

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

Qing Cheng<sup>1</sup>, Yi Yang<sup>1</sup>, Xingjie Shi<sup>1,2</sup>, Kar-Fu Yeung<sup>1</sup>, Can Yang<sup>3</sup>, Heng Peng<sup>4</sup> and Jin Liu<sup>1,\*</sup>

<sup>1</sup>Centre for Quantitative Medicine, Health Services & Systems Research, Duke-NUS Medical School, Singapore 169857, Singapore, <sup>2</sup>Department of Statistics, Nanjing University of Finance and Economics, Nanjing, 210023, China, <sup>3</sup>Department of Mathematics, The Hong Kong University of Science and Technology, Kowloon, Hong Kong and <sup>4</sup>Department of Mathematics, Hong Kong Baptist University, Kowloon, Hong Kong

Received October 17, 2019; Revised February 27, 2020; Editorial Decision April 09, 2020; Accepted April 14, 2020

## 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 (IVs) 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 proposed a probabilistic model for MR analysis in identifying the causal 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 (MR-LDP) and develop a computationally efficient algorithm to make the causal inference. We then conducted comprehensive simulation studies to demonstrate the advantages of MR-LDP over the existing methods. Moreover, we used two real exposure–outcome pairs 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 on peripheral vascular disease and a positive causal effect of BMI on hemorrhoids.

## 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 for evaluating 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, body mass index (BMI) and type 2 diabetes (T2D), and most of completed GWAS are only 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

\*To whom correspondence should be addressed. Tel: +65 65767376; Fax: +65 62251244; Email: jin.liu@duke-nus.edu.sg

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). Despite their successes, they have several limitations. First, they must obey the assumption that IVs affect the outcome exclusively through the risk exposures, which 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 pervasive 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 stepwise 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 these 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 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). RAPS addresses measurement error issues, while MR-Egger and other existing methods applicable to GWAS summary statistics assume that sampling error from SNP exposure is negligible (20).

On the other hand, the classical MR methods (e.g. IVW, MR-Egger) only work for independent IVs and further assume no measurement errors. Among the methods mentioned earlier, only GSMR is capable of accounting weak or moderate linkage disequilibrium (LD) among SNPs, while others demand all-instrumental SNPs to be independent, which is typically achieved by SNP pruning and thus reducing the number of instrumental variants for follow-up MR analysis. As SNPs in close proximity tend to be highly correlated, MR methods that do not account for LD structure may substantially lose statistical power due to the pruning process. Moreover, GSMR is a stepwise method that removes 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 an MR-LD accounting for horizontal pleiotropy (MR-LDP). Sim-

ilar to RAPS, MR-LDP does not require any 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 test 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 (EM) 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 the 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 the 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 the peripheral vascular disease (PVD) and a positive causal effect of BMI on hemorrhoids.

## MATERIALS AND METHODS

### 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 a 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 ( $N = 4284$ ), 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 Europeans, we conducted strict quality control for UK10K data using PLINK (21). First, SNPs were excluded from the analysis if their calling rates were  $<95\%$ , minor allele frequencies were  $<0.01$  or  $P$ -values were  $<1 \times 10^{-6}$  in the Hardy–Weinberg equilibrium test. We then removed the individuals with genotype missing rates  $>5\%$ . To further remove individuals with high relatedness in all samples, we used GCTA (22) to first identify those individual pairs with estimated genetic relatedness  $>0.05$  and then randomly remove one from each pair. Additionally, we carried out the principal component analysis on the individuals to identify the population stratification (23). In this way, we extracted the clustering subgroup representing the primary European ancestry using hierarchical clustering on princi-

pal component approach (24). After data pre-processing, there were 3764 individuals remaining with 989 932 SNPs.

### Choice of LD matrix

Since the LD between two SNPs decays exponentially concerning 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:

$$\widehat{\mathbf{R}}^{(l)} = \arg \min_{\mathbf{R}^{(l)} > 0} (\|\mathbf{R}^{(l)} - \widehat{\mathbf{R}}_{\text{emp}}^{(l)}\|_F^2 / 2 - \tau \log |\mathbf{R}^{(l)}| + \lambda \|\mathbf{R}^{(l)} - \mathbf{I}\|_1), \quad (1)$$

where  $\widehat{\mathbf{R}}_{\text{emp}}^{(l)}$  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 fixed the shrinkage parameter  $\lambda$  to be 0.055 in simulation studies and vary  $\lambda \in \{0.1, 0.15\}$  in real data analysis.

### 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  $\{\widehat{\boldsymbol{\gamma}}_k, \widehat{\mathbf{s}}_k^2\}_{k=1, \dots, p}$  from univariate linear regression are available:

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

where  $\mathbf{g}_k$  is the  $k$ th column of  $\mathbf{G}$ , and  $\widehat{\boldsymbol{\gamma}}_k$  and  $\widehat{\mathbf{s}}_k^2$  are the estimated effect sizes and their variance for SNP  $k$ , respectively.  $\widehat{\mathbf{R}}$  denotes the correlation among all genotyped SNPs and  $\widehat{\mathbf{S}} = \text{diag}(\widehat{\mathbf{s}}_1, \dots, \widehat{\mathbf{s}}_p)$  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 (27–30):

$$\widehat{\boldsymbol{\gamma}} | \boldsymbol{\gamma}, \widehat{\mathbf{R}}, \widehat{\mathbf{S}} \sim \mathcal{N}(\widehat{\mathbf{S}} \widehat{\mathbf{R}} \widehat{\mathbf{S}}^{-1} \boldsymbol{\gamma}, \widehat{\mathbf{S}} \widehat{\mathbf{R}} \widehat{\mathbf{S}}), \quad (2)$$

where  $\widehat{\boldsymbol{\gamma}} = [\widehat{\boldsymbol{\gamma}}_1, \dots, \widehat{\boldsymbol{\gamma}}_p]^T$ . Analogously, we apply this distribution to the two-sample MR analysis. The summary statistics for SNP exposure and SNP outcome are denoted by  $\{\widehat{\boldsymbol{\gamma}}_k, \widehat{\mathbf{s}}_{\boldsymbol{\gamma}_k}^2\}_{k=1, \dots, p}$  and  $\{\widehat{\boldsymbol{\Gamma}}_k, \widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_k}^2\}_{k=1, \dots, p}$ , respectively. Therefore,

