A Novel Penalized Inverse-Variance Weighted Estimator for Mendelian Randomization with Applications to COVID-19 Outcomes

Siqi Xu[1], Peng Wang[2], Wing Kam Fung[1,*] and Zhonghua Liu[1,]**

[1]Department of Statistics and Actuarial Science, The University of Hong Kong,

Pokfulam Road, Hong Kong, China

[2]Department of Epidemiology and Biostatistics, Tongji Medical College,

Huazhong University of Science and Technology, Wuhan, China

*email: wingfung@hku.hk

**email: zhhliu@hku.hk

Summary: Mendelian randomization (MR) utilizes genetic variants as instrumental variables (IVs) to estimate the causal effect of an exposure variable on an outcome of interest even in the presence of unmeasured confounders. However, the popular inverse-variance weighted (IVW) estimator could be biased in the presence of weak IVs, a common challenge in MR studies. In this article, we develop a novel penalized inverse-variance weighted (pIVW) estimator, which adjusts the original IVW estimator to account for the weak IV issue by using a penalization approach to prevent the denominator of the pIVW estimator from being close to zero. Moreover, we adjust the variance estimation of the pIVW estimator to account for the presence of horizontal pleiotropy. We show that the recently proposed debiased IVW (dIVW) estimator is a special case of our proposed pIVW estimator. We further prove that the pIVW estimator has smaller bias and variance than the dIVW estimator under some regularity conditions. We also conduct extensive simulation studies to demonstrate the performance of the proposed pIVW estimator. Furthermore, we apply the pIVW estimator to estimate the causal effects of five obesity-related exposures on three coronavirus disease 2019 (COVID-19) outcomes. Notably, we find that hypertensive disease is associated with an increased risk of hospitalized COVID-19, and peripheral vascular disease and higher body mass index are associated with increased risks of COVID-19 infection, hospitalized COVID-19 and critically ill COVID-19. The R package for the pIVW method is publicly available at https://github.com/siqixu/mr.pivw.

Key words: COVID-19; Horizontal pleiotropy; Instrumental variables; Mendelian randomization; Penalization; Weak instruments.

1. Introduction

It is of scientific interest to estimate the causal effects of modifiable risk factors on various health outcomes in epidemiological studies. For example, estimating the causal effects of modifiable risk factors on the coronavirus disease 2019 (COVID-19) outcomes is currently one of the most pressing global public health problems (Jordan et al., 2020; Zheng et al., 2020). The COVID-19 pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has posed a serious threat to human health all over the world (Pascarella et al., 2020). As of July 1, 2021, the COVID-19 pandemic has led to over 180 million cases and more than 3.9 million deaths, which has brought unprecedented medical and economic burdens worldwide (Dong et al., 2020). It is crucial to identify causal risk factors associated with COVID-19 incidence and mortality so that we can develop more effective prevention and intervention strategies. One major challenge is the unmeasured confounding bias for the exposure-outcome relationship in observational epidemiological studies.

To address the unmeasured confounding issue in observational studies, Mendelian randomization (MR) utilizes genetic variants as instrumental variables (IVs) to estimate the causal effect of an exposure variable on an outcome of interest even in the presence of unmeasured confounders (Smith and Ebrahim, 2003, 2004; Sheehan et al., 2008). With the increasing availability of summary-level data from genome-wide association studies (GWASs), many MR methods have been developed based on GWAS summary-level data (Lawlor, 2016; Zheng et al., 2017). However, the validity of MR analysis critically depends on the following three core assumptions defining a valid IV (Didelez and Sheehan, 2007; Lawlor et al., 2008):

(1) IV Relevance: the IV is associated with the exposure, which is empirically testable;

(2) IV Independence: the IV is independent of any confounders of the exposure-outcome relationship;

(3) Exclusion Restriction: the IV only affects the outcome via the exposure.

When one of these IV assumptions is violated, the MR analysis may yield biased estimation of the causal effect. In particular, the first IV assumption (IV Relevance) can be nearly violated when the IVs are only weakly associated with the exposure variable (Burgess and Thompson, 2011; Burgess et al., 2011; Davies et al., 2015). In MR studies, the weak IV bias may occur when the genetic variants only explain a small proportion of variance for the exposure variable. On the other hand, the widespread horizontal pleiotropy in human genome can also lead to the violation of the third IV assumption (Exclusion Restriction) (Verbanck et al., 2018; Hemani et al., 2018), which is a phenomenon that the genetic variants directly affect the outcome not mediated by the exposure variable (see Figure 1 for a graphical illustration).

The inverse-variance weighted (IVW) estimator is one of the most popular MR methods that has been widely used in health studies (Burgess et al., 2013). It has a simple and explicit expression, which combines the estimated causal effects from multiple IVs into a weighted average with the idea borrowed from the fixed-effect meta-analysis literature. Despite its widespread popularity, recent studies pointed out that the IVW estimator can be seriously biased in the presence of weak IVs (Zhao et al., 2020; Ye et al., 2021). MR-RAPS is a maximum profile likelihood estimator, which was shown to be robust to weak IVs (Zhao et al., 2020). However, MR-RAPS has no closed-form solution and might have multiple roots. Recently, the debiased IVW (dIVW) estimator was proposed to account for the weak IV issue by a simple modification to the IVW estimator (Ye et al., 2021). The dIVW estimator has been proved to be consistent even in the presence of many weak IVs under certain conditions. Nevertheless, as a ratio estimator, the dIVW estimator is still likely to yield a biased estimate when its denominator is close to zero. In fact, when the denominator is close to zero, a ratio estimator may have heavy-tailed distribution and thus may not even have finite moments (Press, 1969; Piegorsch and Casella, 1985).

In this article, we develop a novel penalized inverse-variance weighted (pIVW) estimator, where the original IVW estimator is adjusted by a proposed penalized log-likelihood function. Through the penalization, we can prevent the denominator in the ratio estimator from being close to zero and thus provide improved estimation in the presence of weak IVs. Moreover, we account for the horizontal pleiotropy by adjusting the variance estimation of the pIVW estimator. The proposed pIVW estimator has some nice features. First, our theoretical and numerical results show that the proposed pIVW estimator has smaller bias and variance than the dIVW estimator under some regularity conditions. Second, it is consistent and asymptotically normal even in the presence of many weak IVs, and requires no more assumptions than the dIVW estimator. Third, it has a unique and closed-form solution, whereas some other robust MR methods like MR-RAPS might have multiple roots and diverging solutions.

We demonstrate the improved performance of the proposed pIVW estimator compared to the other competing MR methods in extensive simulation studies. Furthermore, we apply the pIVW estimator to estimate the causal effects of five obesity-related exposures (i.e., peripheral vascular disease, dyslipidemia, hypertensive disease, type 2 diabetes and body mass index (BMI)) on three COVID-19 outcomes (i.e., COVID-19 infection, hospitalized COVID-19 and critically ill COVID-19). We find that hypertensive disease is significantly associated with an increased risk of hospitalized COVID-19, and peripheral vascular disease and higher BMI are significantly associated with increased risks of COVID-19 infection, hospitalized COVID-19 and critically ill COVID-19.

