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

Qing Cheng[1], Yi Yang[1], Xingjie Shi[1,2], 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 MR methods account for both linkage disequilibrium (LD) and ubiquitously existing horizontal pleiotropy among complex traits, which is a phenomenon that a variant has a direct effect on the outcome other than exclusively affects the outcome 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 that leverages GWAS summary statistics in the presence of LD, as well as properly accounts for horizontal Pleiotropy among genetic variants (MR-LDP). MR-LDP utilizes a computationally efficient parameter-expanded variational Bayes expectation-maximization (PX-VBEM) algorithm, calibrating the evidence lower bound (ELBO) for a likelihood ratio test. We further conducted comprehensive

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

simulation studies to demonstrate the advantages of MR-LDP over existing methods in terms of both type-I error control and point estimates. Moreover, we used two real exposure-outcome pairs (CAD-CAD and BMI-BMI; CAD for coronary artery disease and BMI for body mass index) to validate results from MR-LDP in comparison with alternative methods, particularly showing our method is more efficient using all instrumental variants in LD. By further applying MR-LDP to lipid traits and BMI as risk factors on complex diseases, we identified multiple pairs of significant causal relationships, including protective effect of high-density lipoprotein cholesterol (HDL-C) on peripheral vascular disease (PVD), and positive causal effect of body mass index (BMI) on haemorrhoids.

INTRODUCTION

Epidemiological studies tremendously contributed to better understand 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 as a golden standard to evaluate the causality in a exposure-outcome pair, RCTs have certain limitations including impracticality (no intervention may exist), high expense, and unethicality [5]. Fortunately, as germline genetic variants (single nucleotide polymorhpisms, 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 interference from unobserved confounding factors [3]. In the past decade, a large number of genome-wide association studies (GWAS) have been successfully completed in identifying genetic variants associated with complex traits at genome-wide significance level including both health factors and diseases, e.g., lipids, body mass index (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 promoted the wide use of two-sample MR as an efficient and cost-effective way to interrogate the causal relationships among many health risk factors and disease outcomes.

MR is closely related to the instrumental variable (IV) methods, which have a long history 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 GWAS context, instrumental variants are not in LD.

The first one 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 auto-immune 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 relax this assumption, various statistical methods have been proposed for two-sample MR analysis 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 conduct MR analysis in the subsequent step, including Q test [12], Cook's distance [13], Studentized residuals [13], GSMR [14], and MR-PRESSO [15]. The drawback of this type of methods is that as abundant pleiotropy exists among complex traits, the number of SNPs after removal is limited, 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 [16] and RAPS [17]. Comparing 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 [18].

On the other hand, there is infertile literature regarding the relaxation of the second assumption above. Among the methods mentioned above, only GSMR is capable of accounting SNPs with weak or moderate linkage disequilibrium (LD) while others demand all instrumental SNPs be independent, which is usually done by conducting SNP pruning and thus reduces the number of instrumental variants for the follow-up MR analysis. As SNPs within proximity tend to be highly correlated, MR methods not accounting 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 horizontal Pleiotropy (MR-LDP). Similar to RAPS, MR-LDP does not require 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 conduct a formal hypothesis testing to make inference for 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. The 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 in comparison with alternative methods, particularly showing our methods are more efficient using all SNPs in LD. By further applying MR-LDP to summary statistics from GWAS, we identified multiple pairs of significant causal relationships, including protective effect of high-density lipoprotein cholesterol (HDL-C) on peripheral vascular disease (PVD), and positive causal effect of body mass index (BMI) on haemorrhoids.

MATERIALS AND METHODS

GWAS summary statistics

In this article, we used lipids including high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C) and total cholesterol (TC), and body mass index (BMI) as health risk factors. The disease traits include coronary artery disease (CAD), asthma, allergic rhinitis (AR), cancer, major depression disorder (MDD), type 2 diabetes (T2D), dyslipidemia, hypertensive disease (Hyper), hemorrhoids, hernia abdominopelvic cavity, insomnia, iron deficiency anemias (IDA), irritable bowel syndrome (IBS), macular degeneration (MD), osteoarthritis (Osteoa), osteoporosis (Osteop), peripheral vascular disease (PVD), peptic ulcer(PU), psychiatric disorder, acute reaction to stress(Stress), varicose veins(VV), and disease count(SumDis) that is the number of diseases affecting each individual. Table S6 summarizes the total number of SNPs and sample sizes for each trait in each health risk factor or disease outcome. Details for the sources of these GWAS summary statistics can be found in Table S7 from the supplementary document.

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, as we primarily focus on European populations, we choose to use samples from two resources as the external reference panel: 1000 Genome Project Phase 1 ($N = 379$), denoted as 1GP1, and UK10K Project (Avon Longitudinal Study of Parents and Children (ALSPAC) and TwinsUK) merged with 1000 Genome Project Phase 3 ($N = 4,284$), denoted abusively as 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 [19]. First, SNPs were excluded from the analysis if their calling rates were less than 95%, minor allele frequencies less than 0.01, or p-values less than 1×10^{-6} in the Hardy-Weinberg equilibrium test. We then removed the individuals with genotype missing rates more than 5%. To further remove individuals with high relatedness in all samples, we used GCTA [20] to first identify those individual pairs with estimated genetic relatedness more than 0.05 and then randomly remove one from such a pair. After QC, there were 3,764 individuals remaining with 989,932 SNPs.

Choice of LD matrix

Since the LD between two SNPs decays exponentially with respect to their distance, we use LDetect [21] 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 [22]. In particular, the correlation matrix estimator $\widehat{\mathbf{R}}^{(l)}$ in each block is obtained by optimizing

$$\widehat{\mathbf{R}}^{(l)} = \underset{\mathbf{R}^{(l)} \succ 0}{\arg\min}(\|\mathbf{R}^{(l)} - \widehat{\mathbf{R}}^{(l)}_{\text{emp}}\|_F^2/2 - \tau \log|\mathbf{R}^{(l)}| + \lambda\|\mathbf{R}^{(l)-}\|_1), \tag{1}$$

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, the lasso-type penalty ensures a sparse solution. $\tau > 0$ is fixed at a small value and the logarithmic barrier term is used to enforce a positive-definite solution, more detailed can be found in [22]. A corresponding R package named *PDSCE* is available to complete the estimate process. In addition, we fix shrinkage parameter λ to be 0.055 in simulation studies and vary $\lambda \in \{0.1, 0.15\}$ in real data analysis.

Methods for comparison

We compared the performance of six methods in the main text: (1) our MR-LD and MR-LDP implemented in the R package *MR.LDP*; (2) the least squares implemented in the *lm* function in R; (3) GSMR implemented in the R package *gsmr*; (4) RAPS implemented in the R package *mr.raps*; (5) IVW implemented in the R package *MendelianRandomization*; (6) 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/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 by removing invalid instruments first, 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.