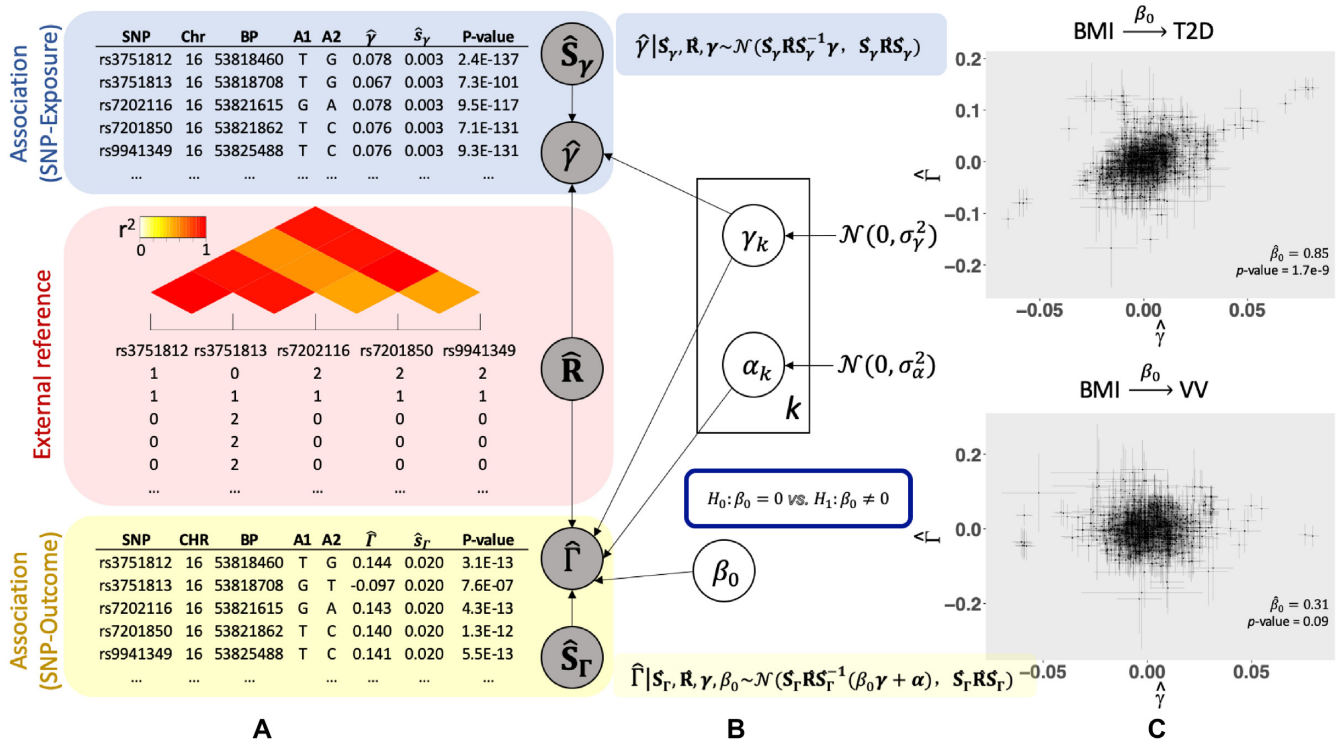
the likelihood for two-sample summary statistics can be written as

$$\begin{aligned} \widehat{\boldsymbol{\gamma}} | \boldsymbol{\gamma}, \widehat{\mathbf{R}}, \widehat{\mathbf{S}}_{\boldsymbol{\gamma}} &\sim \mathcal{N}(\widehat{\mathbf{S}}_{\boldsymbol{\gamma}} \widehat{\mathbf{R}} \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\gamma}, \widehat{\mathbf{S}}_{\boldsymbol{\gamma}} \widehat{\mathbf{R}} \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}), \\ \widehat{\boldsymbol{\Gamma}} | \boldsymbol{\Gamma}, \widehat{\mathbf{R}}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}} &\sim \mathcal{N}(\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}} \widehat{\mathbf{R}} \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} \boldsymbol{\Gamma}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}} \widehat{\mathbf{R}} \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}), \end{aligned} \quad (3)$$

where  $\widehat{\mathbf{S}}_{\boldsymbol{\gamma}} = \text{diag}(\widehat{\mathbf{s}}_{\boldsymbol{\gamma}_1}, \dots, \widehat{\mathbf{s}}_{\boldsymbol{\gamma}_p})$  and  $\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}} = \text{diag}(\widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_1}, \dots, \widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_p})$  are both diagonal matrices, and  $\widehat{\boldsymbol{\Gamma}} = [\widehat{\boldsymbol{\Gamma}}_1, \dots, \widehat{\boldsymbol{\Gamma}}_p]^T$ . 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, the distributions for summary statistics [Equation (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 the ‘Choice of LD matrix’ section.

### MR-LDP model overview

The fundamental assumptions for two-sample MR analysis include the independence among IVs, and three IV assumptions for a genetic instrument: (i) associated with health risk factors ( $\gamma \neq 0$ ); (ii) independent of unobserved confounding factors between the risk factors and the disease outcomes; and (iii) independent of the outcome given risk factors and confounders. Given the strong LD structure among SNPs and abundant horizontal pleiotropy in GWAS, these unique features invalidate the independence assumption for genetic variants and IV assumption (iii). Our proposed MR-LDP aims to make the 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 in Equation (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 of  $\boldsymbol{\alpha}$ , 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 (20,31). 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  $\alpha_k$  and  $\boldsymbol{\gamma}_k$  follow two independent Gaussian distributions. The latent variable  $\boldsymbol{\gamma}_k$  and parameter  $\beta_0$  jointly assist with formu-



**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,  $k = 1, \dots, p$ . The circles are either variables or parameters. The circle outside the box is the primary parameter of interest. The variables in shaded circles are observed (i.e. GWAS summary statistics  $\{\hat{\gamma}_k, \hat{S}_{\gamma_k}\}_{k=1, \dots, p}$  and  $\{\hat{\Gamma}_k, \hat{S}_{\Gamma_k}\}_{k=1, \dots, p}$ , and the estimated  $\hat{\mathbf{R}}$  for these  $p$  SNPs from a reference panel) and variables in unshaded circles are latent variables (i.e.  $\gamma_k$  and  $\alpha_k$ ,  $k = 1, \dots, p$ ). The primary goal is to conduct a formal hypothesis testing for  $\mathcal{H}_0: \beta_0 = 0$  versus  $\mathcal{H}_1: \beta_0 \neq 0$ . (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  $\beta_0$  and their  $P$ -values from MR-LDP are shown in each part.

lating the distribution for SNP outcome. Then, we can formalize the hypothesis testing for  $\beta_0$ , 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 mainly due to LD, and methods that do not account for LD tend to inflate findings.

## 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:

$$\boldsymbol{\Gamma}_j = \beta_0 \boldsymbol{\gamma}_j \quad \text{or} \quad \boldsymbol{\Gamma}_j = \boldsymbol{\alpha}_j + \beta_0 \boldsymbol{\gamma}_j, \quad (4)$$

where  $j = 1, \dots, p$ , considering without/with horizontal pleiotropy, respectively (12,32). Note that  $\beta_0$  is the effect size of the exposure on the outcome and  $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_p]^T$  is the vector of effects of genetic variants on the outcome due to horizontal pleiotropy. Importantly,  $\beta_0$  can be interpreted as the causal effect between exposure and outcome in the study (32). More details regarding linear structural models incorporating the relationship provided in Equation (4) are

available in the Supplementary Data. 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 Data 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:

$$\begin{aligned} & \Pr(\hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \hat{\mathbf{S}}_\gamma, \hat{\mathbf{S}}_\Gamma, \hat{\mathbf{R}}; \boldsymbol{\theta}) \\ &= \Pr(\hat{\boldsymbol{\Gamma}} | \boldsymbol{\gamma}, \boldsymbol{\alpha}, \hat{\mathbf{R}}, \hat{\mathbf{S}}_\Gamma; \beta_0) \Pr(\hat{\boldsymbol{\gamma}} | \boldsymbol{\gamma}, \hat{\mathbf{R}}, \hat{\mathbf{S}}_\gamma) \Pr(\boldsymbol{\alpha} | \sigma_\alpha^2) \Pr(\boldsymbol{\gamma} | \sigma_\gamma^2), \end{aligned} \quad (5)$$

