standard EM algorithm, the crux of VBEM is to optimize q within a factorizable family of distributions by the mean-field assumption [35], which assumes that $q(\boldsymbol{\gamma}, \boldsymbol{\alpha})$ can be factorized as

$$q(\boldsymbol{\gamma}, \boldsymbol{\alpha}) = \prod_{j=1}^{p} q_{\gamma_j}(\gamma_j) \prod_{k=1}^{p} q_{\alpha_k}(\alpha_k). \tag{2.9}$$

This only assumption in variational inference promotes computational efficiency and scalability in large-scale computational problems given that a coordinate descent algorithm is commonly used to identify the optimal distribution q^*. To briefly show this, we first note that this factorization (2.9) is used as an approximation for the posterior distribution $p(\boldsymbol{\gamma}, \boldsymbol{\alpha} | \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\Gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}, \widehat{\mathbf{R}}; \boldsymbol{\theta})$. In the VB E-step, given the latent variables $\boldsymbol{\gamma}_{-k}$ and $\boldsymbol{\alpha}$, the terms with γ_k have a quadratic form, where $\boldsymbol{\gamma}_{-k}$ is the $\boldsymbol{\gamma}$ vector removing the k-th element. Similarly, when all other latent variables fixed, we can show that the terms with α_k also take a quadratic form. Thus, the variational posterior distribution for γ_k and α_k are both from Gaussian distributions, $\mathcal{N}(\mu_{\gamma_k}, \sigma^2_{\gamma_k})$ and $\mathcal{N}(\mu_{\alpha_k}, \sigma^2_{\alpha_k})$, respectively, where we call $\{\mu_{\gamma_k}, \sigma^2_{\gamma_k}, \mu_{\alpha_k}, \sigma^2_{\alpha_k}\}_{k=1,\ldots,p}$ variational parameters. The details of derivations for updating these variational parameters, and the ELBO $\mathcal{L}(q)$ in the marginal log-likelihood (2.7) at the old parameter $\boldsymbol{\theta}^{old}$ can be found in the supplementary

document. After updating variational parameters in the VB E-step, model parameters ($\boldsymbol{\theta}$) can be updated by setting the derivative of the ELBO to zero. Derivation details can be found in supplementary document, where we summarize the PX-VBEM algorithms for MR-LD and MR-LDP in Algorithms 1 and 2, respectively.

Inference for causality We can easily formulate the problem (2.5) as a statistical test for the null hypothesis that the health risk factor is not associated with the disease of interest, or $\mathcal{H}_0 : \beta_0 = 0$. Testing this hypothesis requires evaluating the marginal log-likelihood of observed data in MR-LD or MR-LDP similar to what has been done previously in [36, 37]; details are given in supplementary document. As VB searches within a factorizable family for the posterior distribution, one can only obtain an approximation for the posterior distribution of latent variables. Earlier works showed that VBEM provides useful and accurate posterior mean estimates [38]. Despite its computational efficiency and accuracy for estimating posterior mean, VB suffers from under-estimating the variance of the target distribution [25, 39, 40]. Thus, the evidence lower bound (ELBO) from VB-type algorithm cannot be directly used to conduct a likelihood-based test. In this paper, we follow Yang et al. [37] and adopt the similar strategy to calibrate ELBO as well as mitigate the bias of variance. Details for the PX-VBEM algorithm and the calibration of ELBO can be found in the supplementary document.

Relationship between MR-LD and TWAS Using transcriptome data as risk factors, MR-LD can be viewed as a TWAS-type analysis using summary-level data from both expression quantitative trait loci (eQTL) and GWAS, where eQTL and GWAS summary statistics are used for SNP-exposure and SNP-outcome, respectively. Since TWAS-type analysis only seeks genes that are significantly associated with the outcome of interest at the genome-wide level, one cannot infer causality without excluding other potential associations, e.g., horizontal pleiotropy. We note that PMR-Egger [41] was recently proposed to calibrate the type-I error control by using a burden test assumption to infer causal relationship. However, this assumption depends heavily on the fact that all effect sizes from horizontal pleiotropy are the same. Therefore, MR-LDP can also be viewed as a relaxation of the burden assumption that is more powerful to account for horizontal pleiotropy with more general patterns.

3 Results

3.1 Simulations

Methods for comparison We compared the performance of five methods in the main text: (1) our MR-LD and MR-LDP implemented in the R package *MR.LDP*; (2) GSMR implemented in the R package *gsmr*; (3) RAPS implemented in the R package *mr.raps*; (4) IVW implemented in the R package *MendelianRandomization*; (5) MR-Egger implemented in the R package *MendelianRandomization*. All methods were used with default settings. We conducted comprehensive simulation studies to better gauge the performance of each method in simulation studies in terms of type-I error control and point estimates.

In simulation studies, we considered genetic instruments both without and with horizontal pleiotropy. In the scenario that genetic instruments have horizontal pleiotropy, we further considered two cases: sparse and dense horizontal pleiotropy, i.e., sparse horizontal pleiotropy indicates that only a proportion of genetic instruments have direct effects on the outcome while dense horizontal pleiotropy indicates that all genetic instruments have direct effects. As GSMR is a step-wise method that first removes invalid instruments, dense horizontal pleiotropy theoretically implies that all genetic instruments are invalid. To make fair comparisons, we considered sparse horizontal pleiotropy with sparsity at 0.2 or 0.4. In addition, as RAPS, IVW, and MR-Egger tend to inflate type-I error in the presence of LD, we conducted SNP pruning for a fair comparison of point estimates.

Simulation settings To make our simulations as realistic as possible, we started by generating the individual-level two-sample data as follows

$$\mathbf{x} = \mathbf{G}_1\boldsymbol{\gamma} + \mathbf{U}_x\boldsymbol{\eta}_x + \mathbf{e}_1, \quad \mathbf{y} = \beta_0\mathbf{x} + \mathbf{G}_2\boldsymbol{\alpha} + \mathbf{U}_y\boldsymbol{\eta}_y + \mathbf{e}_2,$$

where $\mathbf{G}_1 \in \mathbb{R}^{n_1\times p}$ and $\mathbf{G}_2 \in \mathbb{R}^{n_2\times p}$ were both genotype matrices, $\mathbf{U}_x \in \mathbb{R}^{n_1\times q}$ and $\mathbf{U}_y \in \mathbb{R}^{n_2\times q}$ were matrices for confounding variables, n_1 and n_2 were the corresponding sample sizes, p was the number of genetic variants, $\mathbf{x} \in \mathbb{R}^{n_1\times 1}$ was the exposure vector, $\mathbf{y} \in \mathbb{R}^{n_2\times 1}$ was the outcome vector, and the error terms $\mathbf{e}_1$ and $\mathbf{e}_2$ were obtained from $\mathcal{N}(\mathbf{0}, \sigma^{\mathbf{2}}_{\mathbf{e_1}}\mathbf{I}_{\mathbf{n_1}})$ and $\mathcal{N}(\mathbf{0}, \sigma^{\mathbf{2}}_{\mathbf{e_2}}\mathbf{I}_{\mathbf{n_2}})$, respectively. In this generative model, β_0 was the true causal effect while $\boldsymbol{\alpha}$ exhibited the direct effects on the disease. We considered two cases: dense and sparse horizontal pleiotropy. For the dense case, we assumed that α_ks was independent and identically distributed as $\mathcal{N}(0, \sigma^2_{\boldsymbol{\alpha}})$. However, for the sparse case, we assumed that only a fraction of α_ks was from a Gaussian

