Supplementary Materials for "Robust Mendelian Randomization Analysis by Automatically Selecting Valid Genetic Instruments with Applications to Identify Plasma Protein Biomarkers for Alzheimer's Disease"

Minhao Yao[1], Gary W. Miller[2], Badri N. Vardarajan[3], Andrea A. Baccarelli[2], Zijian Guo[4*], Zhonghua Liu[5*]

[1] *Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong SAR, China.*
[2] *Department of Environmental Health Sciences, Columbia University, New York, NY, USA.*
[3] *Taub Institute on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University, New York, NY, USA.*
[4] *Department of Statistics, Rutgers University, Piscataway, NJ, USA.*
[5] *Department of Biostatistics, Columbia University, New York, NY, USA.*

[*] *Correspondence to: Zijian Guo (zijguo@stat.rutgers.edu) or Zhonghua Liu (zl2509@cumc.columbia.edu)*

S1 The Potential Outcome Model for Instrumental Variables

Let $Y^{(d,\boldsymbol{z})}$ be the potential outcome if we set the exposure $D = d \in \mathbb{R}$ and the instruments $\boldsymbol{Z} = \boldsymbol{z} \in \mathbb{R}^p$. For two possible values of the exposure $d, d' \in \mathbb{R}$ and the instruments $\boldsymbol{z}, \boldsymbol{z}' \in \mathbb{R}^p$, we consider the following potential outcome model:

$$Y^{(d',\boldsymbol{z}')} - Y^{(d,\boldsymbol{z})} = (\boldsymbol{z}' - \boldsymbol{z})^\intercal \boldsymbol{\kappa} + (d' - d)\beta, \tag{S1}$$

$$\mathbb{E}(Y^{(0,\boldsymbol{0})}|\boldsymbol{Z}) = \boldsymbol{Z}^\intercal \boldsymbol{\eta}, \tag{S2}$$

where $\beta \in \mathbb{R}$ is the causal effect of interest, $\boldsymbol{\kappa} \in \mathbb{R}^p$ represents the violation level of assumption (A2), and $\boldsymbol{\eta} \in \mathbb{R}^p$ represents the violation level of assumption (A3), i.e., the presence of unmeasured confounding between the instruments and the outcome. If the exclusion restriction assumption holds, then we have $Y^{(d,\boldsymbol{z})} = Y^{(d,\boldsymbol{z}')}$ for all $\boldsymbol{z}, \boldsymbol{z}' \in \mathbb{R}^p$, which implies $\boldsymbol{\kappa} = \boldsymbol{0}$. If there is no unmeasured confounding between the instruments and the outcome, then $Y^{(d,\boldsymbol{z})}$ should be independent of $\boldsymbol{Z}$, which implies $\boldsymbol{\eta} = \boldsymbol{0}$.

Let $\boldsymbol{\pi} = \boldsymbol{\kappa} + \boldsymbol{\eta}$, $e = Y^{(0,\boldsymbol{0})} - \mathbb{E}(Y^{(0,\boldsymbol{0})}|\boldsymbol{Z})$, and $\mathrm{Var}(e|\boldsymbol{Z}) = \sigma_e^2$. Then we have the following linear model for the observational data:

$$Y = D\beta + \boldsymbol{Z}^\intercal \boldsymbol{\pi} + e, \quad \mathbb{E}(e|\boldsymbol{Z}) = 0. \tag{S3}$$

Therefore, the parameters $\boldsymbol{\pi}$ in model (S3) encode the violation of assumptions (A2) and (A3): if $\kappa_j = \eta_j = \boldsymbol{0}$, i.e., no violations of (A2) and (A3) for SNP j, then $\pi_j = 0$; otherwise $\pi_j \neq 0$.

S2 Asymptotic Distributions for Marginal Effect Estimates

We denote $\epsilon = \beta\delta + e$, then the exposure model and the outcome model can be written as the following reduced form:

$$D = \boldsymbol{Z}^\intercal \boldsymbol{\gamma} + \delta, \tag{S4}$$

$$Y = \boldsymbol{Z}^\intercal \boldsymbol{\Gamma} + \epsilon. \tag{S5}$$

In two-sample MR design, summary statistics $\{\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\sigma}}_\gamma\}$ and $\{\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\sigma}}_\Gamma\}$ are calculated with two

non-overlapping populations $\{D_{i_1}, \boldsymbol{Z}_{i_1}\}_{1\le i_1\le n_1}$ and $\{Y_{i_2}, \boldsymbol{X}_{i_2}\}_{1\le i_2\le n_2}$, where $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_1, \cdots, \widehat{\gamma}_p)^\intercal$ and $\widehat{\boldsymbol{\sigma}}_\gamma = (\widehat{\sigma}_{\gamma_1}, \cdots, \widehat{\sigma}_{\gamma_p})^\intercal$ are the genetic associations and the corresponding standard errors of the exposure, and $\widehat{\boldsymbol{\Gamma}} = (\widehat{\Gamma}_1, \cdots, \widehat{\Gamma}_p)^\intercal$ and $\widehat{\boldsymbol{\sigma}}_\Gamma = (\widehat{\sigma}_{\Gamma_1}, \cdots, \widehat{\sigma}_{\Gamma_p})^\intercal$ are the genetic associations and the corresponding standard errors of the outcome. We assume that $\{\boldsymbol{Z}_{i_1} = (Z_{i_1 1}, \cdots, Z_{i_1 p})^\intercal\}_{1\le i_1\le n_1}$ and $\{\boldsymbol{X}_{i_2} = (X_{i_2 1}, \cdots, X_{i_2 p})^\intercal\}_{1\le i_2\le n_2}$ follow the same distribution and have been standardized. Note that in two-sample setting we also have the reduced structural equation models:

$$D_{i_1} = \boldsymbol{Z}_{i_1}^\intercal \boldsymbol{\gamma} + \delta_{i_1},\ 1 \le i_1 \le n_1, \tag{S6}$$

$$Y_{i_2} = \boldsymbol{X}_{i_2}^\intercal \boldsymbol{\Gamma} + \epsilon_{i_2},\ 1 \le i_2 \le n_2. \tag{S7}$$

Let $\boldsymbol{Z} = (\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_{n_1})^\intercal$ and $\boldsymbol{X} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_{n_2})^\intercal$ be the genotype matrices of the two samples, respectively. If $\boldsymbol{Z}$ and $\boldsymbol{X}$ are properly standardized, then the summary statistics $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\Gamma}}$ can be expressed explicitly as:

$$\widehat{\boldsymbol{\gamma}} = \frac{1}{n_1}\boldsymbol{Z}^\intercal \boldsymbol{D} \quad \text{and} \quad \widehat{\boldsymbol{\Gamma}} = \frac{1}{n_2}\boldsymbol{X}^\intercal \boldsymbol{Y}, \tag{S8}$$

where $\boldsymbol{D} = (D_1, \cdots, D_{n_1})^\intercal$ and $\boldsymbol{Y} = (Y_1, \cdots, Y_{n_2})^\intercal$. We shall use $\boldsymbol{Z}_{\cdot j} \in \mathbb{R}^n$ to denote the j-th column of the matrix $\boldsymbol{Z}$. Then for $1 \le j \le p$,

$$\begin{aligned}
\sqrt{n_1}\,(\widehat{\gamma}_j - \gamma_j) &= \sqrt{n_1}\left[\left(\frac{1}{n_1}\boldsymbol{Z}_{\cdot j}^\intercal \boldsymbol{Z}_{\cdot j} - 1\right)\gamma_j + \frac{1}{n_1}\boldsymbol{Z}_{\cdot j}^\intercal\left(\sum_{l\ne j}\boldsymbol{Z}_{\cdot l}\gamma_l + \boldsymbol{\delta}\right)\right] \\
&= \frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\left[(Z_{ij}^2 - 1)\gamma_j + Z_{ij}(\sum_{l\ne j} Z_{il}\gamma_l + \delta_i)\right].
\end{aligned} \tag{S9}$$

Since $\mathbb{E}(\widehat{\gamma}_j - \gamma_j) = 0$, by independence we have:

$$\begin{aligned}
&\mathrm{Var}\left((Z_{ij}^2 - 1)\gamma_j + Z_{ij}(\sum_{l\ne j} Z_{il}\gamma_l + \delta_i)\right) \\
&= \mathbb{E}(Z_{ij}^2 - 1)^2\gamma_j^2 + \mathbb{E}Z_{ij}^2(\sum_{l\ne j} Z_{il}\gamma_l + \delta_i)^2 \\
&= \mathrm{Var}(Z_{ij}^2)\gamma_j^2 + \sum_{l\ne j}\gamma_l^2 + \sigma_\delta^2.
\end{aligned} \tag{S10}$$

On the other hand, for $1 \leq j_1 \neq j_2 \leq p$,

$$\begin{aligned}
&\mathrm{Cov}\left((Z_{ij_1}^2-1)\gamma_{j_1}+Z_{ij_1}(\sum_{l\neq j_1}Z_{il}\gamma_l+\delta_i),(Z_{ij_2}^2-1)\gamma_{j_2}+Z_{ij_2}(\sum_{l\neq j_2}Z_{il}\gamma_l+\delta_i)\right)\\
&=\mathrm{Cov}\left((Z_{ij_1}^2-1)\gamma_{j_1},(Z_{ij_2}^2-1)\gamma_{j_2}\right)+\mathrm{Cov}\left(Z_{ij_1}(\sum_{l\neq j_1,j_2}Z_{il}\gamma_l+\delta_i),Z_{ij_2}(\sum_{l\neq j_1,j_2}Z_{il}\gamma_l+\delta_i)\right)\\
&+\mathrm{Cov}\left(Z_{ij_1}Z_{ij_2}\gamma_{j_2},Z_{ij_2}(\sum_{l\neq j_1,j_2}Z_{il}\gamma_l+\delta_i)\right)+\mathrm{Cov}\left(Z_{ij_1}(\sum_{l\neq j_1,j_2}Z_{il}\gamma_l+\delta_i),Z_{ij_1}Z_{ij_2}\gamma_{j_1}\right)\\
&+\mathrm{Cov}\left((Z_{ij_1}^2-1)\gamma_{j_1},Z_{ij_2}(\sum_{l\neq j_2}Z_{il}\gamma_l+\delta_i)\right)+\mathrm{Cov}\left(Z_{ij_1}(\sum_{l\neq j_1}Z_{il}\gamma_l+\delta_i),(Z_{ij_2}^2-1)\gamma_{j_2}\right)\\
&+\mathrm{Cov}\left(Z_{ij_1}Z_{ij_2}\gamma_{j_2},Z_{ij_1}Z_{ij_2}\gamma_{j_1}\right).
\end{aligned} \tag{S11}$$

By independence and standardization, we have:

- $\mathrm{Cov}\left((Z_{ij_1}^2-1)\gamma_{j_1},(Z_{ij_2}^2-1)\gamma_{j_2}\right)=0$,
- $\mathrm{Cov}\left(Z_{ij_1}(\sum_{l\neq j_1,j_2}Z_{il}\gamma_l+\delta_i),Z_{ij_2}(\sum_{l\neq j_1,j_2}Z_{il}\gamma_l+\delta_i)\right)=0$
- $\mathrm{Cov}\left(Z_{ij_1}Z_{ij_2}\gamma_{j_2},Z_{ij_2}(\sum_{l\neq j_1,j_2}Z_{il}\gamma_l+\delta_i)\right)=\mathrm{Cov}\left(Z_{ij_1}(\sum_{l\neq j_1,j_2}Z_{il}\gamma_l+\delta_i),Z_{ij_1}Z_{ij_2}\gamma_{j_1}\right)=0$,
- $\mathrm{Cov}\left((Z_{ij_1}^2-1)\gamma_{j_1},Z_{ij_2}(\sum_{l\neq j_2}Z_{il}\gamma_l+\delta_i)\right)=\mathrm{Cov}\left(Z_{ij_1}(\sum_{l\neq j_1}Z_{il}\gamma_l+\delta_i),(Z_{ij_2}^2-1)\gamma_{j_2}\right)=0$,
- $\mathrm{Cov}\left(Z_{ij_1}Z_{ij_2}\gamma_{j_2},Z_{ij_1}Z_{ij_2}\gamma_{j_1}\right)=\gamma_{j_1}\gamma_{j_2}\left(\mathbb{E}(Z_{ij_1}^2Z_{ij_2}^2)-(\mathbb{E}Z_{ij_1}Zij_2)^2\right)=\gamma_{j_1}\gamma_{j_2}$.

Therefore,

$$\mathrm{Cov}\left((Z_{ij_1}^2-1)\gamma_{j_1}+Z_{ij_1}(\sum_{l\neq j_1}Z_{il}\gamma_l+\delta_i),(Z_{ij_2}^2-1)\gamma_{j_2}+Z_{ij_2}(\sum_{l\neq j_2}Z_{il}\gamma_l+\delta_i)\right)=\gamma_{j_1}\gamma_{j_2}, \tag{S12}$$

and similarly for $\widehat{\mathbf{\Gamma}}$. Since we use two non-overlapping samples, the covariance between $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\mathbf{\Gamma}}$ should be 0. Therefore, we establish the joint limiting distribution of $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\mathbf{\Gamma}}$.

$$\begin{pmatrix}\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\\ \widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}\end{pmatrix}\xrightarrow{d}N\left[\mathbf{0},\begin{pmatrix}\frac{1}{n_1}\mathbf{V}_\gamma & \mathbf{0}\\ \mathbf{0} & \frac{1}{n_2}\mathbf{V}_\Gamma\end{pmatrix}\right],$$

where the diagonal entries of $\mathbf{V}_\gamma$ and $\mathbf{V}_\Gamma$ are $\mathbf{V}_{\gamma,jj}=\mathrm{Var}(Z_{ij}^2)\gamma_j^2+\sum_{l\neq j}\gamma_l^2+\sigma_\delta^2$ and $\mathbf{V}_{\Gamma,jj}=\mathrm{Var}(Z_{ij}^2)\Gamma_j^2+\sum_{l\neq j}\Gamma_l^2+\sigma_\epsilon^2$, respectively, and the off-diagonal entries of $\mathbf{V}_\gamma$ and $\mathbf{V}_\Gamma$ are $\mathbf{V}_{\gamma,j_1j_2}=$

$\gamma_{j_1}\gamma_{j_2}$ and $\mathbf{V}_{\Gamma,j_1j_2} = \Gamma_{j_1}\Gamma_{j_2}$ $(j_1 \neq j_2)$, respectively.