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

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

*Algorithm.* The standard EM algorithm is a common choice to find the maximum likelihood for probabilistic models in the presence of latent variables (33). The conventional EM algorithm involves the inverse of large matrix  $\hat{\mathbf{R}}$

that is at order  $O(p^3)$  computationally, making it computational infeasible when a large number of instrumental variants are used. To address this issue, 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 [Equation (5)] as follows:

$$\widehat{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \widehat{\mathbf{R}}, \widehat{\mathbf{S}}_{\boldsymbol{\gamma}} \sim \mathcal{N}(\xi \widehat{\mathbf{S}}_{\boldsymbol{\gamma}} \widehat{\mathbf{R}} \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\gamma}, \widehat{\mathbf{S}}_{\boldsymbol{\gamma}} \widehat{\mathbf{R}} \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}). \quad (6)$$

Next, we sketch the VBEM algorithm using the parameter expanded in Equation (6) for MR-LDP and algorithmic details for MR-LD can be found in the Supplementary Data. The model parameters for MR-LDP after parameter expansion become  $\boldsymbol{\theta} = \{\beta_0, \sigma_{\boldsymbol{\gamma}}^2, \sigma_{\boldsymbol{\alpha}}^2, \xi\}$ . 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 ELBO and the Kullback–Leibler (KL) divergence, which is denoted as follows:

$$\log \Pr(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\Gamma}}|\widehat{\mathbf{S}}_{\boldsymbol{\gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}, \widehat{\mathbf{R}}; \boldsymbol{\theta}) = \mathcal{L}(q) + \mathbb{KL}(q||p), \quad (7)$$

where

$$\begin{aligned} \mathcal{L}(q) &= \iint_{\boldsymbol{\gamma}, \boldsymbol{\alpha}} q(\boldsymbol{\gamma}, \boldsymbol{\alpha}) \log \frac{\Pr(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\Gamma}}, \boldsymbol{\gamma}, \boldsymbol{\alpha}|\widehat{\mathbf{S}}_{\boldsymbol{\gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}, \widehat{\mathbf{R}}; \boldsymbol{\theta})}{q(\boldsymbol{\gamma}, \boldsymbol{\alpha})} d\boldsymbol{\gamma} d\boldsymbol{\alpha}, \\ \mathbb{KL}(q||p) &= \iint_{\boldsymbol{\gamma}, \boldsymbol{\alpha}} q(\boldsymbol{\gamma}, \boldsymbol{\alpha}) \log \frac{q(\boldsymbol{\gamma}, \boldsymbol{\alpha})}{p(\boldsymbol{\gamma}, \boldsymbol{\alpha}|\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\Gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}, \widehat{\mathbf{R}}; \boldsymbol{\theta})} d\boldsymbol{\gamma} d\boldsymbol{\alpha}. \end{aligned} \quad (8)$$

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

$$q(\boldsymbol{\gamma}, \boldsymbol{\alpha}) = \prod_{j=1}^p q_{\boldsymbol{\gamma}_j}(\boldsymbol{\gamma}_j) \prod_{k=1}^p q_{\boldsymbol{\alpha}_k}(\boldsymbol{\alpha}_k). \quad (9)$$

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 [Equation (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  $\boldsymbol{\gamma}_k$  have a quadratic form, where  $\boldsymbol{\gamma}_{-k}$  is the  $\boldsymbol{\gamma}$  vector removing the  $k$ th element. Similarly, when all other latent variables are fixed, we can show that the terms with  $\boldsymbol{\alpha}_k$  also take a quadratic form. Thus, the variational posterior distributions for  $\boldsymbol{\gamma}_k$  and  $\boldsymbol{\alpha}_k$  are both from Gaussian distributions,  $\mathcal{N}(\mu_{\boldsymbol{\gamma}_k}, \sigma_{\boldsymbol{\gamma}_k}^2)$  and  $\mathcal{N}(\mu_{\boldsymbol{\alpha}_k}, \sigma_{\boldsymbol{\alpha}_k}^2)$ , respectively, where we call  $\{\mu_{\boldsymbol{\gamma}_k}, \sigma_{\boldsymbol{\gamma}_k}^2, \mu_{\boldsymbol{\alpha}_k}, \sigma_{\boldsymbol{\alpha}_k}^2\}_{k=1, \dots, p}$  variational parameters. The details of derivations for updating these variational parameters and the ELBO  $\mathcal{L}(q)$  in the marginal log-likelihood [Equation (7)] at the old parameter  $\boldsymbol{\theta}^{\text{old}}$  can be found in the

Supplementary Data. 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 the Supplementary Data, 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 provided in Equation (5) as a statistical test for the null hypothesis that the health risk factor is not associated with the disease of interest, or  $H_0 : \beta_0 = 0$ . 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 the Supplementary Data. 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 the posterior mean, VB suffers from underestimating the variance of the target distribution (25,39–40). Thus, the ELBO from the 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 Data.

*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 as SNP exposure and SNP outcome in the analysis, 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, which makes it more powerful in accounting for horizontal pleiotropy with more general patterns.

## RESULTS

### Simulations

*Methods for comparison.* We compared the performance of five methods in the main text: (i) our MR-LD and MR-LDP implemented in the R package *MR.LDP*; (ii) GSMR implemented in the R package *gsmr*; (iii) RAPS implemented in the R package *mr.raps*; (iv) IVW implemented in the R package *MendelianRandomization*; and (v) MR-Egger implemented in the R package *MendelianRandomization*. All methods were used with default settings. Note that we performed GSMR analysis with removing outliers first when there exists horizontal pleiotropy ( $h_{\boldsymbol{\alpha}}^2 \neq 0$ ).

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: the sparse and dense horizontal pleiotropy. The sparse horizontal pleiotropy indicates that only a proportion of genetic instruments have direct effects ( $\alpha$  is sparse) on the outcome, while the dense horizontal pleiotropy indicates that all genetic instruments have direct effects ( $\alpha$  is dense). As GSMR is a stepwise method that first removes invalid instruments, the dense horizontal pleiotropy theoretically implies that all genetic instruments are invalid. To make fair comparisons, we considered the 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  $\mathcal{N}(\mathbf{0}, \sigma_{\mathbf{e}_1}^2 \mathbf{I}_{n_1})$  and  $\mathcal{N}(\mathbf{0}, \sigma_{\mathbf{e}_2}^2 \mathbf{I}_{n_2})$ , respectively. In this generative model,  $\beta_0$  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  $\alpha_k$  was independent and identically distributed as  $\mathcal{N}(0, \sigma_{\alpha}^2)$ . However, for the sparse case, we assumed that only a fraction of  $\alpha_k$  was from a Gaussian distribution and remaining were zero. In simulations, we considered sparsity at both 0.2 and 0.4. Note that  $\sigma_{\alpha}^2$  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  $n_3 \in \{500, 2500, 4000\}$ . In detail, 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 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 the 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  $1000$ . 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