2. The Two-Sample MR Design and Prior Work

2.1 *Linear Structural Models*

Suppose that there are p independent genetic variants $\{G_j\}_{j=1}^p$. When there is no horizontal pleiotropy, the relationships among the genetic variants G_js, the exposure X, the outcome Y and the unmeasured confounder U depicted in Figure 1 can be formulated by the linear structural models as follows (Bowden et al., 2015):

$$X = \sum_{j=1}^{p} \gamma_j G_j + U + \epsilon_X, \tag{1}$$

$$Y = \beta X + U + \epsilon_Y, \tag{2}$$

where γ_j is the genetic effect of G_j on X, β is the causal effect of our interest, and ϵ_X and ϵ_Y are mutually independent random errors. Let Γ_j denotes the effect of G_j on Y, then we have $\Gamma_j = \beta\gamma_j$ by substituting Equation (1) for X in Equation (2). When Y is a binary outcome and we replace Equation (2) by a logistic model, then $\Gamma_j = \beta\gamma_j$ can still be justified by a probit approximation of the logistic model (Vansteelandt et al., 2011; Zhao et al., 2020).

Let $\hat{\gamma}_j$ and $\hat{\Gamma}_j$ be the estimates of γ_j and Γ_j with the variances $\sigma^2_{\hat{\gamma}_j}$ and $\sigma^2_{\hat{\Gamma}_j}$, respectively. In the two-sample MR design, $\left\{\hat{\gamma}_j, \sigma_{\hat{\gamma}_j}\right\}_{j=1}^p$ and $\left\{\hat{\Gamma}_j, \sigma_{\hat{\Gamma}_j}\right\}_{j=1}^p$ can be obtained from two independent GWASs (Lawlor, 2016). Since the GWASs generally involve large sample sizes, it is common to assume that $\hat{\gamma}_j$ and $\hat{\Gamma}_j$ are independently distributed as $\hat{\gamma}_j \sim N\left(\gamma_j, \sigma^2_{\hat{\gamma}_j}\right)$ and $\hat{\Gamma}_j \sim N\left(\Gamma_j, \sigma^2_{\hat{\Gamma}_j}\right)$ with known $\sigma^2_{\hat{\gamma}_j}$ and $\sigma^2_{\hat{\Gamma}_j}$, respectively (Zhao et al., 2020).

[Figure 1 about here.]

2.2 *IVW Estimator and Debiased IVW (dIVW) Estimator*

The popular inverse-variance weighted (IVW) estimator combines the estimated causal effects $\hat{\beta}_j = \hat{\Gamma}_j/\hat{\gamma}_j$ from multiple genetic variants with the weights $w_j = \sigma^{-2}_{\hat{\Gamma}_j}\hat{\gamma}_j^2$ as follows:

$$\hat{\beta}_{IVW} = \frac{\sum_{j=1}^p w_j \hat{\beta}_j}{\sum_{j=1}^p w_j} = \frac{\sum_{j=1}^p \sigma^{-2}_{\hat{\Gamma}_j}\hat{\gamma}_j\hat{\Gamma}_j}{\sum_{j=1}^p \sigma^{-2}_{\hat{\Gamma}_j}\hat{\gamma}_j^2}.$$

Let $\mu_1 = \sum_{j=1}^{p} \sigma_{\hat{\Gamma}_j}^{-2} \gamma_j \Gamma_j$ and $\mu_2 = \sum_{j=1}^{p} \sigma_{\hat{\Gamma}_j}^{-2} \gamma_j^2$. We have $\beta = \mu_1/\mu_2$ since $\Gamma_j = \beta\gamma_j$ under models (1)-(2). As shown by Zhao et al. (2020), the IVW estimator can be approximated by

$$\hat{\beta}_{IVW} \approx \frac{E\left[\sum_{j=1}^{p} \sigma_{\hat{\Gamma}_j}^{-2} \hat{\gamma}_j \hat{\Gamma}_j\right]}{E\left[\sum_{j=1}^{p} \sigma_{\hat{\Gamma}_j}^{-2} \hat{\gamma}_j^2\right]} = \frac{\mu_1}{\mu_2 + \sum_{j=1}^{p} \sigma_{\hat{\Gamma}_j}^{-2} \sigma_{\hat{\gamma}_j}^2} = \frac{\beta}{1 + \frac{\sum_{j=1}^{p} \sigma_{\hat{\Gamma}_j}^{-2} \sigma_{\hat{\gamma}_j}^2}{\sum_{j=1}^{p} \sigma_{\hat{\Gamma}_j}^{-2} \gamma_j^2}}.$$

When there is no measurement error for γ_j (i.e., $\sigma_{\hat{\gamma}_j}^2 = 0$), we have $\hat{\beta}_{IVW} \approx \mu_1/\mu_2 = \beta$. However, some recent studies have shown that the IVW estimator can be seriously biased toward zero for ignoring the measurement errors of γ_j especially in the presence of many weak IVs that have small $\sigma_{\hat{\gamma}_j}^{-2} \gamma_j^2$ (Zhao et al., 2020; Ye et al., 2021) .

To handle the bias due to weak IVs, the debiased IVW (dIVW) estimator replaces the denominator in the IVW estimator with an unbiased estimator $\hat{\mu}_2$ of μ_2 as

$$\hat{\beta}_{dIVW} = \frac{\hat{\mu}_1}{\hat{\mu}_2} = \frac{\sum_{j=1}^{p} \sigma_{\hat{\Gamma}_j}^{-2} \hat{\gamma}_j \hat{\Gamma}_j}{\sum_{j=1}^{p} \sigma_{\hat{\Gamma}_j}^{-2} \left(\hat{\gamma}_j^2 - \sigma_{\hat{\gamma}_j}^2\right)}.$$

The dIVW estimator was shown to be consistent and asymptotically normal under weaker conditions than the IVW estimator. However, the dIVW estimator is more likely to yield extreme estimates in the presence of weak IVs, because its denominator $\hat{\mu}_2$ has a larger coefficient of variation than that of the IVW estimator and thus has a higher probability of being close to zero. To overcome the limitations of the IVW estimator and dIVW estimator, we adjust the IVW estimator to account for the weak IVs by using a penalized log-likelihood function for μ_1 and μ_2, which can prevent the estimator of μ_2 from being close to zero.

3. Method

3.1 *The Penalized IVW (pIVW) Estimator*

Assume that the estimators $\hat{\mu}_1 = \sum_{j=1}^{p} \sigma_{\hat{\Gamma}_j}^{-2} \hat{\gamma}_j \hat{\Gamma}_j$ and $\hat{\mu}_2 = \sum_{j=1}^{p} \sigma_{\hat{\Gamma}_j}^{-2} \left(\hat{\gamma}_j^2 - \sigma_{\hat{\gamma}_j}^2\right)$ follow a bivariate normal distribution as

$$\begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} v_1 & v_{12} \\ v_{12} & v_2 \end{pmatrix}\right).$$

