Methods

Causal Mediation Analysis with Multiple Causally Ordered and Non-ordered Mediators based on Summarized Genetic Data

Lei Hou[1,2], Yuanyuan Yu[1,2], Xiaoru Sun[1,2], Xinhui Liu[1,2], Yifan Yu[1,2], Ran Yan[1,2], Hongkai Li[1,2*†], Fuzhong Xue[1,2*†]

[†] Fuzhong Xue and Hongkai Li contributed equally to this work.

1. Institute for Medical Dataology, Cheeloo College of Medicine, Shandong University, Jinan, People's Republic of China, 250000
2. Department of Epidemiology and Health Statistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, People's Republic of China, 250000

Corresponding author:

1. Fuzhong Xue, E-mail: xuefzh@sdu.edu.cn, Telephone: 13906405997, Address: School of public health, Shandong University, 44 Wenhua West Road, Jinan, Shandong province, China
2. Hongkai Li, E-mail: lihongkaiyouxiang@163.com, Telephone: 18310601363, Address: School of public health, Shandong University, 44 Wenhua West Road, Jinan, Shandong province, China

Short title: Causal Mediation Analysis with Multiple Mediators based on Summarized Genetic Data

Conflicts of Interest

None declared

Source of Funding

FX was supported by the National Natural Science Foundation of China (Grant 81773547) and Shandong Provincial Key Research and Development project (2018CXGC1210). HL was supported by the National Natural Science Foundation of China (Grant 82003557).

Availability of data and materials

GWAS summary data for BMI, Lipids and CVD are publicly available at https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files, http://lipidgenetics.org/ and http://www.cardiogramplusc4d.org/, respectively. Code to implement the method and reproduce all simulations and analyses is available on Github (https://github.com/hhoulei/PSEMR).

Ethics approval and consent to participate

Ethical approval was not sought, because this study involved analysis of publicly available summary-level data from GWASs, and no individual-level data were used.

Authors' contributions

HL and FX conceived the study. LH, HL contributed to theoretical derivation with assistance from YY, XL and XS. LH, RY and YY contributed to the data simulation. LH, HL and SS contributed to the application. LH and HL wrote the manuscript with input from all other authors. All authors reviewed and approved the final manuscript.

Acknowledgements

We would like to thank Editage (www.editage.com) for English language editing.

Abstract

Causal mediation analysis aims to investigate the mechanism linking an exposure and an outcome. Dealing with the impact of unobserved confounders among the exposure, mediator and outcome has always been an issue of great concern. Moreover, when multiple mediators exist, this causal pathway intertwines with other causal pathways, making it more difficult to estimate of path-specific effects (PSEs). In this article, we propose a method (PSE-MR) to identify and estimate PSEs of an exposure on an outcome through multiple causally ordered and non-ordered mediators using Mendelian Randomization, when there are unmeasured confounders among the exposure, mediators and outcome. Additionally, PSE-MR can be used when pleiotropy exists, and can be implemented using only summarized genetic data. We also conducted simulations to evaluate the finite sample performances of our proposed estimators in different scenarios. The results show that the causal estimates of PSEs are almost unbiased with good coverage and Type I error properties. We illustrate the utility of our method through a study of exploring the mediation effects of lipids in the causal pathways from body mass index to cardiovascular disease.

Key words: mediation analysis, multiple mediators, causally ordered mediators, causally non-ordered mediators, Mendelian randomization, summarized genetic data

Author summary

A new method (PSE-MR) is proposed to identify and estimate PSEs of an exposure on an outcome through multiple causally ordered and non-ordered mediators using summarized genetic data, when there are unmeasured confounders among the exposure, mediators and outcome. Lipids play important roles in the causal pathways from body mass index to cardiovascular disease

1 Introduction

Mediation analyses help to uncover the mechanisms underlying causal relationships between an exposure and an outcome by using mediator variables [1]. In mediation analyses, the total effect of an exposure on an outcome is partitioned into indirect and direct effects. Indirect effects act through mediators of interest, whereas direct effects are determined by fixing the mediator at a specified level. Estimating direct and indirect effects via existing methods typically requires a stringent sequential ignorability assumption [2] that no unmeasured confounders exist among the exposure, mediators and outcome [3]. However, this assumption may not hold in practice and omitting important confounders will necessarily bias results [4]. When multiple intermediate variables (M_1 and M_2) are involved in a study, three types of mediators with respect to M_1 and M_2 may arise, as shown in Figure 1. In Figure 1A, M_1 is conditionally independent of M_2 given the treatment (X) and measured covariates [5]. In Figure 1B, M_1 and M_2 are not causally ordered because they are independent of each other, conditional upon the treatment (X) and measured covariates [6]. In Figure 1C, mediators are causally ordered, and M_1 is treated as a mediator-outcome confounder affected by the treatment. If we are interested in the mediator M_2, we get a two-way decomposition into an indirect effect through M_2 and a direct effect (not through M_2). Imai and Yamamoto [7] proposed an approach for all the three types of mediators under a linear structural equation model. Daniel et al. [8] considered the finest possible decomposition of the total effect when there are two causally ordered mediators, and evaluated each path-specific effect (PSE) under

the counterfactual framework. Additionally, VanderWeele and Vansteelandt [9] regarded the multiple mediators simultaneously as joint mediators, and defined the "joint" natural direct and indirect effects as extensions of the usual two-way decomposition of the total effect using regression-based approach and weighting approach. Several methods [10-16] have been developed to relax the sequential ignorability assumption. However, none of them allowed for the simultaneous existence of unmeasured confounders among the exposure, mediators and the outcome.

Mendelian randomization (MR) analyses [17] using summarized data have recently become popular due to the increase in public availability of suitable data in large sample sizes from recently published genome-wide association studies [18]. For instance, Tikkanen E et al. (2019) performed a two-sample MR to evaluate independent causal roles of body components (fat-free mass and fat mass) on atrial fibrillation (AF) [19]. Firstly, univariate MR was used to estimate the causal effect of fat-free mass on AF by leveraging genetic variants (instrumental variables). Some genetic variants may be associated with both fat-free mass and fat mass, which is problematic because fat mass is also associated with AF. These genetic variants are invalid because they violate the assumption of exclusion restriction, since – they unlock the pathway from genetic variants to AF not via fat-free mass. This phenomenon is called horizontal pleiotropy, and fat mass is considered a pleiotropic trait [20]. In order to eliminate the effect of pleiotropy on causal estimation, multivariable MR [21] was performed to evaluate the causal role of fat-free mass on AF independent of fat mass. Similarly, we can

obtain the causal effect of fat mass on AF independent of fat-free mass.

Risk factors associated with genetic variants may not always be pleiotropic traits, rather they may be mediators in the causal pathway from the exposure to the outcome (Figure 1D). In this case, these genetic variants are still valid instruments and MR can be used for mediation analysis. Burgess S et al. (2017) showed that total and direct effects in a single mediator setting can be estimated by univariate and multivariable MR analyses, respectively [22]. We will review this in Section 2.1. In Section 2.2, we extend the analysis from a single mediator setting to a multiple mediators setting (PSE-MR) for both causally ordered and non-ordered mediators. Then in Section 3, we apply our method to estimate PSEs from body mass index (BMI) to cardiovascular disease (CVD) through lipids mediators. In Section 4, we conduct simulations to compare the performance of PSE-MR in different scenarios. Finally, we discuss the methods and results of this study and its potential for application. R package *PSEMR* for implementing PSE-MR is provided in Github (https://github.com/hhoulei/PSEMR).

2 Methods

Throughout, we let X, Y, M and G denote the exposure, outcome, mediator and genetic variant, respectively. U denotes a set of baseline covariates and potential confounders of the mediators, exposure and outcome relationships. We also let θ_0, α_1 and δ_1 denote the effect of X on Y, X on M and M on Y, respectively. The subscript j ($j = 1,\ldots, J$), denotes the j-th genetic variant. Increasingly, MR analyses are implemented using summarized data on the

associations of each genetic variant with the exposure, mediator and outcome, obtained from linear regressions on non-overlapping data consortia. This included the beta-coefficients ($\hat{\beta}_{Xj}$, $\hat{\beta}_{Mj}$, $\hat{\beta}_{Yj}$) and their standard errors (se($\hat{\beta}_{Xj}$), se($\hat{\beta}_{Mj}$), se($\hat{\beta}_{Yj}$)). If the exposure *X* or the outcome *Y* is binary, then these summarized association estimates may be replaced with association estimates (log(OR)) obtained from logistic regression.

Initially, we consider the indirect (through *M*) and direct (not through the above mediators) effects of an exposure *X* on an outcome *Y* using genetic variants *G*. Then we declare several assumptions. We assume all genetic variants are uncorrelated (not in linkage disequilibrium). We also assume all variables are continuous, and relationships between variables (the genetic associations with the exposure *X*, mediator *M*, and outcome *Y*, and the causal effects of *X* and *M* on *Y* as well as *X* on the *M*) are linear with homogeneity across the population. In other words, interactions between the exposure (*X*) and mediator (*M*) are not allowed unless individual data is available. We also assume that the consistency and composition assumptions in causal mediation analyses hold [24] (see S1 Appendix, Section 1). Note that we relax the assumption of no unmeasured confounders among the exposure *X*, mediator *M*, and outcome *Y*, which is required in most studies.

2.1 PSE-MR in one mediator setting

In a single mediator setting (Figure 1D), a valid instrumental variable G_j must satisfy the following three assumptions:

Assumption I. For each $j(j=1,...,J)$, the instrumental variable G_j is associated with the

146 exposure *X*.

147 This assumption requires that G_j should be strongly associated with *X*, otherwise,
148 weak instrumental variable bias will exist [25]. The "rule of thumb" advocates that the *F*
149 statistic of each instrumental variable should be at least 10 to avoid this bias [26-27] (see S1
150 Appendix, Section 2.3).