S3 Tail Bound of Normal Distribution and the Multiplicity Adjustment

Let $W \sim N(\mu, \sigma^2)$, then by applying the Markov inequality to $e^{\lambda(W-\mu)}$ for $\lambda \geq 0$, we have:

$$\mathbb{P}((W-\mu) \geq t) = \mathbb{P}\left(e^{\lambda(W-\mu)} \geq e^{\lambda t}\right) \leq \frac{\mathbb{E}e^{\lambda(W-\mu)}}{e^{\lambda t}} = e^{\frac{\lambda^2\sigma^2}{2} - \lambda t}.$$

Therefore, we have:

$$\log \mathbb{P}((W-\mu) \geq t) \leq -\sup_{\lambda \geq 0}\left\{\lambda t - \frac{\lambda^2\sigma^2}{2}\right\} = -\frac{t^2}{2\sigma^2}.$$

By the symmetry of W, we have:

$$\mathbb{P}((|W-\mu| \geq t) \leq 2e^{-\frac{t^2}{2\sigma^2}}. \tag{S13}$$

Apply equation (S13) to $W = \widehat{\pi}_k^{[j]}$ and $t = \widehat{\mathrm{SE}}(\widehat{\pi}_k^{[j]})\sqrt{\log \min\{n_1, n_2\}}$, then for k, j such that $\mathbb{E}\widehat{\pi}_k^{[j]} = 0$, we have:

$$\mathbb{P}\left(|\widehat{\pi}_k^{[j]}| \leq \widehat{\mathrm{SE}}(\widehat{\pi}_k^{[j]})\sqrt{\log \min\{n_1, n_2\}}\right) \geq 1 - \mathbb{P}\left(|\widehat{\pi}_k^{[j]}| \geq \widehat{\mathrm{SE}}(\widehat{\pi}_k^{[j]})\sqrt{\log \min\{n_1, n_2\}}\right) \geq 1 - \frac{2}{\sqrt{\min\{n_1, n_2\}}} \to 1.$$

S4 Derivation of the Variance of MR-SPI Estimator

From Section S2, the joint limit distribution of $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\Gamma}}$ are:

$$\begin{pmatrix} \widehat{\boldsymbol{\gamma}} \\ \widehat{\boldsymbol{\Gamma}} \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\Gamma} \end{pmatrix}, \begin{pmatrix} \mathbf{V}_\gamma/n_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_\Gamma/n_2 \end{pmatrix}\right). \tag{S14}$$

For simplicity of notation, we assume that $\widehat{\mathcal{V}}$ contains all candidate IVs. Then $\widehat{\beta}_{\text{SPI}} = \widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}}/\widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}}$. Apply delta method with $g(u,v) = u/v$, then we have:

$$\text{Var}\left(\widehat{\beta}_{\text{SPI}}\right) \to \left(\frac{\mathbb{E}\widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}}}{\mathbb{E}\widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}}}\right)^2 \left(\frac{\text{Var}(\widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}})}{(\mathbb{E}\widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}})^2} + \frac{\text{Var}(\widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}})}{(\mathbb{E}\widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}})^2} - 2\frac{\text{Cov}(\widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}})}{(\mathbb{E}\widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}})(\mathbb{E}\widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}})}\right). \tag{S15}$$

Due to the independence between $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\mathbf{\Gamma}}$, we have $\mathbb{E}\widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}} = \mathbf{\Gamma}^\intercal\boldsymbol{\gamma}$ and $\mathbb{E}\widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^\intercal\boldsymbol{\gamma} + tr(\mathbf{V}_\gamma)/n_1$. As for the variance and covariance terms in equation (S15), we can apply delta method again with function $h(u,v) = (vu, u^2)$. First, we can calculate the gradient:

$$\nabla h(\boldsymbol{\gamma}, \mathbf{\Gamma}) = \begin{pmatrix} \mathbf{\Gamma} & 2\boldsymbol{\gamma} \\ \boldsymbol{\gamma} & \mathbf{0} \end{pmatrix}. \tag{S16}$$

Then we have:

$$\begin{aligned} \text{Var}(h(\widehat{\boldsymbol{\gamma}}, \widehat{\mathbf{\Gamma}})) &= \begin{pmatrix} \text{Var}(\widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}}) & \text{Cov}(\widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}}) \\ \text{Cov}(\widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}}) & \text{Var}(\widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}}) \end{pmatrix} \\ &\to \begin{pmatrix} \mathbf{\Gamma} & 2\boldsymbol{\gamma} \\ \boldsymbol{\gamma} & \mathbf{0} \end{pmatrix}^\intercal \begin{pmatrix} \mathbf{V}_\gamma/n_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_\Gamma/n_2 \end{pmatrix} \begin{pmatrix} \mathbf{\Gamma} & 2\boldsymbol{\gamma} \\ \boldsymbol{\gamma} & \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{\Gamma}^\intercal\mathbf{V}_\gamma\mathbf{\Gamma}/n_1 + \boldsymbol{\gamma}^\intercal\mathbf{V}_\Gamma\boldsymbol{\gamma}/n_2 & 2\boldsymbol{\gamma}^\intercal\mathbf{V}_\gamma\mathbf{\Gamma}/n_1 \\ 2\boldsymbol{\gamma}^\intercal\mathbf{V}_\gamma\mathbf{\Gamma}/n_1 & 4\boldsymbol{\gamma}^\intercal\mathbf{V}_\gamma\boldsymbol{\gamma}/n_1 \end{pmatrix}. \end{aligned} \tag{S17}$$

Combining Equations (S15) and (S17), and using plug-in method to replace the values with their estimates, then we can get the estimate of the variance of $\widehat{\beta}_{\text{SPI}}$:

$$\begin{aligned} \widehat{\text{Var}}(\widehat{\beta}_{\text{SPI}}) = &\left(\frac{\widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}}}{\widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}} + tr(\widehat{\mathbf{V}}_\gamma)/n_1}\right)^2 \\ &\times \left\{\frac{\widehat{\boldsymbol{\gamma}}^\intercal\widehat{\mathbf{V}}_\Gamma\widehat{\boldsymbol{\gamma}}/n_2 + \widehat{\mathbf{\Gamma}}^\intercal\widehat{\mathbf{V}}_\gamma\widehat{\mathbf{\Gamma}}/n_1}{(\widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}})^2} + 4\frac{\widehat{\boldsymbol{\gamma}}^\intercal\widehat{\mathbf{V}}_\gamma\widehat{\boldsymbol{\gamma}}/n_1}{(\widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}} + tr(\widehat{\mathbf{V}}_\gamma)/n_1)^2} - 4\frac{\widehat{\mathbf{\Gamma}}^\intercal\widehat{\mathbf{V}}_\gamma\widehat{\boldsymbol{\gamma}}/n_1}{(\widehat{\mathbf{\Gamma}}^\intercal\widehat{\boldsymbol{\gamma}})(\widehat{\boldsymbol{\gamma}}^\intercal\widehat{\boldsymbol{\gamma}} + tr(\widehat{\mathbf{V}}_\gamma)/n_1)}\right\}. \end{aligned} \tag{S18}$$

S5 The Coverage Probability of MR-SPI

We can decompose $\widehat{\beta}_{\text{SPI}}$ as:

$$\widehat{\beta}_{\text{SPI}} = \frac{\widehat{\mathbf{\Gamma}}_{\widehat{\mathcal{V}}}^\intercal\widehat{\boldsymbol{\gamma}}_{\widehat{\mathcal{V}}}}{\widehat{\boldsymbol{\gamma}}_{\widehat{\mathcal{V}}}^\intercal\widehat{\boldsymbol{\gamma}}_{\widehat{\mathcal{V}}}} = \frac{\widehat{\mathbf{\Gamma}}_{\mathcal{V}}^\intercal\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}{\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}^\intercal\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}\mathbf{1}_{\widehat{\mathcal{V}}=\mathcal{V}} + \sum_{\mathcal{V}'\neq\mathcal{V}} \frac{\widehat{\mathbf{\Gamma}}_{\mathcal{V}'}^\intercal\widehat{\boldsymbol{\gamma}}_{\mathcal{V}'}}{\widehat{\boldsymbol{\gamma}}_{\mathcal{V}'}^\intercal\widehat{\boldsymbol{\gamma}}_{\mathcal{V}'}}\mathbf{1}_{\widehat{\mathcal{V}}=\mathcal{V}'}. \tag{S19}$$

And thus

$$\sqrt{n_1}\left(\widehat{\beta}_{\mathrm{SPI}}-\beta\right)=\sqrt{n_1}\left(\frac{\widehat{\mathbf{\Gamma}}_{\mathcal{V}}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}{\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}-\beta\right)\mathbf{1}_{\widehat{\mathcal{V}}=\mathcal{V}}+\sum_{\mathcal{V}'\neq\mathcal{V}}\left(\frac{\widehat{\mathbf{\Gamma}}_{\mathcal{V}'}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}'}}{\widehat{\boldsymbol{\gamma}}_{\mathcal{V}'}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}'}}-\beta\right)\mathbf{1}_{\widehat{\mathcal{V}}=\mathcal{V}'}. \tag{S20}$$

According to equation (S18), if $n_1/n_2=O(1)$, then we also have that $\mathrm{v}:=n_1\mathrm{Var}\left(\widehat{\beta}_{\mathrm{SPI}}\right)=O(1)$. Therefore, we have:

$$\sqrt{n_1}\left(\frac{\widehat{\mathbf{\Gamma}}_{\mathcal{V}}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}{\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}-\beta\right)\xrightarrow{d}N(0,\mathrm{v}). \tag{S21}$$

According to Lemma 1 in [S1], we have $\mathbb{P}(\widehat{\mathcal{V}}=\mathcal{V})=1$. Therefore, combining (S21) with Slutsky's theorem and the fact that $\mathbf{1}_{\widehat{\mathcal{V}}=\mathcal{V}}\xrightarrow{\mathbb{P}}1$, we have:

$$\sqrt{n_1}\left(\frac{\widehat{\mathbf{\Gamma}}_{\mathcal{V}}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}{\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}-\beta\right)\mathbf{1}_{\widehat{\mathcal{V}}=\mathcal{V}}\xrightarrow{d}N(0,\tau). \tag{S22}$$

On the other hand, for any $\varepsilon>0$, we have:

$$\left|\sqrt{n_1}\left(\widehat{\beta}_{\mathrm{SPI}}-\beta\right)-\sqrt{n_1}\left(\frac{\widehat{\mathbf{\Gamma}}_{\mathcal{V}}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}{\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}-\beta\right)\mathbf{1}_{\widehat{\mathcal{V}}=\mathcal{V}}\right|\geq\varepsilon\Longrightarrow\widehat{\mathcal{V}}\neq\mathcal{V}. \tag{S23}$$

Therefore, we can get that

$$\mathbb{P}\left(\left|\sqrt{n_1}\left(\widehat{\beta}_{\mathrm{SPI}}-\beta\right)-\sqrt{n_1}\left(\frac{\widehat{\mathbf{\Gamma}}_{\mathcal{V}}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}{\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}-\beta\right)\mathbf{1}_{\widehat{\mathcal{V}}=\mathcal{V}}\right|\geq\varepsilon\right)\leq\mathbb{P}\left(\widehat{\mathcal{V}}\neq\mathcal{V}\right)\to0. \tag{S24}$$

And thus

$$\sqrt{n_1}\left(\widehat{\beta}_{\mathrm{SPI}}-\beta\right)-\sqrt{n_1}\left(\frac{\widehat{\mathbf{\Gamma}}_{\mathcal{V}}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}{\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}^{\intercal}\widehat{\boldsymbol{\gamma}}_{\mathcal{V}}}-\beta\right)\mathbf{1}_{\widehat{\mathcal{V}}=\mathcal{V}}\xrightarrow{\mathbb{P}}0. \tag{S25}$$

Combining (S22) and (S25) with the asymptotic equivalence lemma, and we have:

$$\sqrt{n_1}\left(\widehat{\beta}_{\mathrm{SPI}}-\beta\right)\xrightarrow{d}N(0,\mathrm{v}). \tag{S26}$$

And the asymptotic coverage probability of conditional confidence interval is guaranteed by (S26).

S6 Comparing MR-SPI to other competing MR methods in simulation studies

We conduct extensive simulations to evaluate the performance of MR-SPI in the presence of invalid IVs. We simulate data in a two-sample setting under four setups: (**S1**) majority rule condition holds, and no locally invalid IVs exist; (**S2**) plurality rule condition holds, and no locally invalid IVs exist; (**S3**) majority rule condition holds, and locally invalid IVs exist; (**S4**) plurality rule condition holds, and locally invalid IVs exist. More detailed simulation settings are described in Online Methods.

We compare the percent bias, empirical coverage and average lengths of 95% confidence intervals of MR-SPI to the following competing MR methods: (i) the random-effects IVW method that performs random-effects meta-analysis to account for pleiotropy [S2], (ii) MR-RAPS that assumes pleiotropic effects are normally distributed and applies the maximum profile likelihood estimation to obtain the causal effect estimate [S3], (iii) MR-PRESSO that detects the SNPs that substantially reduce the residual sum of squares of the regression when omitted from the analysis as outliers [S4], (iv) the weighted median method that takes the weighted median of the ratio estimates as the causal effect [S5], (v) the mode-based estimation that takes the mode of the smoothed empirical density function of the ratio estimates as the causal effect [S6], (vi) MRMix that models the SNP-exposure and SNP-outcome effects with a bivariate normal mixture distribution [S7], and (vii) the contamination mixture method that models the ratio estimates of SNPs with a normal mixture distribution [S8]. We exclude MR-Egger in this simulation since it is heavily biased under our simulation settings. Among those methods, the random-effects IVW method and MR-RAPS require the InSIDE assumption, MR-PRESSO and the weighted median method require the majority rule condition, while MR-SPI, the mode-based estimation, MRMix and the contamination mixture method require the plurality rule condition (or the ZEMPA assumption). For simplicity, we shall use IVW to represent the random-effects IVW method hereafter.

In Figure 2(a), we present the percent bias of those MR methods in simulated data with a sample size of 5,000 for both the exposure and the outcome. Moreover, Supplementary Figure S1(a) and Supplementary Table S1 provide a comparison of percent bias of those MR methods across different sample sizes (n = 5,000, 10,000, 20,000, 40,000, 80,000). Generally, the proposed MR-SPI has small bias in all four settings. IVW and MR-RAPS are biased since the InSIDE

assumption does not hold in our simulation settings. When the InSIDE assumption holds and the pleiotropic effects have zero mean, both IVW and MR-RAPS can give a nearly unbiased estimate of the causal effect[S3, S2]. Besides, the biases of these two methods are smaller in settings (**S3**) and (**S4**) compared to (**S1**) and (**S2**), as the degree of violation of (A2) and (A3) is generally smaller when some of the candidate IVs are only locally invalid. MR-PRESSO yields biased estimates as it fails to remove outliers in most of our settings. As discussed in [S4], MR-PRESSO performs well when (1) both the majority rule condition and the InSIDE assumption hold, and (2) the pleiotropic effects have zero mean. However, the InSIDE assumption does not hold in all four settings, and the mean of pleiotropic effects is nonzero in settings (**S1**), (**S3**) and (**S4**). The weighted median estimator is biased when only the plurality rule condition holds, since it requires more than half of the candidate IVs to be valid (majority rule). The mode-based estimation, MRMix and the contamination mixture method are all nearly unbiased, because these three methods only require the plurality rule condition to hold, which is satisfied in all the four simulation settings.

Figure 2(b) reports the empirical coverage of the confidence intervals of those methods in simulated data with a sample size of 5,000. Additional results for empirical coverage of those methods under different sample sizes (n = 5,000, 10,000, 20,000, 40,000, 80,000) can be found in Supplementary Figure S1(b) and Supplementary Table S2. Under settings (**S1**) and (**S2**) where locally invalid IV does not exist, the confidence interval of MR-SPI can attain 95% coverage level even when the sample sizes are small (e.g., 5,000). In the presence of locally invalid IVs, i.e., under settings (**S3**) and (**S4**), the empirical coverage of the confidence interval of MR-SPI can still attain the nominal level when sample sizes are 80,000, because MR-SPI can correctly distinguish locally invalid IVs from valid IVs when sample sizes are large enough. However, it is difficult for MR-SPI to identify those locally invalid IVs under (**S3**) and (**S4**) if the sample sizes are not large (e.g., 5,000), and therefore the empirical coverage of MR-SPI is lower than 95%. In such cases, our proposed robust confidence interval constructed by Algorithm 2 in Online Methods, can still attain the 95% coverage level and is immune to the IV selection error in finite samples. The empirical coverage of the weighted median method is lower than MR-SPI even in setting (**S1**) where the majority rule condition holds. For example, when the sample size is 20,000, the empirical coverage of the weighted median method is 0.638. Compared to the confidence interval of MR-SPI, the confidence interval of the mode-based estimation is generally more conservative with coverage above the nominal level in our simulation settings, which is the price to pay for being

less affected by the invalid instruments[S6]. Both MRMix and the contamination mixture method cannot attain the 95% coverage level in all the four simulation settings. This is because these two methods assume a normal mixture model for either the SNP-trait associations or the ratio estimates, which is violated in our simulation settings, and thus their coverage levels are below the nominal level. However, when the underlying distributional assumption is satisfied, both MRMix and the contamination mixture method can attain the nominal level under the plurality rule condition, as shown in additional simulations in Supplementary Section S8.