distribution and remainders were zero. In simulations, we considered sparsity both at 0.2 and 0.4. Note that $\sigma^2_{\boldsymbol{\alpha}}$ was set by controlling the heritability due to horizontal pleiotropy. Moreover, to mimic the real applications where an external reference panel was applied to estimate the correlation among SNPs, another genotype matrix $\mathbf{G}_3 \in \mathbb{R}^{n_3 \times p}$ was generated as the reference panel data to estimate the correlation matrix, where n_3 was the sample size in the reference panel. We fixed $n_1 = n_2 = 20,000$ but varied $n_3 \in \{500, 2,500, 4,000\}$. In details, we first generated a data matrix from multivariate normal distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\rho))$, where $\boldsymbol{\Sigma}(\rho)$ is a block autoregressive (AR) with $\rho = 0, 0.4$, or 0.8 representing weak, moderate or strong LD, respectively. We then generated minor allele frequencies from a uniform distribution $\mathbb{U}(0.05, 0.5)$ and categorized the data matrix into dosage values $\{0, 1, 2\}$ according to Hardy-Weinberg equilibrium under the generated minor allele frequencies. The number of blocks was $M = 10$ or 20 and the number of SNPs within each block was 50. Correspondingly, $p = 500$ or $1,000$. For confounding variables, we sampled each column of $\mathbf{U}_x$ and $\mathbf{U}_y$ from a standard normal distribution with fixed $q = 50$ while $\boldsymbol{\eta}_x \in \mathbb{R}^{q\times 1}$ and $\boldsymbol{\eta}_y \in \mathbb{R}^{q\times 1}$ were the corresponding coefficients of confounding factors. Each row of $(\boldsymbol{\eta}_x, \boldsymbol{\eta}_y)$ was generated from a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$, and $\boldsymbol{\Sigma}_\eta$ is a two-by-two matrix with diagonal elements set as 1 and off-diagonal elements set as 0.8.

We then conducted single-variant analysis to obtain the summary statistics for SNP-exposure and SNP-outcome, $\{\widehat{\gamma}_k, \widehat{s}^2_{\boldsymbol{\gamma}k}\}_{k=1,\cdots,p}$ and $\{\widehat{\Gamma}_k, \widehat{s}^2_{\boldsymbol{\Gamma}k}\}_{k=1,\cdots,p}$, respectively. In simulations, we controlled the signal magnitude for both $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$ using their corresponding heritability, $h^2_{\boldsymbol{\gamma}} = \frac{\text{var}(\beta_0 \mathbf{G}_1 \boldsymbol{\gamma})}{\text{var}(\mathbf{y})}$ and $h^2_{\boldsymbol{\alpha}} = \frac{\text{var}(\mathbf{G}_2 \boldsymbol{\alpha})}{\text{var}(\mathbf{y})}$, respectively. Thus, we could control $h^2_{\boldsymbol{\alpha}}$ and $h^2_{\boldsymbol{\gamma}}$ at any value by controlling confounding variables, and the error terms, $\sigma^2_{\mathbf{e}_1}$ and $\sigma^2_{\mathbf{e}_2}$. In all settings, we fixed $h^2_{\boldsymbol{\gamma}} = 0.1$ and varied $h^2_{\boldsymbol{\alpha}} \in \{0, 0.05, 0.1\}$.

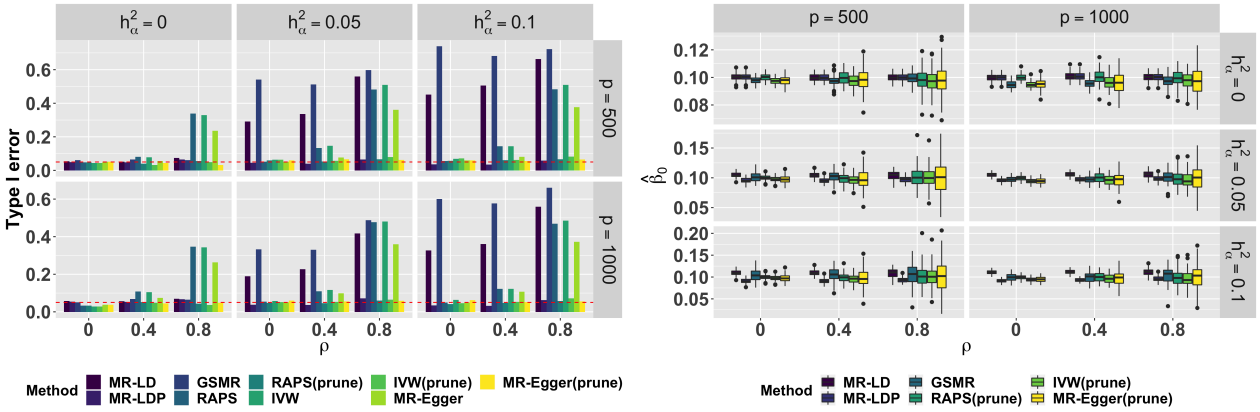
Simulation results: Type-I error control and point estimates We conducted various simulation studies to make comparisons of MR-LD and MR-LDP with other four commonly used alternative methods: (1) IVW; (2) MR-Egger; (3) GSMR; (4) RAPS. We first compared the type-I error rate for MR-LD and MR-LDP together with other alternative methods based on 1,000 replications. The simulation results for dense pleiotropy and sparse pleiotropy with sparsity at 0.2 and 0.4 are shown in Figures 2, and S2 - S8, respectively with $n_3 = 500; 2,500; 4,000$, respectively. Note that when $h^2_{\boldsymbol{\alpha}} = 0$, there was no difference between dense and sparse pleiotropy. As shown in the left column of Figure 2A, in the case of no horizontal pleiotropy ($h^2_{\boldsymbol{\alpha}} = 0$), all methods

could control type-I error at the nominal level 0.05 generally well when genetic variants were independent ($\rho = 0$). However, as LD become stronger ($\rho = 0.4$ or 0.8), alternative methods failed to control type-I error without SNP pruning. In this setting ($h^2_{\boldsymbol{\alpha}} = 0$), MR-LD and MR-LDP performed equally well in type-I error control. In the presence of horizontal pleiotropy ($h^2_{\boldsymbol{\alpha}} = 0.05$ or 0.1), as shown in the middle and right columns of Figure 2A, MR-LD failed to control type-I error for all ρ values while type-I error rates of alternative methods without SNP pruning were not controlled in the case of moderate or strong LD. However, MR-LDP could still control type-I error at its nominal level. The similar patterns could be observed for settings under sparse horizontal pleiotropy with sparsity at 0.2 and 0.4 as shown in Figures 2C, and S4 - S8, where the settings was not in favor of MR-LDP. Note that after SNP pruning, genetic variants that remained could be taken as independent. Thus, alternative methods after SNP pruning could control type-I error in all settings. However, this is achieved at the expense of losing weak instruments in LD.