151 **Assumption II.** For each $j(j=1,...,J)$, $G_j \perp U$, and these three unmeasured confounders
152 satisfy the following criteria:

153 1) There is no additive $X-U$ interaction on *M* and *Y*.

154 2) There is no additive $M-U$ interaction on *Y*.

155 3) There is no confounders of *M*-*Y* relationship induced by *X*..

156 In this assumption, we posit that there is no confounders of *M*-*Y* relationship induced by
157 *X*, nor any interactions between *X* (or *M*) and these confounders [17]. When the interactions
158 between *M* and *U* exist, the direct effect of *X* on *Y* can be identified (see S1 Appendix,
159 Section 4). Swanson S and VanderWeele T [28] suggested that the E-value can be used to
160 examine the independence between G_j and *U*, that is, to evaluate the sensitivity of estimates
161 to confounders between G_j and *Y* (see S1 Appendix, Section 5).

162 **Assumption III.** For each $j(j=1,...,J)$, $G_j \perp Y \mid (X,U), G_j \perp M \mid (X,U)$.

163 This assumption means that there is no pleiotropy. In other words, G_j must affect *Y*
164 through *X*, and the pathways $G_j \to M \to Y$ or $G_j \to Y$ (not via *X*) are not allowed. We
165 examine and relax this assumption in Section 2.1.2.

2.1.1 PSE-MR based on IVW (PSE-IVW)

For each $j(j=1,...,J)$, we do not allow for direct effects between G_j and M ($\gamma_{1j}=0$) as well as G_j and Y ($\gamma_{0j}=0$) (Figure 1D). Based on above three assumptions, the inverse-variance weighting method (IVW) can provide an estimate of the total effect θ_T of X on Y by the following weighted regression with the intercept set to zero

$$\hat{\beta}_{Yj} = \theta_T \hat{\beta}_{Xj} + \varepsilon_{Tj}, \quad \varepsilon_{Tj} \sim N\left(0, se(\hat{\beta}_{Yj})^2\right). \tag{1}$$

The total effect θ_T between X and Y can be decomposed into a direct effect ($\theta_T=\theta_I+\theta_D=\alpha_1\times\delta_1+\theta_0$) and an indirect effect via M.

Under the framework of multivariable MR, the weighted regression model can be expanded by including genetic associations with the mediator

$$\hat{\beta}_{Yj} = \theta_0 \hat{\beta}_{Xj} + \delta_1 \hat{\beta}_{Mj} + \varepsilon_{Dj}, \quad \varepsilon_{Dj} \sim N\left(0, se(\hat{\beta}_{Yj})^2\right) \tag{2}$$

where $\hat{\theta}_0$ provides an estimate of the direct effect θ_D. The indirect effect θ_I of exposure on the outcome can be calculated as $\theta_{ID}=\theta_T-\theta_D$ (difference indirect effect). It is equivalent to $\theta_{IP}=\alpha_1\times\delta_1$ (product indirect effect), where δ_1 can be estimated by equation (2) and α_1 can be estimated by the following weighted regression with the intercept set to zero

$$\hat{\beta}_{Mj} = \alpha_1 \hat{\beta}_{Xj} + \varepsilon_{Mj}, \quad \varepsilon_{Mj} \sim N\left(0, se(\hat{\beta}_{Mj})^2\right). \tag{3}$$

The standard error of the difference and product indirect effects are presented in S1 Appendix, Section 5. The total effect can also be estimated from individual-level data using the two-stage least squares (2SLS) method. The direct effect can also be estimated using 2SLS by regressing the outcome on fitted values of the exposure, and further on fitted values of the

mediator [22].

2.1.2 PSE-MR of a single mediator based on MR-Egger (PSE-Egger)

The method proposed by Burgess et al. (2017) has some limitations. This method cannot be used if Assumption III is violated, that is, direct effects of G_j on M (M simultaneously plays the role of a pleiotropic trait) or G_j on Y (pleiotropic pathway) exist. Thus, we relax the Assumption III by allowing for direct effects between G_j and M ($\gamma_{1j} \neq 0$) as well as G_j and Y ($\gamma_{0j} \neq 0$) (Figure 1D). Without the limitation of intercept set to zero, the causal effect of X on Y can be obtained by MR-Egger regression. To satisfy the InSIDE assumption [23] for MR-Egger, we require

$$\beta_{Xj} \perp \gamma_{0j} \perp \gamma_{1j}. \tag{4}$$

The total effect θ_T can be estimated by the following weighted linear regression

$$\hat{\beta}_{Yj} = \gamma_T + \theta_T \hat{\beta}_{Xj} + \varepsilon_{Tj}, \quad \varepsilon_{Tj} \sim N\left(0, se(\hat{\beta}_{Yj})^2\right). \tag{5}$$

θ_T can also be decomposed into the direct effect $\theta_D = \theta_0$ and the product indirect effect θ_{IP}, where θ_D can be obtained by multivariable MR-Egger regression:

$$\hat{\beta}_{Yj} = \gamma_{0j} + \theta_0 \hat{\beta}_{Xj} + \delta_1 \hat{\beta}_{Mj} + \varepsilon_{Dj}, \quad \varepsilon_{Dj} \sim N\left(0, se(\hat{\beta}_{Yj})^2\right) \tag{6}$$

The intercept term γ_{0j} that differs from zero is an indicator of direct effect between G_j and Y, which is called directional pleiotropy. For product indirect effect θ_{IP}, δ_1 can be estimated by above equation (6), and α_1 can also be obtained by the following multivariable MR-Egger regression:

$$\hat{\beta}_{Mj} = \gamma_{1j} + \alpha_1 \hat{\beta}_{Xj} + \varepsilon_{Mj}, \quad \varepsilon_{Mj} \sim N\left(0, se(\hat{\beta}_{Mj})^2\right) \tag{7}$$

where γ_{1j} that differs from zero is an indicator of direct effect between G_j and *M*. The estimation of standard error for difference and product indirect effect is presented in the S1 Appendix.

2.2 *Extending PSE-MR to multiple mediators setting*

In this section, we extend the PSE-MR method to a multiple mediators setting. If there are *n* mediators $M_1, M_2, ..., M_n$ in the causal pathway from *X* to *Y*, PSEs can be identified. In the multiple mediators setting, we consider two relationships among mediators: causally non-ordered and causally ordered, respectively. In both cases, a valid instrumental variable must satisfy Assumption Ⅰ mentioned in Section 2.1, and the following Assumption Ⅱ* and Ⅲ*, which extend from the Assumption Ⅱ and Ⅲ.

Assumption Ⅱ*. For each $i, j (i = 1, ..., n, j = 1, ..., J)$, $G_j \perp U$.

1) There is no additive $X - U$ interaction on M_i and *Y*.

2) There is no additive $M_i - U$ interaction on *Y*.

3) There is no confounders of $M_i - Y$ relationship induced by *X*.

Assumption Ⅲ*. For each $i, j (i = 1, ..., n, j = 1, ..., J)$, $G_j \perp Y \mid (X, U), G_j \perp M_i \mid (X, U)$.

The illustrations and examinations for Assumptions Ⅱ and Ⅲ can also be extended to the multiple mediators setting.

2.2.1 PSE-MR for causally non-ordered mediators

Firstly, we consider causally non-ordered mediators (Figure 2A, B), where *n* mediators are independent of each other, conditional on *X*. Total effect θ_T can also be estimated by

equation (1). The direct effect ($\theta_D=\theta_0$) and product indirect effect $\theta_{IP}=\sum_{i=1}^{n}\alpha_i\delta_i$ can be estimated by the following weighted regressions with the intercept set to zero:

$$\mathrm{B}=\Psi\mathrm{A}+\boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon}\sim N(\mathbf{0},\Sigma),\ \ \Sigma=diag(se(\hat{\beta}_{Yj})^2,se(\hat{\beta}_{M_1j})^2,\cdots,se(\hat{\beta}_{M_nj})^2) \tag{8}$$

where $B=\begin{bmatrix}\hat{\beta}_{Yj}\\ \hat{\beta}_{M_1j}\\ \hat{\beta}_{M_2j}\\ \vdots\\ \hat{\beta}_{M_nj}\end{bmatrix},\Psi=\begin{bmatrix}\theta_0 & \delta_1 & \delta_2 & \cdots & \delta_n\\ \alpha_1 & 0 & 0 & \cdots & 0\\ \alpha_2 & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \cdots & \vdots\\ \alpha_n & 0 & 0 & \cdots & 0\end{bmatrix},A=\begin{bmatrix}\hat{\beta}_{Xj}\\ \hat{\beta}_{M_1j}\\ \hat{\beta}_{M_2j}\\ \cdots\\ \hat{\beta}_{M_nj}\end{bmatrix},\boldsymbol{\varepsilon}=\begin{bmatrix}\varepsilon_{0j}\\ \varepsilon_{1j}\\ \varepsilon_{2j}\\ \vdots\\ \varepsilon_{nj}\end{bmatrix}$. These estimations can also be obtained from individual-level data using 2SLS method.