We propose a penalized log-likelihood function to adjust the estimates of μ_1 and μ_2 as

$$l_p(\mu_1, \mu_2) = \log f(\mu_1, \mu_2) + \lambda \log |\mu_2|,$$

where $f(\mu_1, \mu_2)$ denotes the bivariate normal density function of $(\hat{\mu}_1, \hat{\mu}_2)'$ and $\lambda > 0$ is the penalty parameter. The penalized log-likelihood $l_p(\mu_1, \mu_2)$ becomes small when μ_2 approaches zero due to the penalty term $\lambda \log |\mu_2|$. Therefore, an estimate of μ_2 being close to zero is less preferable by $l_p(\mu_1, \mu_2)$. Specifically, we obtain the adjusted estimators of μ_1 and μ_2 by maximizing $l_p(\mu_1, \mu_2)$ as follows:

$$\tilde{\mu}_1 = \hat{\mu}_1 + \frac{\hat{v}_{12}}{\hat{v}_2}(\tilde{\mu}_2 - \hat{\mu}_2) \quad \text{and} \quad \tilde{\mu}_2 = \left(\frac{1}{2} + \sqrt{\frac{1}{4} + \lambda \frac{\hat{v}_2}{\hat{\mu}_2^2}}\right)\hat{\mu}_2, \tag{3}$$

where $\hat{v}_{12} = 2\sum_{j=1}^{p} \sigma_{\hat{\gamma}_j}^2 \sigma_{\hat{\Gamma}_j}^{-4} \hat{\Gamma}_j \hat{\gamma}_j$ and $\hat{v}_2 = \sum_{j=1}^{p} \sigma_{\hat{\Gamma}_j}^{-4} \left[4\left(\hat{\gamma}_j^2 - \sigma_{\hat{\gamma}_j}^2\right)\sigma_{\hat{\gamma}_j}^2 + 2\sigma_{\hat{\gamma}_j}^4\right]$ are the estimators of v_{12} and v_2, respectively. Then, we propose the penalized IVW (pIVW) estimator as a ratio of $\tilde{\mu}_1$ and $\tilde{\mu}_2$, given by

$$\hat{\beta}_{pIVW} = \frac{\tilde{\mu}_1}{\tilde{\mu}_2} = \frac{\hat{\mu}_1}{\tilde{\mu}_2} + \frac{\hat{v}_{12}}{\hat{v}_2}\left(1 - \frac{\hat{\mu}_2}{\tilde{\mu}_2}\right). \tag{4}$$

Note that when the penalty parameter $\lambda = 0$, we have $\tilde{\mu}_2 = \hat{\mu}_2$ and therefore $\hat{\beta}_{pIVW}$ reduces to $\hat{\beta}_{dIVW}$. When $\lambda > 0$, $\hat{\beta}_{pIVW}$ will also be close to $\hat{\beta}_{dIVW}$ when $\hat{v}_2/\hat{\mu}_2^2$ approaches zero. Following Ye et al. (2021), we define the IV strength as

$$\kappa = \frac{\sum_{j=1}^{p} \gamma_j^2 \sigma_{\hat{\gamma}_j}^{-2}}{p},$$

which can be estimated by $\hat{\kappa} = \frac{1}{p}\sum_{j=1}^{p}\left(\hat{\gamma}_j^2 - \sigma_{\hat{\gamma}_j}^2\right)\sigma_{\hat{\gamma}_j}^{-2}$. To study the asymptotic properties of the pIVW estimator, we make the following Assumptions 1 and 2 which are also required for the consistency of the dIVW estimator (Ye et al., 2021).

Assumption 1: The number of IVs p diverges to infinity.

Assumption 2: $\left\{\hat{\gamma}_j, \hat{\Gamma}_j\right\}_{j=1}^{p}$ are independently distributed as $\hat{\gamma}_j \sim N\left(\gamma_j, \sigma_{\hat{\gamma}_j}^2\right)$ and $\hat{\Gamma}_j \sim N\left(\beta\gamma_j, \sigma_{\hat{\Gamma}_j}^2\right)$ with known variances $\sigma_{\hat{\gamma}_j}^2$ and $\sigma_{\hat{\Gamma}_j}^2$. The ratio of variances $\sigma_{\hat{\gamma}_j}^2/\sigma_{\hat{\Gamma}_j}^2$ is bounded away from zero and infinity for all j.

Assumptions 1 and 2 are reasonable in the two-sample MR design since the GWASs often have large sample sizes and a large number of genetic variants. The independence among $\left\{\hat{\gamma}_j, \hat{\Gamma}_j\right\}_{j=1}^{p}$ can be guaranteed by the linkage-disequilibrium clumping. Then, under Assumptions 1 and 2, $\hat{v}_2/\hat{\mu}_2^2 = O_p\left(\frac{1}{\kappa p} + \frac{1}{\kappa^2 p}\right)$ and hence $\hat{\beta}_{pIVW}$ reduces to $\hat{\beta}_{dIVW}$ as $\kappa\sqrt{p} \to \infty$. More importantly, in the following Theorem 1 (a)-(b), we show that $\hat{\beta}_{pIVW}$ has smaller bias and variance than $\hat{\beta}_{dIVW}$ under a proper choice of the penalty parameter λ. Further, we show that $\hat{\beta}_{pIVW}$ is consistent and asymptotically normal under some regularity conditions in the following Theorem 1 (c), which requires no more assumptions comparing to $\hat{\beta}_{dIVW}$.

THEOREM 1: *Suppose that models (1)-(2) and Assumptions 1-2 hold. Then, as $\kappa\sqrt{p} \to \infty$, we have the following conclusions:*

(a) The bias of $\hat{\beta}_{dIVW}$ is of order $O\left(\frac{1}{\kappa p} + \frac{1}{\kappa^2 p}\right)$, and the bias of $\hat{\beta}_{pIVW}$ is of smaller order $o\left(\frac{1}{\kappa p} + \frac{1}{\kappa^2 p}\right)$ when $\lambda = 1$.

(b) The variance of $\hat{\beta}_{pIVW}$ is smaller than that of $\hat{\beta}_{dIVW}$ when $\lambda > 0$.

(c) Assume that $\max_j \gamma_j^2 \sigma_{\hat{\gamma}_j}^{-2}/(\kappa p + p) \to 0$. Then, $\hat{\beta}_{pIVW}$ is consistent and asymptotically normal, i.e.,

$$\hat{V}^{-\frac{1}{2}}\left(\hat{\beta}_{pIVW} - \beta\right) \xrightarrow{d} N(0,1),$$

where

$$\hat{V} = \tilde{\mu}_2^{-2} \sum_{j=1}^{p} \left[\sigma_{\hat{\Gamma}_j}^{-2}\hat{\gamma}_j^2 + \hat{\beta}_{pIVW}^2 \sigma_{\hat{\gamma}_j}^2 \sigma_{\hat{\Gamma}_j}^{-4}\left(\hat{\gamma}_j^2 + \sigma_{\hat{\gamma}_j}^2\right)\right].$$

The proof of Theorem 1 is provided in the Web Appendix A. In the setting of many weak IVs, we may have $\kappa \to 0$ as $p \to \infty$ because more weak IVs are likely to be included into the analysis as the number of IVs p increases, which may reduce the IV strength κ. The above theorem holds in this case as long as $\kappa\sqrt{p} \to \infty$, which means that it allows the presence of many weak IVs. On the other hand, although many ratio estimators may not have finite moments when their denominators are close to zero (Piegorsch and Casella, 1985; Press,

1969; Marsaglia et al., 2006), for the denominator $\hat{\mu}_2$ of $\hat{\beta}_{dIVW}$, we have $P(|\hat{\mu}_2| > \epsilon) \to 1$ for any $\epsilon > 0$ under the regularity conditions in Theorem 1. We have the same result for the denominator $\tilde{\mu}_2$ of $\hat{\beta}_{pIVW}$, since $\tilde{\mu}_2$ is in larger magnitude than $\hat{\mu}_2$ by Equation (3). Therefore, we can still approximate the biases and variances of $\hat{\beta}_{dIVW}$ and $\hat{\beta}_{pIVW}$ in this case.