We report the average lengths of 95% confidence intervals of MR-SPI and the other competing methods with sample size 5,000 in Figure 2(c), and additional results for other sample sizes are provided in Supplementary Figure S1(c) and Supplementary Table S3. Although MR-RAPS generally has the shortest confidence interval, it is biased and the coverage level is close to zero in our simulation settings, because the required InSIDE assumption does not hold here. Compared to other competing MR methods, MR-SPI generally has the shortest confidence interval under all the four simulation settings. The average length of confidence interval of IVW is not decreasing as the sample sizes increase, since we apply the random-effects IVW method here, which scales up the standard error of the causal effect estimate when there exists heterogeneity in the ratio estimates[S2]. In setting (**S4**), MR-PRESSO has longer confidence interval as the sample sizes increase, since it tends to treat none of the candidate SNPs as outlier under this simulation setting, i.e., when the majority rule does not hold and locally invalid IV exists. When no outlier is identified, MR-PRESSO uses all candidate SNPs. Thus, the standard error of MR-PRESSO under setting (**S4**) will be close to that of IVW when the sample sizes are large.

In Table S4, we report the following results: (1) the false discovery rate (FDR) that is defined by the expected proportion of invalid IVs in the set of SNPs selected by MR-SPI; and (2) the true positive proportion (TPP) that is defined by the proportion of valid IVs selected by MR-SPI in the set of true valid IVs. In our simulations, we select valid IVs by finding the maximum clique in the voting matrix. The TPP of MR-SPI is close to one under all the four settings (**S1**) - (**S4**), and the FDR of MR-SPI is close to 0 if locally invalid IV does not exist and the plurality rule condition holds. Under settings (**S3**) and (**S4**), MR-SPI might incorrectly select locally invalid IVs when the sample sizes are not large enough (e.g., 5,000). But when sample sizes increase, the FDR of MR-SPI approaches 0 even in the presence of locally invalid IVs. For example, the FDR is 0.005 under setting (**S4**) when the sample size is 80,000. Therefore, even when locally invalid

IV exists, MR-SPI can still correctly identify valid IVs if the sample sizes are large enough.

Our simulation studies demonstrate that MR-SPI performs better compared to the other competing MR methods under the plurality rule condition. When locally invalid IV does not exist, MR-SPI can select valid IVs correctly and provide nearly unbiased estimates of the causal effect, and the confidence interval of MR-SPI attains the nominal coverage level. In practice, we can perform sensitivity analysis of the causal effect estimate by changing the threshold in the voting step (see Online Methods and Supplementary Figure S13). If the causal effect estimate is sensitive to the choice of the threshold, then MR-SPI might suffer from finite-sample IV selection error. In such cases, the proposed robust confidence interval of MR-SPI is recommended for use.

S7 Length of the Robust Confidence Interval

From previous simulation results, we notice that the average length of MR-SPI's robust confidence interval constructed by Algorithm 2 is always longer than the default confidence interval calculated by equation (8). Therefore, we conduct a simulation study to explore in what case the length of the robust confidence interval would be closer to the default confidence interval. Specifically, we set the sample sizes $n_1 = n_2 = 10{,}000$ and the number of SNPs $p = 10$, and fix the IV strength to be $0.2 \cdot \mathbf{1}_{10}^{\intercal}$. We set $\boldsymbol{\pi} = \tau \cdot (\mathbf{0}_4, \mathbf{1}_3, -\mathbf{1}_3)^{\intercal}$, where $\tau \in [0, 0.3]$ is the violation strength to control the degree of violation of (A2) and (A3). We run 1000 replications in total and present the ratio of average lengths of robust confidence interval to default confidence interval in Figure S14. As the violation strength τ increases, the ratio of average lengths of these two confidence intervals decreases at first and then increases to a stable level. When the violation strength τ is small, i.e., the degrees of violation of (A2) and (A3) is small, the default confidence interval would not be much affected even if invalid IVs are incorrectly selected. When the violation strength τ is large, the invalid IVs are more likely to be detected by MR-SPI, and thus the default confidence interval would also be less influenced by the invalid IVs. Compared to the default confidence interval, the robust confidence interval is less affected by the violation strength τ, indicating its robustness to the IV selection error. The minimum of this ratio is around 1.5, which is the price to pay for robustness in this numerical example.

S8 Simulation Studies under the Distributional Assumption of MR-Mix or the Contamination Mixture Method

In this section, we conduct more simulations to compare the performance of MR-SPI and MRMix or the contamination mixture method under the corresponding distributional assumption. MRMix assumes the true SNP-exposure associations γ_j and the degree of violation of assumptions (A2) and (A3) π_j follow the following bivariate normal mixture distribution:

$$\begin{pmatrix}\gamma_j\\ \pi_j\end{pmatrix} \sim p_1 \begin{pmatrix}N(0,\sigma_\gamma^2)\\ \delta_0\end{pmatrix} + p_2 N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_\gamma^2 & \sigma_{\gamma,\pi}\\ \sigma_{\gamma,\pi} & \sigma_\pi^2\end{pmatrix}\right) + p_3 \begin{pmatrix}\delta_0\\ N(0,\sigma_\pi^2)\end{pmatrix} + p_4 \begin{pmatrix}\delta_0\\ \delta_0\end{pmatrix},$$

where p_1, p_2, p_3, p_4 are mixture probabilities of different components and $p_1 + p_2 + p_3 + p_4 = 1$, and δ_0 is the probability mass at 0. In this simulation, we set sample sizes $n_1 = n_2 = 50,000$, the number of candidate SNPs $M = 5,000$, $(p_1, p_2, p_3, p_4) = (0.01, 0.01, 0.01, 0.97)$, $\sigma_\gamma^2 = \sigma_\pi^2 = 5 \times 10^{-5}$, $\sigma_{\gamma,\pi} = 2.5 \times 10^{-5}$, and the true causal effect $\beta = 0.5$. The bias of MRMix is generally small (around 0.004), and the coverage probability attains the 95% level. In comparison, our proposed MR-SPI has relatively large bias (0.185), as $\{\pi_j\}_{j=1}^M$ are generated from a normal mixture model, and thus the absolute value of π_j can be small enough so that SNP j is locally invalid and incorrectly selected into analysis by MR-SPI. However, the robust confidence interval of MR-SPI still attains the norminal level, indicating the robustness of MR-SPI to the finite-sample IV selection error.

The contamination mixture method assumes that $\widehat{\beta}_j$, the ratio estimate of SNP j, follows a normal distribution around the true causal effect β if SNP j is a valid IV, and follows another normal distribution around an unknown center $\beta_{F,j}$ if SNP j is an invalid IV. Furthermore, it also assume that $\beta_{F,j}$ follows a normal distribution centered at 0 with variance equal to ψ. Let ζ_j be the indicator of whether SNP j is a valid IV, i.e., $\zeta_j = 1$ if SNP j is a valid IV; otherwise $\zeta_j = 0$. Then, the distribution of the ratio estimate $\widehat{\beta}_j$ conditional on ζ_j is

$$\widehat{\beta}_j | \zeta_j \sim \zeta_j N(\beta, \mathrm{SE}(\widehat{\beta}_j)^2) + (1 - \zeta_j) N(0, \psi + \mathrm{SE}(\widehat{\beta}_j)^2).$$

In this simulation, we consider the following data generating process:

$$D_i = \sum_{j=1}^{n} \gamma_j Z_{i,j} + U_i + \epsilon_{X,i},$$
$$Y_i = \beta X_i + \sum_{j=1}^{n} \pi_j Z_{i,j} + U_i + \epsilon_{Y,i}.$$

We set sizes $n_1 = n_2 = 100{,}000$, the number of candidate SNPs $M = 20$, the number of valid IVs $V = 8$, the true causal effect $\beta = 0.5$. The SNP-exposure association $\{\gamma_j\}_{j=1}^{M}$ are generated independently from a uniform distribution $U(0.05, 0.2)$. We set the first $V = 8$ SNPs as valid IVs with $\pi_j = 0$. For invalid IVs, the degrees of violation of assumptions (A2) and (A3) $\{\pi_j\}_{j=V+1}^{M}$ are generated independently from a normal distribution $N(0, 1)$. The unmeasured confounding U_i is generated from $N(0, 1)$. The error terms $\epsilon_{X,i}$ and $\epsilon_{Y,i}$ are generated independently from $N(0, 1)$. Under this simulation setting, both the contamination mixture method and our proposed MR-SPI have small bias. Besides, both the contamination mixture and the robust confidence interval of MR-SPI can attain the 95% coverage level.

S9 Evaluating the performance of MR-SPI using two benchmark datasets

In this section, we apply the proposed MR-SPI method to two benchmark datasets to evaluate its performance. These two datasets serve as the benchmark because the exposure and the outcome are the same trait in each dataset, and thus the horizontal pleiotropic effects are expected to be zero. We first apply MR-SPI to the dataset in which both the exposure and the outcome are coronary artery disease (CAD), and we refer to it as the CAD-CAD dataset. Since both the exposure and the outcome are CAD, the causal effect is expected to be one. The exposure data come from the Coronary Artery Disease (C4D) Genetics Consortium[S9], and the outcome data come from the Coronary ARtery DIsease Genome-wide Replication and Meta-analysis (CARDIoGRAM) consortium[S10]. We first perform linkage disequilibrium (LD)-based clumping to the SNPs in the exposure data using the software Plink[S11], with $r^2 < 0.01$ to obtain independent genetic instruments, and then use 1×10^{-6} as the p-value threshold to select relevant instruments. In total, five relevant instruments are included for downstream MR analysis. We compare MR-SPI to the other eight competing MR methods including IVW, MR-Egger, MR-RAPS, MR-PRESSO,

the weighted median method, the mode-based estimation, MRMix and the contamination mixture method. The causal effect estimates and the corresponding 95% confidence intervals using those methods are presented in Figure S2(a). Generally, the confidence intervals of MR-SPI, IVW and MR-Egger all cover one, but MR-SPI provides the shortest confidence interval. In addition, none of the relevant IVs is excluded in the voting step, which is in line with the expectation that horizontal pleiotropy should not exist in this dataset.

Next, we apply MR-SPI to the dataset where the exposure and the outcome data are the body mass index (BMI) GWAS summary-level statistics for physically active men and women respectively, and we refer to this dataset as the BMI-BMI dataset. In this dataset, both the exposure and the outcome data come from the GIANT consortium[S12]. After LD clumping and filtering SNPs using the same setup as in the CAD-CAD dataset, 64 candidate SNPs are selected as relevant IVs and none of them is detected to be invalid by MR-SPI. The point estimates of the causal effect and corresponding 95% confidence intervals using MR-SPI and the other competing methods are shown in Figure S2(b). Overall, all the MR methods (except MR-Egger) provide causal effect estimates that are below one. As discussed in previous studies[S12], some significant loci of BMI might exhibit heterogeneity in genetic effects between men and women. Therefore, the "true" effect might not be equal to one in this dataset due to the difference in the genetic architecture of BMI between men and women.

S10 Learning causal relationships of 146 exposure-outcome pairs

In this section, we examine the causal relationships between complex traits and diseases from four categories including ischemic stroke, cholesterol levels, heart diseases, and coronavirus disease 2019 (COVID-19) related traits. Since MR-SPI requires that the GWAS summary statistics of the exposure and the outcome come from two non-overlapping samples, we exclude the trait pairs whose exposure and outcome are in the same consortium. In addition, we also exclude trait pairs whose exposure and outcome are two similar phenotypes (for example, heart failure and coronary artery disease), and we finally obtain 146 pairwise exposure-outcome combinations. All the GWAS summary statistics used for MR analysis are publicly available with more detailed description of each dataset given in Supplementary Table S5.

We first perform LD clumping using the software Plink[S11] to obtain independent SNPs with

$r^2 < 0.01$, and then use 1×10^{-6} as the p-value threshold to select relevant IVs that are associated with each exposure trait. Among the 146 exposure-outcome pairs, MR-SPI detects invalid IVs for 16 exposure-outcome pairs. For example, MR-SPI detects one invalid SNP (rs616154, marked by red triangle) in the causal relationship from cardioembolic stroke (CES) to SARS-CoV-2 infection, as illustrated in Figure 3(a). SNP rs616154 is identified to be invalid since its ratio estimate of the causal effect is 0.525, which is far away from other SNPs' ratio estimates and thus no other relevant SNP votes for it to be a valid IV. We search for the human phenotypes that are strongly associated with SNP rs616154 using the PhenoScanner tool[S13, S14], and find that this SNP is also associated with the Interleukin-6 (IL-6) levels which is a potential biomarker of COVID-19 progression[S15], indicating that SNP rs616154 might exhibit horizontal pleiotropy in the relationship of cardioembolic stroke on SARS-CoV-2 infection and might be an invalid IV. After excluding SNP rs616154, the point estimate of the causal effect by MR-SPI (represented by the slope of the green solid line in Figure 3(a)) is nearly zero, suggesting that cardioembolic stroke might not be a risk factor for SARS-CoV-2 infection. The causal effect estimate of MR-PRESSO (represented by the slope of the blue dashed line in Figure 3(a)) is also close to zero, as MR-PRESSO also detects SNP rs616154 as an outlier and excludes it from downstream MR analysis. However, IVW and MR-RAPS include SNP rs616154 in the MR analysis, and hence their causal effect estimates (represented by the slopes of the black and orange dashed line in Figure 3(a), respectively) might be biased.

Figure 3(a) illustrates that the inclusion of invalid IVs might lead to misleading scientific findings. MR-SPI selects valid IVs for downstream analysis and can provide reliable causal inference. After excluding those potentially invalid IVs, MR-SPI identifies 27 significant associations after Bonferroni correction for multiple comparison[S16], with results summarized in Figure 3(c). We also apply the other eight competing MR methods including IVW, MR-Egger, MR-RAPS, MR-PRESSO, the weighted median method, the mode-based estimation, MRMix and the contamination mixture method to infer causal relationships among these exposure-outcome pairs, and the results are presented in Supplementary Figures S3-S10. Among the 146 exposure-outcome pairs, MR-SPI detects invalid IVs for 16 exposure-outcome pairs. Some of our findings are in line with previous studies, for example, an increase in LDL level might be associated with increased risks of CAD and HF[S17, S18]. In addition, MR-SPI also detects significant associations that cannot be discovered by other competing MR methods. For example, MR-SPI suggests that SARS-CoV-2

infection might be a risk factor for HF ($\widehat{\beta} = 0.14, p\text{-value} = 1.43 \times 10^{-4}$), which cannot be discovered by the other competing MR methods considered in this paper. Our finding is consistent with a former study that reported a significant increase in the risk of developing acute heart failure in patients with confirmed COVID-19 infection[S19].