Similarly, we relax Assumption III* by allowing for the direct effect between the instrumental variable G_j and mediators M_i (γ_{1j}, γ_{2j}, …, γ_{nj}), as well as G_j and Y ($\gamma_{0j}\neq 0$) (Figure 2B). Under the InSIDE assumption $\beta_{Xj}\perp\gamma_{1j}\perp\gamma_{2j}\perp\ldots\perp\gamma_{nj}\perp\gamma_{0j}$, the total effect θ_T can also be estimated by equation (5). The direct effect ($\theta_D=\theta_0$) and product indirect effect (θ_{IP}) can also be estimated by the following linear regression equations:

$$\mathrm{B}=\boldsymbol{\gamma}+\Psi\mathrm{A}+\boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon}\sim N(\mathbf{0},\Sigma),\ \ \Sigma=diag(se(\hat{\beta}_{Yj})^2,se(\hat{\beta}_{M_1j})^2,\cdots,se(\hat{\beta}_{M_nj})^2) \tag{9}$$

where $\boldsymbol{\gamma}=\begin{bmatrix}\gamma_{0j} & \gamma_{1j} & \gamma_{2j} & \cdots & \gamma_{nj}\end{bmatrix}^T$. Intercept terms γ_{0j} and $\gamma_{ij}(i=1,...,n)$ that differ from zero are indicators of direct effect between G_j and Y, as well as G_j and M_i, respectively. Detailed theoretical derivations are presented in S1 Appendix, section 3.

2.2.2 PSE-MR for causally ordered mediators

When all the mediators are causally ordered (Figure 2C, D), we let r_{pq} denote the direct effect of M_p on $M_q, p,q\in(1,2,...,n), p\neq q$. The total effect θ_T can also be estimated by

equation (5). The direct effect ($\theta_D = \theta_0$) and product indirect effect

$$\theta_{IP} = \sum_{i=1}^{n} \alpha_i \delta_i + \sum_{i=2}^{n} \sum_{j=1}^{i-1} \alpha_i r_{ij} \delta_j + \sum_{i=3}^{n} \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \alpha_i r_{ij} r_{jk} \delta_k + \ldots + \sum_{i=n-1}^{n} \sum_{j=n-2}^{i-1} \ldots \sum_{h=1}^{l-1} \alpha_i r_{ij} \ldots r_{lh} \delta_h + \alpha_n r_{n(n-1)} \ldots r_{21} \delta_1 \quad (10)$$

can be estimated by the weighted regressions in equation (8) and (9) by substituting Ψ^* for Ψ, where

$$\Psi^* = \begin{bmatrix} \theta_0 & \delta_1 & \delta_2 & \cdots & \delta_n \\ \alpha_1 & 0 & 0 & \cdots & 0 \\ \alpha_2 & r_{12} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \alpha_n & r_{1n} & r_{2n} & \cdots & 0 \end{bmatrix}.$$

The causal effect r_{pq} from M_p to M_q, $p, q \in (1, 2, \ldots, n)$, $p \neq q$ can be identified. Details of theoretical derivation are presented in S1 Appendix, section 3. In practice, we can use Mendelian randomization to justify the causal direction of any two mediators. Then we combine the results of causal relationships of any two mediators to obtain the ordering of multiple mediators.

3 Application

We attempted to reveal the causal mechanism from body mass index (BMI) to cardiovascular disease (CVD) as an illustrative example. CVD, which includes coronary heart disease, stroke and heart failure, is the leading cause of death worldwide [29]. High BMI is an important risk factor of CVD [30]. Furthermore, dyslipidaemia in obesity is characterized by increased levels of very low density lipoprotein (VLDL) cholesterol, triacylglycerols (TG) and total cholesterol (TC), and lower high density lipoprotein (HDL) cholesterol levels levels [31].

Previous studies suggested that a variety of alterations in cardiac structure and function occur in the individual as adipose tissue accumulates excessively [32]. However, Van Gaal LF et al. found little evidence that LDL cholesterol is enhanced in obesity [31]. Hence, we aim to examine whether BMI affects CVD through its influence on HDL and TG.

Genetic associations with BMI in 694,649 participants from European were obtained from the Genetic Investigation of ANthropometric Traits (GIANT) [33]. Genetic associations with TG and HDL in 188,577 participants were obtained from the Global Lipids Genetics Consortium (GLGC) [34]. Genetic associations with CVD risk in 22,233 cases and 64,762 controls of European descent were obtained from the CARDIoGRAMplusC4D Consortium [35]. We identified 285 single-nucleotide polymorphisms (SNPs) associated with BMI as a genetic instrument with *F* statistics greater than 10 (explaining 2.89% of exposure variance), by extracting the effect sizes for SNP associated with BMI ($P \leq 5\times10^{-8}$) from summary statistics. As the extracted SNPs for BMI might be correlated with each other, we pruned the variants by linkage disequilibrium (LD) ($r^2 < 0.01$, clumping window = 10000 kbp). Then we tested whether these SNPs violate the exclusion restriction assumption. Firstly we plotted funnel plot (Figure 3) and found three SNPs were outliers. After removing them, the funnel plots were more symmetric. The Egger test revealed no significant effects of the mediators, HDL (P = 0.204), TG (P = 0.349) and the outcome CVD (P = 0.071). These results indicate the absence of directional pleiotropy. Details of the SNPs are listed in ***S1 Appendix***.

Firstly, we performed a single mediator analysis for the mediators (HDL and TG) via

PSE-MR. Table 1 suggests TG and HDL are mediators in the causal pathway from BMI to CVD. Then we performed PSE-MR analysis with multiple mediators to test whether BMI has indirect effects on CVD risk through HDL and TG. Although a higher BMI increase the risk of CVD, no significant direct effect was obtained after adjusting for genetic associations with TG and HDL. Indirect effects through TG and HDL explained a large proportion of causal effect from BMI to CVD, and their total mediation proportion (MP) is 93.44%. In conclusion, three pathways exist from BMI to CVD: BMI→HDL→CVD (MP: 27.1% [17.1, 38.2]), BMI→TG→CVD (MP: 24.9% [16.3, 34.7]) and BMI→TG→HDL→CVD (MP: 23.7% [2.5, 49.3]). These results (Figure 4) are consistent with results from a pooled analysis of 97 prospective cohorts with 1.8 million participants [37] and previously described biological mechanisms [36, 38].

4 Simulation

4.1 Settings

To validate the utility of the PSE-MR method for estimating PSEs, we designed six scenarios: when Assumption III is satisfied (PSE-IVW) or violated (PSE-Egger) for settings with one mediator (simulations A, B), multiple causally non-ordered (simulation C, D) and multiple causally ordered mediators (simulation E, F).

We generated data on 25 genetic variants, an exposure (X), mediators (M), and outcome (Y) for 20,000 individuals. Briefly, we specified different values of the parameters θ_D (the direct effect of X on Y) and θ_I (the indirect effect of X on M) to observe performances of

our methods. According to the specification of θ_D and θ_I, simulations from A to F included four settings: no direct effect, no indirect effect, a direct effect along with a directionally concordant indirect effect, and a direct effect and a directionally discordant indirect effect. For PSE-Egger, the data were simulated to consider the following three cases:

Case (a): Balanced pleiotropy, InSIDE assumption satisfied;

Case (b): Directional pleiotropy, InSIDE assumption satisfied;

Case (c): Directional pleiotropy, InSIDE assumption not satisfied.

We also performed additional simulations for sensitivity analyses, where bidirectional causal effects between the exposure and mediators, population homogeneity assumption is violated, the causal order is misspecified and one of the mediators is missing. In addition, we also consider the performance of PSE-MR when the exposure and outcome are time varying. We also find the optimal number of genetic variants when we consider multiple mediators. Details of the simulation are presented in S2 Appendix.

We used the following metrics to evaluate performance of our methods: mean bias, standard errors (SE), mean square error (MSE), type I error rate for a null causal effect and empirical power to detect a non-null effect (i.e., the proportion of confidence intervals excluding zero).

4.2 Results

We varied the sample size, the number of instrumental variables, and simulated four scenarios for different sets of parameter values. We found that causal estimates of direct and indirect

effects were unbiased with good Type I error properties. As the sample size increased, bias and standard errors decreased, while power improved. Higher power and lower bias were observed as the number of instrumental variables increased (see S2 Appendix, Section 1, 3 and 5).

For two non-ordered mediators, PSE-IVW showed good performance of in standard MR when estimating the total, direct and indirect effects as well as three PSEs (Table 2). As the sample size and the number of genetic variants increased, the bias was smaller and the type I error was more stable at approximately 0.05 (see S2 Appendix, Section 3). The performance of PSE-MR based on IVW and MR-Egger with two non-ordered mediators in Case (a) and (b), are listed in eTables 9 to12 (see S2 Appendix, section 4). In Case (a), we observed that the bias was close to zero and Type I error rates was around 0.05 in PSE-MR. PSE-Egger had less bias and more stable Type I error rates than IVW when directional pleiotropy existed in at least one pathway from G to Y (Case (b)). MR-Egger performed better than IVW in term of bias, even when the InSIDE assumption was not satisfied (Case (c)). When the pleiotropic effects through confounders (violating the InSIDE assumption) were 2.5 times larger than the direct pleiotropic effects (satisfying InSIDE), estimates from PSE-Egger were much less biased and rejection rates of the causal null hypothesis were much closer to the nominal 5% rate than those from PSE-IVW were. In all cases, PSE-Egger had smaller MSE and more stable Type I error rates (0.05) than PSE-IVW when the PSE was zero. Estimators of indirect effects based on product method had more stable Type I error rates (0.05) than those based on

the difference method. Results for the two ordered multiple mediators were similar to those of two non-ordered mediators (Table 3 and eTables 17-24 in S2 Appendix, section 6). In addition, the magnitude of r_{qp} does not influence the performances of PSE-MR. Details are presented in eTable 15 (see S2 Appendix, section 5).