Theorem 1 (a) indicates that the bias of $\hat{\beta}_{pIVW}$ with $\lambda = 1$ converges to zero at a faster rate than that of $\hat{\beta}_{dIVW}$. Specifically, we approximate the bias of $\hat{\beta}_{pIVW}$ by the Taylor series expansion. We find that the first term in the bias of $\hat{\beta}_{pIVW}$ is of order $O\left(\frac{1}{\kappa p} + \frac{1}{\kappa^2 p}\right)$ when $\lambda \neq 1$, and the remaining terms are of order $o\left(\frac{1}{\kappa p} + \frac{1}{\kappa^2 p}\right)$. When $\lambda = 1$, we find that the first term in the bias of $\hat{\beta}_{pIVW}$ is exactly equal to zero (more details are given in Web Appendix A). Hence, $\hat{\beta}_{pIVW}$ has the smallest bias when $\lambda = 1$, which is also of smaller order $o\left(\frac{1}{\kappa p} + \frac{1}{\kappa^2 p}\right)$ than that of $\hat{\beta}_{dIVW}$. According to Theorem 1 (b), $\hat{\beta}_{pIVW}$ with $\lambda = 1$ also has smaller variance than $\hat{\beta}_{dIVW}$. Therefore, we recommend $\lambda = 1$ for the pIVW estimator in practice.

3.2 *Selection of Candidate Instruments*

In this section, we extend Theorem 1 to the situation where IV selection is conducted to remove some weak IVs from the analysis, which is a common practice in MR studies to handle the weak IV bias.

Suppose that there is a selection dataset $\left\{\hat{\gamma}_j^*, \sigma_{\hat{\gamma}_j}^*\right\}_{j=1}^p$ independent of the exposure dataset and the outcome dataset. Then, an IV is included into the analysis when $|\hat{\gamma}_j^*| > \delta\sigma_{\hat{\gamma}_j}^*$ with a pre-set threshold $\delta > 0$ (Zhao et al., 2019). Ye et al. (2021) showed that IV selection with an appropriate threshold δ could reduce the bias of the IVW estimator and improve the efficiency of the dIVW estimator. They also recommended a threshold $\delta = \sqrt{2\log p}$ to guarantee a small probability of selecting any null IVs (i.e., $\gamma_j = 0$). When the IV selection is performed at a threshold δ, we define the IV strength as

$$\kappa_\delta = \frac{\sum_{j=1}^p \gamma_j^2 \sigma_{\hat{\gamma}_j}^{-2} q_{\delta,j}}{p_\delta},$$

where $q_{\delta,j} = P\left(|\hat{\gamma}_j^*| > \delta\sigma_{\hat{\gamma}_j}^*\right)$ and $p_\delta = \sum_{j=1}^p q_{\delta,j}$. Let $S_\delta = \left\{j : |\hat{\gamma}_j^*| > \delta\sigma_{\hat{\gamma}_j}^*\right\}$ be the set of selected IVs. Then, we can estimate κ_δ by $\hat{\kappa}_\delta = \hat{p}_\delta^{-1}\sum_{j\in S_\delta}\left(\hat{\gamma}_j^2 - \sigma_{\hat{\gamma}_j}^2\right)\sigma_{\hat{\gamma}_j}^{-2}$, where $\hat{p}_\delta$ denotes the number of selected IVs within S_δ. To study the theoretical properties of the proposed pIVW estimator under IV selection, we have the following Assumption 3 for the summary-level data in the selection dataset.

Assumption 3: $\left\{\hat{\gamma}_j^*, \hat{\gamma}_j, \hat{\Gamma}_j\right\}_{j=1}^p$ are mutually independent and $\hat{\gamma}_j^* \sim N\left(\gamma_j, \sigma_{\hat{\gamma}_j}^{*2}\right)$ with known variance $\sigma_{\hat{\gamma}_j}^{*2}$ for every j. The ratio of variances $\sigma_{\hat{\gamma}_j}^2/\sigma_{\hat{\gamma}_j}^{*2}$ is bounded away from zero and infinity for all j.

Given a selection threshold δ, we evaluate the dIVW estimator $\hat{\beta}_{\delta,dIVW} = \hat{\mu}_{1,\delta}/\hat{\mu}_{2,\delta}$ and the proposed pIVW estimator $\hat{\beta}_{\delta,pIVW} = \tilde{\mu}_{1,\delta}/\tilde{\mu}_{2,\delta}$ using the selected IVs within the set S_δ. In the following Theorem 2, we show that $\hat{\beta}_{\delta,pIVW}$ still has smaller bias when $\lambda = 1$ and has smaller variance when $\lambda > 0$ comparing to $\hat{\beta}_{\delta,dIVW}$, and $\hat{\beta}_{\delta,pIVW}$ is consistent and asymptotically normal under some regularity conditions.

Theorem 2: *Suppose that models (1)-(2) and Assumptions 1-3 hold, and* $\kappa_\delta\sqrt{p_\delta}/\max(1,\delta^2) \to \infty$. *Then, we have the following conclusions:*

(a) The bias of $\hat{\beta}_{\delta,dIVW}$ *is of order* $O\left(\frac{1}{\kappa_\delta p_\delta} + \frac{1+\delta^4}{\kappa_\delta^2 p_\delta}\right)$, *and the bias of* $\hat{\beta}_{\delta,pIVW}$ *is of smaller order* $o\left(\frac{1}{\kappa_\delta p_\delta} + \frac{1+\delta^4}{\kappa_\delta^2 p_\delta}\right)$ *when* $\lambda = 1$.

(b) The variance of $\hat{\beta}_{\delta,pIVW}$ *is smaller than that of* $\hat{\beta}_{\delta,dIVW}$ *when* $\lambda > 0$.

(c) Assume that $\max_j \gamma_j^2\sigma_{\hat{\gamma}_j}^{-2}q_{\delta,j}/(\kappa_\delta p_\delta + p_\delta) \to 0$. *Then,* $\hat{\beta}_{\delta,pIVW}$ *is consistent and asymptotically normal, i.e.,*

$$\hat{V}_\delta^{-\frac{1}{2}}\left(\hat{\beta}_{\delta,pIVW} - \beta\right) \xrightarrow{d} N(0,1),$$

where

$$\hat{V}_\delta = \tilde{\mu}_{2,\delta}^{-2}\sum_{j\in S_\delta}\left[\sigma_{\hat{\Gamma}_j}^{-2}\hat{\gamma}_j^2 + \hat{\beta}_{\delta,pIVW}^2\sigma_{\hat{\gamma}_j}^2\sigma_{\hat{\Gamma}_j}^{-4}\left(\hat{\gamma}_j^2 + \sigma_{\hat{\gamma}_j}^2\right)\right].$$

The proof of Theorem 2 is provided in the Web Appendix B. Theorem 2 is an extension of

Theorem 1 to the situation under the IV selection. When the selection threshold $\delta = 0$ (i.e., no IV selection), Theorem 2 reduces to Theorem 1. It shows that the pIVW estimator still has smaller bias and variance than the dIVW estimator, and it is consistent and asymptotically normal after accounting for the IV selection. We extend the results for the dIVW estimator in Ye et al. (2021) to the pIVW estimator.

