Received: 8 October 2021 | Revised: 14 June 2022 | Accepted: 27 July 2022

DOI: 10.1111/biom.13732

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

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] | **Zhonghua Liu**[3]

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

[2]Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[3]Department of Biostatistics, Columbia University, New York, New York, USA

Correspondence
Zhonghua Liu, Department of Biostatistics, Columbia University, New York, NY 10032, USA.
Email: zl2509@cumc.columbia.edu

Abstract

Mendelian randomization 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 balanced 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.

KEYWORDS

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). It is crucial to identify causal risk factors associated with COVID-19 incidence and mortality so that we

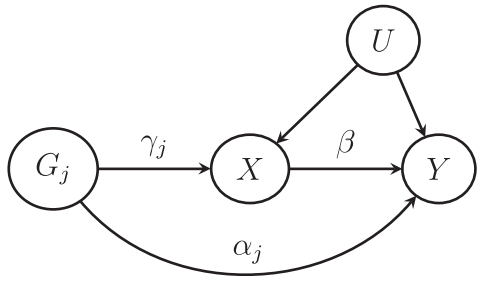

FIGURE 1 The relationships among the jth genetic variant G_j, the exposure X, the outcome Y, and the unmeasured confounder U. The effect of G_j on X is γ_j, the direct effect (pleiotropic effect) of G_j on Y is α_j, and the causal effect of X on Y is β

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 this challenge, 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 & 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 & Sheehan, 2007; Lawlor et al., 2008):

(1) IV relevance: the IV must be associated with the exposure;
(2) IV independence: the IV is independent of any confounder of the exposure-outcome relationship;
(3) Exclusion restriction: the IV affects the outcome only through the exposure.

When any one of these three IV assumptions is violated, conventional MR analysis may yield biased estimation of the causal effect. In particular, the IV relevance assumption can be nearly violated when the IVs are only weakly associated with the exposure variable (Burgess & 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 exclusion restriction assumption (Hemani et al., 2018; Verbanck 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 (Brockwell & Gordon, 2001). 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 not have unique estimates. 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 a heavy-tailed distribution and thus may not even have finite moments (Press, 1969; Piegorsch & 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 balanced horizontal pleiotropy by adjusting the variance estimation of the pIVW estimator. The proposed pIVW estimator has some attractive 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 (e.g., MR-RAPS) do not have a closed-form solution and might not have unique estimates in practice.

We demonstrate the improved performance of the proposed pIVW estimator compared to the other competing MR methods via 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_j, the exposure X, the outcome Y, and the unmeasured confounder U (as 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).

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, $\{\hat{\gamma}_j, \sigma_{\hat{\gamma}_j}\}_{j=1}^p$ and $\{\hat{\Gamma}_j, \sigma_{\hat{\Gamma}_j}\}_{j=1}^p$ 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 $\hat{\gamma}_j \sim N(\gamma_j, \sigma^2_{\hat{\gamma}_j})$ and $\hat{\Gamma}_j \sim N(\Gamma_j, \sigma^2_{\hat{\Gamma}_j})$ with known $\sigma^2_{\hat{\gamma}_j}$ and $\sigma^2_{\hat{\Gamma}_j}$, respectively (Zhao et al., 2020).

2.2 | The IVW estimator and debiased IVW (dIVW) estimator

The popular inverse-variance weighted 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 (Burgess et al., 2013):

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

Let $\mu_1 = \sum_{j=1}^p \sigma^{-2}_{\hat{\Gamma}_j}\gamma_j\Gamma_j$ and $\mu_2 = \sum_{j=1}^p \sigma^{-2}_{\hat{\Gamma}_j}\gamma_j^2$. 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

$$\hat{\beta}_{\mathrm{IVW}} \approx \frac{E\left(\sum_{j=1}^p \sigma^{-2}_{\hat{\Gamma}_j}\hat{\gamma}_j\hat{\Gamma}_j\right)}{E\left(\sum_{j=1}^p \sigma^{-2}_{\hat{\Gamma}_j}\hat{\gamma}_j^2\right)} = \frac{\mu_1}{\mu_2 + \sum_{j=1}^p \sigma^{-2}_{\hat{\Gamma}_j}\sigma^2_{\hat{\gamma}_j}} = \frac{\beta}{1 + \mu_2^{-1}\sum_{j=1}^p \sigma^{-2}_{\hat{\Gamma}_j}\sigma^2_{\hat{\gamma}_j}}.$$

When there is no measurement error for γ_j (i.e., $\sigma^2_{\hat{\gamma}_j} = 0$), we have $\hat{\beta}_{\mathrm{IVW}} \approx \beta$. 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 $\sigma^{-2}_{\hat{\gamma}_j}\gamma_j^2$ (Ye et al., 2021; Zhao et al., 2020).