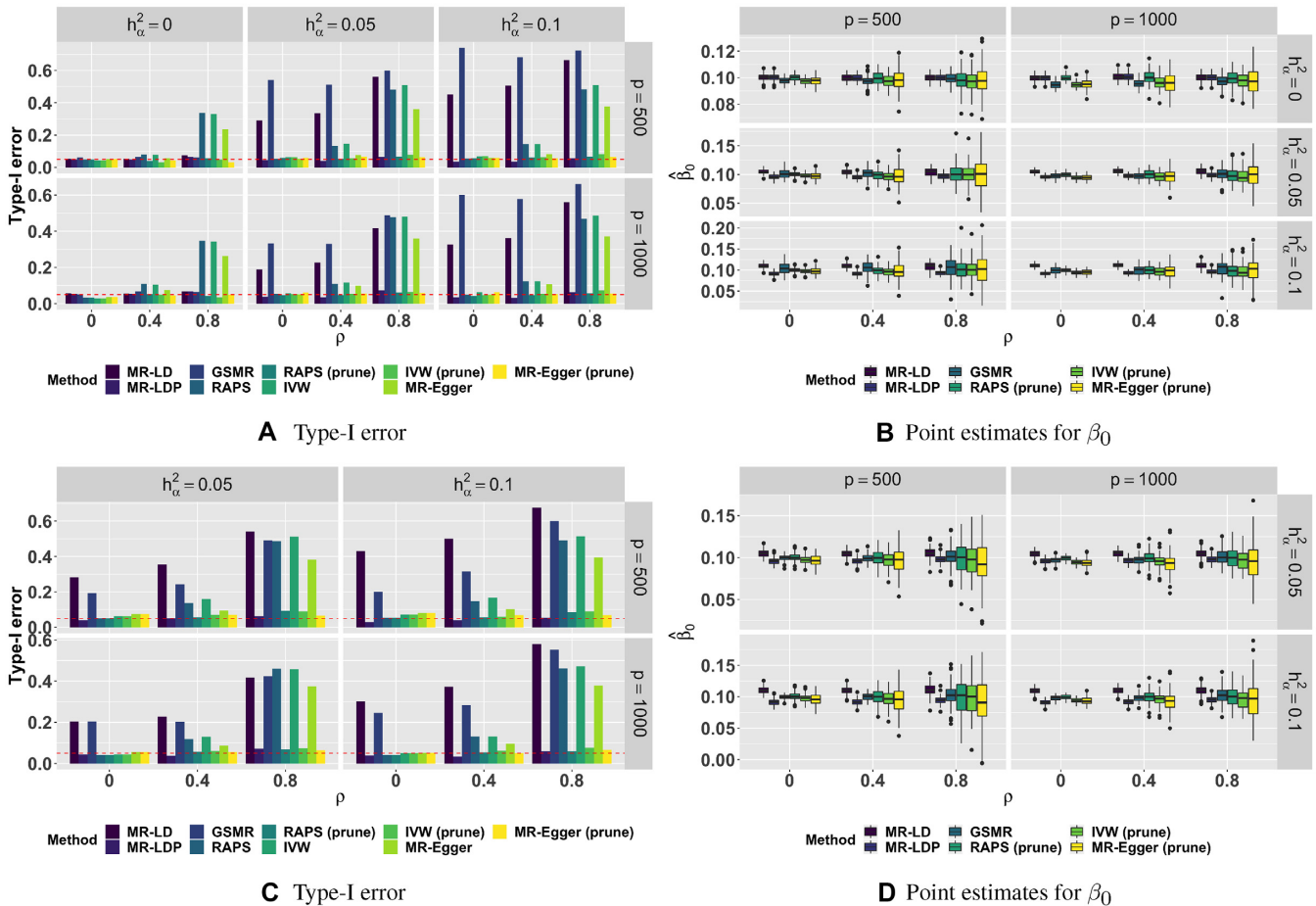
of  $(\boldsymbol{\eta}_x, \boldsymbol{\eta}_y)$  was generated from a multivariate normal distribution  $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\eta})$ , where  $\boldsymbol{\Sigma}_{\eta}$  is a  $2 \times 2$  matrix with diagonal elements set as 1 and off-diagonal elements set as 0.8.

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

*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: (i) IVW; (ii) MR-Egger; (iii) GSMR; and (iv) RAPS. We first compared the type I error rate for MR-LD and MR-LDP together with other alternative methods based on 1000 replications. The simulation results for dense pleiotropy and sparse pleiotropy with sparsity at 0.2 and 0.4 are shown in Figure 2 and Supplementary Figures S2–S8, respectively, with  $n_3 = 500, 2500$  and  $4000$ , respectively. Note that when  $h_{\boldsymbol{\alpha}}^2 = 0$ , 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 ( $h_{\boldsymbol{\alpha}}^2 = 0$ ), all methods could control type I error at the nominal level of 0.05, generally well when genetic variants were independent ( $\rho = 0$ ). However, as LD becomes stronger ( $\rho = 0.4$  or  $0.8$ ), the alternative methods failed to control type I error without SNP pruning. In this setting ( $h_{\boldsymbol{\alpha}}^2 = 0$ ), MR-LD and MR-LDP performed equally well in type I error control. In the presence of horizontal pleiotropy ( $h_{\boldsymbol{\alpha}}^2 = 0.05$  or  $0.1$ ), as shown in the middle and right columns of Figure 2A, MR-LD failed to control type I error for all  $\rho$  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 Figure 2C and Supplementary Figures 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, the proposed methods, MR-LD and MR-LDP, had smaller standard errors than alternative methods when LD was moderate or strong ( $\rho = 0.4$  or  $0.8$ ) (Figure 2B), as SNP pruning causes the alternative methods to use fewer valid instruments. MR-LD and MR-LDP performed equally well in the case of no horizontal pleiotropy, while MR-LD was biased in the presence of horizontal pleiotropy. Similar patterns could be observed for dense and sparse pleiotropy both at sparsity equaling 0.2



**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).  $n_1 = n_2 = 20\,000$ ;  $n_3 = 500$ .

and 0.4, as shown in Figure 2D and Supplementary Figures S4–S8.