3.3 *Accounting for Horizontal Pleiotropy*

When there exists horizontal pleiotropy (i.e., direct effect of G_j on Y not mediated by X), the linear structural model (2) can be modified as follows (Bowden et al., 2015):

$$Y = \beta X + \sum_{j=1}^{p} \alpha_j G_j + U + \epsilon_Y, \tag{5}$$

where α_j denotes the direct genetic effect of G_j on the outcome Y (i.e., pleiotropic effect). In this case, we have $\Gamma_j = \beta\gamma_j + \alpha_j$. We follow a common practice in many MR methods to assume that the horizontal pleiotropy is balanced (i.e., the pleiotropic effect has mean zero) and treat α_j as random effect following $\alpha_j \sim N(0, \tau^2)$ (Bowden et al., 2017; Zhao et al., 2020; Ye et al., 2021). Then, $\hat{\Gamma}_j \sim N\left(\beta\gamma_j, \sigma^2_{\hat{\Gamma}_j} + \tau^2\right)$ has larger variance due to the presence of horizontal pleiotropy. To account for the horizontal pleiotropy, we estimate the variance of $\hat{\beta}_{\delta,pIVW}$ by

$$\hat{V}^*_\delta = \tilde{\mu}^{-2}_{2,\delta} \sum_{j \in S_\delta} \left[\sigma^{-2}_{\hat{\Gamma}_j} \hat{\gamma}^2_j \left(1 + \hat{\tau}^2 \sigma^{-2}_{\hat{\Gamma}_j}\right) + \hat{\beta}^2_{\delta,pIVW} \sigma^2_{\hat{\gamma}_j} \sigma^{-4}_{\hat{\Gamma}_j} \left(\hat{\gamma}^2_j + \sigma^2_{\hat{\gamma}_j}\right) \right],$$

where $\hat{\tau}$ is the estimator of τ^2. Following Ye et al. (2021), we derive $\hat{\tau}^2$ as

$$\hat{\tau}^2 = \frac{\sum_{j=1}^{p} \left[\left(\hat{\Gamma}_j - \hat{\beta}_{pIVW}\hat{\gamma}_j\right)^2 - \sigma^2_{\hat{\Gamma}_j} - \hat{\beta}^2_{pIVW}\sigma^2_{\hat{\gamma}_j} \right] \sigma^{-2}_{\hat{\Gamma}_j}}{\sum_{j=1}^{p} \sigma^{-2}_{\hat{\Gamma}_j}}. \tag{6}$$

To establish the theoretical results for $\hat{\beta}_{\delta,pIVW}$ in the presence of horizontal pleiotropy, we replace $\hat{\Gamma}_j \sim N\left(\Gamma_j, \sigma^2_{\hat{\Gamma}_j}\right)$ in Assumption 2 by $\hat{\Gamma}_j \sim N\left(\Gamma_j, \sigma^2_{\hat{\Gamma}_j} + \tau^2\right)$. We further assume that $\tau^2 < c_1 \sigma^2_{\hat{\Gamma}_j}$ with a constant c_1 for all j, and that $\max_j \sigma^{-2}_{\hat{\Gamma}_j} < c_2 p^{-1} \sum_{j=1}^{p} \sigma^{-2}_{\hat{\Gamma}_j}$ for a constant c_2 in Theorem 2 (c). Then, Theorem 2 with $\hat{V}_\delta$ replaced by $\hat{V}^*_\delta$ is extended to the situation with horizontal pleiotropy .

4. Simulation Study

4.1 *Simulation Settings*

We generate the summary-level data for 1000 IVs from $\hat{\gamma}_j \sim N\left(\gamma_j, \sigma^2_{\hat{\gamma}_j}\right)$ and $\hat{\Gamma}_j \sim N\left(\Gamma_j, \sigma^2_{\hat{\Gamma}_j}\right)$ independently. For the true γ_j, we consider three scenarios similar to those in Ye et al. (2021):

(1) Scenario A (some strong IVs and many null IVs): 10 IVs have $\gamma_j \sim N\left(0, 0.03^2\right)$ and the rest of IVs have $\gamma_j = 0$. As such, the IV strength of 10 non-null IVs is around 15.54 and that of all 1000 IVs is around 0.16.

(2) Scenario B (many weak IVs and many null IVs): 100 IVs have $\gamma_j \sim N\left(0, 0.02^2\right)$ and the rest of IVs have $\gamma_j = 0$. As such, the IV strength of 100 non-null IVs is around 3.11 and that of all 1000 IVs is around 0.31.

(3) Scenario C (all weak IVs): all IVs have $\gamma_j \sim N\left(0, 0.02^2\right)$. The IV strength is around 4.07.

Then, we let $\Gamma_j = \beta\gamma_j + \alpha_j$, where $\alpha_j \sim N\left(0, \tau^2\right)$. We set $\beta = 0.5$, and $\tau = 0$ and 0.01 which represent the absence and the presence of horizontal pleiotropy, respectively. The variances $\sigma^2_{\hat{\gamma}_j}$ and $\sigma^2_{\hat{\Gamma}_j}$ are given by $\sigma^2_{\hat{\gamma}_j} = \frac{\mathrm{var}(X) - \gamma_j^2 \mathrm{var}(G_j)}{n_X \mathrm{var}(G_j)}$ and $\sigma^2_{\hat{\Gamma}_j} = \frac{\mathrm{var}(Y) - \Gamma_j^2 \mathrm{var}(G_j)}{n_Y \mathrm{var}(G_j)}$, where n_X and n_Y denote the sample sizes of the GWASs for the exposure and the outcome, respectively. We set $n_X = n_Y = 100,000$. For $\mathrm{var}(G_j)$, we let $G_j \sim Bin(2, \mathrm{MAF}_j)$ and randomly generate the minor allele frequencies from $\mathrm{MAF}_j \sim U(0.1, 0.5)$. For $\mathrm{var}(X)$ and $\mathrm{var}(Y)$, we calculate them from Equations (1) and (5) with the variances of U, ϵ_X and ϵ_Y being 2, respectively. Furthermore, we generate an independent dataset with $\hat{\gamma}_j^* \sim N\left(\gamma_j, \sigma^2_{\hat{\gamma}_j}/2\right)$ for the IV selection at threshold $\delta = \sqrt{2\log p} = 3.72$. The simulation is based on 10,000 replicates.

We first investigate the impact of the penalty parameter λ on the performance of the proposed pIVW estimator, where λ increases from 0 to 2.5 by 0.5. Then, we compare the proposed pIVW estimator with $\lambda = 1$ to other competing MR methods, including the IVW,

MR-Egger (Bowden et al., 2015), MR-Median (Bowden et al., 2016), MR-RAPS (Zhao et al., 2020) and dIVW estimators. The performances among various methods are compared in terms of the relative bias (bias divided by the true β) and the empirical standard error of the estimated causal effect, as well as the coverage probability of nominal 95% confidence interval.

4.2 *Simulation Results*

[Table 1 about here.]