Likelihood for summary statistics

Before elaborating 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. Denote $\widehat{\mathbf{R}}$ the correlation among all genotyped SNPs and $\widehat{\mathbf{S}} = \mathrm{diag}([\widehat{\mathbf{s}}_1, \cdots, \widehat{\mathbf{s}}_p])$

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 with [23, 24, 25, 26]:

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

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:

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

where $\widehat{\mathbf{S}}_\gamma = \mathrm{diag}([\widehat{\mathbf{s}}_{\gamma_1}, \cdots, \widehat{\mathbf{s}}_{\gamma_p}])$ and $\widehat{\mathbf{S}}_\Gamma = \mathrm{diag}([\widehat{\mathbf{s}}_{\Gamma_1}, \cdots, \widehat{\mathbf{s}}_{\Gamma_p}])$ 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 [25] showed that $\widehat{\mathbf{R}}$ can be replaced with $\widehat{\mathbf{R}}^{\mathrm{ref}}$ 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 sample size in 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 $\widehat{\mathbf{R}}^{\mathrm{ref}}$ and details on estimating $\widehat{\mathbf{R}}$ can be found in Section Choice of LD matrix.

Details of MR-LDP

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

$$\Gamma_j = \beta_0\gamma_j, \quad \text{or} \quad \Gamma_j = \alpha_j + \beta_0\gamma_j, \tag{4}$$

where $j = 1, \ldots, p$, considering without/with horizontal pleiotropy, respectively [27]. 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 direct effects of genetic variants on the outcome. Importantly, β_0 can be interpreted as the causal effect between exposure and outcome in the study [27]. More details are available in the supplementary document. As MR-LD can be taken as a special case of MR-LDP by taking

all $\boldsymbol{\alpha}$ as zeros, we focus on deriving MR-LDP in the main text, resorting the supplementary document for details in 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:

$$\Pr(\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}, \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}, \widehat{\mathbf{R}}; \boldsymbol{\theta}) = \Pr(\widehat{\boldsymbol{\Gamma}} | \boldsymbol{\gamma}, \boldsymbol{\alpha}, \widehat{\mathbf{R}}, \widehat{\mathbf{S}}_{\Gamma}; \beta_0) \Pr(\widehat{\boldsymbol{\gamma}} | \boldsymbol{\gamma}, \widehat{\mathbf{R}}, \widehat{\mathbf{S}}_{\gamma}) \Pr(\boldsymbol{\alpha} | \sigma^2_{\boldsymbol{\alpha}}) \Pr(\boldsymbol{\gamma} | \sigma^2_{\boldsymbol{\gamma}}), \tag{5}$$

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

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

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 [28]. But it may cause instability or numerical failure as $\widehat{\mathbf{R}}$ can be non-positive definite due to relative small sample size in the reference panel. To address these issues, we develop an accelerated variational Bayes (VB) EM algorithm in light of [29], namely, PX-VBEM. To start with the algorithm, we expand the original MR-LD/MR-LDP model (5) as follows:

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

Next, we sketch the VBEM algorithm with parameter expanded in Equation (6) for MR-LDP while 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 a 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,

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

where

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

where $\mathcal{L}(q)$ is the ELBO of the marginal log-likelihood, and $\mathbb{KL}(q \| p)$ is the KL divergence between two distributions and we have $\mathbb{KL}(q \| p) \geq 0$ 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. Comparing with the standard EM algorithm, the crux of VBEM is to optimize q within a factorizable family of distributions by mean-field assumption [30], which assumes that $q(\boldsymbol{\gamma}, \boldsymbol{\alpha})$ can be factorized as

$$q(\boldsymbol{\gamma}, \boldsymbol{\alpha}) = \prod_{j=1}^{p} q_{\gamma_j}(\gamma_j) \prod_{k=1}^{p} q_{\alpha_k}(\alpha_k). \tag{9}$$

This only assumption in variational inference promotes the computational efficiency and scalability in large-scale computational problems as a coordinate descent algorithm is commonly used to find the optimal distribution q^*. To briefly show this, we first note that this factorization (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, we can show that the terms with α_k, given all other latent variables fixed, 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 (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 as follows:

$$\begin{aligned}
\beta_0 &= \left\{\boldsymbol{\mu}_{\boldsymbol{\gamma}}^T \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} \widehat{\mathbf{R}} \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\gamma}} + Tr(\widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} \widehat{\mathbf{R}} \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})\right\}^{-1} \left(\widehat{\boldsymbol{\Gamma}}^T \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-2} \boldsymbol{\mu}_{\boldsymbol{\gamma}} - \boldsymbol{\mu}_{\boldsymbol{\alpha}}^T \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} \widehat{\mathbf{R}} \widehat{\mathbf{S}}_{\boldsymbol{\Gamma}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\gamma}}\right), \\
\sigma_\gamma^2 &= \left\{\boldsymbol{\mu}_{\boldsymbol{\gamma}}^T \boldsymbol{\mu}_{\boldsymbol{\gamma}} + Tr(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}})\right\}/p, \\
\sigma_\alpha^2 &= \left\{\boldsymbol{\mu}_{\boldsymbol{\alpha}}^T \boldsymbol{\mu}_{\boldsymbol{\alpha}} + Tr(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}})\right\}/p, \\
\xi &= \left\{\boldsymbol{\mu}_{\boldsymbol{\gamma}}^T \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}^{-1} \widehat{\mathbf{R}} \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\gamma}} + Tr(\widehat{\mathbf{S}}_{\boldsymbol{\gamma}}^{-1} \widehat{\mathbf{R}} \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})\right\}^{-1} \left(\widehat{\boldsymbol{\gamma}}^T \widehat{\mathbf{S}}_{\boldsymbol{\gamma}}^{-2} \boldsymbol{\mu}_{\boldsymbol{\gamma}}\right).
\end{aligned}$$

where $\boldsymbol{\mu}_{\boldsymbol{\gamma}} = (\mu_{\gamma_1}, \cdots, \mu_{\gamma_p})^T, \boldsymbol{\mu}_{\boldsymbol{\alpha}} = (\mu_{\alpha_1}, \cdots, \mu_{\alpha_p})^T$, $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \text{diag}([\sigma^2_{\gamma_1}, \cdots, \sigma^2_{\gamma_p}])$, and $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \text{diag}([\sigma^2_{\alpha_1}, \cdots, \sigma^2_{\alpha_p}])$, 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 (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 as has previously been done in [31, 32]; details are given in supplementary document. As VB searches within a factorizable family for 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 [33]. Despite its computational efficiency and accuracy for estimating posterior mean, VB suffers from under-estimating the variance of target distribution [21, 34, 35]. 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. [32] 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.