The estimation of direct effect is unbiased regardless of whether bidirectional causal effects between exposure and mediators exist, or the causal order is misspecified, though the estimation of PSEs is biased. Heterogeneous populations sometimes introduce bias of causal estimation for non-ordered and ordered mediators. Note that if we are missing upstream mediators (e.g. M_1), M_1 is the confounder of M_2 and Y and it is affected by X (i.e. X–induced unmeasured confounder of M_2 and Y). Thus the assumption of cross-world independence is violated. In addition, if we can obtain the information in each time points, PSE-MR can be applied into time varying exposure and mediators and it can also deal with the bi-directional relationship between exposure and mediators (see S1 Appendix, section 7-13). Performance of PSE-MR with different number of SNPs and mediators are listed in the eTable 41-42 and eFigure 9-10.

5 Discussion

In this paper, we develop a method PSE-MR to identify and estimate PSEs from an exposure on an outcome through the mediator(s) using MR when there are unmeasured confounders among the exposure, mediators and the outcome. We extend PSE-MR from a single mediator setting to the multiple mediator setting for both causally ordered and non-ordered mediators,

and outline the assumptions required to obtain causal effect. PSE-IVW can be used to explore the role of multiple mediators in the causal pathways between the exposure and outcome. The PSE-Egger can be viewed as a sensitivity analysis to provide robustness against both measured and unmeasured pleiotropy and to strengthen the evidence from the PSE-IVW analysis.

PSE-MR can estimate the direct effects between the exposure and outcome and indirect effects through mediators when the sequential ignorability assumption [39] in mediation analyses is relaxed. We compared the assumptions of PSE-MR with traditional mediation analysis methods in Table 4. Our method requires other independent assumptions. While Assumptions Ⅰ and Ⅲ are testable, there is no accepted method to test for the Assumption Ⅱ. Several sensitivity analyses can be performed to examine this assumption, such as the E-value [28] and heterogeneity test. The validity of multiple mediators PSE-Egger and its ability to estimate consistent causal effects rely on the InSIDE assumption [21] being satisfied. When the direct genetic associations with the exposure are independent of the direct genetic associations with mediators and outcome, the InSIDE assumption is satisfied. Whereas the InSIDE assumption is plausible in some cases, it sometimes will not always be valid. For example, heterogeneous populations and misspecification of the multiple mediators would bias the mediation effect estimation. When γ_{kj} is not independent from each other or γ_{0j} is not independent with γ_{kj} for $k = 1,...,n$ (e.g. we are missing one of multiple mediators), the direct effect is downward-biased and the indirect effect is upward-biased.

According to our simulation, we find that PSE-IVW is more robust in estimating causal effect than PSE-Egger for heterogeneous populations and misspecified multiple mediators. However, PSE-Egger can be applied to test directional pleiotropy, and it can give less biased estimates when the InSIDE assumption is violated.

For the multiple causally ordered mediator settings, PSE-MR can be widely used in time-varying exposure and mediators. Labrecque and Swanson (2019) [40] suggested that if the genetic associations of the exposure and mediators were time-varying, the lifetime effect estimate could be biased if we obtained the information of the exposure and mediators only at one time point. However, if we can obtain the information of the exposure and mediators at different time points, PSE-MR can provide unbiased estimates of the lifetime effects of the exposure and mediators on the outcome and other PSEs (see S2 Appendix, section 9). Thus PSE-MR can estimate each PSEs, including the causal relationships (which may potentially be bi-directional) in a non-experimental setting.

In conclusion, we propose a method of causal mediation analysis with causally ordered and non-ordered mediators based on summarized genetic data and provides a new perspective for mediation analysis.

Reference

[1]. Lee, H., Herbert, R. D., & McAuley, J. H. (2019). Mediation Analysis. JAMA, 321(7), 697–698. https://doi.org/10.1001/jama.2018.21973

[2]. Kosuke, Imai, Luke, Keele, Teppei, & Yamamoto. (2010). Identification, inference and sensitivity analysis for causal mediation effects. Statistical Science.

[3]. Fulcher, I. R., Shi, X., & Tchetgen Tchetgen, E. J. (2019). Estimation of Natural Indirect Effects Robust to Unmeasured Confounding and Mediator Measurement Error. Epidemiology (Cambridge, Mass.), 30(6), 825–834. https://doi.org/10.1097/EDE.0000000000001084

[4]. Mccandless, L. C. , & Somers, J. M. . (2017). Bayesian sensitivity analysis for unmeasured confounding in causal mediation analysis. Statistical Methods in Medical Research, 962280217729844.

[5]. MacKinnon, D. P. (2000). Contrasts in multiple mediator models. Multivariate applications in substance use research: New methods for new questions, 141-160.

[6]. Avin, C., Shpitser, I., & Pearl, J. (2005). Identifiability of path-specific effects.

[7]. Imai, K., & Yamamoto, T. (2013). Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. Political Analysis, 141-171.

[8]. Daniel, R. M., De Stavola, B. L., Cousens, S. N., & Vansteelandt, S. (2015). Causal mediation analysis with multiple mediators. Biometrics, 71(1), 1-14.

[9]. VanderWeele, T., & Vansteelandt, S. (2014). Mediation analysis with multiple mediators. Epidemiologic methods, 2(1), 95-115.

[10]. Tchetgen, E. J. T., & Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. Annals of statistics, 40(3), 1816.

[11]. Luo, P., & Geng, Z. (2016). Causal mediation analysis for survival outcome with unobserved mediator–outcome confounders. Computational Statistics & Data Analysis, 93, 336-347.

[12]. VanderWeele, T. J., & Chiba, Y. (2014). Sensitivity analysis for direct and indirect effects in the presence of exposure-induced mediator-outcome confounders. Epidemiology, biostatistics, and public health, 11(2).

[13]. Miles, C. H., Shpitser, I., Kanki, P., Meloni, S., & Tchetgen Tchetgen, E. J. (2020). On semiparametric estimation of a path-specific effect in the presence of mediator-outcome confounding. Biometrika, 107(1), 159-172.

[14]. Smith, L. H., & VanderWeele, T. J. (2019). Mediational E-values: approximate sensitivity analysis for unmeasured mediator–outcome confounding. Epidemiology, 30(6), 835-837.

[15]. Fulcher, I. R., Shi, X., & Tchetgen, E. J. T. (2019). Estimation of natural indirect effects robust to unmeasured confounding and mediator measurement error. *Epidemiology*, *30*(6), 825-834.

[16]. Ding P, Vanderweele TJ. Sharp sensitivity bounds for mediation under unmeasured

436 mediator-outcome confounding. Biometrika. 2016;103(2):483-490.
437 doi:10.1093/biomet/asw012

438 [17]. Burgess, S., Small, D. S., & Thompson, S. G. (2017). A review of instrumental
439 variable estimators for Mendelian randomization. Statistical methods in medical research,
440 26(5), 2333-2355.

441 [18]. Burgess, S., Butterworth, A., & Thompson, S. G. (2013). Mendelian randomization
442 analysis with multiple genetic variants using summarized data. Genetic epidemiology,
443 37(7), 658-665.

444 [19]. Tikkanen, E., Gustafsson, S., Knowles, J. W., Perez, M., Burgess, S., & Ingelsson, E.
445 (2019). Body composition and atrial fibrillation: a Mendelian randomization study.
446 European heart journal, 40(16), 1277-1282.

447 [20]. Burgess, S., & Thompson, S. G. (2015). Multivariable Mendelian randomization: the
448 use of pleiotropic genetic variants to estimate causal effects. American journal of
449 epidemiology, 181(4), 251-260.

450 [21]. Rees, J. M., Wood, A. M., & Burgess, S. (2017). Extending the MR-Egger method
451 for multivariable Mendelian randomization to correct for both measured and unmeasured
452 pleiotropy. Statistics in medicine, 36(29), 4705-4718.

453 [22]. Burgess, S., Thompson, D. J., Rees, J. M., Day, F. R., Perry, J. R., & Ong, K. K.
454 (2017). Dissecting causal pathways using Mendelian randomization with summarized
455 genetic data: application to age at menarche and risk of breast cancer. Genetics, 207(2),

481-487.

[23]. Bowden, J., Davey Smith, G., & Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International journal of epidemiology, 44(2), 512-525.

[24]. Taguri, M., Featherstone, J., & Cheng, J. (2018). Causal mediation analysis with multiple causally non-ordered mediators. Statistical methods in medical research, 27(1), 3-19.

[25]. Angrist, J. D., & Krueger, A. B. (1995). Split-sample instrumental variables estimates of the return to schooling. Journal of Business & Economic Statistics, 13(2), 225-235.

[26]. Martens, E. P., Pestman, W. R., de Boer, A., Belitser, S. V., & Klungel, O. H. (2006). Instrumental variables: application and limitations. Epidemiology, 260-267.

[27]. Staiger, D., & Stock, J. H. (1994). Instrumental variables regression with weak instruments (No. t0151). National Bureau of Economic Research.

[28]. Swanson, S. A., & VanderWeele, T. J. (2020). E-Values for Mendelian Randomization. Epidemiology, 31(3), e23-e24.

[29]. Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., ... & AlMazroa, M. A. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. The lancet, 380(9859), 2095-2128.

476 [30]. Prospective Studies Collaboration. (2009). Body-mass index and cause-specific
477 mortality in 900 000 adults: collaborative analyses of 57 prospective studies. The Lancet,
478 373(9669), 1083-1096.

479 [31]. Van Gaal, L. F., Mertens, I. L., & Christophe, E. (2006). Mechanisms linking obesity
480 with cardiovascular disease. Nature, 444(7121), 875-880.

481 [32]. Poirier, P., Giles, T. D., Bray, G. A., Hong, Y., Stern, J. S., Pi-Sunyer, F. X., & Eckel,
482 R. H. (2006). Obesity and cardiovascular disease: pathophysiology, evaluation, and effect
483 of weight loss: an update of the 1997 American Heart Association Scientific Statement
484 on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition,
485 Physical Activity, and Metabolism. Circulation, 113(6), 898-918.