To demonstrate the similarities and differences in the results of MR-SPI and other MR methods, we plot the Venn diagrams to show the number of significant associations that are either shared or uniquely detected by these methods. We present the Venn diagram of the significant pairs using MR-SPI, the mode-based estimation, MRMix and the contamination mixture method in Figure 3(b), as these four methods are all based on the plurality rule condition. More Venn diagrams that compare MR-SPI and the other competing MR methods can be found in Supplementary Figures S11 and S12. From Figure 3(b), we can see that MR-SPI detects more significant associations than the mode-based estimation, MRMix and the contamination mixture method among these 146 exposure-outcome pairs. Indeed, these three competing MR methods fail to discover some causal relationships that have been supported from previous literature. For example, the mode-based estimation, MRMix and the contamination mixture method fail to detect that an increased HDL level might be associated with a decreased risk of CAD, which is identified by MR-SPI ($\widehat{\beta} = -0.18, p\text{-value} = 3.73 \times 10^{-17}$) and has been supported with evidence by previous epidemiological studies[S20, S21]. Supplementary Figure S11 compares the significant relationships detected by MR-SPI and three MR methods that require InSIDE assumption (IVW, MR-Egger and MR-RAPS). MR-RAPS detects 17 significant associations that are not identified by MR-SPI, of which some associations might be spurious. For example, MR-RAPS suggests significant associations of AIS on low-density lipoprotein (LDL), high-density lipoprotein (HDL) and total cholesterol (TC) level. However, the reverse association, i.e., cholesterol level on the risk of stroke, has been reported in previous epidemiological studies[S17, S22]. In Supplementary Figure S12, we compare MR-SPI with MR-PRESSO and the weighted median method that both require the majority rule condition. MR-PRESSO and the weighted median detect 14 and 10 significant associations, respectively, all of which are also identified by MR-SPI. Besides, MR-SPI identifies 11 more significant associations, most of which are in line with previous epidemiological studies, for example, HF might be a risk factor for ischemic stroke[S23, S24].

To deal with the issue of potential IV selection error in finite samples, we also construct robust confidence intervals for these exposure-outcome pairs using MR-SPI Algorithm 2 in Online

Methods. MR-SPI discovers four significant associations whose robust confidence intervals do not include zero (CES on HF, LDL on CAD, TC on CAD, and atrial fibrillation (AF) on CES), and we compare the robust confidence intervals (represented by dark blue bars) with the default confidence intervals (assuming no finite-sample IV selection error) calculated by equation (8) in Online Methods (represented by light blue bars) in Figure 3(d). As shown in Figure 3(d), the robust confidence intervals that allow for finite-sample IV selection error are longer than the default confidence intervals that do not consider finite-sample IV selection error, indicating that locally invalid IVs might exist and might be incorrectly selected as valid IVs in these datasets. Therefore, we suggest using the robust confidence intervals for these four relationships to provide more reliable causal findings.

S11 Supplementary Tables

Table S1: Bias of methods under different scenarios and sample sizes over 1,000 replications. The true causal effect $\beta = 1$.

Scenario	Method	n				
		5,000	10,000	20,000	40,000	80,000
S1 (majority rule condition without locally invalid IVs)	MR-SPI	-0.003	0.000	0.000	0.000	0.000
	IVW	-0.398	-0.397	-0.396	-0.396	-0.396
	MR-RAPS	-0.316	-0.315	-0.316	-0.316	-0.317
	MR-PRESSO	-0.248	-0.389	-0.397	-0.396	-0.395
	Weighted Median	-0.124	-0.084	-0.062	-0.043	-0.030
	Mode-based Estimation	-0.034	-0.019	-0.020	-0.016	-0.025
	MRMix	-0.040	-0.009	-0.007	0.000	0.000
	Contamination Mixture	0.004	0.003	0.001	0.001	0.000
S2 (plurality rule condition without locally invalid IVs)	MR-SPI	-0.003	-0.002	-0.002	0.000	0.000
	IVW	-0.691	-0.693	-0.690	-0.691	-0.690
	MR-RAPS	-0.351	-0.354	-0.352	-0.355	-0.354
	MR-PRESSO	-0.835	-0.727	-0.689	-0.690	-0.690
	Weighted Median	-0.524	-0.509	-0.488	-0.488	-0.475
	Mode-based Estimation	0.002	0.004	0.004	0.001	0.000
	MRMix	-0.055	-0.021	-0.007	-0.003	-0.001
	Contamination Mixture	-0.010	-0.016	-0.021	-0.014	-0.014
S3 (majority rule condition with locally invalid IVs)	MR-SPI	-0.048	-0.034	-0.015	0.000	0.000
	IVW	-0.251	-0.249	-0.248	-0.247	-0.248
	MR-RAPS	-0.195	-0.195	-0.195	-0.194	-0.195
	MR-PRESSO	-0.129	-0.172	-0.238	-0.249	-0.249
	Weighted Median	-0.109	-0.085	-0.062	-0.044	-0.031
	Mode-based Estimation	-0.035	-0.012	-0.001	0.000	0.000
	MRMix	-0.053	-0.043	-0.023	-0.001	0.000
	Contamination Mixture	-0.028	-0.014	-0.003	0.001	0.000
S4 (plurality rule condition with locally invalid IVs)	MR-SPI	-0.014	-0.004	-0.002	0.001	-0.001
	IVW	-0.195	-0.190	-0.191	-0.190	-0.189
	MR-RAPS	-0.050	-0.050	-0.051	-0.050	-0.050
	MR-PRESSO	-0.083	-0.106	-0.156	-0.225	-0.225
	Weighted Median	-0.083	-0.058	-0.043	-0.030	-0.022
	Mode-based Estimation	0.006	0.000	0.000	0.000	0.001
	MRMix	-0.057	-0.054	-0.057	-0.054	-0.033
	Contamination Mixture	0.008	0.012	0.003	0.000	-0.001

Table S2: Empirical coverage of methods under different scenarios and sample sizes over 1,000 replications.

Scenario	Method	n				
		5,000	10,000	20,000	40,000	80,000
S1 (majority rule condition without locally invalid IVs)	MR-SPI	0.949	0.947	0.949	0.946	0.946
	MR-SPI (robust CI)	0.998	0.999	0.999	1.000	0.999
	IVW	0.068	0.022	0.006	0.000	0.000
	MR-RAPS	0.000	0.000	0.000	0.000	0.000
	MR-PRESSO	0.463	0.066	0.003	0.000	0.000
	Weighted Median	0.590	0.633	0.600	0.632	0.638
	Mode-based Estimation	0.966	0.966	0.959	0.967	0.944
	MRMix	0.892	0.934	0.977	0.915	0.903
	Contamination Mixture	0.867	0.839	0.842	0.828	0.810
S2 (plurality rule condition without locally invalid IVs)	MR-SPI	0.962	0.953	0.955	0.953	0.942
	MR-SPI (robust CI)	0.993	0.995	0.993	0.993	0.997
	IVW	0.284	0.229	0.156	0.070	0.025
	MR-RAPS	0.000	0.000	0.000	0.000	0.000
	MR-PRESSO	0.125	0.199	0.171	0.072	0.024
	Weighted Median	0.027	0.000	0.000	0.000	0.000
	Mode-based Estimation	0.980	0.964	0.957	0.953	0.936
	MRMix	0.929	0.925	0.964	0.900	0.854
	Contamination Mixture	0.676	0.659	0.669	0.674	0.670
S3 (majority rule condition with locally invalid IVs)	MR-SPI	0.780	0.777	0.818	0.930	0.948
	MR-SPI (robust CI)	0.997	0.997	0.997	0.999	0.998
	IVW	0.571	0.584	0.601	0.635	0.635
	MR-RAPS	0.009	0.000	0.000	0.000	0.000
	MR-PRESSO	0.339	0.088	0.015	0.010	0.015
	Weighted Median	0.581	0.544	0.579	0.639	0.643
	Mode-based Estimation	0.976	0.986	0.992	0.990	0.977
	MRMix	0.807	0.845	0.920	0.851	0.913
	Contamination Mixture	0.817	0.842	0.840	0.844	0.827
S4 (plurality rule condition with locally invalid IVs)	MR-SPI	0.902	0.845	0.797	0.885	0.954
	MR-SPI (robust CI)	0.998	0.997	0.993	0.993	0.995
	IVW	1.000	1.000	1.000	1.000	1.000
	MR-RAPS	0.693	0.554	0.341	0.137	0.012
	MR-PRESSO	0.736	0.596	0.611	0.911	0.998
	Weighted Median	0.794	0.805	0.815	0.828	0.831
	Mode-based Estimation	0.979	0.948	0.949	0.957	0.965
	MRMix	0.907	0.820	0.860	0.868	0.836
	Contamination Mixture	0.811	0.800	0.763	0.764	0.754

Table S3: Average length of 95% confidence intervals of methods under different scenarios and sample sizes over 1,000 replications.

Scenario	Method	n				
		5,000	10,000	20,000	40,000	80,000
S1 (majority rule condition without locally invalid IVs)	MR-SPI	0.230	0.161	0.115	0.081	0.057
	MR-SPI (robust CI)	0.467	0.326	0.229	0.161	0.114
	IVW	0.653	0.648	0.642	0.641	0.640
	MR-RAPS	0.144	0.102	0.072	0.051	0.036
	MR-PRESSO	0.536	0.644	0.642	0.641	0.639
	Weighted Median	0.279	0.201	0.144	0.103	0.073
	Mode-based Estimation	0.404	0.237	0.162	0.115	0.081
	MRMix	0.261	0.177	0.131	0.105	0.091
	Contamination Mixture	0.231	0.156	0.111	0.079	0.054
S2 (plurality rule condition without locally invalid IVs)	MR-SPI	0.296	0.202	0.142	0.100	0.070
	MR-SPI (robust CI)	0.508	0.360	0.249	0.177	0.125
	IVW	1.324	1.325	1.323	1.321	1.321
	MR-RAPS	0.132	0.093	0.066	0.046	0.033
	MR-PRESSO	1.126	1.315	1.322	1.323	1.322
	Weighted Median	0.312	0.220	0.156	0.112	0.080
	Mode-based Estimation	0.345	0.207	0.140	0.097	0.068
	MRMix	2.119	0.256	0.170	0.133	0.101
	Contamination Mixture	0.460	0.431	0.391	0.357	0.329
S3 (majority rule condition with locally invalid IVs)	MR-SPI	0.221	0.169	0.125	0.085	0.057
	MR-SPI (robust CI)	0.471	0.348	0.243	0.166	0.115
	IVW	0.522	0.514	0.509	0.507	0.505
	MR-RAPS	0.154	0.108	0.077	0.054	0.038
	MR-PRESSO	0.212	0.205	0.151	0.098	0.073
	Weighted Median	0.246	0.183	0.138	0.103	0.074
	Mode-based Estimation	0.487	0.348	0.221	0.129	0.081
	MRMix	0.298	0.254	0.238	0.127	0.111
	Contamination Mixture	0.232	0.172	0.112	0.073	0.047
S4 (plurality rule condition with locally invalid IVs)	MR-SPI	0.277	0.216	0.167	0.113	0.070
	MR-SPI (robust CI)	0.542	0.408	0.290	0.186	0.124
	IVW	0.816	0.806	0.804	0.803	0.803
	MR-RAPS	0.162	0.115	0.081	0.057	0.040
	MR-PRESSO	0.261	0.245	0.391	0.742	0.802
	Weighted Median	0.290	0.213	0.156	0.113	0.080
	Mode-based Estimation	0.444	0.306	0.213	0.138	0.088
	MRMix	0.476	0.316	0.247	0.134	0.098
	Contamination Mixture	0.323	0.254	0.204	0.173	0.146

Table S4: The FDR and TPP of valid IV selection by MR-SPI under different settings and sample sizes. The FDR is close to 0 in the absence of locally invalid IV or when the sample sizes are large. The TPP is close to 1 under all settings.

Sample Sizes	FDR				TPP			
	S1	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
5,000	0.000	0.018	0.225	0.298	0.996	0.998	0.987	0.996
10,000	0.000	0.009	0.198	0.251	0.998	1.000	0.981	0.997
20,000	0.000	0.005	0.124	0.195	0.999	1.000	0.987	0.999
40,000	0.000	0.006	0.022	0.072	1.000	1.000	0.997	0.999
80,000	0.000	0.005	0.000	0.005	0.999	0.999	1.000	1.000

Table S5: Description of GWAS datasets included in our study.

Category	Phenotype	Abbriviation	Sample Size	Year	PMID	Consortium
Ischemic stroke	Any ischemic stroke	AIS	440,328	2018	29531354	MEGASTROKE
	Large-artery atherosclerotic stroke	LAS	150,765	2018	29531354	MEGASTROKE
	Small vessel stroke	SVS	198,048	2018	29531354	MEGASTROKE
	Cardioembolic stroke	CES	211,763	2018	29531354	MEGASTROKE
Cholesterol Level	Low-density lipoprotein	LDL	173,082	2013	24097068	GLGC
	High-density lipoprotein	HDL	187,167	2013	24097068	GLGC
	Total cholesterol	TC	187,365	2013	24097068	GLGC
	Triglycerides	TG	177,861	2013	24097068	GLGC
Heart disease	Heart failure	HF	977,323	2020	31919418	HERMES
	Coronary artery disease	CAD	141,217	2015	26343387	CARDIoGRAMplusC4D
	Atrial fibrillation	AF	1,030,836	2018	30061737	NA
COVID-19	Severe COVID-19	SEVERE	1,086,211	2022	NA	NA
	COVID-19 hospitalization	HOSPITAL	2,095,324	2022	NA	NA
	SARS-CoV2 infection	INFECTION	2,597,856	2022	NA	NA

Table S6: Detailed information of the SNP IVs for the 7 proteins associated with the risk of AD detected by MR-SPI. CHR: chromosome; POS: genomic position; BETA: effect size estimates in UK Biobank proteomics data; SE: standard error of BETA; A1: effect allele; A2: other allele.