Simulations

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 from $\mathcal{N}(\mathbf{0}, \sigma^2_{\mathbf{e_1}}\mathbf{I_{n_1}})$ and $\mathcal{N}(\mathbf{0}, \sigma^2_{\mathbf{e_2}}\mathbf{I_{n_2}})$, respectively. In this generative model, β_0 was the true causal effect while $\boldsymbol{\alpha}$ exhibited the direct effects on the disease, where we considered two cases: dense and sparse horizontal pleiotropy. For dense case, we assumed that α_ks were independent and identically distributed as $\mathcal{N}(0, \sigma^2_{\boldsymbol{\alpha}})$ while for sparse case, we assumed that only a fraction of α_ks was from a Gaussian distribution and rest were zero. In simulations, we consider 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 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 auto-regressive(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 block 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 $(\boldsymbol{\eta}_x, \boldsymbol{\eta}_y)$ were generated from a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$, $\boldsymbol{\Sigma}_\eta$ is a two-by-two matrix with diagonal elements as 1 and off-diagonal elements 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, $h^2_{\boldsymbol{\gamma}} = \frac{\mathrm{var}(\beta_0 \mathbf{G}_1 \boldsymbol{\gamma})}{\mathrm{var}(\mathbf{y})}$ and $h^2_{\boldsymbol{\alpha}} = \frac{\mathrm{var}(\mathbf{G}_2 \boldsymbol{\alpha})}{\mathrm{var}(\mathbf{y})}$, 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 $h^2_{\boldsymbol{\alpha}} \in \{0, 0.05, 0.1\}$.

Software

We implemented MR-LD and MR-LDP as two functions in R package *MR-LDP* with documented functions and toy examples described in the vignette. All functions were implemented in C++ integrating *Rcpp* and *RcppArmadillo* libraries [36]. These libraries simplify the use of *Armadillo* library for C++ language to integrate with R, and enable us to access the high quality linear algebra library for matrix operations within R. *MR-LDP* is available at Github ().

Results

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. As there exist strong LD structure among SNPs and abundant horizontal pleiotropy in GWAS, these unique features invalidate two assumptions easily, i.e., independence among genetic instruments and conditional independence given risk factors and confounders. Our proposed MR-LDP aims to make causal inference of the risk factors on a disease outcome using a probabilistic model by accounting 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 Equation (3). Given p instrumental variants, the inputs to 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 to generate GWAS summary statistics are large [37, 18]. 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-outocme, 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 middle that is largely due to LD, and methods that do not account LD tend to inflate findings.

Simulations: type-I error control and point estimates

We conducted varied simulation studies to make comparisons of MR-LD and MR-LDP with other four commonly used alternative methods:(1) a linear regression; (2) IVW implemented in R packages *MendelianRandomization*; (3) MR-Egger implemented in R package *Mendelian-Randomization*; (4) GSMR implemented in R packages *gsmr*; (5) RAPS implemented in R package *mr.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, 3, S2 - S8, respectively with $n_3 = 500, 2,500, 4,000$. 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 pretty well when genetic variants were independent ($\rho = 0$). But as LD become stronger

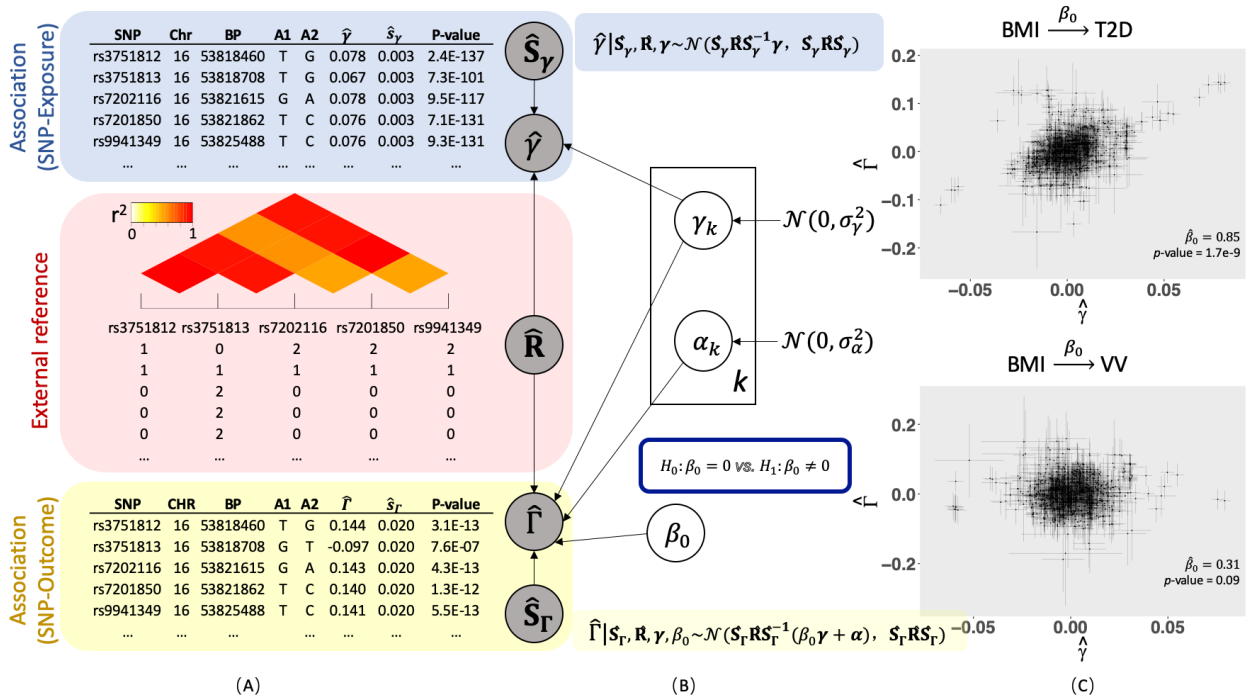
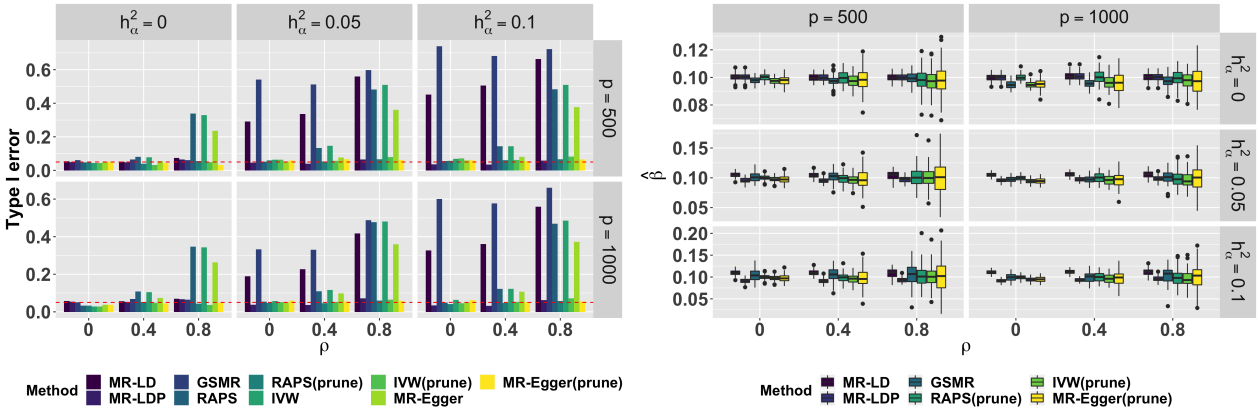