Next, we made comparisons of point estimates for MR-LD and MR-LDP together with alternative methods, where SNP pruning was performed for analysis using alternative methods. In this simulation, $\beta_0 = 0.1$ and results were based on 100 replications. Clearly, as shown in Figure 2B, the proposed methods, MR-LD and MR-LDP, had narrower standard errors than alternative methods when LD was moderate or strong ($\rho = 0.4$ or 0.8) as the number of valid instruments were less after SNP pruning for alternative methods. MR-LD and MR-LDP performed equally well in the case of no horizontal pleiotropy, while MR-LD that did not account for horizontal pleiotropy was biased. Similar patterns could be observed for dense and sparse pleiotropy both at sparsity equaling 0.2 and 0.4, as shown in Figures 2D, and S4 - S8.

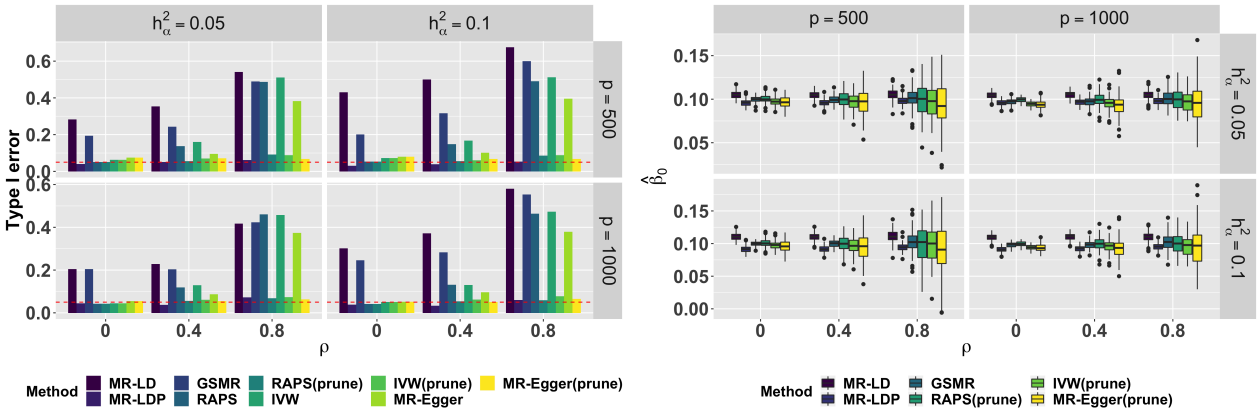
3.2 CAD-CAD and Height-Height studies

In addition, we used real datasets, i.e., CAD-CAD and Height-Height pairs, to compare the estimates from MR-LD and MR-LDP with those from other four alternative methods, where the causal effect β_0 can be taken as known, i.e., $\beta_0 = 1$. In these two examples, we used GWAS summary statistics for the same traits (i.e., CAD and BMI, respectively) from three datasets – selection, exposure and outcome [42]. The first two datasets are non-overlapping GWAS for the same trait. The exposure dataset and outcome dataset are non-overlapping individuals from European ancestry. Since IVW, MR-Egger, and RAPS are designed for independent or weak-LD SNPs and GSMR only works for SNPs with moderate LD, we conducted the LD-based



(A) Type-I error

(B) Point estimates for β_0



(C) Type-I error

(D) Point estimates for β_0

Figure 2: Simulation of type-I error control and point estimates under the dense horizontal pleiotropy (A, B) and the sparse (0.2) horizontal pleiotropy (C, D). $n_1 = n_2 = 20,000, n_3 = 500$.

clumping to obtain the near-independent SNPs based on PLINK [43]. Individual-level genotype data from UK10K projects was served as the reference panel in this study.

For CAD-CAD analysis, the selection dataset is myocardial infarction (MI) data from UK Biobank (UKB), the exposure data is obtained from the C4D Genetics Consortium [44], and the outcome data is obtained from the transatlantic Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) [45]. We first selected instrumental variants using MI from UKB under different p-value thresholds and then conducted MR analysis between the exposure and the outcome using MR-LD, MR-LDP, least squares (LS), IVW, MR-Egger, Raps and GSMR . First, the scatter plots of $\widehat{\boldsymbol{\gamma}}$ (C4D) against $\widehat{\boldsymbol{\Gamma}}$ (CAD1) are shown in Figure S9 in the supplementary document, where we found that when a large threshold, e.g., p-value=0.001, is applied to select more genetic variants, the points in the center make the inference for causality difficult. We reported the point estimates with its 95% corresponding confidence intervals for all methods in Figures 3 and S10 for $\lambda = 0.1$ and 0.15, respectively.

Clearly, MR-LD and MR-LDP were superior to other methods in terms of smaller bias and shorter confidence interval when the number of instrumental variants is large. Moreover, the estimates from MR-LD and MR-LDP also exhibited statistical significance consistently, while the coverage of $\beta_0 = 1$ from other methods was incorrect under small thresholds except for RAPS with larger standard errors due to the SNP pruning. Additionally, estimates from GSMR, IVW, and MR-Egger were always biased when the threshold was small.