Protein ID	SNP	CHR	POS	cis/trans	BETA	SE	A1	A2
CD33	rs62165726	2	134208991	trans	-0.101	0.010	A	C
	rs56094988	3	98771605	trans	0.073	0.007	C	G
	rs7865362	9	33117967	trans	0.029	0.004	T	C
	rs10823374	10	69441960	trans	0.049	0.004	G	A
	rs139130389	11	72139110	trans	-0.051	0.007	CTA	C
	rs74612335	11	126368738	trans	-0.040	0.006	C	T
	rs186021206	17	7166093	trans	0.633	0.026	A	G
	rs2455069	19	51225385	cis	1.170	0.008	G	A
CD55	rs1583090	1	207262438	cis	-0.481	0.009	C	T
	rs1260326	2	27508073	trans	0.049	0.007	C	T
	rs13107325	4	102267552	trans	0.136	0.014	T	C
	-	7	7216532	trans	0.061	0.007	C	CT
	rs2978024	8	133536866	trans	0.078	0.008	A	C
	rs774510679	10	63304076	trans	0.055	0.007	A	ACTTTGCC
	rs777834943	16	20355409	trans	-0.066	0.009	CT	C
	rs186021206	17	7166093	trans	1.068	0.049	A	G
	rs3814995	19	35851310	trans	0.059	0.008	T	C
EPHA1	rs1497406	1	16178825	trans	0.061	0.007	G	A
	rs61747728	1	179557079	trans	0.128	0.018	T	C
	rs75045569	7	143412115	cis	0.251	0.009	G	T
	rs10740131	10	63511728	trans	0.096	0.007	T	A
	rs28929474	14	94378610	trans	0.287	0.024	T	C
	rs34882080	16	20350119	trans	-0.060	0.009	G	A
	rs3814995	19	35851310	trans	0.058	0.007	T	C
PILRA	rs61747728	1	179557079	trans	0.100	0.015	T	C
	rs10085857	7	45216839	trans	0.189	0.008	T	C
	rs1859788	7	100374211	cis	0.994	0.008	G	A
	rs3184504	12	111446804	trans	-0.063	0.006	C	T
	rs186021206	17	7166093	trans	0.548	0.039	A	G
	rs55951667	17	76598422	trans	-0.040	0.006	T	C
	rs8077394	17	81284987	trans	-0.051	0.006	G	A
PILRB	rs10085857	7	45216839	trans	0.215	0.008	T	C
	rs1859788	7	100374211	cis	1.136	0.008	G	A
	rs7310615	12	111427245	trans	-0.056	0.005	G	C
	rs186021206	17	7166093	trans	0.465	0.035	A	G
	rs8077394	17	81284987	trans	-0.046	0.005	G	A
RET	rs6747755	2	100961996	trans	0.060	0.008	A	G
	rs6762552	3	98712363	trans	-0.082	0.008	T	C
	rs7856189	9	98878916	trans	-0.049	0.007	C	T
	rs189906184	10	38695562	trans	0.388	0.024	T	A
	rs2795507	10	42857446	cis	-0.442	0.009	C	T
	rs35887873	11	126349501	trans	0.085	0.008	C	T
	rs186021206	17	7166093	trans	0.454	0.048	A	G
	rs9952412	18	48939170	trans	0.047	0.007	C	A
TREM2	rs10919543	1	161538827	trans	0.052	0.007	G	A
	rs1410996	1	196727803	trans	-0.054	0.007	A	G
	rs531950386	6	32518763	trans	-0.067	0.007	T	G
	rs143332484	6	41161469	cis	-1.474	0.033	T	C
	rs913914	9	99495968	trans	-0.052	0.007	G	C
	rs7232	11	60173126	trans	0.397	0.007	A	T
	rs11057840	12	124831509	trans	0.100	0.010	C	A
	rs738409	22	43928847	trans	0.063	0.008	G	C

Table S7: Results of gene ontology (GO) enrichment analysis using genes encoding the 7 proteins identified by MR-SPI.

ID	Description	Source	Adjusted P-value
GO:0010562	positive regulation of phosphorus metabolic process	Biological Process	1.77E-03
GO:0019220	regulation of phosphate metabolic process	Biological Process	1.67E-02
GO:0030449	regulation of complement activation	Biological Process	2.26E-02
GO:0031399	regulation of protein modification process	Biological Process	2.64E-02
GO:0032691	negative regulation of interleukin-1 beta production	Biological Process	4.07E-02
GO:0033674	positive regulation of kinase activity	Biological Process	6.43E-03
GO:0043085	positive regulation of catalytic activity	Biological Process	8.18E-03
GO:0043549	regulation of kinase activity	Biological Process	4.51E-02
GO:0044093	positive regulation of molecular function	Biological Process	3.89E-02
GO:0045937	positive regulation of phosphate metabolic process	Biological Process	1.77E-03
GO:0051174	regulation of phosphorus metabolic process	Biological Process	1.67E-02
GO:0051347	positive regulation of transferase activity	Biological Process	1.41E-02
GO:0005886	plasma membrane	Cellular Component	5.43E-03
GO:0045121	membrane raft	Cellular Component	5.01E-03
GO:0071944	cell periphery	Cellular Component	9.60E-03
GO:0098857	membrane microdomain	Cellular Component	5.06E-03
GO:0004714	transmembrane receptor protein tyrosine kinase activity	Molecular Function	1.62E-02
GO:0019199	transmembrane receptor protein kinase activity	Molecular Function	2.62E-02
GO:0042287	MHC protein binding	Molecular Function	6.50E-03
GO:0042288	MHC class I protein binding	Molecular Function	1.28E-03

S12 Supplementary Figures

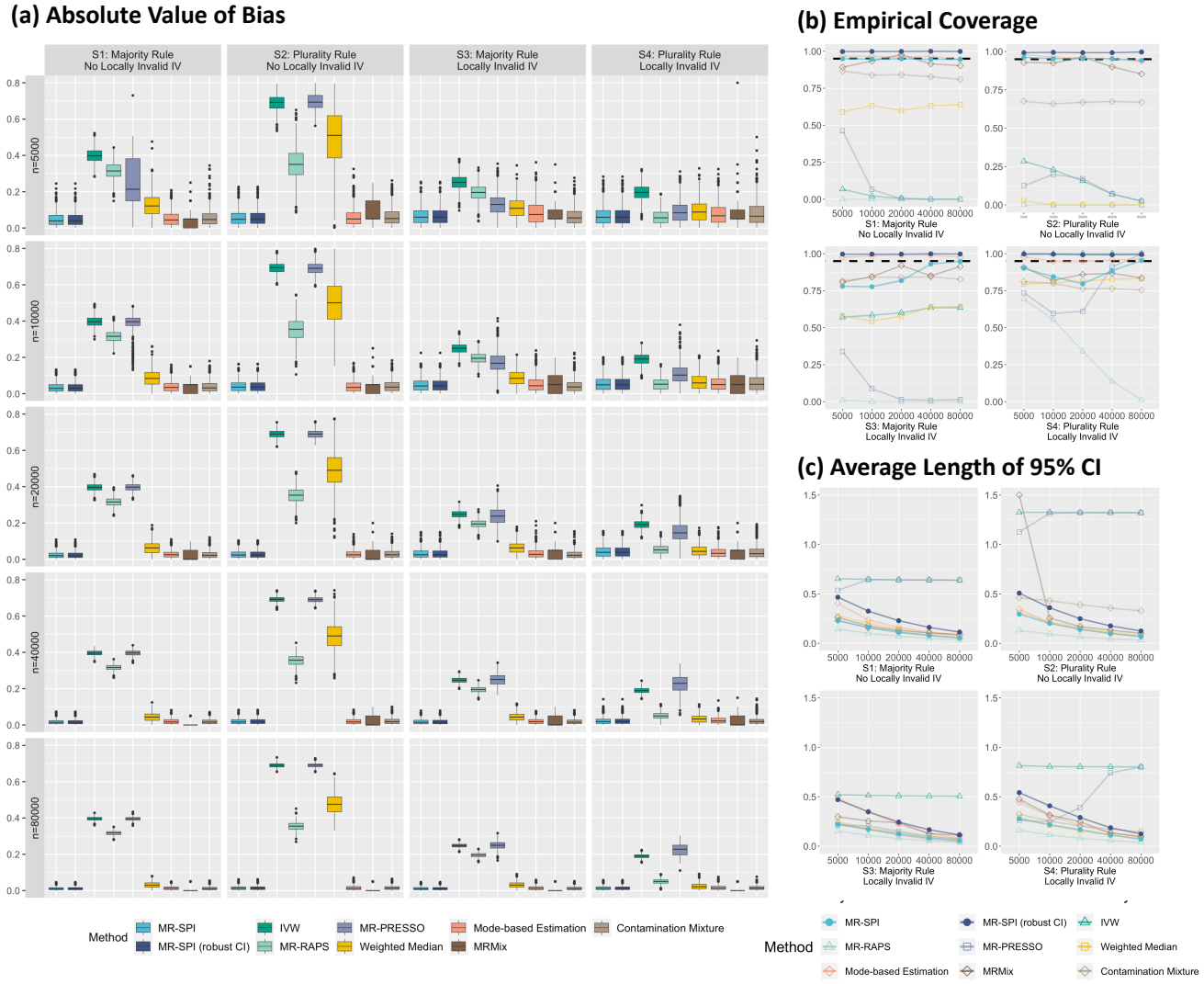

Figure S1: Comparison of MR-SPI to the other competing MR methods in simulated data under different scenarios and different values of sample sizes. (a) Boxplot of the absolute value of bias in causal effect estimates. (b) Empirical coverage of 95% confidence intervals. The black dashed line in (b) represents the nominal level (95%). (c) Average lengths of 95% confidence intervals.

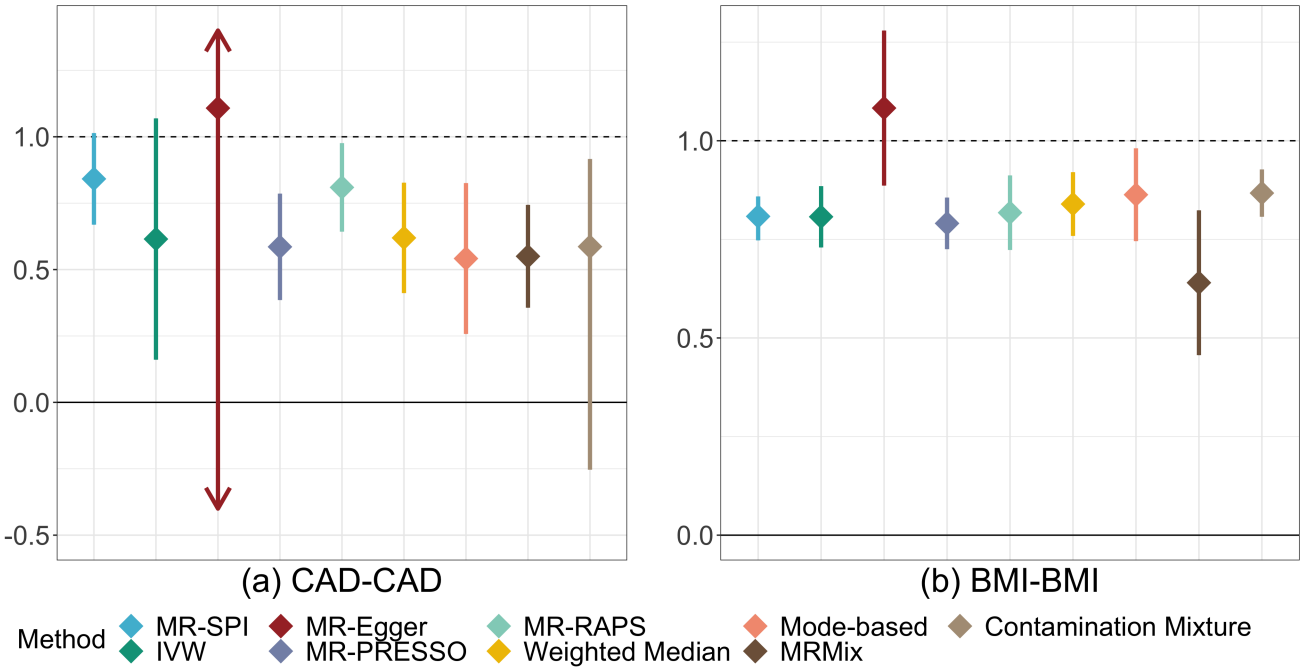

Figure S2: Point estimates and 95% confidence intervals for the causal effects of (a) CAD-CAD dataset and (b) BMI-BMI dataset using different MR methods. Confidence intervals are clipped to arrows if they exceed axis limits. CAD: coronary artery disease; BMI: body mass index.

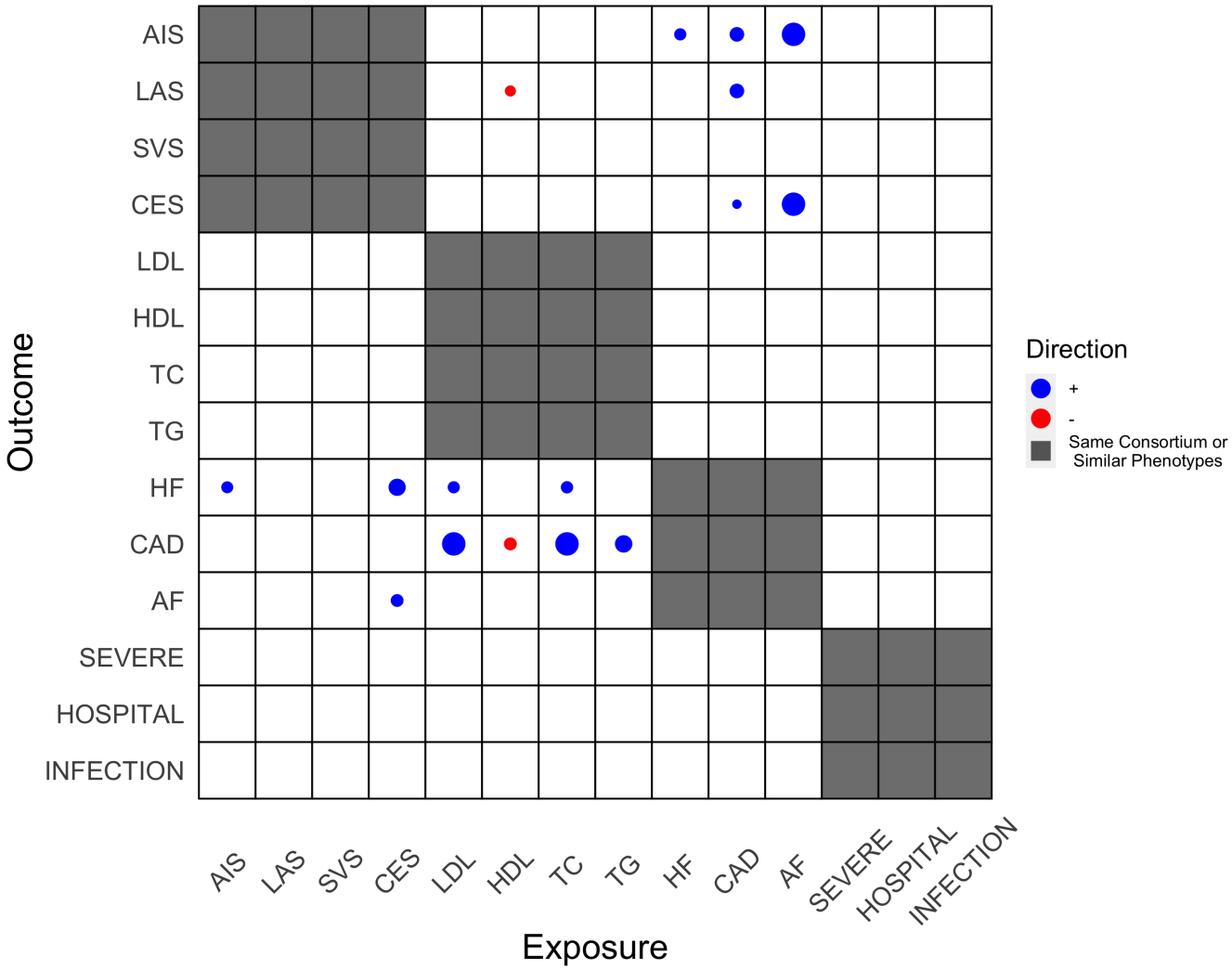

Figure S3: Estimates of causal effects using IVW. The significantly positive and negative causal relationships after Bonferroni correction are marked by blue filled circles and red filled circles, respectively. The radius of a circle is proportional to the $-\log_{10}(p\text{-value})$ of the corresponding exposure-outcome pair. Those pairs whose exposure and outcome come from the same consortium or are two similar phenotypes are marked as grey cells.