Figure 1: MR-LDP model overview: (A) Inputs to MR-LDP are GWAS summary statistics from both the risk factor (Blue) and the disease outcome (Yellow), and an external reference panel data (Red). (B) The probabilistic graphical model representation of MR-LDP. The box is the "plate" representing SNPs, $k = 1, \ldots, p$. 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 interest 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.

($\rho = 0.4$ 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 controllable in the case of moderate or strong LD. But 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 3 and S4 - S8, where the settings was not in favor of MR-LDP. Note that after SNP pruning, genetic variants 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 not accounting horizontal pleiotropy were biased. The similar patterns could be observed for dense and sparse pleiotropy both at sparsity equaling 0.2, 0.4.



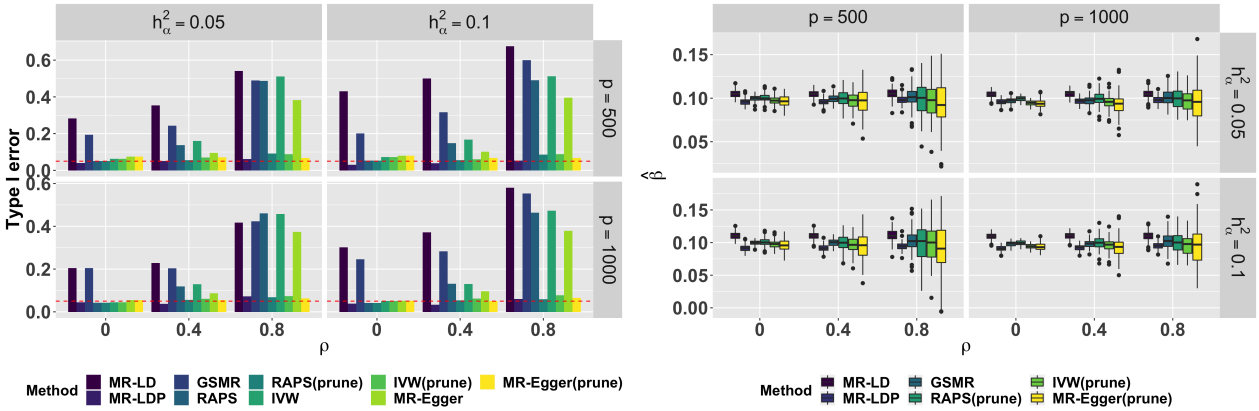
(A) Type-I error

(B) Point estimates for β_0

Figure 2: Simulation of type-I error control and point estimates under the dense horizontal pleiotropy. $n_1 = n_2 = 20,000, n_3 = 500$, and the number of replication is 1000.

CAD-CAD and BMI-BMI studies

In addition, we used real datasets, i.e., CAD-CAD and BMI-BMI 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. 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 [38]. The first two datasets are non-overlapping GWAS for the same or similar traits. The exposure dataset and outcome dataset are non-overlapping from European ancestry. Since IVW, MR-Egger, 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 [39]. Individual-level genotype data from UK10K



(A) Type-I error

(B) Point estimates for β_0

Figure 3: Simulation of type-I error control and point estimates under the sparse(0.2) horizontal pleiotropy. $n_1 = n_2 = 20,000, n_3 = 500$, and the number of replication is 1000.

projects was served as reference panel in this study.

For CAD-CAD analysis, the selection dataset is myocardial infarction (MI) in UK Biobank (UKB), the exposure data is from C4D Genetics Consortium [40], and the outcome data is from the transatlantic Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) [41]. We first selected instrumental variants using MI from UKB under different 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 . We reported the point estimates with its 95% corresponding confidence interval for all methods in Figure 4A. 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 threshold was small.

Next, we investigated the case that both the exposure and outcome were from waist-to-hip ratio adjusted Body Mass Index(WHRadjBMI). In particular, we treated the WHRadjBMI for female in a European population-based study [42] as the screening dataset. The exposure data is from WHRadjBMI for male in a European population-based study and the outcome data is from WHRadjBMI for female in EUR population [43]. The results of point estimates with their 95% confidence intervals were illustrated in Figure 4B. As one can see, similar patterns were observed in all cases, in particular, RAPS only offered a better performance with larger

instrumental SNPs 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 $\beta_0 = 1$.

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, hypertensive disease (Hyper), hemorrhoids, hernia abdominopelvic cavity, 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(SumDis). 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).

First, we applied MR-LDP together with other alternative methods to analyze to the exposure-outcome pairs by taking lipids as the exposure, i.e., high-density lipoprotein cholesterol(HDL-C), low-density lipoprotein cholesterol(LDL-C), and total cholesterol(TC). Specifically, the selection and exposure datasets were from [44] and [45], respectively. 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 [14].

In practice, HDL-C and LDL-C are often referred as "good" and "bad" cholesterol. 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 ($\widehat{\beta} = -0.08$), Type 2 diabetes ($\widehat{\beta} = -0.09$), Dyslipidemia ($\widehat{\beta} = -0.14$), Hypertensive disease ($\widehat{\beta} = -0.05$) , PVD ($\widehat{\beta} = -0.11$) and disease count ($\widehat{\beta} = -0.04$), which is consistent with known epidemiological associations in the same direction [46, 47, 48]. Particularly, even though HDL-C was found to be associated with CAD in multiple observational studies [49, 50, 51], the role of HDL-C in CAD was overturned by later studies [52, 53]. Recently, Zhao et al. [38] 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 of previous studies [54, 55]. On the other hand, MR-LDA identified the significant positive

causality between LDL-C and CAD which is consistent of the fact that LDL-C narrows the arteries and increases the chances of getting heart diseases. As for TC, MR-LDP identified the significant risk effects for cardiovascular disease as confirmed by RCTs .