The pIVW estimator has the smallest bias at $\lambda = 1$ under Scenarios A and B as summarized in Table 1. The empirical standard error of the pIVW estimator decreases as λ increases. In Scenario C with a larger IV strength, the pIVW estimators with various λ have nearly the same performance. We find similar results in the presence of horizontal pleiotropy and IV selection (see Web Tables 1-3). Therefore, we recommend $\lambda = 1$ for the pIVW estimator in practice.

[Table 2 about here.]

We next compare the pIVW estimator ($\lambda = 1$) against the other five competing MR methods without horizontal pleiotropy and IV selection with results summarized in Table 2. Specifically, the pIVW estimator has negligible bias which is the smallest among all six methods in Scenarios A and B with many null IVs. In contrast, the IVW, MR-Egger and MR-Median estimators have serious biases and poor coverage probabilities. The MR-RAPS and dIVW estimators have relatively large empirical standard errors in Scenario A, and we find that they yield extreme estimates in this case (see Web Figure 1 for the box plot). In Scenario C with many weak IVs, the proposed pIVW estimator performs similarly to the MR-RAPS and dIVW estimators, because the IV strength is larger in this case. However, the

IVW, MR-Egger and MR-Median estimators still have substantial biases since they cannot handle the weak IV bias.

[Table 3 about here.]

The results with IV selection at threshold $\delta = \sqrt{2 \log p}$ are given in Table 3. The proposed pIVW estimator still has the smallest bias among six methods and has smaller empirical standard error than the dIVW estimator. We find that, in Scenarios A and B with many null IVs, the IV selection improves the efficiencies of the MR-RAPS and dIVW and pIVW estimators. We obtain similar results in the presence of horizontal pleiotropy (Web Tables 4-5).

5. Real Data Applications to COVID-19 Outcomes

In this section, we focus on estimating the causal effects of five obesity-related exposures (i.e., peripheral vascular disease, dyslipidemia, hypertensive disease, type 2 diabetes and BMI) on three COVID-19 outcomes: (1) COVID-19 infection, (2) hospitalized COVID-19 and (3) critically ill COVID-19 (COVID-19 Host Genetics Initiative, 2021). The GWAS summary-level data for the three COVID-19 outcomes is obtained from the COVID-19 Host Genetics Initiative (COVID-19 Host Genetics Initiative, 2020), which includes up to 49,562 cases and two million controls from 47 distinct studies. For BMI, the selection dataset is from Akiyama et al. (2017) with 173,430 individuals and the exposure dataset is from UK BioBank with 359,983 individuals (Abbott et al., 2018). For the other four obesity-related exposures, the selection datasets are from the GWAS meta-analysis of Genetic Epidemiology Research on Adult Health and Aging (GERA) with 53,991 individuals (Zhu et al., 2018), and the exposure datasets are from the GWAS meta-analysis of UK BioBank with 108,039 individuals (Zhu et al., 2018). To exclude correlated IVs, we perform the linkage-disequilibrium clumping to remove the correlated genetic variants within 10Mb pairs and with the linkage disequilibrium $r^2 < 0.001$. More detailed data description is provided in the Web Table 6.

[Figure 2 about here.]

We find significant positive causal effects of hypertensive disease on hospitalized COVID-19, and BMI on the three COVID-19 outcomes at significance level 0.05. Our findings are concordant with some recent epidemiological studies (Popkin et al., 2020; Nakeshbandi et al., 2020). Some previous MR studies have also found significant causal effects of BMI on the COVID-19 outcomes (Ponsford et al., 2020; Leong et al., 2021), but there is still no MR analysis on the hypertensive disease to the best of our knowledge. Additionally, the pIVW estimator suggests that peripheral vascular disease is significantly associated with higher risks of three COVID-19 outcomes under the IV selection at threshold $\delta = \sqrt{2 \log p}$ (see Web Figure 2). To our knowledge, there is a lack of epidemiological studies and MR studies on the associations between peripheral vascular disease and the COVID-19 outcomes, despite a high incidence of peripheral vascular disease in COVID-19 patients (Hanff et al., 2020). For type 2 diabetes and dyslipidemia, the pIVW estimator does not find any evidence of associations with the three COVID-19 outcomes.

The other competing MR methods provide very different causal effect estimates, likely due to weak IV issues. For peripheral vascular disease with a relatively small IV strength ($\hat{\kappa} = 0.04$), the IVW, MR-Egger and MR-Median estimators have much smaller estimates than the pIVW estimator, since they might be biased toward zero in the presence of many weak IVs (Zhao et al., 2020; Ye et al., 2021). The dIVW estimator generally has much larger estimates than the pIVW estimator. In particular, the dIVW estimator yields a relatively large estimate (4.413) with an extreme estimated standard error (95.942) for the causal effect of peripheral vascular disease on hospitalized COVID-19. The inflated estimate of the dIVW estimator is likely due to the reason that its denominator is very close to zero in the presence of many weak IVs. For BMI with a larger IV strength ($\hat{\kappa} = 5.02$), there is a smaller discrepancy among these methods, and the pIVW estimator and the dIVW estimator provide similar results in this case. More detailed results can be found in Web Figure 2.

6. Discussion

The popular IVW estimator suffers from substantial bias in the presence of weak IVs, a common challenge in MR studies. In this paper, we develop a novel penalized IVW (pIVW) estimator to prevent the denominator of the ratio from being close to zero to reduce the bias due to the presence of many weak IVs. The proposed penalization method is inspired by Wang et al. (2020), which was used to derive a penalized confidence interval of a ratio estimator. Moreover, we allow for the horizontal pleiotropy by adjusting the variance estimation of the pIVW estimator.

The proposed pIVW estimator has some nice features. First, our theoretical and numerical results show that the proposed pIVW estimator has smaller bias and variance compared to the dIVW estimator (Ye et al., 2021). Second, the proposed pIVW estimator is consistent and asymptotically normal even in the presence of many weak IVs, and requires no more assumptions than the dIVW estimator. As the number of IVs and the IV strength increase, the pIVW estimator reduces to the dIVW estimator. Therefore, the dIVW estimator can be viewed as a special case of our proposed pIVW estimator. Third, the pIVW has a unique and closed-form solution, whereas many competing MR methods that are robust to the weak IVs do not have a closed-form solution and may have multiple diverging roots. In future work, we plan to extend the proposed penalization approach to other MR estimators to handle the weak IV bias.

Acknowledgements

The authors thank Professor LIU Lin for very useful comments that improved the presentation of the paper.

References

Abbott, L., Bryant, S., Churchhouse, C., et al. (2018). Round 2 GWAS results of thousands of phenotypes in the UK biobank. http://www.nealelab.is/uk-biobank/ Accessed

November 14, 2018.

Akiyama, M., Okada, Y., Kanai, M., et al. (2017). Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nature Genetics* **49,** 1458.

Bowden, J., Del Greco M, F., Minelli, C., et al. (2017). A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Statistics in Medicine* **36,** 1783–1802.

Bowden, J., Smith, G. D., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* **44,** 512–525.

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

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

Burgess, S. and Thompson, S. G. (2011). Bias in causal estimates from Mendelian randomization studies with weak instruments. *Statistics in Medicine* **30,** 1312–1323.

Burgess, S., Thompson, S. G., and CRP CHD Genetics Collaboration (2011). Avoiding bias from weak instruments in Mendelian randomization studies. *International Journal of Epidemiology* **40,** 755–764.