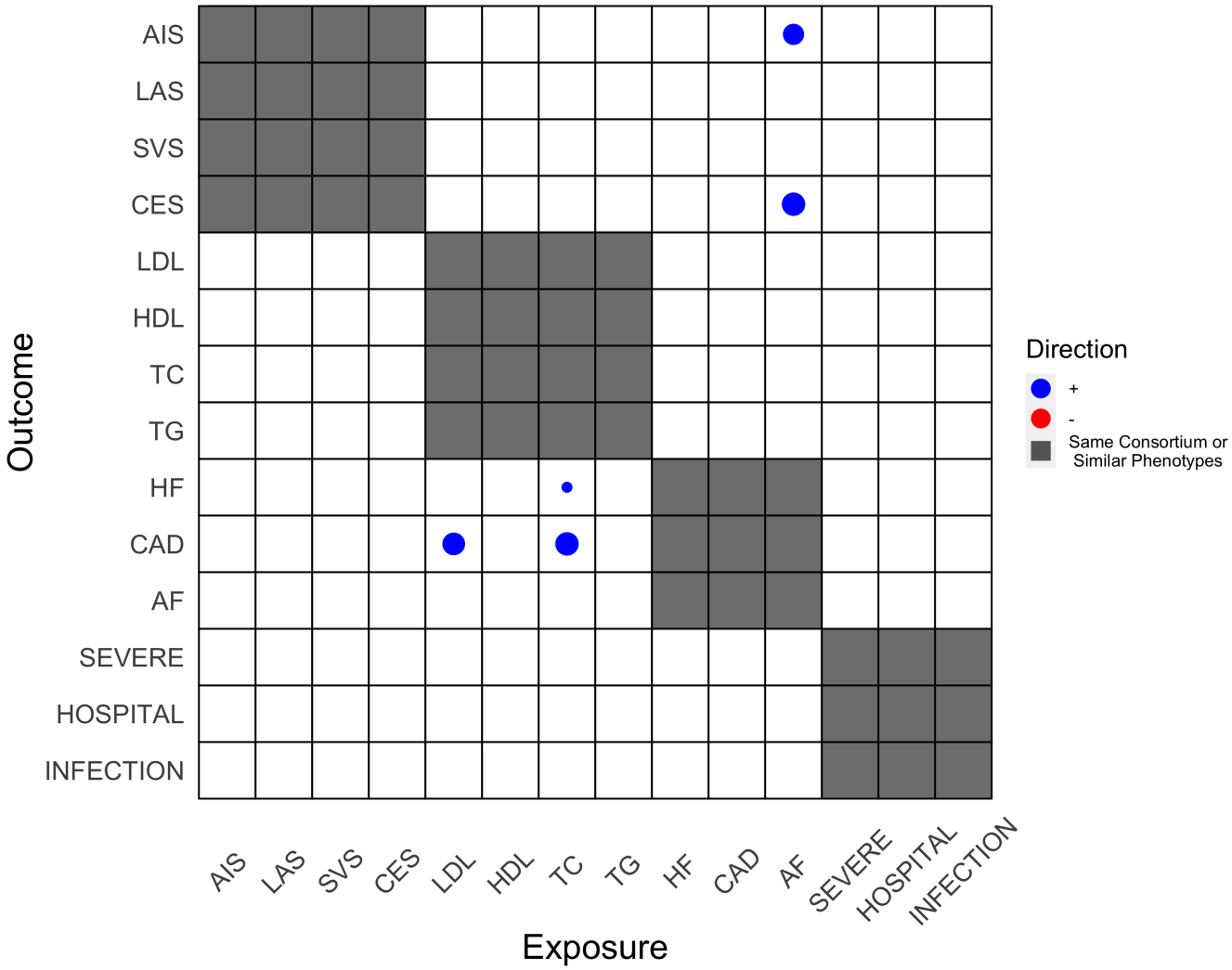

Figure S4: Estimates of causal effects using MR-Egger. The significantly positive and negative causal relationships after Bonferroni correction are marked by blue filled circles and red filled circles, respectively. The radius of a circle is proportional to the $-\log_{10}(p\text{-value})$ of the corresponding exposure-outcome pair. Those pairs whose exposure and outcome come from the same consortium or are two similar phenotypes are marked as grey cells.

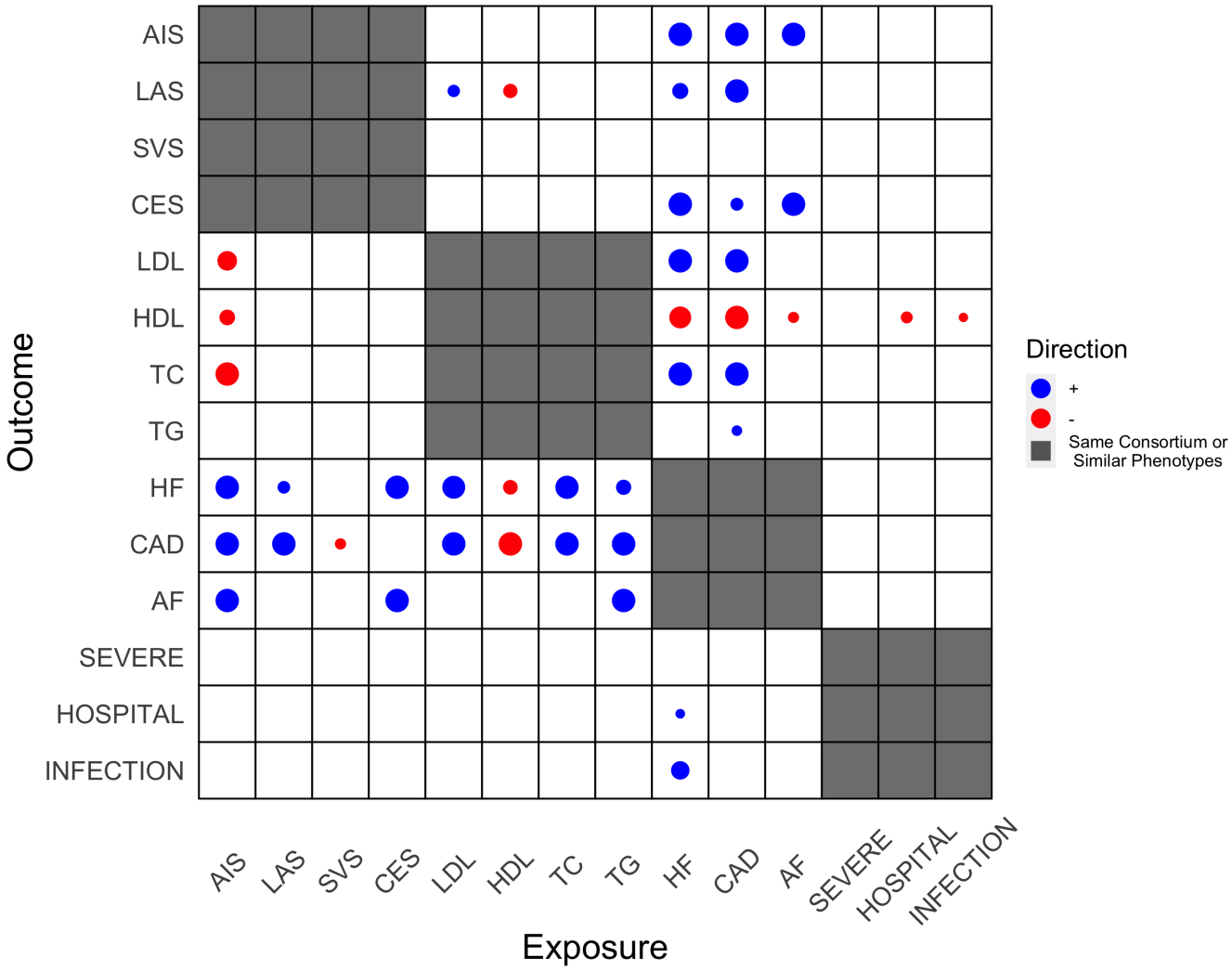

Figure S5: Estimates of causal effects using MR-RAPS. The significantly positive and negative causal relationships after Bonferroni correction are marked by blue filled circles and red filled circles, respectively. The radius of a circle is proportional to the $-\log_{10}(p\text{-value})$ of the corresponding exposure-outcome pair. Those pairs whose exposure and outcome come from the same consortium or are two similar phenotypes are marked as grey cells.

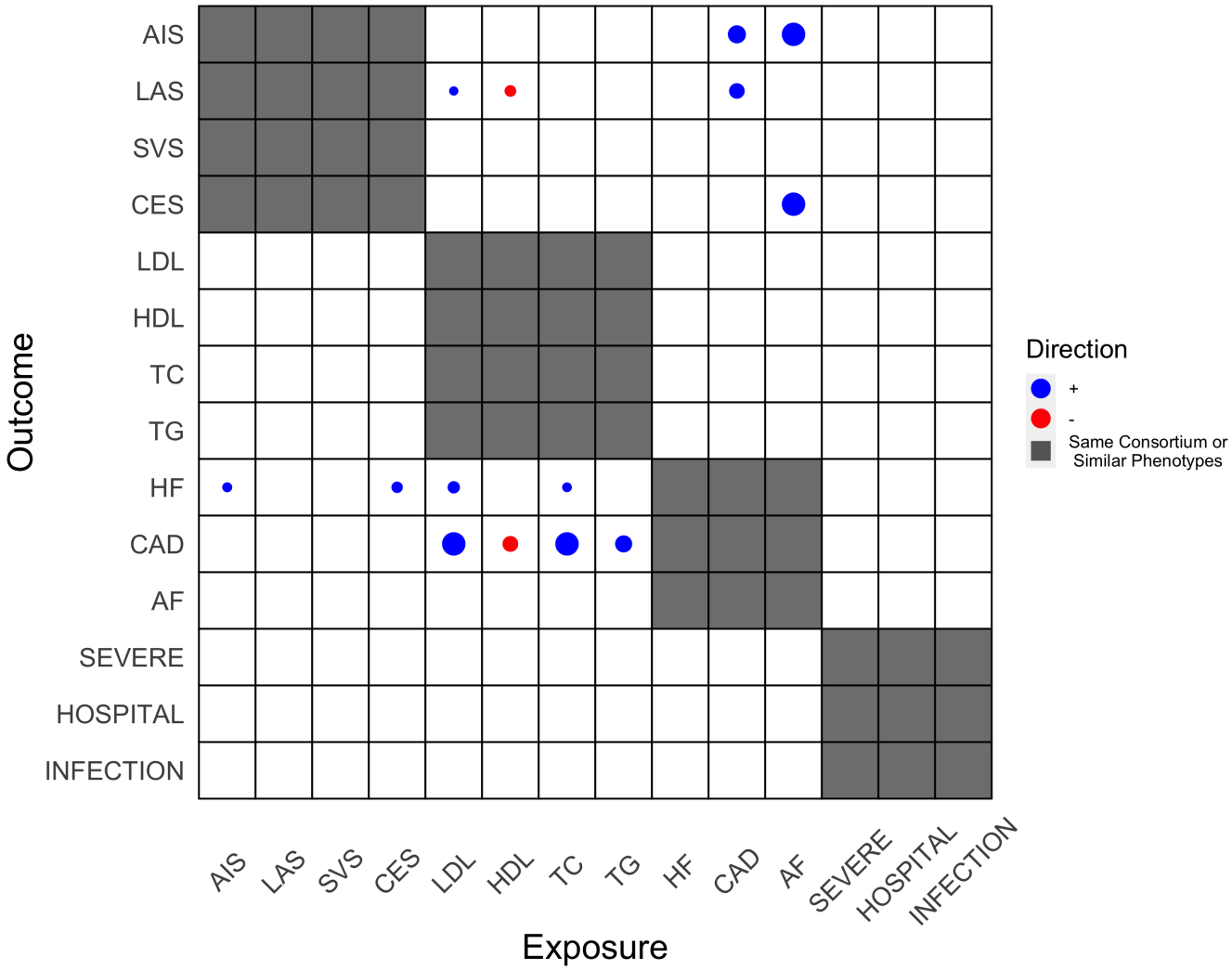

Figure S6: Estimates of causal effects using MR-PRESSO. The significantly positive and negative causal relationships after Bonferroni correction are marked by blue filled circles and red filled circles, respectively. The radius of a circle is proportional to the $-\log_{10}(p\text{-value})$ of the corresponding exposure-outcome pair. Those pairs whose exposure and outcome come from the same consortium or are two similar phenotypes are marked as grey cells.

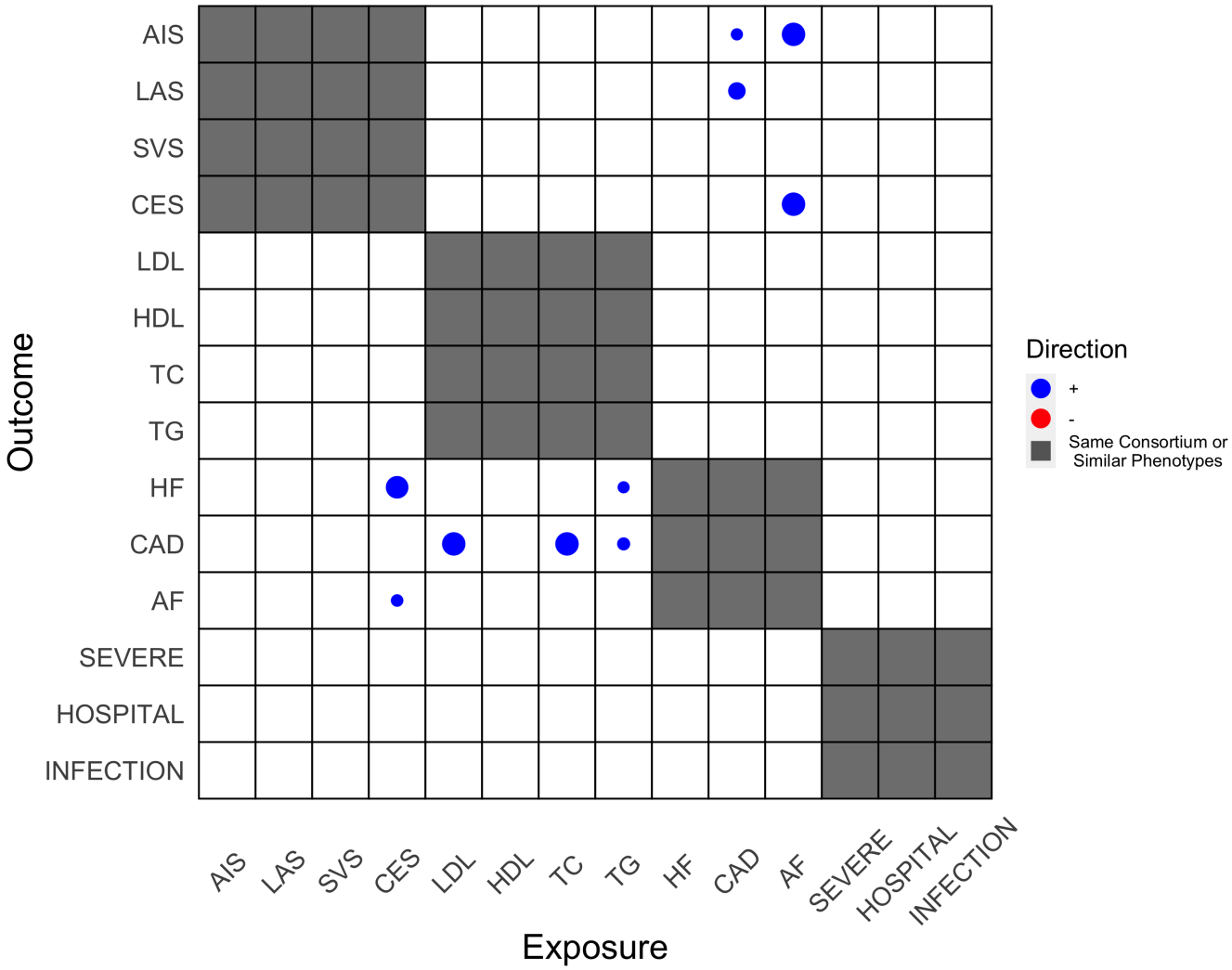

Figure S7: Estimates of causal effects using the weighted median method. The significantly positive and negative causal relationships after Bonferroni correction are marked by blue filled circles and red filled circles, respectively. The radius of a circle is proportional to the $-\log_{10}(p\text{-value})$ of the corresponding exposure-outcome pair. Those pairs whose exposure and outcome come from the same consortium or are two similar phenotypes are marked as grey cells.