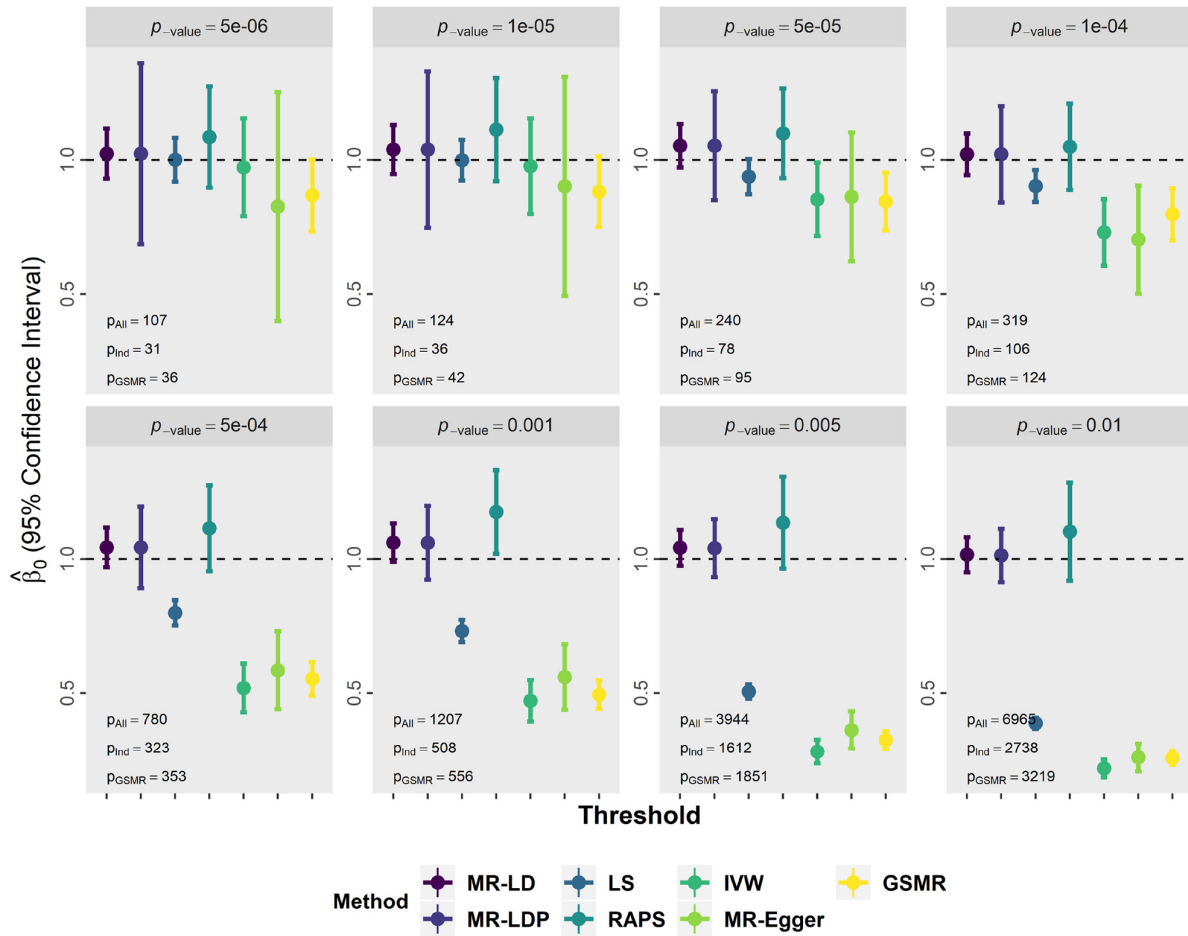
More simulation settings and their corresponding results can be found in the Supplementary Data, including simulations for robustness and power analysis, simulations using a screening dataset and simulations for binary outcome.

### 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 the other four alternative methods, where the causal effect  $\beta_0$  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—screening, 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 clumping to obtain the near-independent SNPs based on PLINK (43). Individual-level genotype data from UK10K projects served as the reference panel in this study.

We used high-performance computing from National SuperComputing Centre, Singapore (<https://www.nsc.sg>) to accomplish our computational work. For example, if the threshold of 0.1 for  $P$ -values was applied in the height–height study, there remained 40 521 SNPs and it took  $\sim 2$  and  $\sim 8$  min, respectively, for MR-LD and MR-LDP to complete the analysis on a Linux platform with a 2.60 GHz Intel Xeon CPU E5-2690 v3 with 30720 KB cache and 96 GB RAM (only 30 GB RAM used). The demo for using *MR.LDP* package can be found both in the Supplementary Data and at GitHub website. In addition, all codes for simulation studies and real data analysis can be found at GitHub website.

For CAD–CAD analysis, the screening dataset is myocardial infarction (MI) data from UK Biobank (UKB), the exposure data are obtained from the C4D Genetics Consortium (44) and the outcome data are 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  $\hat{\gamma}$  (C4D) against  $\hat{\Gamma}$  (CAD1) are shown in Supplementary Figure S13, where we



**Figure 3.** The result of estimates and confidence intervals for CAD–CAD using UK10K as the reference panel with the shrinkage parameter  $\lambda = 0.1$  under different  $P$ -value thresholds to choose genetic variants in the screening dataset, e.g.  $P$ -value =  $5e-6$  and  $1e-05$ . MR-LD, MR-LDP and LS methods use all SNPs selected by the screening dataset (denoted as  $P_{All}$ ), but IVW, MR-Egger, RAPS and GSMR use pruned SNPs, where the default value of  $r^2$  is used for GSMR (the number of SNP used:  $P_{GSMR}$ ) and  $r^2 = 0.001$  is used for VW, MR-Egger and RAPS (the number of SNP used:  $P_{Ind}$ ).

found that when a larger thresholding value, e.g.  $P$ -value = 0.001, is applied in the screening dataset (in other words, more genetic variants would be selected for MR analysis), the points crowded in the center make the inference for causality difficult as shown in Supplementary Figure S13. We reported the point estimates with its 95% corresponding confidence intervals for all methods in Figure 3 and Supplementary Figure S14 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 intervals 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.

Next, we investigated the case that both the exposure and outcome were from human height. In particular, we treated the height in UKB (46) as the screening dataset. The exposure data are from the height for males in a European population-based study, and the outcome data are from the

height for females in a European population (47). First, the scatter plot of  $\hat{y}$  (height for males) against  $\hat{F}$  (height for females) is shown in Supplementary Figure S15. Since height is highly polygenic and the sample size is very large in (47) (around 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 Supplementary Figures S16 and S17 for  $\lambda = 0.1$  and 0.15, respectively. Similar patterns were observed in all cases. In particular, RAPS only offered 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 large standard errors. MR-LD and MR-LDP used all SNPs passing a certain thresholding value and thus provided more accurate estimates of  $\beta_0 = 1$ .

### 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, in-



**Table 1.** Causal associations of lipids with common diseases using UK10K as the reference panel with the shrinkage parameter  $\lambda = 0.1$

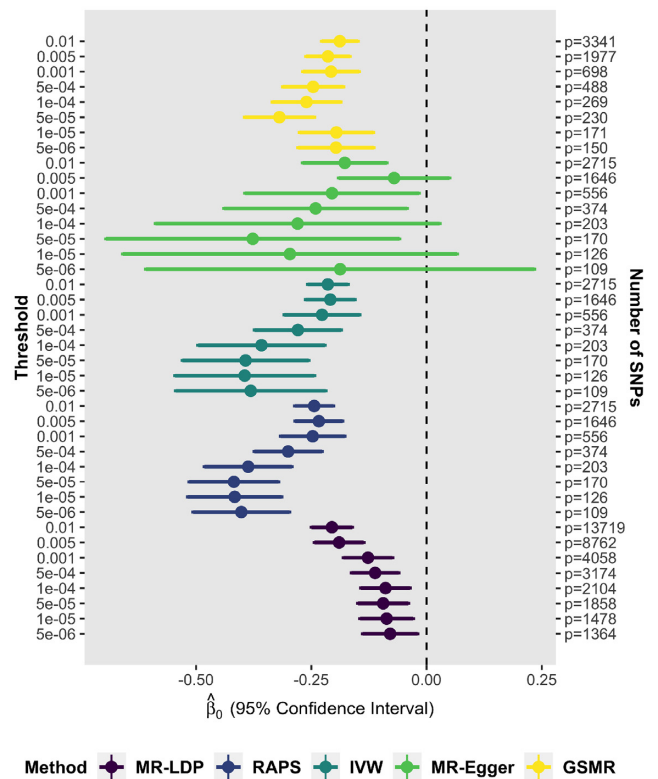
Lipids	Outcome	$P_{All}$	MR-LDP	$P_{GSMR}$	GSMR (prune)	$P_{Ind}$	RAPS	IVW	MR-Egger
HDL-C	CAD1	2104	<b>-0.09 (0.027)</b>	269	<b>-0.26 (0.038)</b>	203	<b>-0.38 (0.07)</b>	<b>-0.36 (0.07)</b>	-0.28 (0.157)
	CAD2	2071	<b>-0.08 (0.02)</b>	277	<b>-0.07 (0.03)</b>	206	<b>-0.15 (0.047)</b>	<b>-0.15 (0.047)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