COVID-19 Host Genetics Initiative (2020). The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *European Journal of Human Genetics* **28,** 715.

COVID-19 Host Genetics Initiative (2021). Mapping the human genetic architecture of COVID-19 by worldwide meta-analysis. *MedRxiv* .

Davies, N. M., Scholder, S. V. K., Farbmacher, H., et al. (2015). The many weak instruments problem and Mendelian randomization. *Statistics in Medicine* **34,** 454–468.

Didelez, V. and Sheehan, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research* **16,** 309–330.

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* **20,** 533–534.

Hanff, T. C., Mohareb, A. M., Giri, J., et al. (2020). Thrombosis in COVID-19. *American Journal of Hematology* **95,** 1578–1589.

Hemani, G., Bowden, J., and Davey Smith, G. (2018). Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human Molecular Genetics* **27,** R195–R208.

Jordan, R. E., Adab, P., and Cheng, K. K. (2020). COVID-19: risk factors for severe disease and death. *BMJ* **368,** m1198.

Lawlor, D. A. (2016). Commentary: Two-sample Mendelian randomization: opportunities and challenges. *International Journal of Epidemiology* **45,** 908.

Lawlor, D. A., Harbord, R. M., Sterne, J. A., et al. (2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine* **27,** 1133–1163.

Leong, A., Cole, J. B., Brenner, L. N., et al. (2021). Cardiometabolic risk factors for COVID-19 susceptibility and severity: A Mendelian randomization analysis. *PLoS Medicine* **18,** e1003553.

Marsaglia, G. et al. (2006). Ratios of normal variables. *Journal of Statistical Software* **16,** 1–10.

Nakeshbandi, M., Maini, R., Daniel, P., et al. (2020). The impact of obesity on COVID-19 complications: a retrospective cohort study. *International Journal of Obesity* **44,**

1832–1837.

Pascarella, G., Strumia, A., Piliego, C., et al. (2020). COVID-19 diagnosis and management: a comprehensive review. *Journal of Internal Medicine* **288,** 192–206.

Piegorsch, W. W. and Casella, G. (1985). The existence of the first negative moment. *The American Statistician* **39,** 60–62.

Ponsford, M. J., Gkatzionis, A., Walker, V. M., et al. (2020). Cardiometabolic traits, sepsis, and severe COVID-19: A Mendelian randomization investigation. *Circulation* **142,** 1791–1793.

Popkin, B. M., Du, S., Green, W. D., et al. (2020). Individuals with obesity and COVID-19: A global perspective on the epidemiology and biological relationships. *Obesity Reviews* **21,** e13128.

Press, S. J. (1969). The t-ratio distribution. *Journal of the American Statistical Association* **64,** 242–252.

Sheehan, N. A., Didelez, V., Burton, P. R., et al. (2008). Mendelian randomisation and causal inference in observational epidemiology. *PLoS Medicine* **5,** e177.

Smith, G. D. and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* **32,** 1–22.

Smith, G. D. and Ebrahim, S. (2004). Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology* **33,** 30–42.

Vansteelandt, S., Bowden, J., Babanezhad, M., et al. (2011). On instrumental variables estimation of causal odds ratios. *Statistical Science* **26,** 403–422.

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

Wang, P., Xu, S., Wang, Y., et al. (2020). Penalized Fieller's confidence interval for the ratio of bivariate normal means. *Biometrics* .

Ye, T., Shao, J., and Kang, H. (2021). Debiased inverse-variance weighted estimator in two-sample summary-data Mendelian randomization. *Annals of Statistics (In press)* .

Zhao, Q., Chen, Y., Wang, J., et al. (2019). Powerful three-sample genome-wide design and robust statistical inference in summary-data Mendelian randomization. *International Journal of Epidemiology* **48,** 1478–1492.

Zhao, Q., Wang, J., Hemani, G., et al. (2020). Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Annals of Statistics* **48,** 1742–1769.

Zheng, J., Baird, D., Borges, M. C., et al. (2017). Recent developments in Mendelian randomization studies. *Current Epidemiology Reports* **4,** 330–345.

Zheng, Z., Peng, F., Xu, B., et al. (2020). Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *Journal of Infection* .

Zhu, Z., Zheng, Z., Zhang, F., et al. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications* **9,** 1–12.

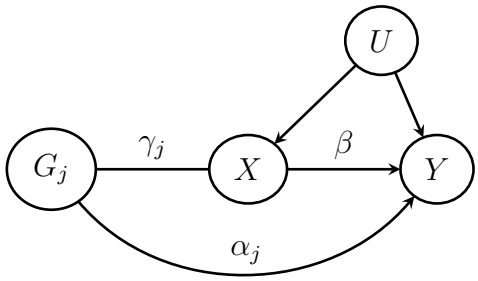

Figure 1. The relationships among the jth genetic variant G_j, the exposure X, the outcome Y and the unmeasured confounder U. The effect of G_j on X is γ_j, the direct effect (pleiotropic effect) of G_j on Y is α_j, and the causal effect of X on Y is β.

(a) Outcome: COVID-19 infection

	IVW	MR-Egger	MR-Median	MR-RAPS	dIVW	pIVW
Peripheral vascular disease (no. of IVs:1783, IV strength:0.04)	3e-03 (0.007) 0.672	3e-03 (0.01) 0.754	5e-03 (0.011) 0.610	0.097 (0.223) 0.664	0.102 (0.27) 0.706	0.057 (0.14) 0.685
Dyslipidemia (no. of IVs:2163, IV strength:0.82)	7e-03 (0.013) 0.567	1e-02 (0.017) 0.579	-4e-03 (0.025) 0.861	0.017 (0.033) 0.600	0.018 (0.032) 0.569	0.018 (0.032) 0.569
Hypertensive disease (no. of IVs:2149, IV strength:0.62)	0.025 (0.018) 0.148	0.032 (0.025) 0.195	0.019 (0.031) 0.532	0.083 (0.055) 0.131	0.074 (0.052) 0.151	0.074 (0.052) 0.151
Type 2 diabetes (no. of IVs:2155, IV strength:0.31)	-5e-03 (0.01) 0.586	-6e-03 (0.013) 0.634	-0.012 (0.018) 0.521	-0.023 (0.042) 0.593	-0.023 (0.042) 0.588	-0.022 (0.041) 0.587
Body mass index (no. of IVs:1893, IV strength:5.01)	0.095 (0.041) 0.020 *	0.119 (0.055) 0.031 *	0.038 (0.082) 0.641	0.116 (0.05) 0.021 *	0.116 (0.05) 0.021 *	0.115 (0.05) 0.021 *