To handle the bias due to weak IVs, the debiased IVW (dIVW) estimator (Ye et al., 2021) replaces the denominator in the IVW estimator by an unbiased estimator $\hat{\mu}_2$ of μ_2 as

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

The dIVW estimator has been shown to be consistent and asymptotically normal under weaker conditions than the IVW estimator. However, we find that the dIVW estimator is more likely to yield extreme estimates in the presence of weak IVs (as shown in Web Figure 1 under the simulation study in Section 4). It can be shown that the denominator of the dIVW estimator has the same variance as the denominator of the IVW estimator, but the expectation of the former is closer to zero than that of the latter (see Web Appendix A for details). Because zero is a singular point for the denominator of a ratio, it may result in the extreme estimates of the dIVW estimator in the presence of weak IVs. To overcome the limitations of the IVW estimator and the 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 estimator

Assume that the estimators $\hat{\mu}_1 = \sum_{j=1}^p \sigma^{-2}_{\hat{\Gamma}_j}\hat{\gamma}_j\hat{\Gamma}_j$ and $\hat{\mu}_2 = \sum_{j=1}^p \sigma^{-2}_{\hat{\Gamma}_j}(\hat{\gamma}_j^2 - \sigma^2_{\hat{\gamma}_j})$ jointly follow the following bivariate normal distribution

$$\begin{pmatrix}\hat{\mu}_1\\ \hat{\mu}_2\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}v_1 & v_{12}\\ v_{12} & v_2\end{pmatrix}\right). \tag{3}$$

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

$$l_p(\mu_1, \mu_2) = \log f(\mu_1, \mu_2) + \lambda \log |\mu_2|, \tag{4}$$

where $f(\mu_1, \mu_2)$ denotes the bivariate normal density function of $\hat{\boldsymbol{\mu}}=(\hat{\mu}_1, \hat{\mu}_2)'$ 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 $\lambda \log |\mu_2|$ (see Web Figure 2 for a graphical illustration). Therefore, an estimate of μ_2 being close to zero is less preferable by $l_p(\mu_1, \mu_2)$. Specifically, we obtain the following estimators of μ_1 and μ_2 by maximizing $l_p(\mu_1, \mu_2)$ with closed-form expressions:

$$\tilde{\mu}_1 = \hat{\mu}_1 + \frac{\hat{\upsilon}_{12}}{\hat{\upsilon}_2}(\tilde{\mu}_2 - \hat{\mu}_2),$$

$$\tilde{\mu}_2 = \left(\frac{1}{2} + \sqrt{\frac{1}{4} + \lambda \hat{c}_\upsilon^2}\right)\hat{\mu}_2 = r_\lambda \hat{\mu}_2,$$

where $r_\lambda = 1/2 + \sqrt{1/4 + \lambda \hat{c}_\upsilon^2}$, $\hat{c}_\upsilon = \sqrt{\hat{\upsilon}_2}/\hat{\mu}_2$ is the estimated coefficient of variation of $\hat{\mu}_2$, $\hat{\upsilon}_{12} = 2\sum_{j=1}^p \sigma_{\hat{\Gamma}_j}^{-4}\sigma_{\hat{\gamma}_j}^2 \hat{\Gamma}_j \hat{\gamma}_j$ and $\hat{\upsilon}_2 = 2\sum_{j=1}^p \sigma_{\hat{\Gamma}_j}^{-4}\sigma_{\hat{\gamma}_j}^4 \{2(\hat{\gamma}_j^2 - \sigma_{\hat{\gamma}_j}^2)\sigma_{\hat{\gamma}_j}^{-2} + 1\}$ are the estimators of υ_{12} and υ_2, respectively (see Web Appendix B for detailed derivation). Note that the bivariate normality assumption of $\hat{\boldsymbol{\mu}}$ can be relaxed. In fact, we can take $\tilde{\mu}_1$ and $\tilde{\mu}_2$ as the estimators that minimize the squared error loss function $||\boldsymbol{\Sigma}^{-1/2}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})||_2^2$ with a penalty term $-2\lambda \log |\mu_2|$, that is, $||\boldsymbol{\Sigma}^{-1/2}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})||_2^2 - 2\lambda \log |\mu_2|$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the mean and the covariance matrix of $\hat{\boldsymbol{\mu}}$, respectively. Then we propose the following penalized IVW (pIVW) estimator as a ratio of $\tilde{\mu}_1$ and $\tilde{\mu}_2$

$$\hat{\beta}_{\text{pIVW}} = \frac{\tilde{\mu}_1}{\tilde{\mu}_2} = \frac{1}{r_\lambda}\hat{\beta}_{\text{dIVW}} + \frac{\hat{\upsilon}_{12}}{\hat{\upsilon}_2}\left(1 - \frac{1}{r_\lambda}\right). \tag{5}$$