MR-LDP uses all SNPs selected by the screening dataset (denoted as  $P_{All}$ ), but IVW, MR-Egger, RAPS and GSMR use pruned SNPs, where the default value of  $r^2$  is used for GSMR (the number of SNP used:  $P_{GSMR}$ ) and  $r^2 = 0.001$  is used for VW, MR-Egger and RAPS (the number of SNP used:  $P_{Ind}$ ). Statistically significant results are indicated in bold.

cluding coronary artery disease (CAD1 and CAD2 from CARDIoGRAM and UKB, respectively), asthma, allergic rhinitis, cancer, major depression disorder, T2D, dyslipidemia (Dyslid), hypertensive disease (Hyper), hemorrhoids, hernia abdominopelvic cavity, insomnia, iron deficiency anemias, irritable bowel syndrome, macular degeneration, osteoarthritis, osteoporosis, PVD, peptic ulcer, psychiatric disorder, acute reaction to stress, VV and disease count (DC). The summary statistics for risk factors include HDL-C, low-density lipoprotein cholesterol (LDL-C), total cholesterol (TC) and BMI. Supplementary Tables S3 and S4 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 the exposure–outcome pairs using lipids as the exposure, i.e. HDL-C, LDL-C and TC. Specifically, the screening and exposure datasets were obtained from (48,49), respectively, where the threshold for selecting instrumental variants in the screening dataset is set to  $1 \times 10^{-4}$ . 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 to 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 ( $\hat{\beta} = -0.09$ ), CAD2 ( $\hat{\beta} = -0.08$ ), T2D ( $\hat{\beta} = -0.09$ ), Dyslid ( $\hat{\beta} = -0.14$ ), Hyper ( $\hat{\beta} = -0.05$ ), PVD ( $\hat{\beta} = -0.11$ ) and DC ( $\hat{\beta} = -0.04$ ), which is consistent with known epidemiological associations in the same direction (50–52). Moreover, MR-LDP identified the significant negative causality between HDL-C and PVD, which is consistent with previous studies (53,54). On the other hand, MR-LDP 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



**Figure 4.** Causal associations of HDL-C on CAD1 under different  $P$ -value thresholds in the screening dataset, where UK10K was used as the reference panel and the shrinkage parameter  $\lambda = 0.1$ .

of developing heart diseases. Regarding TC, MR-LDP identified the significant risk effects for cardiovascular disease, as confirmed by RCTs.

To better understand the impact of different thresholds, we repeated the analysis for HDL-C on CAD1, CAD2 and PVD, separately, using a sequence of thresholds as shown in Figure 4 and Supplementary Figures S18–S22. Several patterns can be observed: (i) methods taking into account LD have small standard errors; (ii) by using more SNPs un-

**Table 2.** Causal associations of BMI with common diseases using UK10K as the reference panel with the shrinkage parameter  $\lambda = 0.1$ 

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

MR-LDP uses all SNPs selected by the screening dataset (denoted as  $P_{All}$ ), but IVW, MR-Egger, RAPS and GSMR use pruned SNPs, where the default value of  $r^2$  is used for GSMR (the number of SNP used:  $P_{GSMR}$ ) and  $r^2 = 0.001$  is used for VW, MR-Egger and RAPS (the number of SNP used:  $P_{Ind}$ ). Statistically significant results are indicated in bold.

der larger thresholds, the standard errors become smaller; and (iii) 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 induces 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 screening datasets were obtained from GIANT (55,56), respectively. We chose the threshold to be  $1 \times 10^{-4}$  for selecting the instrumental variants from the screening dataset. The association results from the analysis are summarized in Table 2. Overall, our MR-LDP detected a relatively more significant causality between BMI and complex diseases in this study. Our findings are consistent with RCTs, which indicate that obesity increases the risk of diseases such as heart disease, T2D and hypertensive disease (57).

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 a significant risk factor for hemorrhoids (58).

In addition, MR-Egger is too conservative in identifying the causal relationship between BMI and common diseases, and the same conclusion can be found in (18). Similar to lipid studies, we repeated the analysis for BMI on hemorrhoids and PVD, respectively, using a sequence of thresholds, as shown in Supplementary Figures S23–S26. The patterns are similar to those in Figure 4 and Supplementary Figures S18–S22.

## 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 estimates the causal effect through an approximated likelihood of two sets of GWAS summary statistics for both the risk factor and disease outcome using an additional variance component to eliminate the impact of horizontal pleiotropy. Thus, the type I error can be well controlled even if horizontal pleiotropy and LD structure exist among instrumental variants. MR-LDP is indeed statistically more powerful than other methods in identifying causal effects as illustrated in our simulations (see the ‘Simulations’ section).

Unlike other MR methods, MR-LDP is particularly suitable to analyze complex traits that have multiple instrumental variants within LD. This is primarily accomplished by jointly modeling the distributions for summary statistics and the causal relationship between the risk factor and disease outcome. These summary statistics distributions rely on the polygenicity in complex traits. Moreover, we model the causality by Equation (4) as the average of ‘local’ causal effect, which can be treated similarly as linear structural model in the Supplementary Data. Similar to RAPS, MR-LDP further eliminates the impact of horizontal pleiotropy using a random component. Consequently, MR-LDP is invariant to the orientation of genetic variants, while the results from MR-Egger depend on this orientation as MR-Egger uses a fixed intercept. We notice that a Gaussian distribution with a mean of zero is generally robust even in the case that the underlying horizontal pleiotropy is sparse. In the framework of EM algorithm, the complete-data likelihood for MR-LDP can be written as Equation (5). To further speed up the computation, we developed a PX-VBEM algorithm by expanding parameters and using VB. We further accelerated MR-LDP by parallel computing implemented in the package *MR.LDP*. To further conduct hypothesis testing for causal effects, we calibrated the EBLO from the PX-VBEM algorithm. In our numerical studies, we observe that only GSMR can handle genetic variants with weak LD but is not applicable to analyze genetic variants with high LD. We further demonstrate that unlike other methods, 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 ap-

ply MR-LDP on GWAS summary statistics, likely discovering more fruitful and meaningful causal relationships in the future.

We used two pairs (CAD–CAD and height–height) of real data to validate the proposed method partially. As the risk factor and the outcome are the same, we can take the 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  $\beta_0$  at the 95% confidence level, for instrument variants chosen under a wide range 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 a wider confidence interval than MR-LD. This wider confidence interval 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  $1 \times 10^{-4}$  in the screening dataset, we identified multiple pairs of significant causal relationships. For example, the well-known protective effect of HDL-C on the PVD (59) was identified by our model. We also identified significant causal relationships between HDL-C and CAD in two parallel experiments. In particular, although HDL-C was found to be associated with CAD in multiple observational studies (60–62), the role of HDL-C in CAD was overturned by later studies (63,64). Recently, Zhao *et al.* (42) showed that the effect of HDL-C in CAD is heterogeneous using different instruments. For BMI, we identified its positive association with knee osteoarthritis and sleep duration, which is consistent with what have been reported in (65,66), respectively. We also confirmed a protective effect of BMI on osteoporosis as suggested previously by (67,68). Moreover, increased BMI is also considered to be one of the contributing factors for PVD in both our study and other related work (69). We further demonstrated the robustness of MR-LDP using a sequence of threshold values to select instrumental variants. As illustrated in Figure 4, the point estimates for  $P$ -value threshold  $\leq 10^{-4}$  are almost stable; hence, we want to suggest  $10^{-4}$  as the starting  $P$ -value cutoff for users to work with.