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	-0.27(0.037)	203	-0.38(0.07)	-0.36(0.07)	-0.28(0.157)
	CAD2	2071	-0.08(0.02)	277	-0.07(0.029)	206	-0.15(0.047)	-0.15(0.047)	-0.08(0.098)
	T2D	2071	-0.09(0.031)	272	-0.15(0.043)	206	-0.33(0.081)	-0.35(0.082)	0.03(0.17)
	Dyslipidemia	2071	-0.14(0.023)	255	-0.09(0.029)	206	-0.23(0.08)	-0.26(0.076)	-0.17(0.158)
	Hyper	2071	-0.05(0.017)	270	-0.15(0.022)	206	-0.2(0.037)	-0.21(0.038)	-0.09(0.079)
	PVD	2071	-0.11(0.048)	277	-0.14(0.075)	206	-0.19(0.109)	-0.19(0.105)	0.12(0.222)
	DiseSum	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	0.41(0.035)	193	0.34(0.065)	0.32(0.062)	0.33(0.133)
	CAD2	1820	0.11(0.021)	266	0.15(0.026)	199	0.15(0.043)	0.14(0.043)	0.26(0.085)
	Dyslipidemia	1820	0.56(0.03)	258	0.94(0.026)	199	0.9(0.053)	0.86(0.051)	0.93(0.1)
	DiseSum	1820	0.08(0.01)	267	0.13(0.011)	199	0.13(0.019)	0.13(0.019)	0.17(0.037)
TC	CAD1	2546	0.24(0.028)	309	0.46(0.035)	215	0.41(0.061)	0.39(0.062)	0.35(0.146)
	CAD2	2484	0.08(0.02)	314	0.16(0.027)	218	0.15(0.043)	0.14(0.043)	0.22(0.094)
	Dyslipidemia	2484	0.54(0.03)	303	1.08(0.028)	218	0.93(0.055)	0.9(0.051)	0.97(0.111)
	DiseSum	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: The causal associations of lipids on common diseases using UK10K as the reference penal with shrinkage parameter $\lambda = 0.1$. MR-LDP use 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 shown in blue color.

Secondly, we examine the associations between BMI and common diseases where the exposure and the selection dataset were from GIANT [56] and [57], respectively. 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. Particularly, it is well-known that the extent of obesity increase the risk of certain diseases such as heart disease, type 2 diabetes and hypertensive disease identified by RCT [58].

We also estimated some causal effects which are rarely involved in the previous MR analysis but reported in the epidemiologic studies. Such as, Body mass is an important risk factor for hemorrhoids [59]. Highest BMI has a relationship with knee Osteoarthritis and the shortest sleepers, reported by [60] and [61], respectively. We also confirmed the protect effect of BMI on osteoporosis reported by [62] and [63]. Moreover, the increased body mass index is also considered to be one of the contributing factor for peripheral vascular disease [64].

In addition, MR-Egger is too conservative to identify the causal relationship between BMI and common diseases for wide confidence interval, same conclusion can be found in [65].

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	0.34(0.07)	563	0.2(0.121)	0.17(0.091)	0.2(0.129)
Asthma	4428	0.28(0.073)	707	0.23(0.059)	563	0.24(0.107)	0.19(0.08)	0.18(0.115)
CAD2	4428	0.23(0.066)	708	0.2(0.06)	563	0.26(0.105)	0.2(0.079)	0.22(0.113)
T2D	4428	0.85(0.141)	708	0.85(0.088)	563	1.22(0.16)	0.93(0.124)	1.46(0.175)
Dyslipidemia	4428	0.22(0.076)	704	0.29(0.057)	563	0.18(0.133)	0.16(0.086)	0.29(0.124)
Hemorrhoids	4428	0.3(0.135)	709	0.2(0.108)	563	0.15(0.17)	0.11(0.129)	-0.1(0.184)
Hyper	4428	0.47(0.066)	703	0.51(0.045)	563	0.58(0.095)	0.46(0.067)	0.54(0.097)
Insomnia	4428	0.77(0.235)	708	0.83(0.209)	563	1.24(0.325)	0.96(0.246)	0.6(0.353)
Osteoa	4428	0.27(0.078)	709	0.27(0.066)	563	0.26(0.114)	0.2(0.084)	0.39(0.119)
Osteop	4428	-0.44(0.178)	709	-0.38(0.146)	563	-0.62(0.238)	-0.48(0.178)	-0.73(0.254)
PVD	4428	0.35(0.167)	709	0.43(0.154)	563	0.32(0.242)	0.24(0.183)	0.41(0.263)
DisSum	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: The causal associations of BMI on common diseases using UK10K as the reference penal 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 shown in blue color.

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 have implemented our method in an R package *MR-LDP* available to download at Github (). MR-LDP jointly estimated the causal effect through an approximated likelihood of GWAS summary statistics from both the risk factor and the 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 to identify 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 causal relationship between the risk factor and the 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 (4) as the average of 'local' causal effect [27]. The linear model (4) holds in very general situations, beyond the linear structural model presented in the supplementary document; see Appendix A in [27] for details. To further eliminate the impact of horizontal pleiotropy, we used a random effect to control the variation in disease outcome due to it. As horizontal pleiotropy is not an estimate of interest, a Gaussian distribution with mean zero and a variance parameter is generally robust enough even the underlying horizontal pleiotropy is sparse. Therefore, complete-data likelihood for MR-LDP

can be written as Equation (5). As the iteration for standard EM algorithm involves the inverse of $\widehat{\mathbf{R}}$ which may cause numerical failure, we developed a PX-VBEM algorithm by expanding parameters. Our previous works have shown that parameter expansion step is crucial in speeding up, where we resort to the supplementary document in [31] for detailed interpretations. To further conduct hypothesis testing for causal effects, we calibrated the EBLO from the PX-VBEM algorithm. In our numerical studies, we have shown 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 with likely yielding more fruitful and meaningful causal discovery in the future.

We used two pairs (CAD-CAD and BMI-BMI) 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, MR-LDP with alternative methods, we found that estimates from the proposed methods can effectively cover the true β_0 with 95% confidence interval with instrument variants chosen under a wide ranging of threshold. When more instrumental variants come into model under 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. We identified multiple pairs of significant causal relationships, including protective effect of high-density lipoprotein cholesterol (HDL-C) on peripheral vascular disease (PVD), and positive causal effect of body mass index (BMI) on haemorrhoids.

SOFTWARE AVAILABILITY

Our software MR-LDP is publicly available on GitHub repository: .....