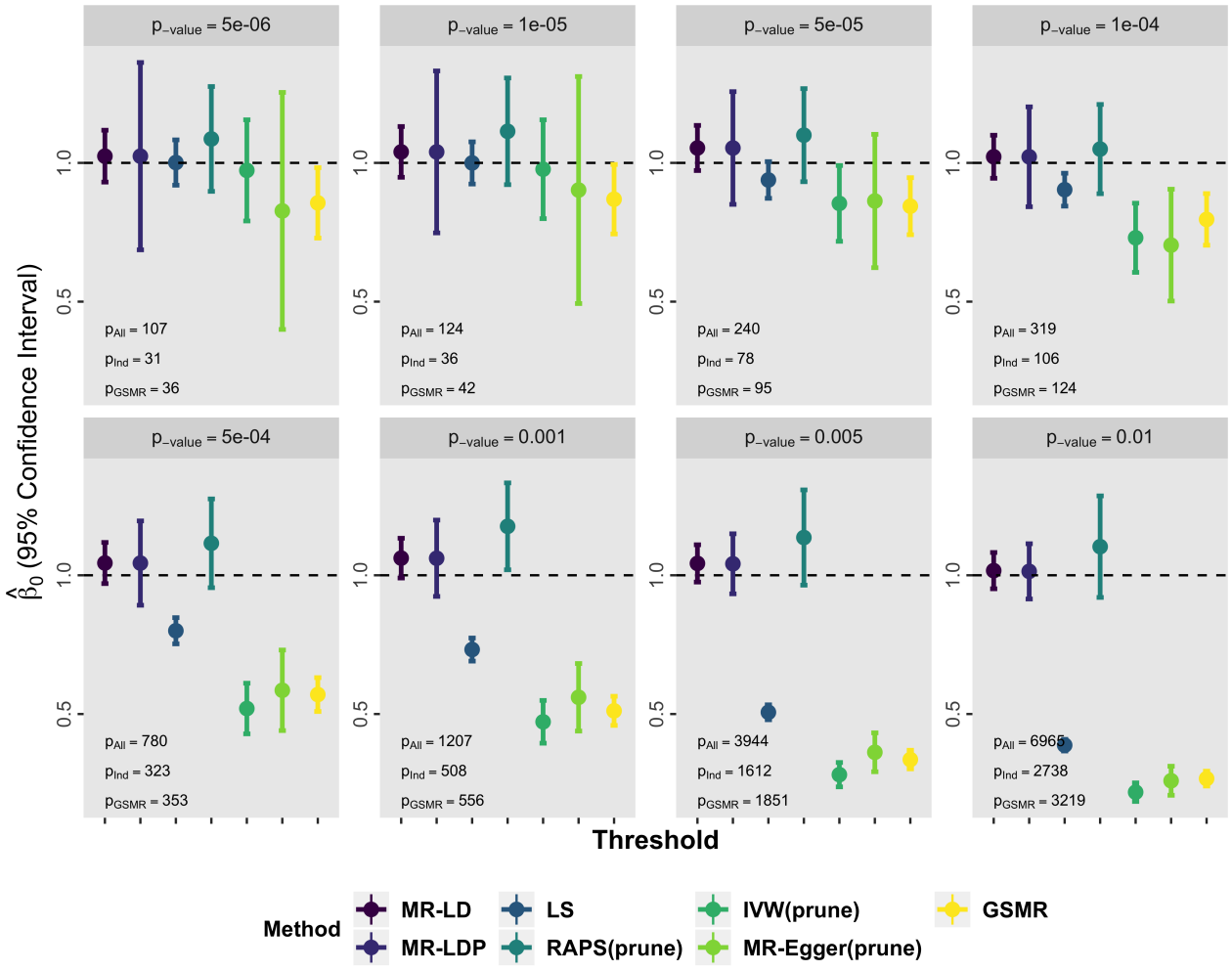


Figure 3: The result of estimates and confidence intervals for CAD-CAD using UK10K as the reference panel with shrinkage parameter $\lambda = 0.1$ under different p-value thresholds to choose genetic variants. MR-LD, MR-LDP and LS methods use all SNPs selected by the screening dataset. Default value is used to choose r^2 in GSMR and the other three methods is 0.001.

Next, we investigated the case that both the exposure and outcome were from human height. In particular, we treated the height in UK Biobank [46] as the screening dataset. The exposure data is from the height for males in a European population-based study and the outcome data is from the height for females in EUR population [47]. First,the scatter plot of $\widehat{\gamma}$ (height for

males) against $\widehat{\mathbf{\Gamma}}$ (height for females) are shown in Figure S11 in the supplementary document. Since height is highly polygenic and sample size is very large in [47] (270,000 individuals), the points are crowded in the middle even with a very small threshold (p-value $= 5 \times 10^{-6}$). The results of point estimates with their 95% confidence intervals were illustrated in Figure S12 - S13 for $\lambda = 0.1$ and 0.15, respectively. Similar patterns were observed in all cases. In particular, RAPS only offered a better performance with larger instrumental variants but did not work for some small thresholds, GSMR failed to estimate the causal effect for this validation study, and other methods underestimated the causal effect with relatively larger standard errors. MR-LD and MR-LDP used all SNPs passing a certain thresholding value and thus provided more accurate estimates of $\beta_0 = 1$.

3.3 The causal effects of lipids and BMI on common human diseases

We further applied our method, MR-LDP, to estimate the causal effects of lipids and BMI on complex diseases including coronary artery disease (CAD1 and CAD2 from CARDIoGRAM and UKB, respectively), asthma, allergic rhinitis (AR), cancer, major depression disorder (MDD), type 2 diabetes (T2D), dyslipidemia (Dyslid), hypertensive disease (Hyper), hemorrhoids, hernia abdominopelvic cavity hernia, insomnia, iron deficiency anemias (IDA), irritable bowel syndrome (IBS), macular degeneration, osteoarthritis, osteoporosis, peripheral vascular disease (PVD), peptic ulcer (PU), psychiatric disorder, acute reaction to stress (Stress), varicose veins (VV), and disease count (DC). The summary statistics for risk factors include lipoprotein cholesterol(HDL-C), low-density lipoprotein cholesterol (LDL-C), total cholesterol (TC), and body mass index (BMI). Tables S6 and S7 in the supplementary document summarize the total number of SNPs and sample sizes for each trait in each health risk factor or disease outcome and the details for the sources of these GWAS summary statistics.

First, we applied MR-LDP together with alternative methods to analyze to the exposure-outcome pairs using lipids as the exposure, i.e., HDL-C, LDL-C, and TC. Specifically, the selection and exposure datasets were obtained from [48] and [49], respectively, where the threshold for selecting instrumental variants in the selection dataset is set to 1×10^{-4}. The association results from the analysis are summarized in Table 1. Note that we did SNP pruning for RAPS, IVW, and MR-Egger and used the default settings in all alternative methods. As GSMR removes SNPs by providing an LD threshold, we chose to use $r^2 = 0.05$ as suggested by its paper [15].

In practice, HDL-C and LDL-C are often referred as "good" and "bad" cholesterol, respectively. HDL-C is known to be inversely correlated with heart and vascular diseases. We found several significant protective effects of HDL-C against CAD1 ($\widehat{\beta} = -0.09$), CAD2 disease ($\widehat{\beta} = -0.08$), T2D ($\widehat{\beta} = -0.09$), Dyslid ($\widehat{\beta} = -0.14$), Hyper ($\widehat{\beta} = -0.05$) , PVD ($\widehat{\beta} = -0.11$) and DC ($\widehat{\beta} = -0.04$), which is consistent with known epidemiological associations in the same direction [50, 51, 52]. In particular, although HDL-C was found to be associated with CAD in multiple observational studies [53, 54, 55], the role of HDL-C in CAD was overturned by later studies [56, 57]. Recently, Zhao et al. [42] showed that the effect of HDL-C in CAD is heterogeneous using different instruments. Moreover, MR-LDP identified the significant negative causality between HDL-C and PVD, which is consistent with previous studies [58, 59]. On the other hand, MR-LDA identified the significant positive causality between LDL-C and CAD which is consistent with the fact that LDL-C narrows the arteries and increases the chance of developing heart diseases. Regarding TC, MR-LDP identified the significant risk effects for cardiovascular disease as confirmed by RCTs .

To better understand of the impact of different thresholds, we re-performed the analysis for HDL-C on CAD1, CAD2, and PVD, separately, using a sequence of thresholds as shown in Figures 4, and S14 - S18. Several patterns can be observed: 1. Methods taking into account LD have small standard errors; 2. Using more SNPs under larger thresholds, the standard errors become smaller; 3. As thresholds become relatively large, e.g., 0.005, the point estimates tend to be biased. The first two patterns are expected. Generally, MR-LDP is robust under different thresholds but shows biasedness when the threshold is too liberal, which is primarily due to the inclusion of invalid variants. As the threshold is relatively large, more genetic variants with no associations to the exposure are included in the analysis, which induce biasedness either upward or downward depending on the directions of effects for invalid instrumental variants.

Second, we examine the associations between BMI and common diseases where the exposure and the selection datasets were obtained from GIANT [60] and [61], respectively. We chose threshold to be 1×10^{-4} for selecting the instrumental variants from the selection dataset. The association results from the analysis are summarized in Table 2. Overall, our MR-LDP detected relatively more significant causality between BMI and complex diseases in this study. The extent of obesity increase the risk of certain diseases, such as heart disease, type 2 diabetes and hypertensive disease identified by RCT [62].