Note that r_λ acts like a correction factor for $\hat{\beta}_{\text{dIVW}}$. When the penalty parameter $\lambda = 0$, the correction factor $r_\lambda = 1$ and then $\hat{\beta}_{\text{pIVW}}$ reduces to $\hat{\beta}_{\text{dIVW}}$. When $\lambda > 0$, we can see that $r_\lambda > 1$ and r_λ increases with the estimated coefficient of variation $\hat{c}_\upsilon$. Therefore, when the estimated coefficient of variation $\hat{c}_\upsilon$ is large (e.g., in the presence of many weak IVs), $\hat{\beta}_{\text{pIVW}}$ adjusts $\hat{\beta}_{\text{dIVW}}$ by r_λ to prevent the denominator $\tilde{\mu}_2$ from being close to zero (see Web Figure 3 for the difference between $\hat{\beta}_{\text{pIVW}}$ and $\hat{\beta}_{\text{dIVW}}$ against various r_λ). A good numerical example can be found in Section 5 (Table 4), where $\hat{\beta}_{\text{dIVW}}$ yields an extreme estimate of the causal effect of peripheral vascular disease on hospitalized COVID-19, and $\hat{\beta}_{\text{pIVW}}$ adjusts it by the correction factor r_λ in the case where no IV selection is performed.

To study the asymptotic properties of the pIVW estimator, we make the following Assumptions 1 and 2, which were 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. *$\{\hat{\gamma}_j, \hat{\Gamma}_j\}_{j=1}^p$ are independently distributed as $\hat{\gamma}_j \sim N(\gamma_j, \sigma_{\hat{\gamma}_j}^2)$ and $\hat{\Gamma}_j \sim N(\beta\gamma_j, \sigma_{\hat{\Gamma}_j}^2)$ 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 $j = 1, \ldots, p$.*

Assumptions 1 and 2 are reasonable in the two-sample MR design settings since the GWASs often have large sample sizes and a large number of genetic variants. The independence of $\{\hat{\gamma}_j, \hat{\Gamma}_j\}_{j=1}^p$ across genetic variants can be achieved by the linkage-disequilibrium clumping (Purcell et al., 2007).

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 is estimated by $\hat{\kappa} = p^{-1}\sum_{j=1}^p (\hat{\gamma}_j^2 - \sigma_{\hat{\gamma}_j}^2)\sigma_{\hat{\gamma}_j}^{-2}$. We also follow Ye et al. (2021) to define the effective sample size $\eta = \kappa\sqrt{p}$. Note that the effective sample size η is determined by the IV strength and the number of IVs in the summary-level data, which is not the sample size of the original individual-level data in GWASs. Under Assumptions 1 and 2, it can be shown that $\hat{c}_\upsilon^2 = O_p(\xi)$ and $\hat{\beta}_{\text{pIVW}} - \hat{\beta}_{\text{dIVW}} = O_p(\xi)$, where $\xi = (\eta\sqrt{p})^{-1} + \eta^{-2}$ converges to zero as $\eta \to \infty$. Therefore, given the consistency of $\hat{\beta}_{\text{dIVW}}$, it is straightforward that $\hat{\beta}_{\text{pIVW}}$ is also consistent as $\eta \to \infty$. In the following Theorem 1(a), we show that the bias of $\hat{\beta}_{\text{pIVW}}$ converges to zero at a faster rate than that of $\hat{\beta}_{\text{dIVW}}$ under the optimal $\lambda_{opt} = 1$. In Theorem 1(b) together with Remark 2, we show that the variance of $\hat{\beta}_{\text{pIVW}}$ is smaller than that of $\hat{\beta}_{\text{dIVW}}$ when $\lambda > 0$. In Theorem 1(c), we also establish the asymptotic normality of $\hat{\beta}_{\text{pIVW}}$, which requires no more assumptions comparing to $\hat{\beta}_{\text{dIVW}}$.