SUPPLEMENTARY DATA

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 National Natural Science Foundation of China, and grant No. 22302815, No. 12316116 and No. 12301417 from the Hong Kong Research Grant Council. The computational work for this article was partially performed on resources of the National Supercomputing Centre, Singapore (https://www.nscc.sg).

References

[1] J Michael Gaziano, Robert J Glynn, William G Christen, Tobias Kurth, Charlene Belanger, Jean MacFadyen, Vadim Bubes, JoAnn E Manson, Howard D Sesso, and Julie E Buring. Vitamins e and c in the prevention of prostate and total cancer in men: the physicians' health study ii randomized controlled trial. *Jama*, 301(1):52–62, 2009.

[2] Eric A Klein, Ian M Thompson, Catherine M Tangen, John J Crowley, M Scott Lucia, Phyllis J Goodman, Lori M Minasian, Leslie G Ford, Howard L Parnes, J Michael Gaziano, et al. Vitamin e and the risk of prostate cancer: the selenium and vitamin e cancer prevention trial (select). *Jama*, 306(14):1549–1556, 2011.

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

[4] James Yarmolinsky, Kaitlin H Wade, Rebecca C Richmond, Ryan J Langdon, Caroline J Bull, Kate M Tilling, Caroline L Relton, Sarah J Lewis, George Davey Smith, and

Richard M Martin. Causal inference in cancer epidemiology: what is the role of mendelian randomization? *Cancer Epidemiology and Prevention Biomarkers*, 27(9):995–1010, 2018.

[5] Joseph Pickrell. Fulfilling the promise of mendelian randomization. *bioRxiv*, page 018150, 2015.

[6] Roger J. Bowden and D. A. Turkington. *Instrumental variables*. Cambridge university press, 1984.

[7] Stephen Burgess, Adam Butterworth, and Simon G Thompson. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic epidemiology*, 37(7):658–665, 2013.

[8] Shanya Sivakumaran, Felix Agakov, Evropi Theodoratou, James G Prendergast, Lina Zgaga, Teri Manolio, Igor Rudan, Paul McKeigue, James F Wilson, and Harry Campbell. Abundant pleiotropy in human complex diseases and traits. *The American Journal of Human Genetics*, 89(5):607–618, 2011.

[9] Brendan Bulik-Sullivan, Hilary K Finucane, Verneri Anttila, Alexander Gusev, Felix R Day, Po-Ru Loh, Laramie Duncan, John RB Perry, Nick Patterson, Elise B Robinson, et al. An atlas of genetic correlations across human diseases and traits. *Nature genetics*, 47(11):1236, 2015.

[10] Chris Cotsapas, Benjamin F Voight, Elizabeth Rossin, Kasper Lage, Benjamin M Neale, Chris Wallace, Gonçalo R Abecasis, Jeffrey C Barrett, Timothy Behrens, Judy Cho, et al. Pervasive sharing of genetic effects in autoimmune disease. *PLoS genetics*, 7(8):e1002254, 2011.

[11] Cross-Disorder Group of the Psychiatric Genomics Consortium et al. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet*, 381(9875):1371–1379, 2013.

[12] Jack Bowden, M Fabiola Del Greco, Cosetta Minelli, Debbie Lawlor, Qingyuan Zhao, Nuala Sheehan, John Thompson, and George Davey Smith. Improving the accuracy of two-sample summary data mendelian randomization: moving beyond the nome assumption. *BioRxiv*, page 159442, 2018.

[13] Laura J Corbin, Rebecca C Richmond, Kaitlin H Wade, Stephen Burgess, Jack Bowden, George Davey Smith, and Nicholas J Timpson. Bmi as a modifiable risk factor for type 2 diabetes: refining and understanding causal estimates using mendelian randomization. *Diabetes*, 65(10):3002–3007, 2016.

[14] Zhihong Zhu, Zhili Zheng, Futao Zhang, Yang Wu, Maciej Trzaskowski, Robert Maier, Matthew R Robinson, John J McGrath, Peter M Visscher, Naomi R Wray, et al. Causal associations between risk factors and common diseases inferred from gwas summary data. *Nature communications*, 9(1):224, 2018.

[15] Marie Verbanck, Chia-Yen Chen, Benjamin Neale, and Ron Do. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nature genetics*, 50(5):693, 2018.

[16] Jack Bowden, George Davey Smith, and Stephen Burgess. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *International journal of epidemiology*, 44(2):512–525, 2015.

[17] Qingyuan Zhao, Jingshu Wang, Gibran Hemani, Jack Bowden, and Dylan S Small. Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score. *arXiv preprint arXiv:1801.09652*, 2018.

[18] Jack Bowden, Fabiola Del Greco M, Cosetta Minelli, George Davey Smith, Nuala Sheehan, and John Thompson. A framework for the investigation of pleiotropy in two-sample summary data mendelian randomization. *Statistics in medicine*, 36(11):1783–1802, 2017.

[19] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel AR Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul IW De Bakker, Mark J Daly, et al. Plink: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3):559–575, 2007.

[20] Jian Yang, S Hong Lee, Michael E Goddard, and Peter M Visscher. Gcta: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1):76–82, 2011.

[21] Tomaz Berisa and Joseph K Pickrell. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics*, 32(2):283, 2016.

[22] Adam J Rothman. Positive definite estimators of large covariance matrices. *Biometrika*, 99(3):733–740, 2012.

[23] Farhad Hormozdiari, Emrah Kostem, Eun Yong Kang, Bogdan Pasaniuc, and Eleazar Eskin. Identifying causal variants at loci with multiple signals of association. *Genetics*, 198(2):497–508, 2014.

[24] Wenan Chen, Beth R Larrabee, Inna G Ovsyannikova, Richard B Kennedy, Iana H Haralambieva, Gregory A Poland, and Daniel J Schaid. Fine mapping causal variants with an approximate bayesian method using marginal test statistics. *Genetics*, 200(3):719–736, 2015.

[25] Xiang Zhu and Matthew Stephens. Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. *The annals of applied statistics*, 11(3):1561, 2017.

[26] Jian Huang, Yuling Jiao, Jin Liu, and Can Yang. Remi: Regression with marginal information and its application in genome-wide association studies. *arXiv preprint arXiv:1805.01284*, 2018.

[27] Qingyuan Zhao, Jingshu Wang, Jack Bowden, and Dylan S Small. Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score. *arXiv preprint arXiv:1801.09652*, 2018.

[28] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[29] Chuanhai Liu, Donald B Rubin, and Ying Nian Wu. Parameter expansion to accelerate em: the px-em algorithm. *Biometrika*, 85(4):755–770, 1998.

[30] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.