Even though MR-LDP accounts two important issues, it has several limitations. First, MR-LDP cannot be utilized if there are overlapping samples between SNP exposure and SNP outcome. Nowadays, GWAS consortia usually generate summary statistics using the meta-analysis from many completed studies, which makes samples inevitably overlapped. Moreover, different analyses involving UKB (70) samples are largely overlapped with each other. Second, similar to RAPS and other methods, MR-LDP requires an additional independent SNP exposure dataset to select instrumental variants, otherwise existing selection bias would invalidate the inference. Third, similar to many existing methods including RAPS, MR-LDP relies on the ‘Instrument Strength Independent of Direct Effect’ assumption to account for horizontal pleiotropy. Therefore, methods capable of accounting for correlated pleiotropy are highly anticipated.

## WEB RESOURCES

MR.LDP is available at GitHub (<https://github.com/QingCheng0218/MR.LDP>).

BMI-JAP (screen) dataset: [ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/AkiyamaM\\_28892062\\_GCST004904](ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/AkiyamaM_28892062_GCST004904).

Height and BMI (exposure) datasets: [https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT\\_consortium\\_data\\_files#2018\\_GIANT\\_and\\_UK\\_BioBank\\_Meta\\_Analysis\\_for\\_Public\\_Release](https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files#2018_GIANT_and_UK_BioBank_Meta_Analysis_for_Public_Release)

Lipid (screen) datasets: <http://csg.sph.umich.edu/willer/public/lipids2010/>.

Lipid (exposure) datasets: <http://csg.sph.umich.edu/willer/public/lipids2013/>.

CAD1 dataset: <http://www.cardiogramplusc4d.org/data-downloads/>

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

UK10K datasets: [https://www.uk10k.org/data\\_access.html](https://www.uk10k.org/data_access.html).

## SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

## ACKNOWLEDGEMENTS

The authors would like to thank the two anonymous reviewers whose constructive comments have greatly improved this manuscript. The computational work for this article was partially performed using resources from the National SuperComputing Centre, Singapore (<https://www.nsc.sg>).

## FUNDING

Duke-NUS Medical School [R-913-200-098-263]; Ministry of Education, Singapore [MOE2016-T2-2-029, MOE2018-T2-1-046 and MOE2018-T2-2-006]; National Natural Science Foundation of China [71501089, 11501579 and 71472023]; Hong Kong Research Grant Council [22302815, 12316116 and 12301417].

*Conflict of interest statement.* None declared.

## REFERENCES

- 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.
- 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.
- Davey Smith, G. and Ebrahim, S. (2003) Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.*, **32**, 1–22.
- 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 Epidemiol. Biomarkers Prev.*, **27**, 995–1010.

5. Pickrell, J. (2015) Fulfilling the promise of Mendelian randomization. bioRxiv doi: <https://doi.org/10.1101/018150>, 16 April 2015, pre-print: not peer reviewed.
6. Bowden, R.J. and Turkington, D.A. (1984) In: *Instrumental Variables*. Cambridge University Press, Cambridge.
7. Burgess, S., Butterworth, A. and Thompson, S.G. (2013) Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.*, **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. *Am. J. Hum. Genet.*, **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. *Nat. Genet.*, **47**, 1236–1241
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 Genet.*, **7**, e1002254.
11. Cross-Disorder Group of the Psychiatric Genomics Consortium (2013) Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *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. *J. Am. Stat. Assoc.*, **111**, 132–144.
13. Bowden, J., Fabiola Del Greco, M., Minelli, C., Lawlor, D., Zhao, Q., Sheehan, N., Thompson, J. and Smith, G.D. (2019) Improving the accuracy of two-sample summary data Mendelian randomization: moving beyond the NOME assumption. *Int. J. Epidemiol.*, **48**, 728–742.
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. *Nat. Commun.*, **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. *Nat. Genet.*, **50**, 693–698.
17. 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.
18. Qi, G. and Chatterjee, N. (2019) Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. *Nat. Commun.*, **10**, 1941.
19. Zhao, Q., Wang, J., Hemani, G., Bowden, J. and Small, D.S. (2018) Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. arXiv: <https://arxiv.org/abs/1801.09652>, 02 January 2019, pre-print: not peer reviewed.
20. Bowden, J., Del Greco M.F., Minelli, C., Davey Smith, G., Sheehan, N. and Thompson, J. (2017) A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat. Med.*, **36**, 1783–1802.
21. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
22. Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
23. Turner, S., Armstrong, L.L., Bradford, Y., Carlson, C.S., Crawford, D.C., Crenshaw, A.T., De Andrade, M., Doheny, K.F., Haines, J.L., Hayes, G. *et al.* (2011) Quality control procedures for genome-wide association studies. *Curr. Protoc. Hum. Genet.*, doi:10.1002/0471142905.hg0119s68.
24. Husson, F., Josse, J. and Pages, J. (2010) In: *Principal component methods—hierarchical clustering—partitional clustering: why would we need to choose for visualizing data*. Applied Mathematics Department, pp. 1–17.
25. Berisa, T. and Pickrell, J.K. (2016) Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics*, **32**, 283–285.
26. Rothman, A.J. (2012) Positive definite estimators of large covariance matrices. *Biometrika*, **99**, 733–740.
27. Hormozdiari, F., Kostem, E., Kang, E.Y., Pasaniuc, B. and Eskin, E. (2014) Identifying causal variants at loci with multiple signals of association. *Genetics*, **198**, 497–508.
28. Chen, W., Larrabee, B.R., Ovsyannikova, I.G., Kennedy, R.B., Haralambieva, I.H., Poland, G.A. and Schaid, D.J. (2015) Fine mapping causal variants with an approximate 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. *Ann. Appl. Stat.*, **11**, 1561–1592.
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: <https://arxiv.org/abs/1805.01284>, 03 May 2018, pre-print: not peer reviewed.
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<sup>2</sup> statistic. *Int. J. Epidemiol.*, **45**, 1961–1974.
32. Zhao, Q., Wang, J., Hemani, G., Bowden, J. and Small, D. S. (2018) Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. arXiv: <https://arxiv.org/abs/1801.09652>, 02 January 2019, pre-print: not peer reviewed.
33. Bishop, C.M. (2006) In: *Pattern Recognition and Machine Learning*. Springer, Berlin.
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. *Mach. Learn.*, **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. (2020) CoMM-S2: a collaborative mixed model using summary statistics in transcriptome-wide association studies. *Bioinformatics*, **36**, 2009–2016.
38. Blei, D.M., Kucukelbir, A. and McAuliffe, J.D. (2017) Variational inference: a review for statisticians. *J. Am. Stat. Assoc.*, **112**, 859–877.
39. Wang, B. and Titterton, D. (2005) Inadequacy of interval estimates corresponding to variational Bayesian approximations. In: *Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics*. Society Artificial Intelligence and Statistics, Barbados.
40. Turner, R.E. and Sahani, M. (2011) Two problems with variational expectation maximisation for time-series models. In: Barber, D., Cemgil, A.T. and Chiappa, S. (eds). *Bayesian Time Series Models*. Cambridge University Press, Cambridge, pp. 104–124.
41. Yuan, Z., Zhu, H., Zeng, P., Yang, S., Sun, S., Yang, C., Liu, J. and Zhou, X. (2019) Testing and controlling for horizontal pleiotropy with the probabilistic Mendelian randomization in transcriptome-wide association studies. bioRxiv doi: <https://doi.org/10.1101/691014>, 03 July 2019, pre-print: not peer reviewed.
42. Zhao, Q., Chen, Y., Wang, J. and Small, D.S. (2019) Powerful three-sample genome-wide design and robust statistical inference in summary-data Mendelian randomization. *Int. J. Epidemiol.*, **48**, 1478–1492.
43. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, **4**, 7.
44. Coronary Artery Disease (C4D) Genetics Consortium (2011) A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.*, **43**, 339–344.