Theorem 1. *Suppose that Assumptions 1 and 2 hold and the effective sample size $\eta \to \infty$. Then, we have the following results:*

(a) The bias of $\hat{\beta}_{\text{dIVW}}$ is of order $O(\xi)$, and the bias of $\hat{\beta}_{\text{pIVW}}$ is

$$E(\hat{\beta}_{\text{pIVW}} - \beta) = (1 - \lambda)E(\hat{\beta}_{\text{dIVW}} - \beta) + o(\xi).$$

The optimal $\lambda_{opt} = 1$ minimizes the absolute bias of $\hat{\beta}_{\text{pIVW}}$, which is only of order $o(\xi)$.

(b) The variances of $\hat{\beta}_{\text{pIVW}}$ *and* $\hat{\beta}_{\text{dIVW}}$ *are both of order* $O(\xi)$*. But the difference between the variances of* $\hat{\beta}_{\text{dIVW}}$ *and* $\hat{\beta}_{\text{pIVW}}$ *is*

$$\text{Var}(\hat{\beta}_{\text{dIVW}}) - \text{Var}(\hat{\beta}_{\text{pIVW}}) = \frac{2\lambda\beta^2}{\mu_2^4}\Delta + o(\xi^2),$$

$$\text{where } \Delta = \frac{3(\mu_1\upsilon_2 - \mu_2\upsilon_{12})^2}{\mu_2^2\beta^2} + \frac{\upsilon_1\upsilon_2 - \upsilon_{12}^2}{\beta^2}$$

$$+8\mu_2\left(\frac{\upsilon_{12}}{\upsilon_2\beta} - 1\right)\sum_{j=1}^{p}\sigma_{\hat{\gamma}_j}^4\sigma_{\hat{\Gamma}_j}^{-6}(\gamma_j + \sigma_{\hat{\gamma}_j}^2).$$

(c) Further assume that $\max_j \gamma_j^2\sigma_{\hat{\gamma}_j}^{-2}/(\kappa p + p) \to 0$*. Then,* $\hat{\beta}_{\text{pIVW}}$ *is asymptotically normal*

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

where

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

The proof of Theorem 1 is provided in Web Appendix C.

Remark 1. Theorem 1(a) states that $\hat{\beta}_{\text{pIVW}}$ has smaller absolute bias than $\hat{\beta}_{\text{dIVW}}$ when $0 < \lambda < 2$. In particular, the bias of $\hat{\beta}_{\text{pIVW}}$ with $\lambda_{opt} = 1$ converges to zero at a faster rate than that of $\hat{\beta}_{\text{dIVW}}$.

Remark 2. Theorem 1(b) shows that $\text{Var}(\hat{\beta}_{\text{pIVW}})$ is smaller than $\text{Var}(\hat{\beta}_{\text{dIVW}})$ when $\Delta > 0$. In fact, we have shown that $\Delta > 0$ is generally true for complex traits, of which a single genetic variant can only explain a very small amount of total variances (Boyle et al., 2017; Park et al., 2010; Shi et al., 2016). More technical details can be found in Web Appendix D. Therefore, when $\lambda > 0$, $\text{Var}(\hat{\beta}_{\text{pIVW}})$ is smaller than $\text{Var}(\hat{\beta}_{\text{dIVW}})$ in MR settings. Together with Theorem 1(a), $\hat{\beta}_{\text{pIVW}}$ with $\lambda_{opt} = 1$ has smaller bias and variance than $\hat{\beta}_{\text{dIVW}}$.

Remark 3. In Theorem 1(c), we show that $\hat{\beta}_{\text{pIVW}}$ is asymptotically normal as $\eta \to \infty$. Therefore, the confidence interval of β can be derived from the normal approximation of $\hat{\beta}_{\text{pIVW}}$. Alternatively, we can derive the confidence interval of β based on bootstrapping Fieller's method (Fieller, 1954; Hwang & Hwang, 1995), which has been shown to have better coverage level than that based on the normal approximation. More details can be found in Web Appendix E.

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 the effective sample size $\eta \to \infty$, which means that it allows the presence of many weak IVs.