[31] Can Yang, Xiang Wan, Xinyi Lin, Mengjie Chen, Xiang Zhou, and Jin Liu. Comm: a collaborative mixed model to dissecting genetic contributions to complex traits by leveraging regulatory information. *Bioinformatics*, 35(10):1644–1652, 2019.

[32] Yi Yang, Xingjie Shi, Yuling Jiao, Jian Huang, Min Chen, Xiang Zhou, Lei Sun, Xinyi Lin, Can Yang, and Jin Liu. Comm-s2: a collaborative mixed model using summary statistics in transcriptome-wide association studies. *bioRxiv*, 2019.

[33] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

[34] Bo Wang and DM Titterington. Inadequacy of interval estimates corresponding to variational bayesian approximations. In *AISTATS*. Barbados, 2005.

[35] Richard E. Turner and Maneesh Sahani. Two problems with variational expectation maximisation for time-series models. 2011.

[36] Dirk Eddelbuettel, Romain François, J Allaire, Kevin Ushey, Qiang Kou, N Russel, John Chambers, and D Bates. Rcpp: Seamless r and c++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011.

[37] Jack Bowden, Fabiola Del Greco M, Cosetta Minelli, George Davey Smith, Nuala A Sheehan, and John R Thompson. 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(6):1961–1974, 2016.

[38] Qingyuan Zhao, Yang Chen, Jingshu Wang, and Dylan S Small. Powerful genome-wide design and robust statistical inference in two-sample summary-data mendelian randomization. *arXiv preprint arXiv:1804.07371*, 2018.

[39] Christopher C Chang, Carson C Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M Purcell, and James J Lee. Second-generation plink: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1):7, 2015.

[40] Coronary Artery Disease (C4D) Genetics Consortium et al. A genome-wide association study in europeans and south asians identifies five new loci for coronary artery disease. *Nature genetics*, 43(4):339, 2011.

[41] Heribert Schunkert, Inke R König, Sekar Kathiresan, Muredach P Reilly, Themistocles L Assimes, Hilma Holm, Michael Preuss, Alexandre FR Stewart, Maja Barbalic, Christian

Gieger, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature genetics*, 43(4):333, 2011.

[42] Dmitry Shungin, Thomas W Winkler, Damien C Croteau-Chonka, Teresa Ferreira, Adam E Locke, Reedik Mägi, Rona J Strawbridge, Tune H Pers, Krista Fischer, Anne E Justice, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, 518(7538):187, 2015.

[43] Loic Yengo, Julia Sidorenko, Kathryn E Kemper, Zhili Zheng, Andrew R Wood, Michael N Weedon, Timothy M Frayling, Joel Hirschhorn, Jian Yang, Peter M Visscher, et al. Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of european ancestry. *Human molecular genetics*, 27(20):3641–3649, 2018.

[44] Tanya M Teslovich, Kiran Musunuru, Albert V Smith, Andrew C Edmondson, Ioannis M Stylianou, Masahiro Koseki, James P Pirruccello, Samuli Ripatti, Daniel I Chasman, Cristen J Willer, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466(7307):707, 2010.

[45] Cristen J Willer, Ellen M Schmidt, Sebanti Sengupta, Gina M Peloso, Stefan Gustafsson, Stavroula Kanoni, Andrea Ganna, Jin Chen, Martin L Buchkovich, Samia Mora, et al. Discovery and refinement of loci associated with lipid levels. *Nature genetics*, 45(11):1274, 2013.

[46] RELIABLE ASSESSMENT. Major lipids, apolipoproteins, and risk of vascular disease. *Jama*, 302(18):1993–2000, 2009.

[47] Zhao Wenhui, Guan Jing, Horswell Ronald, Li Weiqin, Wang Yujie, Wu Xiaocheng, and Hu Gang. Hdl cholesterol and cancer risk among patients with type 2 diabetes. *Diabetes Care*, 37(12):3196–3203, 2014.

[48] Setor K Kunutsor, Lyanne M Kieneker, Stephan J L Bakker, R. W. James, and Robin P F Dullaart. 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, 2017.

[49] David J Gordon, Jeffrey L Probstfield, Robert J Garrison, James D Neaton, William P Castelli, James D Knoke, David R Jacobs Jr, Shrikant Bangdiwala, and H Alfred Tyroler.

High-density lipoprotein cholesterol and cardiovascular disease. four prospective american studies. *Circulation*, 79(1):8–15, 1989.

[50] G Assmann, H Schulte, Eckardstein A von, and Y Huang. The procam experience and pathophysiological implications for reverse cholesterol transport. *Atherosclerosis*, 124(Suppl):S11–S20, 1996.

[51] Guenther Silbernagel, Ben Schöttker, Sebastian Appelbaum, Hubert Scharnagl, Marcus E Kleber, Tanja B Grammer, Andreas Ritsch, Ute Mons, Bernd Holleczek, Georg Goliasch, et al. High-density lipoprotein cholesterol, coronary artery disease, and cardiovascular mortality. *European heart journal*, 34(46):3563–3571, 2013.

[52] Benjamin F Voight, Gina M Peloso, Marju Orho-Melander, Ruth Frikke-Schmidt, Maja Barbalic, Majken K Jensen, George Hindy, Hilma Hólm, Eric L Ding, Toby Johnson, et al. Plasma hdl cholesterol and risk of myocardial infarction: a mendelian randomisation study. *The Lancet*, 380(9841):572–580, 2012.

[53] Anand Rohatgi, Amit Khera, Jarett D Berry, Edward G Givens, Colby R Ayers, Kyle E Wedin, Ian J Neeland, Ivan S Yuhanna, Daniel R Rader, James A de Lemos, et al. Hdl cholesterol efflux capacity and incident cardiovascular events. *New England Journal of Medicine*, 371(25):2383–2393, 2014.

[54] Mariano Sentí, Xavier Nogués, Juan Pedro-Botet, Juan Rubiés-Prat, and Francesc Vidal-Barraquer. Lipoprotein profile in men with peripheral vascular disease. role of intermediate density lipoproteins and apoprotein e phenotypes. *Circulation*, 85(1):30–36, 1992.

[55] F Gerald R. Fowkes, Edward Housley, Rudolph A Riemersma, Celia CA Macintyre, Elizabeth HH Cawood, Robin J Prescott, and C Vaughan Ruckley. Smoking, lipids, glucose intolerance, and blood pressure as risk factors for peripheral atherosclerosis compared with ischemic heart disease in the edinburgh artery study. *American journal of epidemiology*, 135(4):331–340, 1992.

[56] Adam E Locke, Bratati Kahali, Sonja I Berndt, Anne E Justice, Tune H Pers, Felix R Day, Corey Powell, Sailaja Vedantam, Martin L Buchkovich, Jian Yang, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518(7538):197, 2015.