Lipids	Outcome	$\#SNP_{ALL}$	MR-LDP	$\#SNP_{GSMR}$	GSMR(prune)	$\#SNP_{LD}$	RAPS(prune)	IVW(prune)	MR-Egger(prune)
HDL-C	CAD1	2104	-0.09(0.027)	269	-0.26(0.038)	203	-0.38(0.07)	-0.36(0.07)	-0.28(0.157)
	CAD2	2071	-0.08(0.02)	277	-0.07(0.03)	206	-0.15(0.047)	-0.15(0.047)	-0.08(0.098)
	T2D	2071	-0.09(0.031)	272	-0.16(0.044)	206	-0.33(0.081)	-0.35(0.082)	0.03(0.17)
	Dyslid	2071	-0.14(0.023)	255	-0.1(0.03)	206	-0.23(0.08)	-0.26(0.076)	-0.17(0.158)
	Hyper	2071	-0.05(0.017)	270	-0.14(0.022)	206	-0.2(0.037)	-0.21(0.038)	-0.09(0.079)
	PVD	2071	-0.11(0.048)	277	-0.12(0.077)	206	-0.19(0.109)	-0.19(0.105)	0.12(0.222)
	DC	2071	-0.04(0.01)	270	-0.08(0.013)	206	-0.09(0.025)	-0.1(0.025)	-0.03(0.052)
LDL-C	CAD1	1867	0.27(0.029)	257	0.42(0.037)	193	0.34(0.065)	0.32(0.062)	0.33(0.133)
	CAD2	1820	0.11(0.021)	266	0.16(0.027)	199	0.15(0.043)	0.14(0.043)	0.26(0.085)
	Dyslid	1820	0.56(0.03)	258	0.94(0.027)	199	0.9(0.053)	0.86(0.051)	0.93(0.1)
	DC	1820	0.08(0.01)	267	0.13(0.012)	199	0.13(0.019)	0.13(0.019)	0.17(0.037)
TC	CAD1	2546	0.24(0.028)	309	0.46(0.036)	215	0.41(0.061)	0.39(0.062)	0.35(0.146)
	CAD2	2484	0.08(0.02)	314	0.16(0.029)	218	0.15(0.043)	0.14(0.043)	0.22(0.094)
	Dyslid	2484	0.54(0.03)	303	1.08(0.029)	218	0.93(0.055)	0.9(0.051)	0.97(0.111)
	DC	2484	0.06(0.01)	314	0.13(0.012)	218	0.12(0.019)	0.12(0.019)	0.14(0.041)

Table 1: Causal associations of lipids with common diseases using UK10K as the reference panel with shrinkage parameter $\lambda = 0.1$. MR-LDP uses all SNPs selected by the screening dataset. The thresholds of r^2 for GSMR and the other three methods are 0.05 and 0.001, respectively. Statistically significant results are indicated in blue.

We also estimated some causal effects that are rarely involved in the previous MR analysis but reported in the epidemiological studies. For instance, BMI is an important risk factor for hemorrhoids [63]. BMI is positively associated with knee osteoarthritis and sleep duration reported by [64] and [65], respectively. We also confirmed a protective effect of BMI on osteoporosis reported by [66] and [67]. Moreover, the increased BMI is also considered to be one of the contributing factors for peripheral vascular disease [68].

In addition, MR-Egger is too conservative to identify the causal relationship between BMI and common diseases, and the same conclusion can be found in [18]. Similar to lipids studies, we re-performed the analysis for BMI on hemorrhoids and PVD, respectively, using a sequence of thresholds as shown in Figures S19 - S22. The patterns are similar to those in Figures 4, and S14 - S18

Outcome	$\#SNP_{ALL}$	MR-LDP	$\#SNP_{GSMR}$	GSMR(prune)	$\#SNP_{LD}$	RAPS(prune)	IVW(prune)	MR-Egger(prune)
CAD1	4405	0.2(0.084)	701	0.33(0.07)	563	0.2(0.121)	0.17(0.091)	0.2(0.129)
Asthma	4428	0.28(0.073)	707	0.23(0.061)	563	0.24(0.107)	0.19(0.08)	0.18(0.115)
CAD2	4428	0.23(0.066)	708	0.21(0.062)	563	0.26(0.105)	0.2(0.079)	0.22(0.113)
T2D	4428	0.85(0.141)	708	0.84(0.091)	563	1.22(0.16)	0.93(0.124)	1.46(0.175)
Dyslip	4428	0.22(0.076)	704	0.29(0.059)	563	0.18(0.133)	0.16(0.086)	0.29(0.124)
Hemorrhoids	4428	0.3(0.135)	709	0.2(0.111)	563	0.15(0.17)	0.11(0.129)	-0.1(0.184)
Hyper	4428	0.47(0.066)	703	0.5(0.047)	563	0.58(0.095)	0.46(0.067)	0.54(0.097)
Insomnia	4428	0.77(0.235)	708	0.85(0.215)	563	1.24(0.325)	0.96(0.246)	0.6(0.353)
Osteoa	4428	0.27(0.078)	709	0.27(0.068)	563	0.26(0.114)	0.2(0.084)	0.39(0.119)
Osteop	4428	-0.44(0.178)	709	-0.36(0.15)	563	-0.62(0.238)	-0.48(0.178)	-0.73(0.254)
PVD	4428	0.35(0.167)	709	0.41(0.159)	563	0.32(0.242)	0.24(0.183)	0.41(0.263)
DC	4428	0.27(0.035)	700	0.3(0.027)	563	0.3(0.051)	0.23(0.037)	0.26(0.053)

Table 2: Causal associations of BMI with common diseases using UK10K as the reference panel with shrinkage parameter $\lambda = 0.1$. MR-LDP uses all SNPs selected by the screening dataset. The thresholds of r^2 for GSMR and the other three methods are 0.05 and 0.001, respectively. Statistically significant results are indicated in blue.