3.2 | Selection of candidate instruments

In this section, we extend Theorem 1 to the setting 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 $\{\hat{\gamma}_j^*, \sigma_{\hat{\gamma}_j}^*\}_{j=1}^p$ that is independent of the exposure and the outcome datasets. 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 follow Ye et al. (2021) to 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 $q_{\delta,j} = P(|\hat{\gamma}_j^*| > \delta\sigma_{\hat{\gamma}_j}^*)$ and $p_\delta = \sum_{j=1}^p q_{\delta,j}$. Let $S_\delta = \{j : |\hat{\gamma}_j^*| > \delta\sigma_{\hat{\gamma}_j}^*\}$ be the set of selected IVs, and $\hat{p}_\delta$ be the number of selected IVs within S_δ. Then, we can estimate κ_δ by $\hat{\kappa}_\delta = \hat{p}_\delta^{-1}\sum_{j\in S_\delta}(\hat{\gamma}_j^2 - \sigma_{\hat{\gamma}_j}^2)\sigma_{\hat{\gamma}_j}^{-2}$. In the IV selection setting, we define the effective sample size $\eta_\delta = \kappa_\delta\sqrt{p_\delta}/\max(1,\varphi)$ and $\xi_\delta = (\eta_\delta\sqrt{p}_\delta\max(1,\varphi))^{-1} + \eta_\delta^{-2}$, where $\varphi = \sqrt{p_\delta^{-1}\sum_{j=1}^p \gamma_j^4\sigma_{\hat{\gamma}_j}^{-4}q_{\delta,j}(1-q_{\delta,j})}$. 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. $\{\hat{\gamma}_j^*, \hat{\gamma}_j, \hat{\Gamma}_j\}_{j=1}^p$ *are mutually independent and* $\hat{\gamma}_j^* \sim N(\gamma_j, \sigma_{\hat{\gamma}_j}^{*2})$ *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* $j = 1, \ldots, p$*.*

Given a selection threshold δ, we evaluate the dIVW estimator $\hat{\beta}_{\delta,\text{dIVW}} = \hat{\mu}_{1,\delta}/\hat{\mu}_{2,\delta}$ and the proposed pIVW estimator $\hat{\beta}_{\delta,\text{pIVW}} = \tilde{\mu}_{1,\delta}/\tilde{\mu}_{2,\delta}$ using the selected IVs within the set S_δ. Under Assumptions 1–3, we have $\hat{\beta}_{\delta,\text{pIVW}} - \hat{\beta}_{\delta,\text{dIVW}} = O_p(\xi_\delta)$. The following Theorem 2 shows that the

asymptotic properties of the pIVW estimator in Theorem 1 still hold under the IV selection as the effective sample size $\eta_\delta \to \infty$.

Theorem 2. *Suppose that Assumptions 1–3 hold and the effective sample size $\eta_\delta \to \infty$. Then, we have the following results:*

(a) The bias of $\hat{\beta}_{\delta,\text{dIVW}}$ is of order $O(\xi_\delta)$, and the bias of $\hat{\beta}_{\delta,\text{pIVW}}$ is

$$E(\hat{\beta}_{\delta,\text{pIVW}} - \beta) = (1-\lambda)E(\hat{\beta}_{\delta,\text{dIVW}} - \beta) + o(\xi_\delta).$$

The optimal $\lambda_{opt} = 1$ minimizes the absolute bias of $\hat{\beta}_{\delta,\text{pIVW}}$, which is only of order $o(\xi_\delta)$.

(b) The variances of $\hat{\beta}_{\delta,\text{pIVW}}$ and $\hat{\beta}_{\delta,\text{dIVW}}$ are both of order $O(\xi_\delta)$. But the difference between the variances of $\hat{\beta}_{\delta,\text{dIVW}}$ and $\hat{\beta}_{\delta,\text{pIVW}}$ is

$$\text{Var}(\hat{\beta}_{\delta,\text{dIVW}}) - \text{Var}(\hat{\beta}_{\delta,\text{pIVW}}) = \frac{2\lambda\beta^2}{\mu_{2,\delta}^4}\Delta_\delta + o\left(\xi_\delta^2\right),$$

$$\text{where } \Delta_\delta = \frac{3(\mu_{1,\delta}\upsilon_{2,\delta} - \mu_{2,\delta}\upsilon_{12,\delta})^2}{\mu_{2,\delta}^2\beta^2} + \frac{\upsilon_{1,\delta}\upsilon_{2,\delta} - \upsilon_{12,\delta}^2}{\beta^2}$$

$$+8\mu_{2,\delta}\left(\frac{\upsilon_{12,\delta}}{\upsilon_{2,\delta}\beta} - 1\right)\sum_{j=1}^{p}\sigma_{\hat{\gamma}_j}^4\sigma_{\hat{\Gamma}_j}^{-6}(\gamma_j + \sigma_{\hat{\gamma}_j}^2)q_{\delta,j}.$$