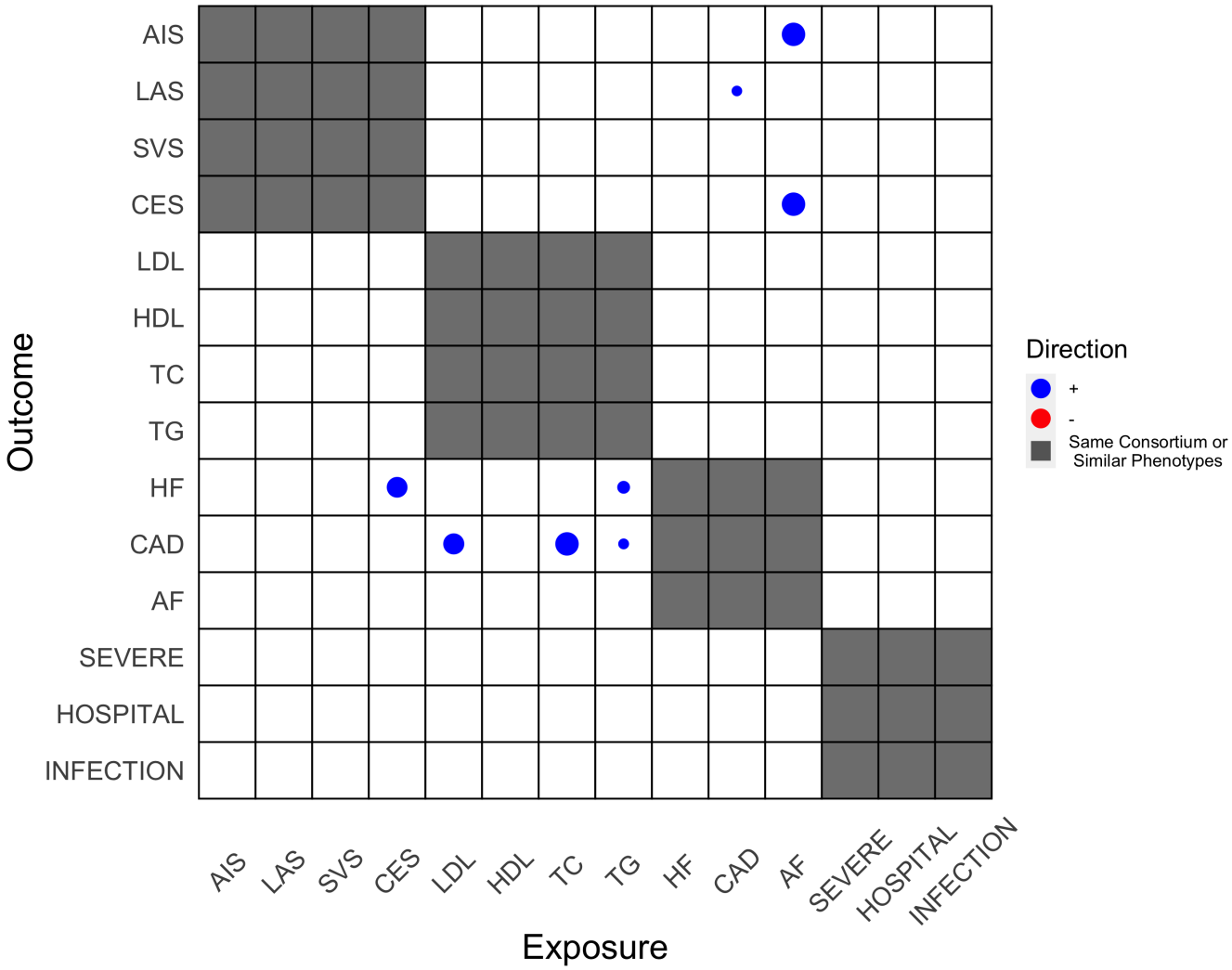

Figure S8: Estimates of causal effects using the mode-based estimation. The significantly positive and negative causal relationships after Bonferroni correction are marked by blue filled circles and red filled circles, respectively. The radius of a circle is proportional to the $-\log_{10}(p\text{-value})$ of the corresponding exposure-outcome pair. Those pairs whose exposure and outcome come from the same consortium or are two similar phenotypes are marked as grey cells.

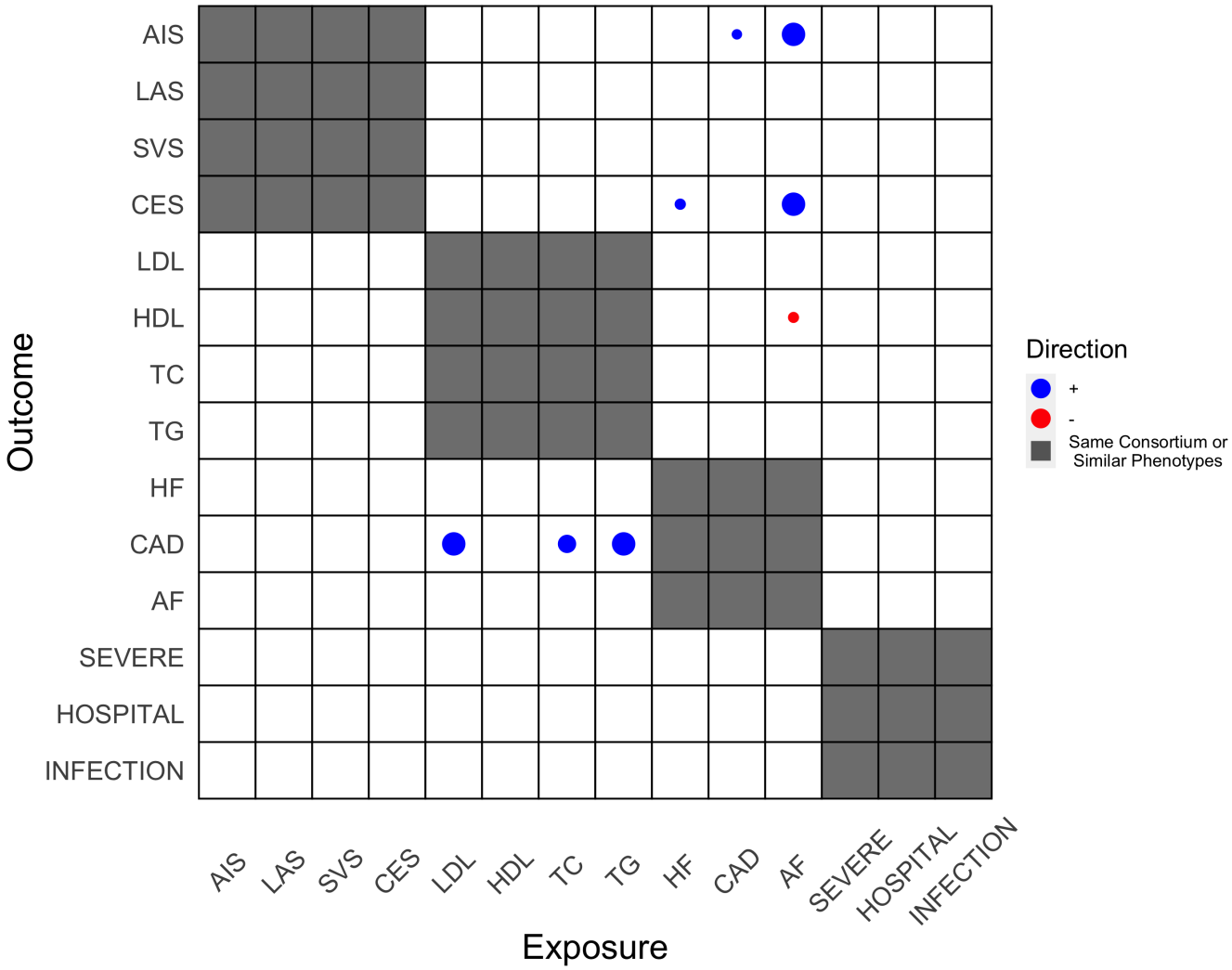

Figure S9: Estimates of causal effects using MRMix. The significantly positive and negative causal relationships after Bonferroni correction are marked by blue filled circles and red filled circles, respectively. The radius of a circle is proportional to the $-\log_{10}(p\text{-value})$ of the corresponding exposure-outcome pair. Those pairs whose exposure and outcome come from the same consortium or are two similar phenotypes are marked as grey cells.

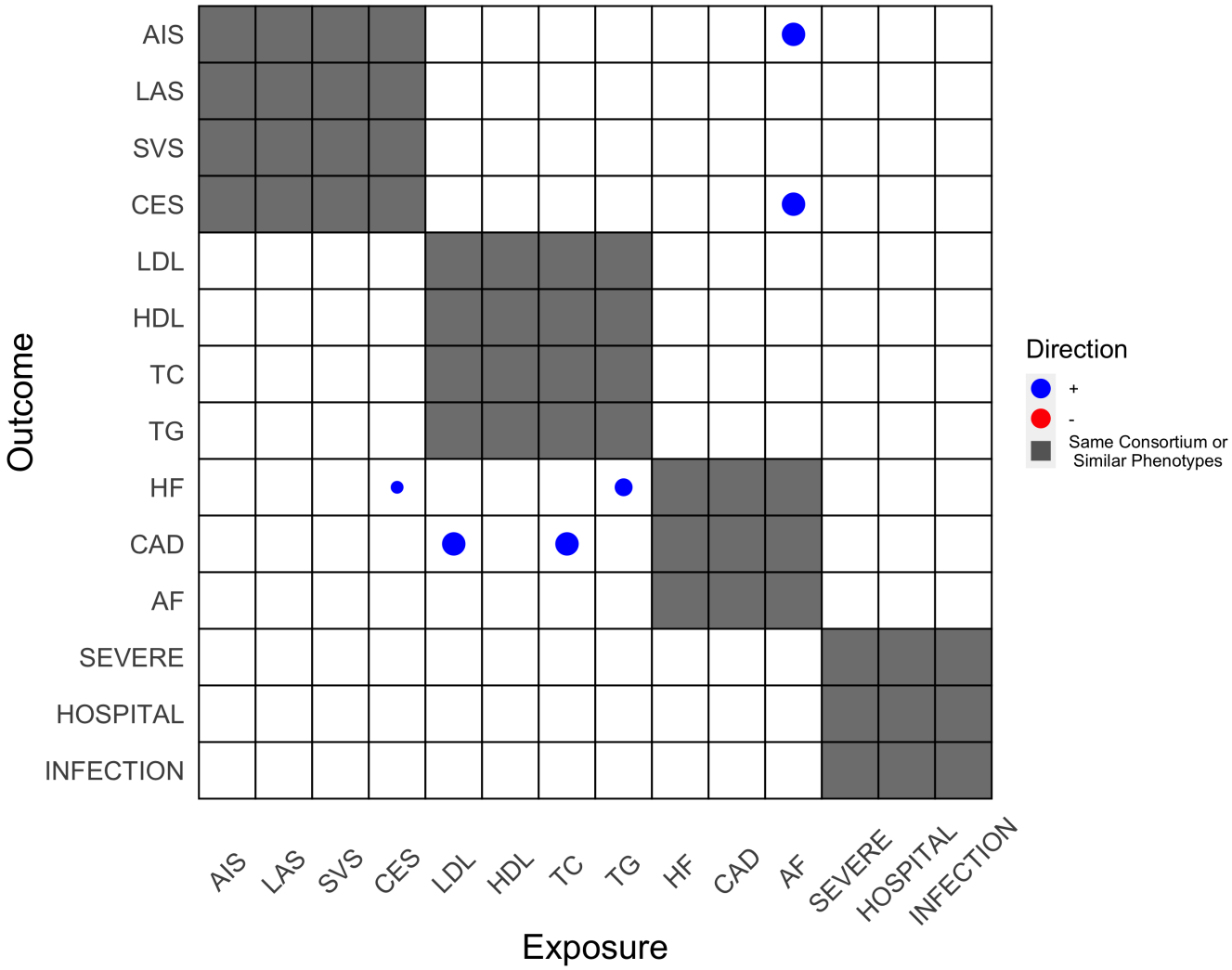

Figure S10: Estimates of causal effects using the contamination mixture method. The significantly positive and negative causal relationships after Bonferroni correction are marked by blue filled circles and red filled circles, respectively. The radius of a circle is proportional to the $-\log_{10}(p\text{-value})$ of the corresponding exposure-outcome pair. Those pairs whose exposure and outcome come from the same consortium or are two similar phenotypes are marked as grey cells.

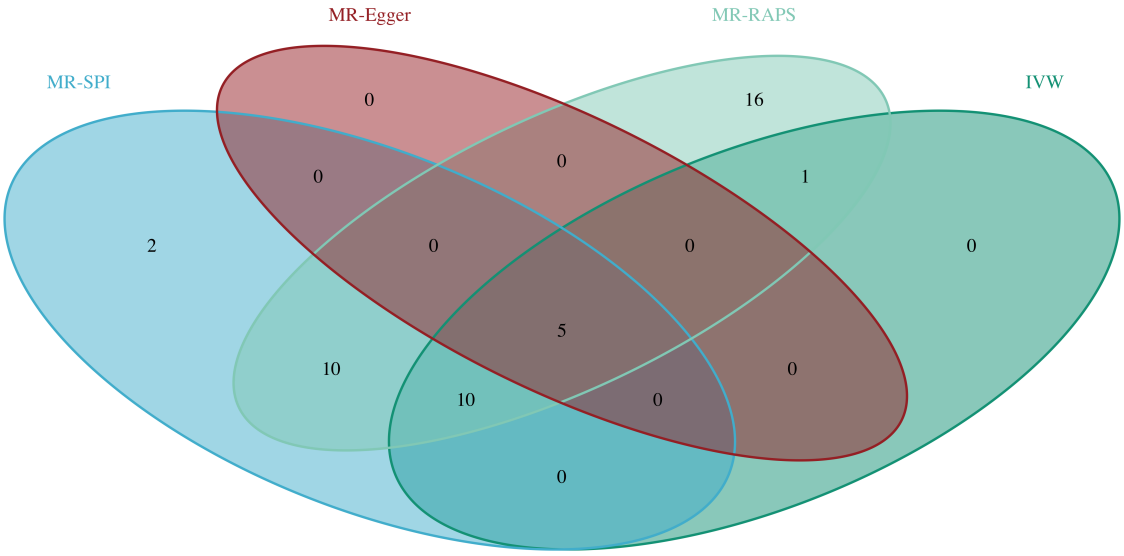

Figure S11: Venn diagram of significant causal relationships detected by MR-SPI, IVW, MR-Egger and MR-RAPS after Bonferroni correction.