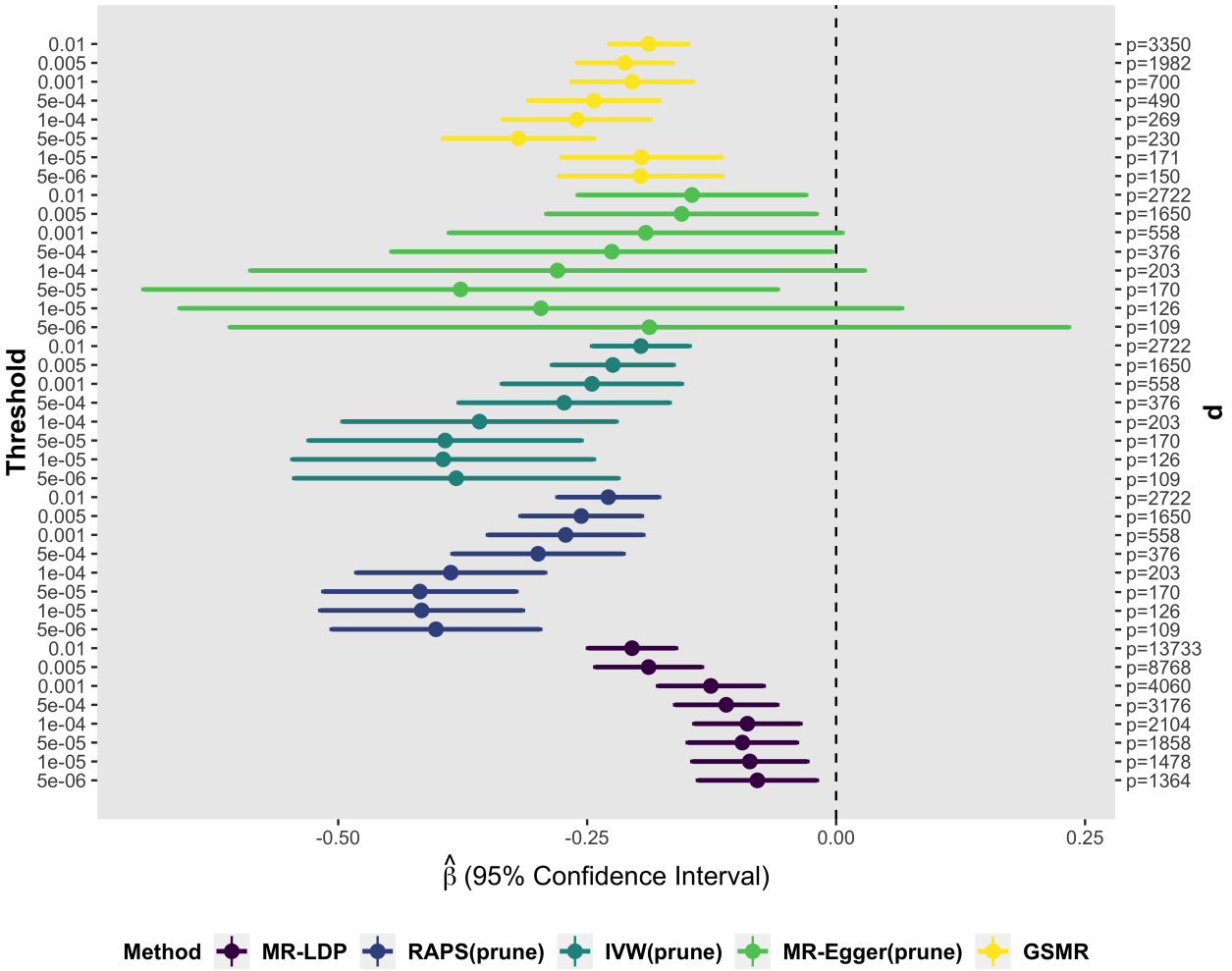


Figure 4: The causal associations of HDL-C on CAD1 under different thresholds using UK10K as the reference panel with $\lambda = 0.1$.

4 Discussion

Here, we proposed a statistically rigorous and efficient approach to perform a two-sample MR analysis that accounts for both LD structure and horizontal pleiotropy using GWAS summary statistics and a genotype reference panel. We implemented our method in the R package *MR.LDP*, which is available for download at Github. MR-LDP jointly estimated the causal effect through an approximated likelihood of GWAS summary statistics from both the risk factor and disease outcome using an additional variance component to eliminate the impact of horizontal pleiotropy. Thus, MR-LDP controls for type-I error in the presence of LD structure among instrumental variants and horizontal pleiotropy and is statistically more powerful in identifying causal effects.

MR-LDP is particularly suited to analyze complex traits that have multiple instrumental

variants within LD. The key is to jointly model the distributions for summary statistics and the causal relationship between the risk factor and disease outcome. The only approximation here uses the fact of polygenicity in complex traits for the distributions of summary statistics. Moreover, we model the causality as Equation (2.4) as the average of 'local' causal effect [32]. The linear model (2.4) holds in very general situations, beyond the linear structural model presented in the supplementary document; see Appendix A in [32] for details. To further eliminate the impact of horizontal pleiotropy, we used a random effect to control the variation in disease outcome. As horizontal pleiotropy is not an estimate of interest, a Gaussian distribution with a mean of zero and a variance parameter is generally robust although the underlying horizontal pleiotropy is sparse. Therefore, the complete-data likelihood for MR-LDP can be written as Equation (2.5). As the iteration for the standard EM algorithm involves inverting $\widehat{\mathbf{R}}$, which may cause numerical failure, we developed a PX-VBEM algorithm by expanding parameters. Our previous works have shown that the parameter expansion step is crucial in speeding up the algorithm, and we refer to the supplementary document in [36] for details. To further conduct hypothesis testing for causal effects, we calibrated the EBLO from the PX-VBEM algorithm. In our numerical studies, we demonstrated that MR-LDP is more effective in controlling type-I error in the presence of LD and either sparse or dense horizontal pleiotropy. These merits enable us to apply MR-LDP using GWAS summary statistics, likely yielding more fruitful and meaningful causal discovery in the future.

We used two pairs (CAD-CAD and Height-Height) of real data to partially validate the proposed method. As the risk factor and the outcome are the same, we can take true causal effect as known ($\beta_0 = 1$). By applying MR-LD and MR-LDP with alternative methods, we found that estimates from the proposed methods can effectively cover the true β_0 with 95% confidence intervals with instrument variants chosen under a wide ranging of thresholds. When more instrumental variants come into the model under a less stringent threshold, the estimates for the causality have narrower confidence intervals or smaller standard errors. We also note that MR-LDP has wider confidence interval. This is because MR-LDP makes additional efforts to model the horizontal pleiotropy.

In this article, we primarily focus on modeling the lipids and BMI as the exposures and complex diseases as the outcomes. Using a threshold of 1×10^{-4} in the selection dataset, we identified multiple pairs of significant causal relationships, including a protective effect

of high-density lipoprotein cholesterol (HDL-C) on peripheral vascular disease (PVD), and a positive causal effect of body mass index (BMI) on hemorrhoids. We further demonstrated the robustness of MR-LDP using a sequence of threshold values to select instrumental variants. The empirical results show that the threshold of 0.001 is optimal to balance the standard error and biasedness. However, MR-LDP is not without limitations. First, MR-LDP cannot be utilized for overlapped samples in SNP-exposure and SNP-outcome. Furthermore, MR-LDP cannot address the selection bias explicitly but uses an extra SNP-exposure summary statistics to select instrumental variants.

Web Resources

MR.LDP is available at Github (`https://github.com/QingCheng0218/MR.LDP`).

BMI(Jap): `ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/AkiyamaM_28892062_GCST004904`.

Other BMI datasets: `https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files#2018_GIANT_and_UK_BioBank_Meta_Analysis_for_Public_Release`

lipids(screen datasets): `http://csg.sph.umich.edu/willer/public/lipids2010/`.

lipids(exposure datasets): `http://csg.sph.umich.edu/willer/public/lipids2013/`.

CAD datasets: `http://www.cardiogramplusc4d.org/data-downloads/`

Common human disease datasets: `http://cnsgenomics.com/data.html`.

UK10K datasets: `https://www.uk10k.org/data_access.html`.

Acknowledgements

This work was supported by grant R-913-200-098-263 from the Duke-NUS Medical School, AcRF Tier 2 (MOE2016-T2-2-029, MOE2018-T2-1-046 and MOE2018-T2-2-006) from the Ministry of Education, Singapore, grant No. 71501089, No. 11501579 and No. 71472023 from the National Natural Science Foundation of China; and grant Nos. 22302815, No. 12316116 and No. 12301417 from the Hong Kong Research Grant Council. The computational work for this article was partially performed using resources from the National Supercomputing Centre, Singapore (https://www.nscc.sg).

References

1. Gaziano, J. M., Glynn, R. J., Christen, W. G., Kurth, T., Belanger, C., MacFadyen, J., Bubes, V., Manson, J. E., Sesso, H. D., and Buring, J. E. (2009). Vitamins e and c in the prevention of prostate and total cancer in men: the physicians' health study ii randomized controlled trial. Jama *301*, 52–62.