(c) Further assume that $\max_j \gamma_j^2\sigma_{\hat{\gamma}_j}^{-2}q_{\delta,j}/(\kappa_\delta p_\delta + p_\delta) \to 0$. Then, $\hat{\beta}_{\delta,\text{pIVW}}$ is asymptotically normal

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

where

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

The proof of Theorem 2 is provided in Web Appendix F. Theorem 2 shows that $\hat{\beta}_{\delta,\text{pIVW}}$ has smaller absolute bias than $\hat{\beta}_{\delta,\text{dIVW}}$ when $0 < \lambda < 2$. The bias of $\hat{\beta}_{\delta,\text{pIVW}}$ with the optimal $\lambda_{opt} = 1$ converges to zero faster than that of $\hat{\beta}_{\delta,\text{dIVW}}$. We also prove that $\Delta_\delta > 0$ generally holds in the genetic studies, and therefore $\text{Var}(\hat{\beta}_{\delta,\text{pIVW}})$ is smaller than $\text{Var}(\hat{\beta}_{\delta,\text{dIVW}})$ when $\lambda > 0$. The pIVW estimator is still 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. Note that, the independent datasets for IV selection might not be available for some traits in practice. However, the pIVW estimator is still useful in this case, since it can handle the weak IV bias even without IV selection as shown in Theorem 1.

3.3 | Accounting for balanced horizontal pleiotropy

When there exists horizontal pleiotropy (i.e., nonzero 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):

$$Y = \beta X + \sum_{j=1}^{p}\alpha_j G_j + U + \epsilon_Y, \tag{6}$$

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, we have $\hat{\Gamma}_j \sim N(\beta\gamma_j, \sigma_{\hat{\Gamma}_j}^2 + \tau^2)$ in the presence of balanced horizontal pleiotropy. To account for the balanced horizontal pleiotropy, we estimate the variance of $\hat{\beta}_{\delta,\text{pIVW}}$ by

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

where we follow Ye et al. (2021) to derive the estimator of τ^2 as

$$\hat{\tau}^2 = \frac{\sum_{j=1}^{p}\left\{\left(\hat{\Gamma}_j - \hat{\beta}_{\text{pIVW}}\hat{\gamma}_j\right)^2 - \sigma_{\hat{\Gamma}_j}^2 - \hat{\beta}_{\text{pIVW}}^2\sigma_{\hat{\gamma}_j}^2\right\}\sigma_{\hat{\Gamma}_j}^{-2}}{\sum_{j=1}^{p}\sigma_{\hat{\Gamma}_j}^{-2}}.$$

To establish the theoretical results for the pIVW estimator in the presence of balanced horizontal pleiotropy, we replace $\hat{\Gamma}_j \sim N(\beta\gamma_j, \sigma_{\hat{\Gamma}_j}^2)$ by $\hat{\Gamma}_j \sim N(\beta\gamma_j, \sigma_{\hat{\Gamma}_j}^2 + \tau^2)$ in Assumption 2, and assume that $\tau^2 < c_1\sigma_{\hat{\Gamma}_j}^2$ with a constant $c_1 > 0$ for all j. We further assume that $\max_j \sigma_{\hat{\Gamma}_j}^{-2} < c_2 p^{-1}\sum_{j=1}^{p}\sigma_{\hat{\Gamma}_j}^{-2}$ for a constant $c_2 > 0$ in Theorems 1(c) and 2(c). Then, Theorems 1 and 2 can be extended to the situation with balanced horizontal pleiotropy. The proofs are provided in Web Appendices C and F, respectively.

4 | SIMULATION STUDY

4.1 | Simulation settings

We generate the summary-level data for 1000 IVs from $\hat{\gamma}_j \sim N(\gamma_j, \sigma_{\hat{\gamma}_j}^2)$ and $\hat{\Gamma}_j \sim N(\Gamma_j, \sigma_{\hat{\Gamma}_j}^2)$ independently. For the true γ_j, we consider a scenario with many weak IVs and many null IVs as in Ye et al. (2021), where we randomly generate $\gamma_j \sim N(0, 0.02^2)$ for the weak IVs and let $\gamma_j = 0$ for the null