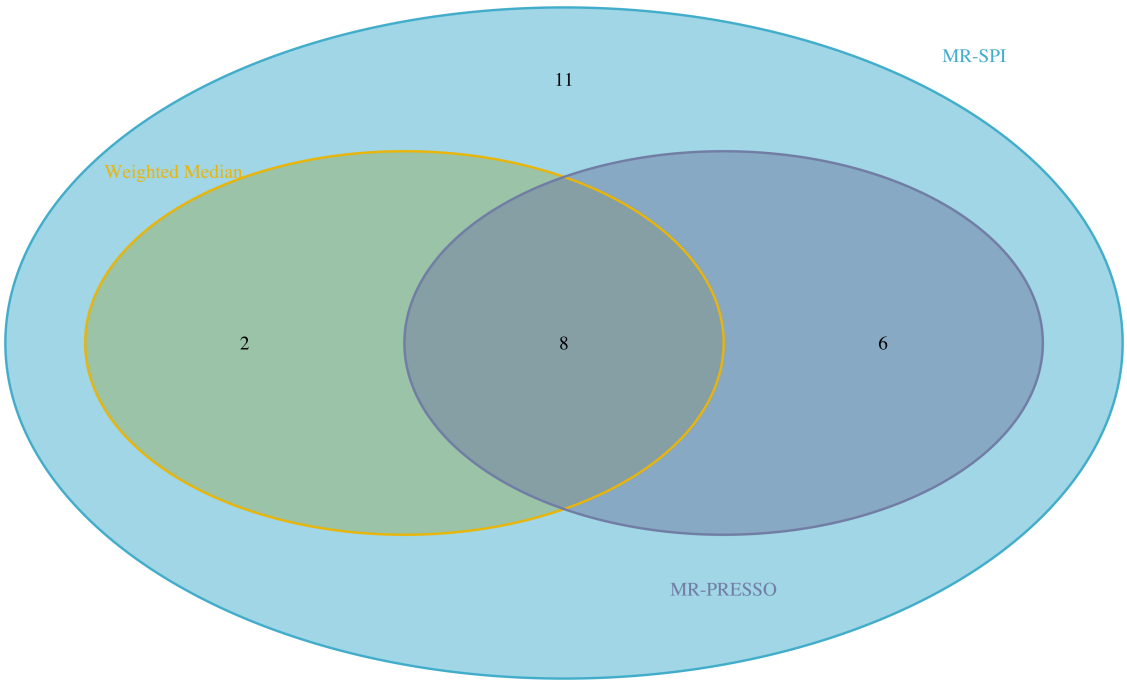

Figure S12: Venn diagram of significant causal relationships detected by MR-SPI, MR-PRESSO and the weighted median method after Bonferroni correction.

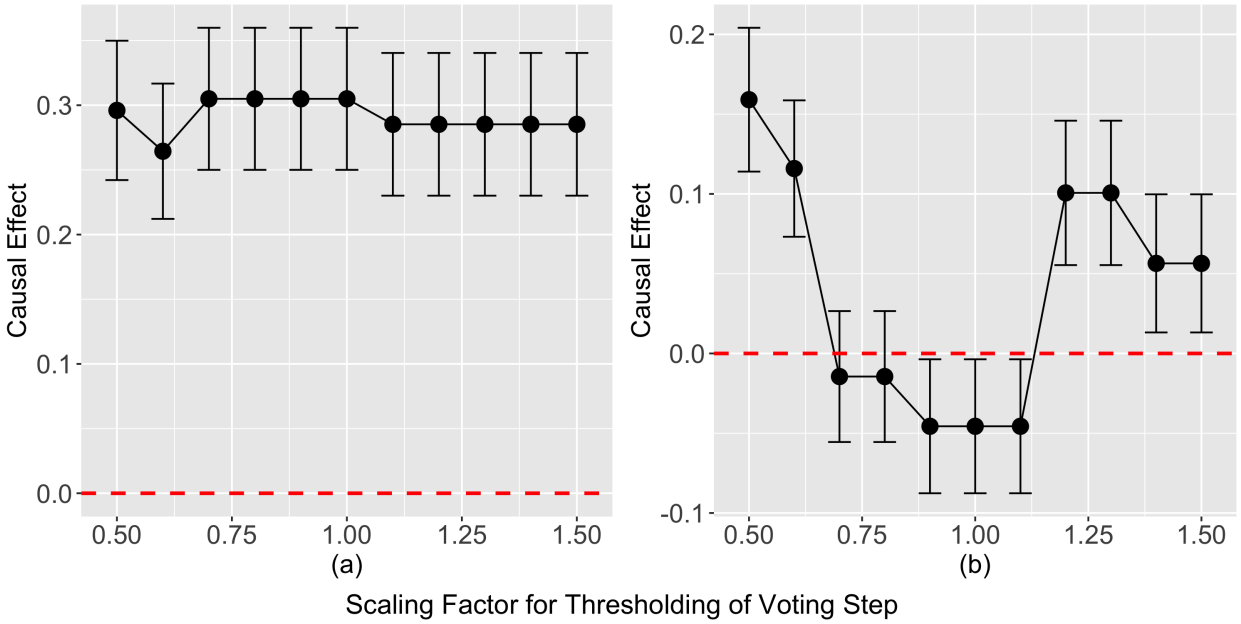

Figure S13: Two examples of sensitivity analysis of the voting procedure of MR-SPI. (a) MR-SPI provides stable inference in this example, as the causal effect estimate and the corresponding confidence interval do not change significantly with the scaling factor η. (b) MR-SPI might suffer from potential IV selection error in this example, since the causal effect estimate and the corresponding confidence interval are susceptible to the choice of η. For example, MR-SPI provides a significantly positive estimate of the causal effect when the scaling factor $\eta = 0.5$, while the estimate becomes negative when $\eta = 1$.

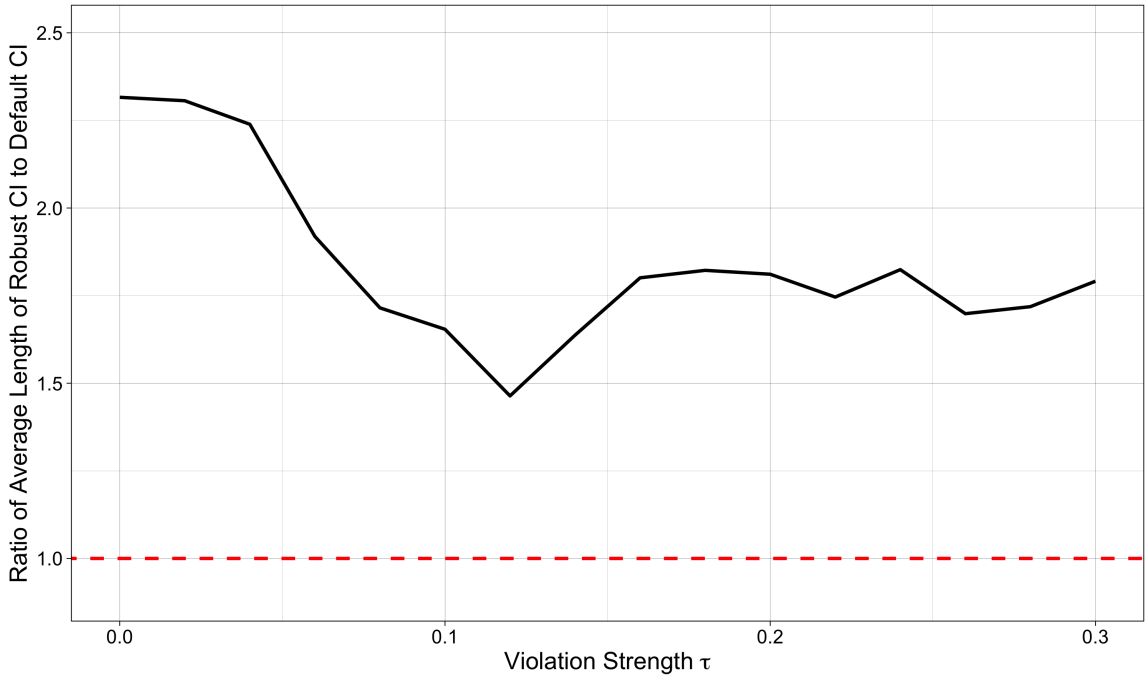

Figure S14: Ratio of the average lengths of 95% confidence interval of the robust confidence interval to the default confidence interval across a range of violation strength τ. The robust confidence interval is calculated by Algorithm 2. The default confidence interval is calculated by the limit distribution of the causal effect estimate using the selected set of valid instruments $\widehat{\mathcal{V}}$.

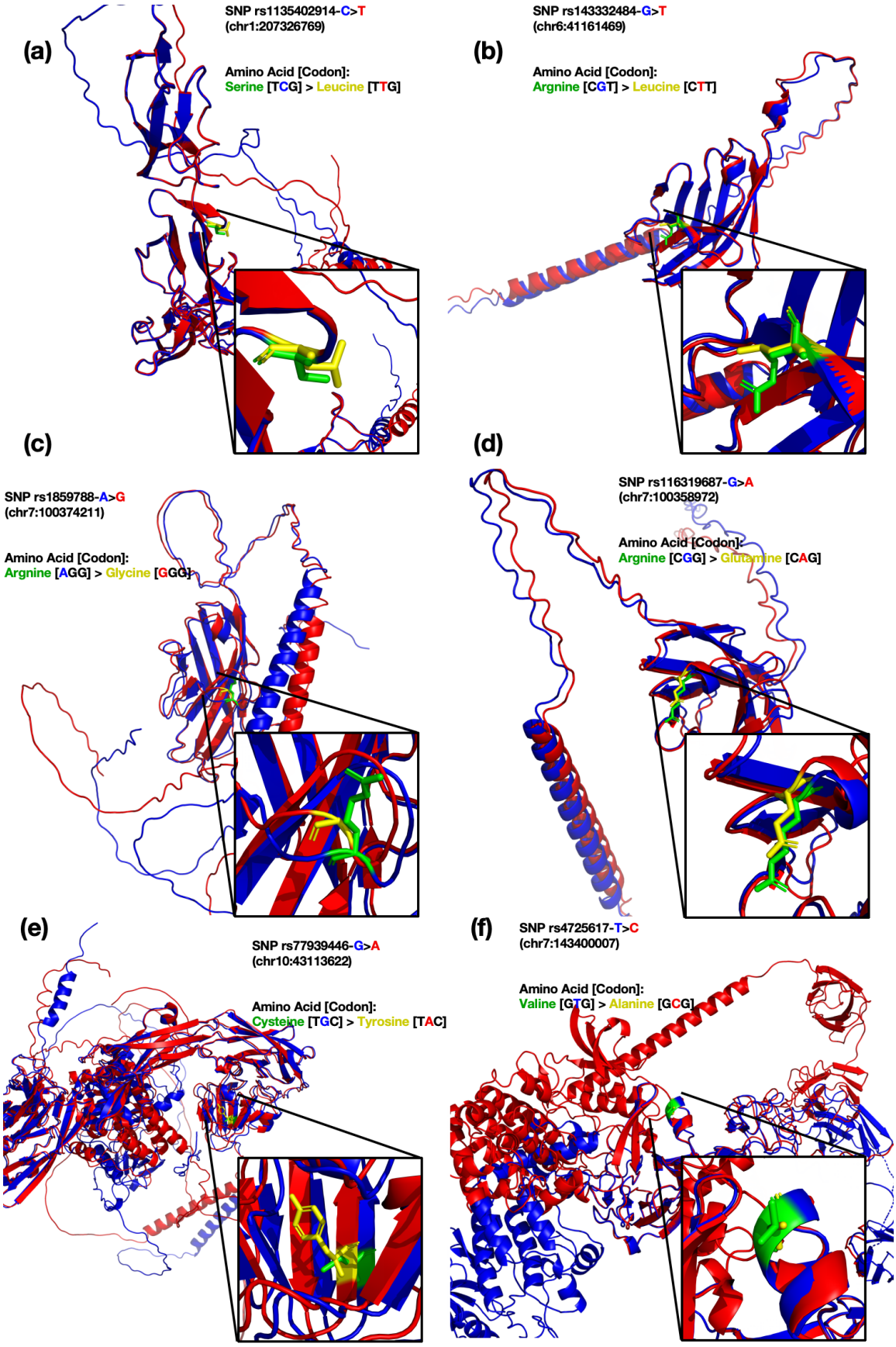

Figure S15: AlphaFold-predicted 3D structural alterations of **(a)** CD55, **(b)** TREM2, **(c)** PILRA, **(d)** PILRB, **(e)** RET, and **(f)** EPHA1.

References

[S1] Guo, Z., Kang, H., Tony Cai, T. & Small, D. S. Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**, 793–815 (2018).

[S2] Bowden, J. *et al.* A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Statistics in Medicine* **36**, 1783–1802 (2017).

[S3] Zhao, Q., Wang, J., Hemani, G., Bowden, J. & Small, D. S. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *The Annals of Statistics* **48**, 1742–1769 (2020).

[S4] Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics* **50**, 693–698 (2018).

[S5] Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology* **40**, 304–314 (2016).

[S6] Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International Journal of Epidemiology* **46**, 1985–1998 (2017).

[S7] Qi, G. & Chatterjee, N. Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. *Nature Communications* **10**, 1–10 (2019).

[S8] Burgess, S., Foley, C. N., Allara, E., Staley, J. R. & Howson, J. M. A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nature Communications* **11**, 1–11 (2020).

[S9] The Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in europeans and south asians identifies five new loci for coronary artery disease. *Nature Genetics* **43**, 339–344 (2011).

[S10] Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics* **43**, 333–338 (2011).

[S11] Purcell, S. *et al.* Plink: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).

[S12] Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

[S13] Staley, J. R. *et al.* PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016).

[S14] Kamat, M. A. *et al.* PhenoScanner V2: an expanded tool for searching human genotype–phenotype associations. *Bioinformatics* **35**, 4851–4853 (2019).

[S15] Ulhaq, Z. S. & Soraya, G. V. Interleukin-6 as a potential biomarker of covid-19 progression. *Medecine et maladies infectieuses* **50**, 382 (2020).

[S16] Dunn, O. J. Multiple comparisons among means. *Journal of the American Statistical Association* **56**, 52–64 (1961).

[S17] Trialists, C. T. *et al.* The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *The Lancet* **380**, 581–590 (2012).

[S18] Gencer, B. *et al.* Efficacy and safety of lowering LDL cholesterol in older patients: a systematic review and meta-analysis of randomised controlled trials. *The Lancet* **396**, 1637–1643 (2020).

[S19] Rey, J. R. *et al.* Heart failure in COVID-19 patients: prevalence, incidence and prognostic implications. *European Journal of Heart Failure* **22**, 2205–2215 (2020).

[S20] Krauss, R. *et al.* Intermediate-density lipoproteins and progression of coronary artery disease in hypercholesterolaemic men. *The Lancet* **330**, 62–66 (1987).

[S21] Rader, D. J. & Hovingh, G. K. HDL and cardiovascular disease. *The Lancet* **384**, 618–625 (2014).

[S22] Collaboration, P. S. *et al.* Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55 000 vascular deaths. *The Lancet* **370**, 1829–1839 (2007).

[S23] Witt, B. J. *et al.* Ischemic stroke after heart failure: a community-based study. *American heart journal* **152**, 102–109 (2006).

[S24] Haeusler, K. G., Laufs, U. & Endres, M. Chronic heart failure and ischemic stroke. *Stroke* **42**, 2977–2982 (2011).