486 [33]. Pulit, S. L., Stoneman, C., Morris, A. P., Wood, A. R., Glastonbury, C. A., Tyrrell,
487 J., ... & Yang, J. (2019). Meta-analysis of genome-wide association studies for body fat
488 distribution in 694 649 individuals of European ancestry. Human molecular genetics,
489 28(1), 166-174.

490 [34]. Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni,
491 S., ... & Beckmann, J. S. (2013). Discovery and refinement of loci associated with lipid
492 levels. Nature genetics, 45(11), 1274.

493 [35]. Schunkert, H., König, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., ...
494 & Absher, D. (2011). Large-scale association analysis identifies 13 new susceptibility
495 loci for coronary artery disease. Nature genetics, 43(4), 333-338.

496 [36]. Lusis, A. J., Attie, A. D., & Reue, K. (2008). Metabolic syndrome: from
497 epidemiology to systems biology. Nature Reviews Genetics, 9(11), 819-830.
498 [37]. Lu, Y., Hajifathalian, K., Ezzati, M., Woodward, M., Rimm, E. B., & Danaei, G.
499 (2013). Metabolic mediators of the effects of body-mass index, overweight, and obesity
500 on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with
501 1· 8 million participants. Lancet (London, England), 383(9921), 970-983.
502 [38]. de Freitas, E. V., Brandão, A. A., Pozzan, R., Magalhães, M. E., Fonseca, F., Pizzi,
503 O., ... & Brandão, A. P. (2011). Importance of high-density lipoprotein-cholesterol
504 (HDL-C) levels to the incidence of cardiovascular disease (CVD) in the elderly. Archives
505 of gerontology and geriatrics, 52(2), 217-222.
506 [39]. Tchetgen, E. J. T., & Shpitser, I. (2012). Semiparametric theory for causal mediation
507 analysis: efficiency bounds, multiple robustness, and sensitivity analysis. Annals of
508 statistics, 40(3), 1816.
509 [40]. Labrecque, J. A., & Swanson, S. A. (2020). Commentary: Mendelian randomization
510 with multiple exposures: the importance of thinking about time. International journal of
511 epidemiology, 49(4), 1158-1162.
512 [41]. Didelez, V., & Sheehan, N. (2007). Mendelian randomization as an instrumental
513 variable approach to causal inference. Statistical methods in medical research, 16(4),
514 309-330.
515 [42]. Burgess, S., & Thompson, S. G. (2017). Interpreting findings from Mendelian

randomization using the MR-Egger method. European journal of epidemiology, 32(5), 377-389.

[43]. Labrecque, J. A., & Swanson, S. A. (2020). Commentary: Mendelian randomization with multiple exposures: the importance of thinking about time. International journal of epidemiology, 49(4), 1158-1162.

[44]. VanderWeele, T. J., Vansteelandt, S., & Robins, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology (Cambridge, Mass.), 25(2), 300.

[45]. Mittinty, M. N., Lynch, J. W., Forbes, A. B., & Gurrin, L. C. (2019). Effect decomposition through multiple causally nonordered mediators in the presence of exposure - induced mediator - outcome confounding. Statistics in medicine, 38(26), 5085-5102.

[46]. VanderWeele, T. J., & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. Statistics and its Interface, 2(4), 457-468.

[47]. Loh, W. W., Moerkerke, B., Loeys, T., & Vansteelandt, S. (2020). Non-linear mediation analysis with high-dimensional mediators whose causal structure is unknown. arXiv preprint arXiv:2001.07147..

[48]. Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychological methods, 18(2), 137.

536 [49]. VanderWeele, T. J. (2014). A unification of mediation and interaction: a four-way
537 decomposition. Epidemiology (Cambridge, Mass.), 25(5), 749.
538 [50]. VanderWeele, T. J., & Tchetgen, E. J. T. (2017). Mediation analysis with time
539 varying exposures and mediators. Journal of the Royal Statistical Society. Series B,
540 Statistical Methodology, 79(3), 917.
541 [51]. Tchetgen, E. J. T., & VanderWeele, T. J. (2014). On identification of natural direct
542 effects when a confounder of the mediator is directly affected by exposure. Epidemiology
543 (Cambridge, Mass.), 25(2), 282.
544 [52]. Vansteelandt, S., Linder, M., Vandenberghe, S., Steen, J., & Madsen, J. (2019).
545 Mediation analysis of time-to-event endpoints accounting for repeatedly measured
546 mediators subject to time-varying confounding. Statistics in medicine, 38(24),
547 4828-4840.
548 [53]. Zheng, W., & van der Laan, M. (2017). Longitudinal mediation analysis with
549 time-varying mediators and exposures, with application to survival outcomes. Journal of
550 causal inference, 5(2).
551

Table 1. Causal effect of each pathway between BMI and CVD

	IVW		MR-Egger	
	OR[95% CI]	Power	OR[95% CI]	Power
Total effect	1.216[1.034,1.431]	1	1.772[1.208,2.600]	1
Direct effect	1.013[0.850,1.208]	-	1.391[0.945,2.048]	-
BMI→HDL→CVD	1.054[1.034,1.101]	0.97	1.097[1.040,1.125]	1
BMI→TG→CVD	1.050[1.033,1.096]	0.97	1.068[1.027,1.137]	1
BMI→TG→HDL→CVD	1.047[1.005,1.101]	0.7	1.052[1.000,1.125]	0.8

BMI: body mass index; CVD: cardiovascular disease; TG: triacylglycerols; HDL: high density lipoprotein; LDL: low density lipoprotein.

Table 2. Simulation of PSE-IVW with two non-ordered mediators in standard MR

True					Estimates						MSE						Power/Type I error					
TE	DE	IE	IE1	IE2	TE	DE	IE_d	IE_p	IE1	IE2	TE	DE	IE_d	IE_p	IE1	IE2	TE	DE	IE_d	IE_p	IE1	IE2
1.6	0.8	0.8	0.5	0.3	1.61	0.75	0.85	0.85	0.53	0.32	0.00	0.00	0.01	0.01	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
0.6	0.8	-0.2	-0.5	0.3	0.60	0.75	-0.15	-0.15	-0.47	0.32	0.00	0.00	0.01	0.01	0.00	0.00	1.00	1.00	0.26	0.29	1.00	1.00
0.7	0.5	0.2	0.5	-0.3	0.70	0.45	0.25	0.25	0.54	-0.29	0.00	0.00	0.01	0.01	0.00	0.00	1.00	1.00	0.74	0.75	1.00	1.00
-0.3	0.5	-0.8	-0.5	-0.3	-0.30	0.46	-0.76	-0.76	-0.47	-0.28	0.00	0.00	0.01	0.01	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
0	-0.8	0.8	0.5	0.3	0.01	-0.84	0.85	0.85	0.53	0.32	0.01	0.00	0.01	0.01	0.00	0.00	0.07	1.00	1.00	1.00	1.00	1.00
-1.6	-0.8	-0.8	-0.5	-0.3	-1.60	-0.85	-0.76	-0.76	-0.47	-0.29	0.00	0.00	0.01	0.01	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
0.8	0	0.8	0.5	0.3	0.81	-0.04	0.85	0.85	0.53	0.32	0.00	0.00	0.01	0.01	0.00	0.00	1.00	0.00	1.00	1.00	1.00	1.00
1.3	0.8	0.5	0.5	0	1.31	0.75	0.55	0.55	0.54	0.02	0.00	0.00	0.01	0.01	0.00	0.00	1.00	1.00	1.00	1.00	1.00	0.00
1.3	0.8	0.5	0.5	0	1.31	0.77	0.54	0.54	0.54	0.00	0.00	0.00	0.01	0.01	0.00	0.00	1.00	1.00	1.00	1.00	1.00	0.07
0.8	0.8	0	0	0	0.80	0.75	0.05	0.05	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.01	0.02	0.01	0.01

TE: total effect; DE: direct effect; IE: indirect effect; IE1: $X \rightarrow M_1 \rightarrow Y$; IE2: $X \rightarrow M_2 \rightarrow Y$; IE_d: indirect effect calculated by difference method; IE_p: indirect effect calculated by product method.

Table 3. Simulation of PSE-IVW with two ordered mediators in standard MR

True						Estimates							MSE							Power/Type I error						
TE	DE	IE	IE1	IE2	IE3	TE	DE	IE_d	IE_p	IE1	IE2	IE3	TE	DE	IE_d	IE_p	IE1	IE2	IE3	TE	DE	IE_d	IE_p	IE1	IE2	IE3
1.8	0.8	1	0.5	0.3	0.2	1.81	0.75	1.06	1.06	0.52	0.28	0.25	0.01	0.00	0.01	0.01	0.00	0.02	0.02	1.00	1.00	1.00	1.00	0.97	0.64	0.62
0.8	0.8	0	-0.5	0.3	0.2	0.81	0.75	0.05	0.05	-0.48	0.29	0.24	0.00	0.00	0.01	0.01	0.00	0.02	0.02	1.00	1.00	0.01	0.00	1.00	0.70	0.60
1.2	0.8	0.4	0.5	-0.3	0.2	1.20	0.75	0.45	0.45	0.54	-0.26	0.16	0.01	0.00	0.01	0.01	0.01	0.01	0.01	1.00	1.00	0.98	0.72	1.00	0.69	0.41
0.2	0.8	-0.6	-0.5	-0.3	0.2	0.20	0.76	-0.56	-0.56	-0.46	-0.25	0.16	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.97	1.00	1.00	1.00	1.00	0.68	0.38
0.2	-0.8	1	0.5	0.3	0.2	0.21	-0.84	1.05	1.06	0.52	0.29	0.25	0.01	0.00	0.01	0.01	0.00	0.02	0.02	0.71	1.00	1.00	1.00	1.00	0.67	0.62
-1.8	-0.8	-1	-0.5	-0.3	-0.2	-1.80	-0.85	-0.96	-0.96	-0.48	-0.25	-0.23	0.01	0.00	0.01	0.01	0.00	0.02	0.01	1.00	1.00	1.00	1.00	1.00	0.66	0.64
1	0	1	0.5	0.3	0.2	1.01	-0.05	1.05	1.05	0.52	0.29	0.25	0.01	0.00	0.01	0.01	0.00	0.02	0.02	1.00	0.07	1.00	1.00	0.99	0.67	0.61
1.3	0.8	0.5	0	0.3	0.2	1.31	0.75	0.56	0.56	0.02	0.29	0.25	0.00	0.00	0.01	0.01	0.00	0.02	0.02	1.00	1.00	1.00	0.89	0.00	0.67	0.62
1.5	0.8	0.7	0.5	0	0.2	1.51	0.77	0.74	0.74	0.52	-0.03	0.25	0.01	0.00	0.01	0.01	0.00	0.02	0.02	1.00	1.00	1.00	0.97	0.99	0.07	0.61
1	0.8	0.2	0	0	0.2	1.01	0.77	0.24	0.24	0.02	-0.03	0.25	0.00	0.00	0.01	0.01	0.00	0.02	0.02	1.00	1.00	0.41	0.01	0.00	0.08	0.61
0	-1	1	0.5	0.3	0.2	0.01	-1.05	1.06	1.06	0.52	0.29	0.25	0.01	0.00	0.01	0.01	0.00	0.02	0.02	0.05	1.00	1.00	1.00	1.00	0.69	0.62