IVs. We set the proportion of null IVs to be 95%, 90%, and 80% corresponding to the effective sample size η around 4.33, 9.52, and 21.85, respectively. Then, we let $\Gamma_j = \beta\gamma_j + \alpha_j$, where $\alpha_j \sim N(0, \tau^2)$. We set $\beta = 0.5$, and $\tau = 0$, and 0.01 that represent the absence and the presence of balanced horizontal pleiotropy, respectively. The variances $\sigma^2_{\hat{\gamma}_j}$ and $\sigma^2_{\hat{\Gamma}_j}$ are given by $\sigma^2_{\hat{\gamma}_j} = \{\text{Var}(X) - \gamma_j^2\text{Var}(G_j)\}/\{n_X\text{Var}(G_j)\}$ and $\sigma^2_{\hat{\Gamma}_j} = \{\text{Var}(Y) - \Gamma_j^2\text{Var}(G_j)\}/\{n_Y\text{Var}(G_j)\}$, where n_X and n_Y denote the sample sizes of the GWASs for the exposure and the outcome, respectively. We set $n_X = 0.5n_Y = 100,000$. For $\text{Var}(G_j)$, we let $G_j \sim Bin(2, \text{MAF}_j)$ and randomly generate the minor allele frequencies from $\text{MAF}_j \sim U(0.1, 0.5)$. For $\text{Var}(X)$ and $\text{Var}(Y)$, we calculate them from Equations (1) and (6) with the variances of U, ϵ_X and ϵ_Y being 2, respectively. Furthermore, we generate an independent dataset with $\hat{\gamma}_j^* \sim N(\gamma_j, 0.5\sigma^2_{\hat{\gamma}_j})$ 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 $\lambda_{opt} = 1$ to other competing MR methods, including the IVW, the MR-Egger (Bowden et al., 2015), the MR-Median (Bowden et al., 2016), the MR-RAPS (Zhao et al., 2020), and the 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. For the pIVW estimator, we present the coverage probability of bootstrapping Fieller's confidence interval in this simulation. In Section 4.3, we also compare bootstrapping Fieller's interval with the confidence interval derived from the normal approximation of the pIVW estimator under a wide range of parameter settings.

4.2 | Simulation results

The pIVW estimator has the smallest bias at $\lambda = 1$ as summarized in Table 1. The empirical standard error of the pIVW estimator decreases as λ increases. As the effective sample size η increases, the value of λ tends to have less influence on the performance of the pIVW estimator. We find similar results in the presence of balanced horizontal pleiotropy and IV selection (see Web Tables 1 and 3). Therefore, we recommend to choose the optimal $\lambda_{opt} = 1$ for the pIVW estimator in practice due to its smallest bias.

We next compare the pIVW estimator ($\lambda_{opt} = 1$) against the other five competing MR methods under the situations without horizontal pleiotropy and IV selection with the results summarized in Table 2. The pIVW estimator has negligible bias that is the smallest among all six methods. In contrast, the IVW, the MR-Egger, and the MR-Median estimators have serious biases and poor coverage probabilities. The MR-RAPS and the dIVW estimators have relatively large empirical standard errors when the effective sample size η is small ($\eta = 4.33$), and we find that they yield some extreme estimates in this case (see Web Figure 1). As η increases, the differences in the performance among the MR-RAPS, the dIVW, and the pIVW estimators become smaller.

TABLE 1 The pIVW estimator with various penalty parameter λ. 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. Bias (%): bias divided by β; CP (%): coverage probability of the 95% confidence interval; SE: empirical standard error

η	λ	Bias	SE	CP
4.33	0	22.4	2.156	94.8
	0.5	3.7	0.341	94.3
	1	−3.3	0.293	94.0
	1.5	−8.3	0.267	93.4
	2	−12.3	0.248	92.8
	2.5	−15.6	0.234	91.9
9.52	0	3.0	0.148	94.8
	0.5	1.3	0.143	94.6
	1	−0.2	0.139	94.6
	1.5	−1.6	0.136	94.4
	2	−2.9	0.133	94.2
	2.5	−4.2	0.130	93.9
21.85	0	0.7	0.069	94.6
	0.5	0.4	0.068	94.5
	1	0.0	0.068	94.5
	1.5	−0.3	0.067	94.4
	2	−0.7	0.067	94.3
	2.5	−1.0	0.067	94.1

The results with IV selection at threshold $\delta = \sqrt{2\log p} = 3.72$ 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 obtain similar results in the presence of balanced horizontal pleiotropy (see Web Tables 4 and 5).