* P<0.05

Estimated causal effect

0.08

0.04

0.00

(b) Outcome: hospitalized COVID-19

	IVW	MR-Egger	MR-Median	MR-RAPS	dIVW	pIVW
Peripheral vascular disease (no. of IVs:1783, IV strength:0.04)	7e-03 (0.013) 0.564	0.022 (0.019) 0.250	0.032 (0.02) 0.108	0.379 (0.55) 0.491	4.413 (95.942) 0.963	0.219 (0.429) 0.610
Dyslipidemia (no. of IVs:2163, IV strength:0.82)	-5e-03 (0.025) 0.850	-8e-03 (0.033) 0.812	-0.013 (0.046) 0.781	-0.013 (0.063) 0.830	-0.011 (0.059) 0.852	-0.011 (0.059) 0.851
Hypertensive disease (no. of IVs:2149, IV strength:0.62)	0.089 (0.033) 7e-03 *	0.068 (0.047) 0.147	0.109 (0.059) 0.064	0.241 (0.095) 0.011 *	0.246 (0.093) 9e-03 *	0.244 (0.093) 8e-03 *
Type 2 diabetes (no. of IVs:2155, IV strength:0.31)	-5e-04 (0.019) 0.981	-0.031 (0.026) 0.244	-0.038 (0.041) 0.352	-3e-03 (0.085) 0.973	-2e-03 (0.085) 0.981	-2e-03 (0.083) 0.977
Body mass index (no. of IVs:1893, IV strength:5.01)	0.382 (0.077) 8e-07 *	0.455 (0.105) 2e-05 *	0.549 (0.141) 1e-04 *	0.466 (0.097) 1e-06 *	0.468 (0.096) 1e-06 *	0.468 (0.096) 1e-06 *

* P<0.05

Estimated causal effect

0.6

0.4

0.2

0.0

(c) Outcome: critically ill COVID-19

	IVW	MR-Egger	MR-Median	MR-RAPS	dIVW	pIVW
Peripheral vascular disease (no. of IVs:1783, IV strength:0.04)	9e-03 (0.019) 0.645	0.014 (0.029) 0.618	0.030 (0.029) 0.305	0.320 (0.578) 0.580	0.464 (1.326) 0.726	0.202 (0.449) 0.653
Dyslipidemia (no. of IVs:2163, IV strength:0.82)	-8e-03 (0.036) 0.829	0.014 (0.047) 0.764	-0.022 (0.066) 0.737	-0.020 (0.078) 0.799	-0.017 (0.077) 0.831	-0.017 (0.077) 0.830
Hypertensive disease (no. of IVs:2149, IV strength:0.62)	0.092 (0.049) 0.062	0.051 (0.07) 0.473	0.081 (0.089) 0.363	0.247 (0.133) 0.064	0.244 (0.132) 0.063	0.243 (0.131) 0.063
Type 2 diabetes (no. of IVs:2155, IV strength:0.31)	-0.033 (0.028) 0.247	-0.065 (0.041) 0.109	-0.042 (0.058) 0.475	-0.146 (0.124) 0.239	-0.139 (0.122) 0.254	-0.138 (0.12) 0.251
Body mass index (no. of IVs:1893, IV strength:5.01)	0.366 (0.115) 1e-03 *	0.431 (0.155) 5e-03 *	0.391 (0.206) 0.057	0.439 (0.141) 2e-03 *	0.441 (0.14) 2e-03 *	0.441 (0.14) 2e-03 *

* P<0.05

Estimated causal effect

0.4

0.2

0.0

Figure 2. Estimated causal effects of five obesity-related exposures on (a) COVID-19 infection (b) hospitalized COVID-19 (c) critically ill COVID-19. No IV selection is conducted. In each cell, the first row displays the estimated causal effect and the estimated standard error (in bracket), and the second row displays the P value which is labeled with "$*$" when less than significance level 0.05. The pIVW estimator with penalty parameter $\lambda = 1$

Table 1

The performance of the pIVW estimator with various penalty parameter λ. The relative bias (bias divided by the true causal effect), the empirical standard error and the coverage probability of the 95% confidence interval. The true causal effect $\beta = 0.5$. No horizontal pleiotropy exists ($\tau = 0$). No IV selection is conducted. The simulation is based on 10,000 replicates

Scenario	Penalty parameter λ	Relative bias (%)	Empirical standard error	Coverage probability (%)
A (some strong and many null IVs)	0.0	15.3	1.272	96.1
	0.5	4.7	0.382	95.6
	1.0	-1.1	0.341	94.9
	1.5	-5.5	0.317	94.2
	2.0	-9.1	0.299	93.6
	2.5	-12.1	0.285	92.8
B (many weak and many null IVs)	0.0	3.4	0.190	96.1
	0.5	1.7	0.185	95.9
	1.0	0.2	0.181	95.7
	1.5	-1.2	0.177	95.3
	2.0	-2.6	0.173	95.0
	2.5	-3.8	0.170	94.6
C (all weak IVs)	0.0	0.0	0.025	95.2
	0.5	0.0	0.025	95.2
	1.0	0.0	0.025	95.2
	1.5	-0.1	0.025	95.1
	2.0	-0.1	0.025	95.1
	2.5	-0.2	0.025	95.1

Table 2

Comparison among six methods. The relative bias (bias divided by the true causal effect), the empirical standard error and the coverage probability of the 95% confidence interval. The true causal effect $\beta = 0.5$. No horizontal pleiotropy exists ($\tau = 0$). No IV selection is conducted. The simulation is based on 10,000 replicates

Scenario	Method	Relative bias (%)	Empirical standard error	Coverage probability (%)
A (some strong and many null IVs)	IVW	-86.5	0.039	0.0
	MR-Egger	-73.1	0.060	0.0
	MR-Median	-70.9	0.061	0.0
	MR-RAPS	9.2	1.595	94.2
	dIVW	15.3	1.272	96.1
	pIVW*	-1.1	0.341	94.9
B (many weak and many null IVs)	IVW	-76.8	0.037	0.0
	MR-Egger	-63.2	0.057	0.0
	MR-Median	-65.7	0.054	0.0
	MR-RAPS	1.6	0.167	95.0
	dIVW	3.4	0.190	96.1
	pIVW*	0.2	0.181	95.7
C (all weak IVs)	IVW	-19.7	0.019	0.1
	MR-Egger	-18.4	0.031	16.5
	MR-Median	-18.8	0.026	9.4
	MR-RAPS	0.0	0.024	95.1
	dIVW	0.0	0.025	95.2
	pIVW*	0.0	0.025	95.2

*The pIVW estimator with penalty parameter $\lambda = 1$

Table 3

Comparison among six methods. The relative bias (bias divided by the true causal effect), the empirical standard error and the coverage probability of the 95% confidence interval. The true causal effect $\beta = 0.5$. No horizontal pleiotropy exists ($\tau = 0$). The IV selection threshold $\delta = \sqrt{2 \log p}$. The simulation is based on 10,000 replicates

Scenario	Method	Relative bias (%)	Empirical standard error	Coverage probability (%)
A (some strong and many null IVs)	IVW	-1.8	0.116	95.3
	MR-Egger	-4.7	0.285	95.4
	MR-Median	-1.8	0.127	96.4
	MR-RAPS	0.8	0.120	95.7
	dIVW	1.5	0.121	95.6
	pIVW*	0.1	0.119	95.1
B (many weak and many null IVs)	IVW	-7.6	0.106	93.8
	MR-Egger	-43.6	0.277	87.0
	MR-Median	-11.6	0.131	95.6
	MR-RAPS	1.1	0.117	95.4
	dIVW	1.8	0.120	95.5
	pIVW*	0.4	0.118	95.2
C (all weak IVs)	IVW	-7.4	0.025	70.3
	MR-Egger	-40.6	0.064	11.3
	MR-Median	-10.8	0.033	74.4
	MR-RAPS	0.0	0.028	95.2
	dIVW	0.0	0.028	95.2
	pIVW*	0.0	0.028	95.3

*The pIVW estimator with penalty parameter $\lambda = 1$