TE: total effect; DE: direct effect; IE: indirect effect; IE1: $X \rightarrow M_1 \rightarrow Y$; IE2: $X \rightarrow M_2 \rightarrow Y$; IE3: $X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$; IE_d: indirect effect calculated by difference method; IE_p: indirect effect calculated by product method.

Table 4. Comparison of the assumption in PSE-MR and typical causal mediation analysis

Methods	PSE-MR	Typical causal mediation analysis	Solution for violating assumptions
Common	(1) Consistency assumption: $M(x)=M$ and $Y(x)=Y$ if $X=x$; $Y(x,m)=Y$ if $X=x$, $M=m$.		
	(2) Composition assumption: $Y(x)=Y(x,M(x))$ if $X=x$.		
	(3) A cross-world independence assumption: $Y(x,m) \perp M(x^*) \mid C$ for all (x, x^*, m).		[10] [44] [45] [51]
	(4) There is no additive interaction of the exposure and the mediators on the outcome (Y).		[10] [46] [48] [49]
	(5) Linearity.		[10] [47]
Different	(1) The IV G_j is associated with the exposure X but independent of all the unmeasured confounders.	(1) No-unmeasured confounders of the X-Y relation, that is, $Y(x,m) \perp X \mid C$ for all (x,m).	[10] [15]
	(2) The Exclusion restriction assumption or the InSIDE assumption.	(2) No-unmeasured confounders of the M-Y relation, that is, $Y(x,m) \perp M(x) \mid X = x, C$ for all (x,m).	[10] [11] [12] [13] [14] [15][16]
	(3) There is no additive interaction of the exposure or the mediators and confounders on the mediator (M) and the outcome (Y).	(3) No-unmeasured confounders of the X-M relation, that is, $M(x) \perp X \mid C$ for all x.	[10] [15]
	(4) Data for exposure, mediators and outcome can from different datasets with homogeneous population.	(4) Data for exposure, mediators and outcome must from the same dataset.	
		(5) Exposure, mediators and confounders not vary with time.	[51] [52] [53]

Figure legends

Figure 1. Three types of settings with two mediators, M_1 and M_2 are shown in (A) where M_1 is independent of M_2; (B) where M_1 is related to M_2, but not causally; and (C) where M_1 is causally related to M_2 (causally-ordered mediators). Graphical diagrams for PSE-MR are given in settings with one mediator (D), two non-ordered mediators (E), and two ordered mediators (F). X: the exposure, M_1 and M_2: two mediators, Y: outcome, G: instrumental variables (genetic variants).

Figure 2. Graphical diagrams of relationships between the exposure (X), causally non-ordered mediators (M_1, …, M_n), outcome (Y), and instrumental variables (G), which omits the confounders among X, M and Y, are shown as analyzed with (A) PSE-IVW and (B) PSE-Egger. Graphical diagrams of relationships between exposure (X), causally ordered mediators (M_1, …, M_n), outcome (Y), and instrumental variables (G), which omits the confounders (U) among X, M_1, …, M_n and Y are shown, as analyzed with (C) PSE-IVW and (D) PSE-Egger.

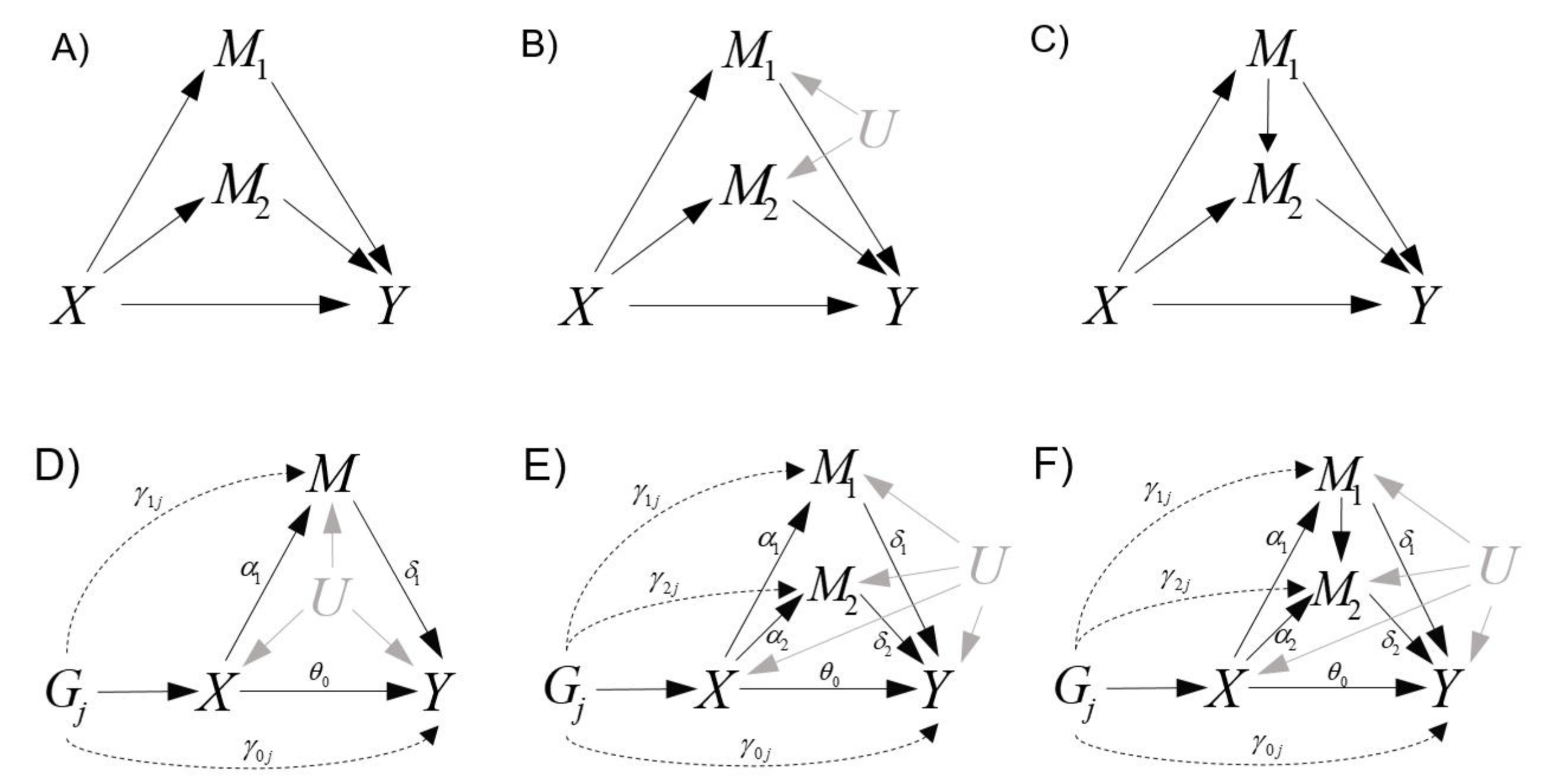
Figure 3. Funnel plots before (A-D) and after (E-H) removing outliers.

Figure 4. Diagrams of the causal pathway from BMI to CVD. BMI, body mass index; CVD, cardiovascular disease; TG, triacylglycerol; HDL, high-density lipoprotein.