2. Klein, E. A., Thompson, I. M., Tangen, C. M., Crowley, J. J., Lucia, M. S., Goodman, P. J., Minasian, L. M., Ford, L. G., Parnes, H. L., Gaziano, J. M., et al. (2011). Vitamin e and the risk of prostate cancer: the selenium and vitamin e cancer prevention trial (select). Jama *306*, 1549–1556.

3. Davey Smith, G. and Ebrahim, S. (2003). mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? International journal of epidemiology *32*, 1–22.

4. Yarmolinsky, J., Wade, K. H., Richmond, R. C., Langdon, R. J., Bull, C. J., Tilling, K. M., Relton, C. L., Lewis, S. J., Smith, G. D., and Martin, R. M. (2018). Causal inference in cancer epidemiology: what is the role of mendelian randomization? Cancer Epidemiology and Prevention Biomarkers *27*, 995–1010.

5. Pickrell, J. (2015). Fulfilling the promise of mendelian randomization. bioRxiv pp. 018150.

6. Bowden, R. J. and Turkington, D. A. (1984). Instrumental variables. (Cambridge university press).

7. Burgess, S., Butterworth, A., and Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. Genetic epidemiology *37*, 658–665.

8. Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J. G., Zgaga, L., Manolio, T., Rudan, I., McKeigue, P., Wilson, J. F., and Campbell, H. (2011). Abundant pleiotropy in human complex diseases and traits. The American Journal of Human Genetics *89*, 607–618.

9. Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., Duncan, L., Perry, J. R., Patterson, N., Robinson, E. B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nature genetics *47*, 1236.

10. Cotsapas, C., Voight, B. F., Rossin, E., Lage, K., Neale, B. M., Wallace, C., Abecasis, G. R., Barrett, J. C., Behrens, T., Cho, J., et al. (2011). Pervasive sharing of genetic effects in autoimmune disease. PLoS genetics *7*, e1002254.

11. of the Psychiatric Genomics Consortium, C.-D. G. et al. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. The Lancet *381*, 1371–1379.

12. 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. Journal of the American Statistical Association *111*, 132–144.

13. Bowden, J., Fabiola Del Greco, M., Minelli, C., Lawlor, D., Zhao, Q., Sheehan, N., Thompson, J., and Smith, G. D. (2018). Improving the accuracy of two-sample summary data mendelian randomization: moving beyond the nome assumption. BioRxiv pp. 159442.

14. Corbin, L. J., Richmond, R. C., Wade, K. H., Burgess, S., Bowden, J., Smith, G. D., and Timpson, N. J. (2016). Bmi as a modifiable risk factor for type 2 diabetes: refining and understanding causal estimates using mendelian randomization. Diabetes *65*, 3002–3007.

15. Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., Robinson, M. R., McGrath, J. J., Visscher, P. M., Wray, N. R., et al. (2018). Causal associations between risk factors and common diseases inferred from gwas summary data. Nature communications *9*, 224.

16. 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. Nature genetics *50*, 693.

17. Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. International journal of epidemiology *44*, 512–525.

18. Qi, G. and Chatterjee, N. (2019). Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. Nature communications *10*, 1941.

574 19. Zhao, Q., Wang, J., Hemani, G., Bowden, J., and Small, D. S. (2018). Statistical inference
575 in two-sample summary-data mendelian randomization using robust adjusted profile score.
576 arXiv preprint arXiv:1801.09652.

577 20. Bowden, J., Del Greco M, F., Minelli, C., Davey Smith, G., Sheehan, N., and Thompson,
578 J. (2017). A framework for the investigation of pleiotropy in two-sample summary data
579 mendelian randomization. Statistics in medicine *36*, 1783–1802.

580 21. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J.,
581 Sklar, P., De Bakker, P. I., Daly, M. J., et al. (2007). Plink: a tool set for whole-genome
582 association and population-based linkage analyses. The American journal of human genetics
583 *81*, 559–575.

584 22. Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). Gcta: a tool for
585 genome-wide complex trait analysis. The American Journal of Human Genetics *88*, 76–82.

586 23. Turner, S., Armstrong, L. L., Bradford, Y., Carlson, C. S., Crawford, D. C., Crenshaw,
587 A. T., De Andrade, M., Doheny, K. F., Haines, J. L., Hayes, G., et al. (2011). Quality
588 control procedures for genome-wide association studies. Current protocols in human genetics
589 *68*, 1–19.

590 24. Husson, F., Josse, J., and Pages, J. (2010). Principal component methods-hierarchical
591 clustering-partitional clustering: why would we need to choose for visualizing data. Applied
592 Mathematics Department pp. 1–17.

593 25. Berisa, T. and Pickrell, J. K. (2016). Approximately independent linkage disequilibrium
594 blocks in human populations. Bioinformatics *32*, 283.

595 26. Rothman, A. J. (2012). Positive definite estimators of large covariance matrices. Biometrika
596 *99*, 733–740.

597 27. Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B., and Eskin, E. (2014). Identifying
598 causal variants at loci with multiple signals of association. Genetics *198*, 497–508.

599 28. Chen, W., Larrabee, B. R., Ovsyannikova, I. G., Kennedy, R. B., Haralambieva, I. H.,
600 Poland, G. A., and Schaid, D. J. (2015). Fine mapping causal variants with an approximate
601 bayesian method using marginal test statistics. Genetics *200*, 719–736.

29. Zhu, X. and Stephens, M. (2017). Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. The annals of applied statistics *11*, 1561.

30. Huang, J., Jiao, Y., Liu, J., and Yang, C. (2018). Remi: Regression with marginal information and its application in genome-wide association studies. arXiv preprint arXiv:1805.01284.

31. Bowden, J., Del Greco M, F., Minelli, C., Davey Smith, G., Sheehan, N. A., and Thompson, J. R. (2016). Assessing the suitability of summary data for two-sample mendelian randomization analyses using mr-egger regression: the role of the i 2 statistic. International journal of epidemiology *45*, 1961–1974.

32. Zhao, Q., Wang, J., Bowden, J., and Small, D. S. (2018). Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score. arXiv preprint arXiv:1801.09652.

33. Bishop, C. M. (2006). Pattern recognition and machine learning. (springer).

34. Liu, C., Rubin, D. B., and Wu, Y. N. (1998). Parameter expansion to accelerate em: the px-em algorithm. Biometrika *85*, 755–770.

35. Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. Machine learning *37*, 183–233.

36. Yang, C., Wan, X., Lin, X., Chen, M., Zhou, X., and Liu, J. (2019). Comm: a collaborative mixed model to dissecting genetic contributions to complex traits by leveraging regulatory information. Bioinformatics *35*, 1644–1652.

37. Yang, Y., Shi, X., Jiao, Y., Huang, J., Chen, M., Zhou, X., Sun, L., Lin, X., Yang, C., and Liu, J. (2019). Comm-s2: a collaborative mixed model using summary statistics in transcriptome-wide association studies. bioRxiv.

38. Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. Journal of the American Statistical Association *112*, 859–877.

39. Wang, B. and Titterington, D. (2005). Inadequacy of interval estimates corresponding to variational bayesian approximations. In AISTATS Barbados.

628 40. Turner, R. E. and Sahani, M. (2011). Two problems with variational expectation maximi-
629 sation for time-series models.

630 41. Yuan, Z., Zhu, H., Zeng, P., Yang, S., Sun, S., Yang, C., Liu, J., and Zhou, X. (2019). Testing
631 and controlling for horizontal pleiotropy with the probabilistic mendelian randomization in
632 transcriptome-wide association studies. bioRxiv pp. 691014.