45. Schunkert, H., König, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F., Barbalic, M., Gieger, C. *et al.* (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.*, **43**, 333–338.
46. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J. *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature*, **562**, 203–209.
47. Randall, J.C., Winkler, T.W., Kutalik, Z., Berndt, S.I., Jackson, A.U., Monda, K.L., Kilpeläinen, T.O., Esko, T., Mägi, R., Li, S. *et al.* (2013) Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.*, **9**, e1003500.
48. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J. *et al.* (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, **466**, 707–713.
49. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S. *et al.* (2013) Discovery and refinement of loci associated with lipid levels. *Nat. Genet.*, **45**, 1274–1283.
50. Emerging Risk Factors Collaboration, Di Angelantonio, E., Sarwar, N., Perry, P., Kaptoge, S., Ray, K.K., Thompson, A., Wood, A.M., Lewington, S., Sattar, N. *et al.* (2009) Major lipids, apolipoproteins, and risk of vascular disease. *JAMA*, **302**, 1993–2000.
51. Wenhui, Z., Jing, G., Ronald, H., Weiqin, L., Yujie, W., Xiaocheng, W. and Gang, H. (2014) HDL cholesterol and cancer risk among patients with type 2 diabetes. *Diabetes Care*, **37**, 3196–3203.
52. Kunutsor, S.K., Kieneker, L.M., Bakker, S.J.L., James, R.W. and Dullaart, R.P.F. (2017) The inverse association of HDL-cholesterol with future risk of hypertension is not modified by its antioxidant constituent, paraoxonase-1: the PREVEND prospective cohort study. *Atherosclerosis*, **263**, 219–226.
53. Sentí, M., Nogués, X., Pedro-Botet, J., Rubiés-Prat, J. and Vidal-Barraquer, F. (1992) Lipoprotein profile in men with peripheral vascular disease. Role of intermediate density lipoproteins and apoprotein E phenotypes. *Circulation*, **85**, 30–36.
54. Gerald, R., Fowkes, F., Housley, E., Riemersma, R.A., Macintyre, C.C., Cawood, E.H., Prescott, R.J. and Ruckley, C.V. (1992) Smoking, lipids, glucose intolerance, and blood pressure as risk factors for peripheral atherosclerosis compared with ischemic heart disease in the Edinburgh Artery Study. *Am. J. Epidemiol.*, **135**, 331–340.
55. 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.
56. Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K. *et al.* (2017) Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.*, **49**, 1458–1467.
57. Look AHEAD Research Group (2010) Long term effects of a lifestyle intervention on weight and cardiovascular risk factors in individuals with type 2 diabetes: four-year results of the Look AHEAD trial. *Arch. Intern. Med.*, **170**, 1566–1575.
58. Ravindranath, G. and Rahul, B. (2018) Prevalence and risk factors of hemorrhoids: a study in a semi-urban centre. *Int. Surg. J.*, **5**, 496–499.
59. Agrawal, R., Sharma, P., Pal, M., Kochar, A. and Kochar, D. (2006) Magnitude of dyslipidemia and its association with micro and macro vascular complications in type 2 diabetes: a hospital based study from Bikaner (Northwest India). *Diabetes Res. Clin. Pract.*, **73**, 211–214.
60. Gordon, D.J., Probstfield, J.L., Garrison, R.J., Neaton, J.D., Castelli, W.P., Knoke, J.D., Jacobs, D.R. Jr, Bangdiwala, S. and Tyroler, H.A. (1989) High-density lipoprotein cholesterol and cardiovascular disease. Four prospective American studies. *Circulation*, **79**, 8–15.
61. Assmann, G., Schulte, H., von Eckardstein, E. and Huang, Y. (1996) High-density lipoprotein cholesterol as a predictor of coronary heart disease risk. The PROCAM experience and pathophysiological implications for reverse cholesterol transport. *Atherosclerosis*, **124**, S11–S20.
62. Silbernagel, G., Schöttker, B., Appelbaum, S., Scharnagl, H., Kleber, M.E., Grammer, T.B., Ritsch, A., Mons, U., Holleczer, B., Goliasch, G. *et al.* (2013) High-density lipoprotein cholesterol, coronary artery disease, and cardiovascular mortality. *Eur. Heart J.*, **34**, 3563–3571.
63. Voight, B.F., Peloso, G.M., Orho-Melander, M., Frikke-Schmidt, R., Barbalic, M., Jensen, M.K., Hindy, G., Hólm, H., Ding, E.L., Johnson, T. *et al.* (2012) Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *Lancet*, **380**, 572–580.
64. Rohatgi, A., Khera, A., Berry, J.D., Givens, E.G., Ayers, C.R., Wedin, K.E., Neeland, I.J., Yuhanna, I.S., Rader, D.R., de Lemos, J.A. *et al.* (2014) HDL cholesterol efflux capacity and incident cardiovascular events. *New Engl. J. Med.*, **371**, 2383–2393.
65. Manek, N.J., Hart, D., Spector, T.D. and MacGregor, A.J. (2003) The association of body mass index and osteoarthritis of the knee joint: an examination of genetic and environmental influences. *Arthritis Rheum.*, **48**, 1024–1029.
66. Grandner, M.A., Schopfer, E.A., Sands-Lincoln, M., Jackson, N. and Malhotra, A. (2015) Relationship between sleep duration and body mass index depends on age. *Obesity*, **23**, 2491–2498.
67. Asomaning, K., Bertone-Johnson, E.R., Nasca, P.C., Hooven, F. and Pekow, P.S. (2006) The association between body mass index and osteoporosis in patients referred for a bone mineral density examination. *J. Womens Health*, **15**, 1028–1034.
68. Barrera, G., Bunout, D., Gattás, V., de la Maza, M.P., Leiva, L. and Hirsch, S. (2004) A high body mass index protects against femoral neck osteoporosis in healthy elderly subjects. *Nutrition*, **20**, 769–771.
69. Ylitalo, K.R., Sowers, M. and Heeringa, S. (2011) Peripheral vascular disease and peripheral neuropathy in individuals with cardiometabolic clustering and obesity: National Health and Nutrition Examination Survey 2001–2004. *Diabetes Care*, **34**, 1642–1647.
70. Collins, R. (2012) What makes UK Biobank special? *Lancet*, **9822**, 1173–1174.