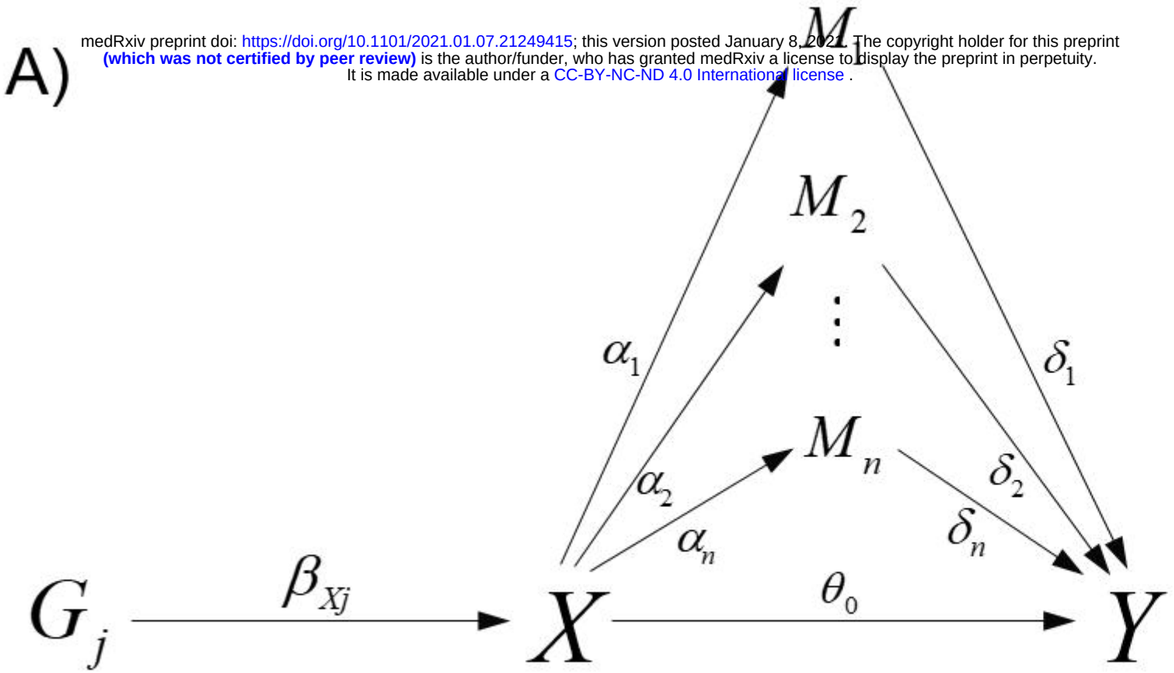
Supplementary Digital Content

S1 Appendix. Supplemental methods.

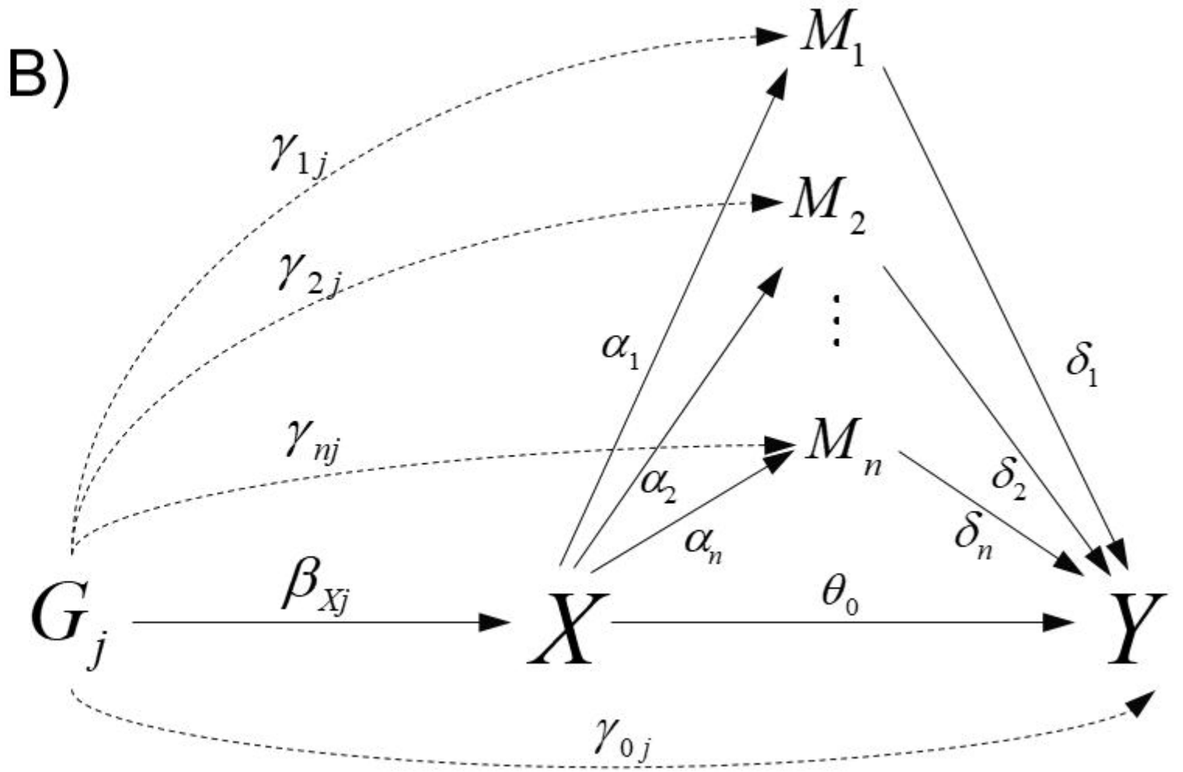
S2 Appendix. Supplemental simulations.


A)
M_1
M_2
X
Y
B)
M_1
U
M_2
X
Y
C)
M_1
M_2
X
Y
D)
M
γ_{1j}
α_1
U
δ_1
G_j
X
θ_0
Y
γ_{0j}
E)
M_1
γ_{1j}
α_1
δ_1
U
γ_{2j}
M_2
α_2
δ_2
G_j
X
θ_0
Y
γ_{0j}
F)
M_1
γ_{1j}
α_1
δ_1
U
γ_{2j}
M_2
α_2
δ_2
G_j
X
θ_0
Y
γ_{0j}

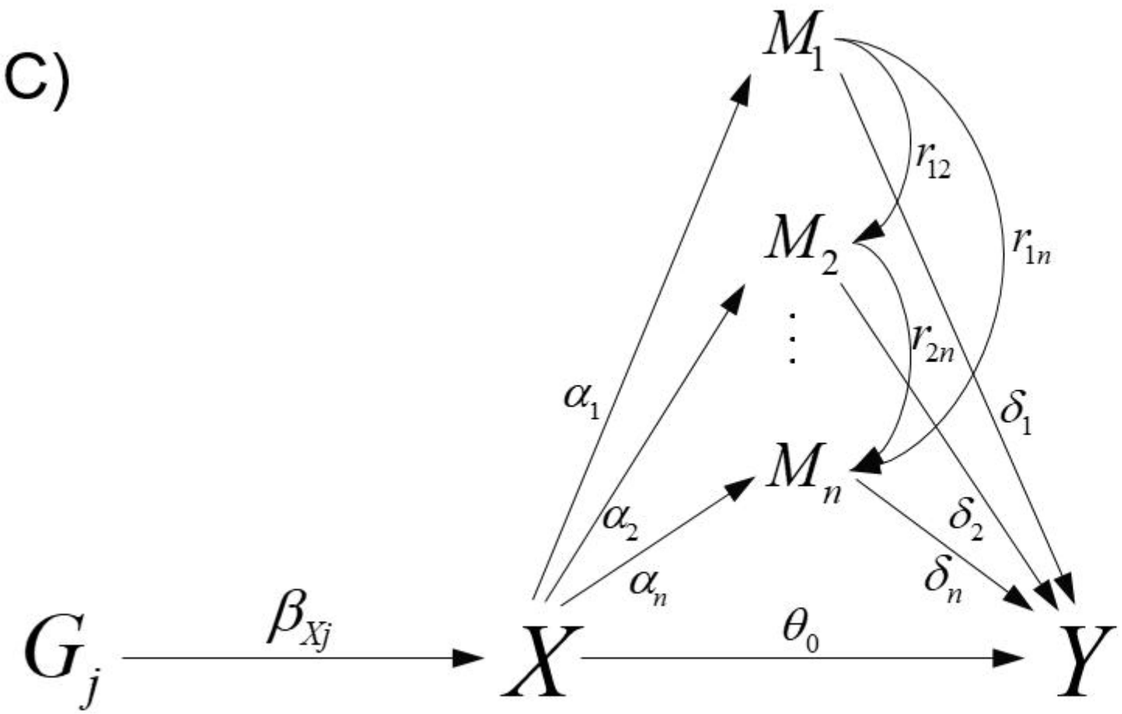
A)



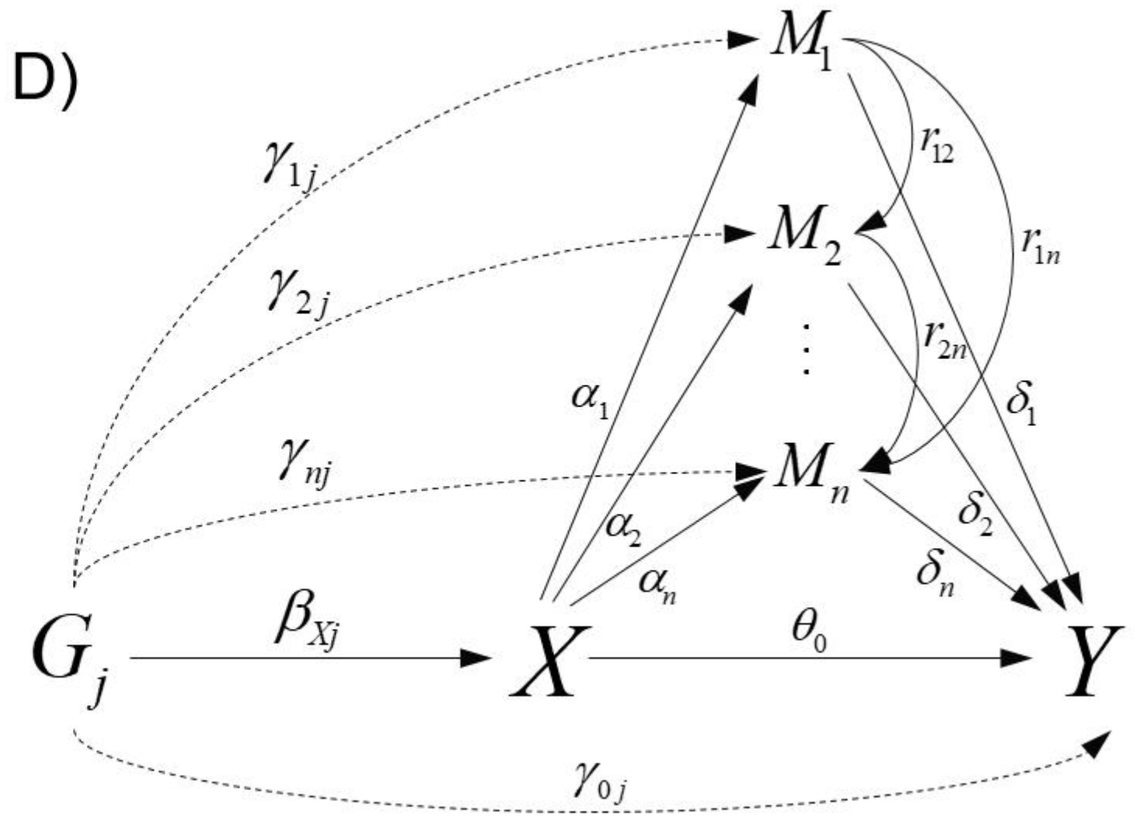
B)