633 42. Zhao, Q., Chen, Y., Wang, J., and Small, D. S. (2018). Powerful genome-wide design and
634 robust statistical inference in two-sample summary-data mendelian randomization. arXiv
635 preprint arXiv:1804.07371.

636 43. Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015).
637 Second-generation plink: rising to the challenge of larger and richer datasets. Gigascience
638 *4*, 7.

639 44. Consortium, C. A. D. C. G. et al. (2011). A genome-wide association study in europeans
640 and south asians identifies five new loci for coronary artery disease. Nature genetics *43*,
641 339.

642 45. Schunkert, H., König, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., Preuss,
643 M., Stewart, A. F., Barbalic, M., Gieger, C., et al. (2011). Large-scale association analysis
644 identifies 13 new susceptibility loci for coronary artery disease. Nature genetics *43*, 333.

645 46. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A.,
646 Vukcevic, D., Delaneau, O., OConnell, J., et al. (2018). The uk biobank resource with
647 deep phenotyping and genomic data. Nature *562*, 203.

648 47. Randall, J. C., Winkler, T. W., Kutalik, Z., Berndt, S. I., Jackson, A. U., Monda, K. L.,
649 Kilpeläinen, T. O., Esko, T., Mägi, R., Li, S., et al. (2013). Sex-stratified genome-wide
650 association studies including 270,000 individuals show sexual dimorphism in genetic loci
651 for anthropometric traits. PLoS genetics *9*, e1003500.

652 48. Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki,
653 M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J., et al. (2010). Biological,
654 clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707.

655 49. Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S.,
656 Ganna, A., Chen, J., Buchkovich, M. L., Mora, S., et al. (2013). Discovery and refinement
657 of loci associated with lipid levels. Nature genetics *45*, 1274.

658 50. ASSESSMENT, R. (2009). Major lipids, apolipoproteins, and risk of vascular disease.
659 Jama *302*, 1993–2000.

660 51. Wenhui, Z., Jing, G., Ronald, H., Weiqin, L., Yujie, W., Xiaocheng, W., and Gang, H.
661 (2014). Hdl cholesterol and cancer risk among patients with type 2 diabetes. Diabetes Care
662 *37*, 3196–3203.

663 52. Kunutsor, S. K., Kieneker, L. M., Bakker, S. J. L., James, R. W., and Dullaart, R. P. F.
664 (2017). The inverse association of hdl-cholesterol with future risk of hypertension is not
665 modified by its antioxidant constituent, paraoxonase-1: The prevend prospective cohort
666 study. Atherosclerosis *263*, 219–226.

667 53. Gordon, D. J., Probstfield, J. L., Garrison, R. J., Neaton, J. D., Castelli, W. P., Knoke,
668 J. D., Jacobs Jr, D. R., Bangdiwala, S., and Tyroler, H. A. (1989). High-density lipoprotein
669 cholesterol and cardiovascular disease. four prospective american studies. Circulation *79*,
670 8–15.

671 54. Assmann, G., Schulte, H., von, E. A., and Huang, Y. (1996). The procam experience
672 and pathophysiological implications for reverse cholesterol transport. Atherosclerosis *124*,
673 S11–S20.

674 55. Silbernagel, G., Schöttker, B., Appelbaum, S., Scharnagl, H., Kleber, M. E., Grammer,
675 T. B., Ritsch, A., Mons, U., Holleczek, B., Goliasch, G., et al. (2013). High-density
676 lipoprotein cholesterol, coronary artery disease, and cardiovascular mortality. European
677 heart journal *34*, 3563–3571.

678 56. Voight, B. F., Peloso, G. M., Orho-Melander, M., Frikke-Schmidt, R., Barbalic, M., Jensen,
679 M. K., Hindy, G., Hólm, H., Ding, E. L., Johnson, T., et al. (2012). Plasma hdl cholesterol
680 and risk of myocardial infarction: a mendelian randomisation study. The Lancet *380*,
681 572–580.

682 57. Rohatgi, A., Khera, A., Berry, J. D., Givens, E. G., Ayers, C. R., Wedin, K. E., Neeland,
683 I. J., Yuhanna, I. S., Rader, D. R., de Lemos, J. A., et al. (2014). Hdl cholesterol

684 efflux capacity and incident cardiovascular events. New England Journal of Medicine *371*,
685 2383–2393.

686 58. Sentí, M., Nogués, X., Pedro-Botet, J., Rubiés-Prat, J., and Vidal-Barraquer, F. (1992).
687 Lipoprotein profile in men with peripheral vascular disease. role of intermediate density
688 lipoproteins and apoprotein e phenotypes. Circulation *85*, 30–36.

689 59. Gerald R. Fowkes, F., Housley, E., Riemersma, R. A., Macintyre, C. C., Cawood, E. H.,
690 Prescott, R. J., and Ruckley, C. V. (1992). Smoking, lipids, glucose intolerance, and blood
691 pressure as risk factors for peripheral atherosclerosis compared with ischemic heart disease
692 in the edinburgh artery study. American journal of epidemiology *135*, 331–340.

693 60. Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C.,
694 Vedantam, S., Buchkovich, M. L., Yang, J., et al. (2015). Genetic studies of body mass
695 index yield new insights for obesity biology. Nature *518*, 197.

696 61. Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., Iwata,
697 N., Ikegawa, S., Hirata, M., Matsuda, K., et al. (2017). Genome-wide association study
698 identifies 112 new loci for body mass index in the japanese population. Nature genetics *49*,
699 1458.

700 62. Group, L. A. R. et al. (2010). Long term effects of a lifestyle intervention on weight and
701 cardiovascular risk factors in individuals with type 2 diabetes: four year results of the look
702 ahead trial. Archives of internal medicine *170*, 1566.

703 63. Ravindranath, G. and Rahul, B. (2018). Prevalence and risk factors of hemorrhoids: a
704 study in a semi-urban centre. International Surgery Journal *5*, 496–499.

705 64. Manek, N. J., Hart, D., Spector, T. D., and MacGregor, A. J. (2003). The association
706 of body mass index and osteoarthritis of the knee joint: an examination of genetic and
707 environmental influences. Arthritis & Rheumatism: Official Journal of the American
708 College of Rheumatology *48*, 1024–1029.

709 65. Grandner, M. A., Schopfer, E. A., Sands-Lincoln, M., Jackson, N., and Malhotra, A. (2015).
710 Relationship between sleep duration and body mass index depends on age. Obesity *23*,
711 2491–2498.

712 66. Asomaning, K., Bertone-Johnson, E. R., Nasca, P. C., Hooven, F., and Pekow, P. S. (2006).
713 The association between body mass index and osteoporosis in patients referred for a bone
714 mineral density examination. Journal of Women's Health *15*, 1028–1034.

715 67. Barrera, G., Bunout, D., Gattás, V., de la Maza, M. P., Leiva, L., and Hirsch, S. (2004). A
716 high body mass index protects against femoral neck osteoporosis in healthy elderly subjects.
717 Nutrition *20*, 769–771.

718 68. Ylitalo, K. R., Sowers, M., and Heeringa, S. (2011). Peripheral vascular disease and
719 peripheral neuropathy in individuals with cardiometabolic clustering and obesity: National
720 health and nutrition examination survey 2001–2004. Diabetes Care *34*, 1642–1647.