[57] Masato Akiyama, Yukinori Okada, Masahiro Kanai, Atsushi Takahashi, Yukihide Momozawa, Masashi Ikeda, Nakao Iwata, Shiro Ikegawa, Makoto Hirata, Koichi Matsuda, et al. Genome-wide association study identifies 112 new loci for body mass index in the japanese population. *Nature genetics*, 49(10):1458, 2017.

[58] Look AHEAD Research Group et al. 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. *Archives of internal medicine*, 170(17):1566, 2010.

[59] GG Ravindranath and BG Rahul. Prevalence and risk factors of hemorrhoids: a study in a semi-urban centre. *International Surgery Journal*, 5(2):496–499, 2018.

[60] Nisha J Manek, Deborah Hart, Timothy D Spector, and Alexander J MacGregor. The association of body mass index and osteoarthritis of the knee joint: an examination of genetic and environmental influences. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, 48(4):1024–1029, 2003.

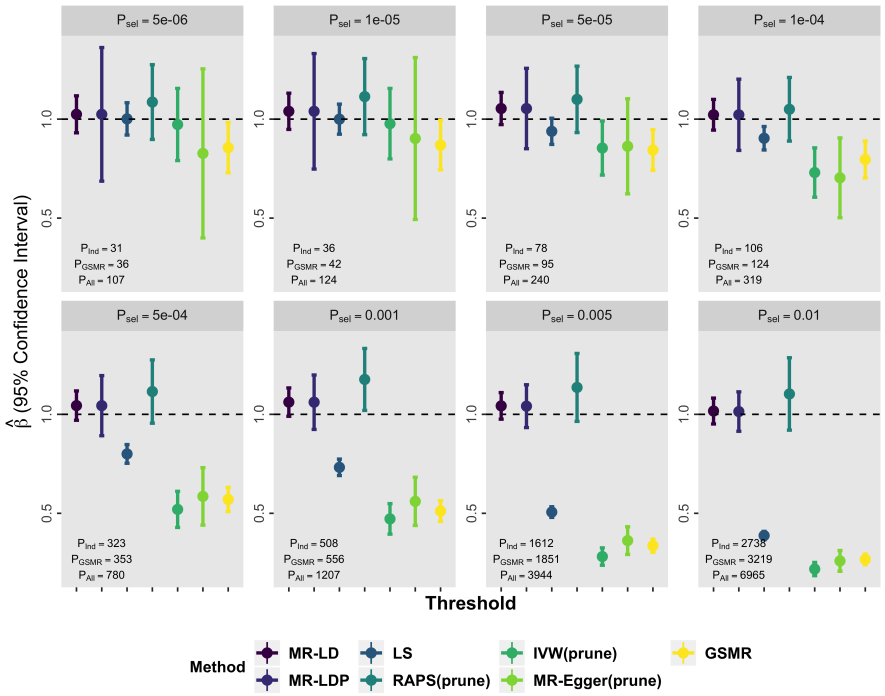
[61] Michael A Grandner, Elizabeth A Schopfer, Megan Sands-Lincoln, Nicholas Jackson, and Atul Malhotra. Relationship between sleep duration and body mass index depends on age. *Obesity*, 23(12):2491–2498, 2015.

[62] Kofi Asomaning, Elizabeth R Bertone-Johnson, Philip C Nasca, Frederick Hooven, and Penelope S Pekow. The association between body mass index and osteoporosis in patients referred for a bone mineral density examination. *Journal of Women's Health*, 15(9):1028–1034, 2006.

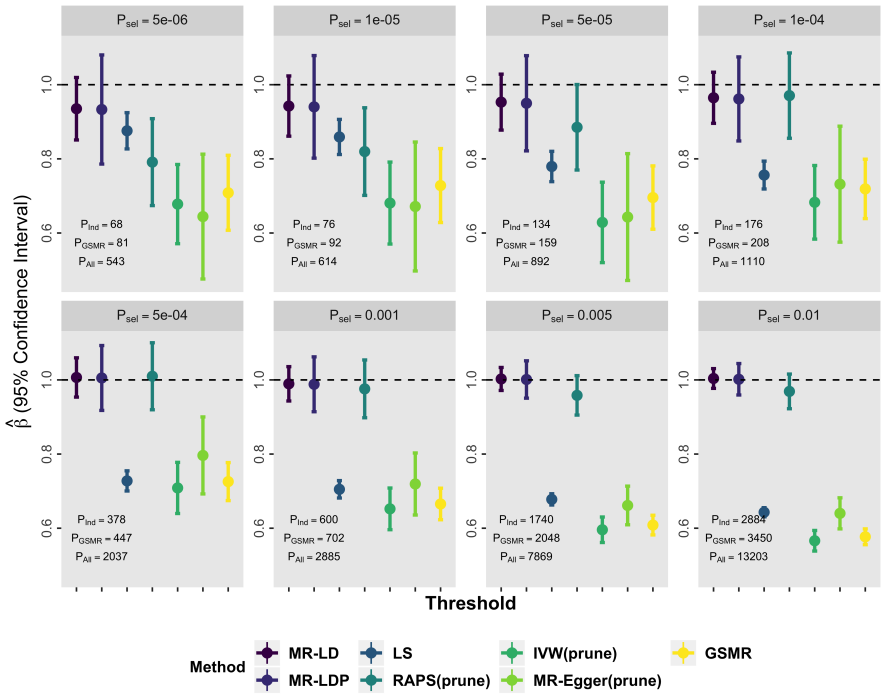
[63] Gladys Barrera, Daniel Bunout, Vivien Gattás, María Pía de la Maza, Laura Leiva, and Sandra Hirsch. A high body mass index protects against femoral neck osteoporosis in healthy elderly subjects. *Nutrition*, 20(9):769–771, 2004.

[64] Kelly R Ylitalo, MaryFran Sowers, and Steven Heeringa. Peripheral vascular disease and peripheral neuropathy in individuals with cardiometabolic clustering and obesity: National health and nutrition examination survey 2001–2004. *Diabetes Care*, 34(7):1642–1647, 2011.

[65] Guanghao Qi and Nilanjan Chatterjee. Mendelian randomization analysis using mixture models (mrmix) for genetic effect-size-distribution leads to robust estimation of causal effects. *bioRxiv*, 2018.



(A) CAD-CAD



(B) BMI-BMI

Figure 4: The result of CAD-CAD and BMI-BMI using UK10K as the reference penal with shrinkage parameter $\lambda = 0.1$. MR-LD, MR-LDP and LS methods use all SNPs selected by the screening dataset. The default value is used for thresholding r^2 in GSMR and the other three methods are 0.001.