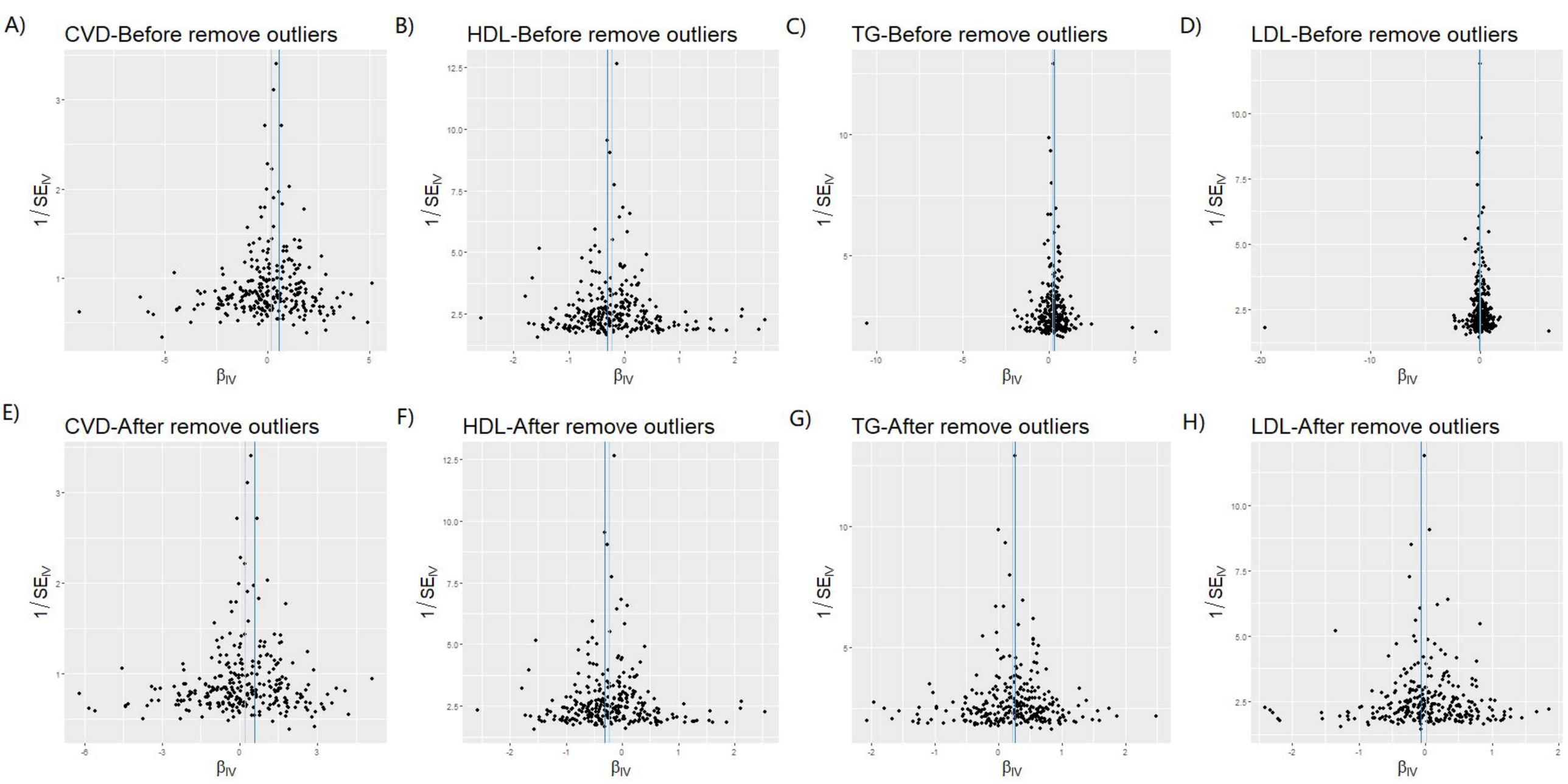


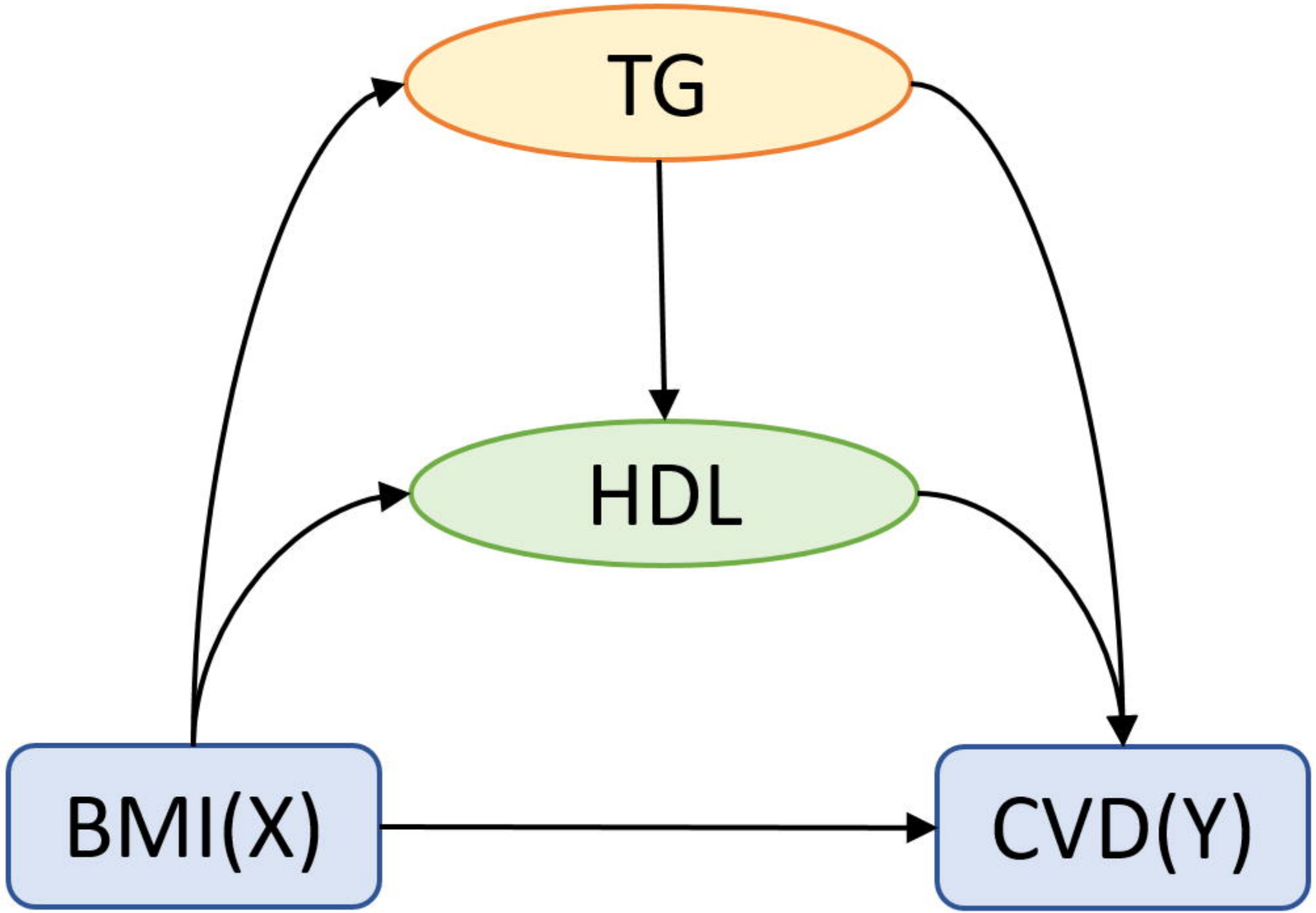
C)



D)




A)
CVD-Before remove outliers
B)
HDL-Before remove outliers
C)
TG-Before remove outliers
D)
LDL-Before remove outliers
E)
CVD-After remove outliers
F)
HDL-After remove outliers
G)
TG-After remove outliers
H)
LDL-After remove outliers
1/SE_IV
β_IV



BMI→HDL→CVD
OR: 1.054[1.034,1.101]
MP: 27.1% [17.1, 38.2])

BMI→TG→CVD
OR: 1.050[1.033,1.096]
MP: 24.9% [16.3, 34.7]

BMI→TG→HDL→CVD
OR: 1.047[1.005,1.101]
MP: 23.7% [2.5, 49.3]