We conduct an additional simulation study to mimic the individual-level data-generating mechanisms in GWASs. We first simulate the individual-level data based on the linear structural models (1) and (6). Then, we obtain the summary-level data by estimating the marginal effects and their standard errors from the linear regressions as in Ye et al. (2021) and Wang et al. (2022). We have

TABLE 2 Comparison of the pIVW estimator ($\lambda_{opt} = 1$) with other competing MR methods. 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. Bias (%): bias divided by β; CP (%): coverage probability of the 95% confidence interval; SE: empirical standard error

η	Method	Bias	SE	CP
4.33	IVW	−88.0	0.028	0.0
	MR-Egger	−80.2	0.045	0.0
	MR-Median	−83.1	0.041	0.0
	MR-RAPS	7.4	0.743	93.5
	dIVW	22.4	2.156	94.6
	pIVW	−3.3	0.293	94.0
9.52	IVW	−76.9	0.027	0.0
	MR-Egger	−63.5	0.042	0.0
	MR-Median	−67.6	0.040	0.0
	MR-RAPS	1.2	0.120	94.7
	dIVW	3.0	0.148	95.5
	pIVW	−0.2	0.139	94.6
21.85	IVW	−59.1	0.024	0.0
	MR-Egger	−42.0	0.036	0.0
	MR-Median	−46.7	0.036	0.0
	MR-RAPS	0.3	0.060	94.8
	dIVW	0.7	0.069	94.8
	pIVW	0.0	0.068	94.5

TABLE 3 Comparison of the pIVW estimator ($\lambda_{opt} = 1$) with other competing MR methods. 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. Bias (%): bias divided by β; CP (%): coverage probability of the 95% confidence interval; SE: empirical standard error

η_δ	Method	Bias	SE	CP
6.76	IVW	−7.3	0.122	93.4
	MR-Egger	−41.9	0.376	89.3
	MR-Median	−11.6	0.140	95.1
	MR-RAPS	1.4	0.137	95.8
	dIVW	3.2	0.144	96.0
	pIVW	−0.1	0.136	95.3
10.26	IVW	−7.9	0.079	91.4
	MR-Egger	−43.6	0.207	79.2
	MR-Median	−12.2	0.097	93.0
	MR-RAPS	0.7	0.088	95.5
	dIVW	1.4	0.090	95.5
	pIVW	0.1	0.088	95.1
17.84	IVW	−7.9	0.049	87.4
	MR-Egger	−48.6	0.130	51.2
	MR-Median	−12.5	0.062	88.3
	MR-RAPS	0.4	0.055	94.9
	dIVW	0.7	0.056	95.0
	pIVW	0.2	0.055	94.7

similar findings in the simulation with individual-level data. More details are given in Web Appendix G and Web Tables 6–9.

4.3 | Empirical guidelines on η for asymptotics

In Theorem 1, the asymptotic properties of $\hat{\beta}_{\text{pIVW}}$ require the effective sample size $\eta \to \infty$. To investigate how large of η is enough for the asymptotics, we conduct further simulations for $\hat{\beta}_{\text{pIVW}}$ with $\lambda_{opt} = 1$ under a wide range of parameter settings, including: (1) $\gamma_j \sim N(0, \sigma^2)$ with σ varying from 0.01 to 0.05, (2) the proportion of null IVs from 0 to 99%, and (3) $n_x = cn_y = 100,000$ with c ranging from 0.1 to 10. The results show that the relative bias of $\hat{\beta}_{\text{pIVW}}$ decreases more rapidly than that of $\hat{\beta}_{\text{dIVW}}$ as η increases (see Figure 2(A)). $\hat{\beta}_{\text{pIVW}}$ is nearly unbiased when $\eta > 5$, while $\hat{\beta}_{\text{dIVW}}$ requires $\eta > 15$ to have a negligible bias. The variance of $\hat{\beta}_{\text{pIVW}}$ is much smaller than that of $\hat{\beta}_{\text{dIVW}}$ when $\eta < 5$, and they get close to each other as η increases (see Figure 2(B)). The confidence interval derived from the normal approximation of $\hat{\beta}_{\text{pIVW}}$ maintains nominal coverage probability when $\eta > 10$, while bootstrapping Fieller's confidence interval maintains nominal coverage probability when $\eta > 5$ (see Figure 2(C)).

We have similar findings for η_δ under the IV selection (see Web Figure 4), where we consider a grid of two additional parameters, including the selection threshold δ from 1 to 4 and $\hat{\gamma}_j^* \sim N(\gamma_j, c\sigma_{\hat{\gamma}_j}^2)$ with c from 0.1 to 10. We also find similar results in the presence of balanced horizontal pleiotropy (see Web Figures 5 and 6). Therefore, we recommend that η (or η_δ) should be larger than 5 for the pIVW estimator to have a negligible bias and a nominal coverage probability of the confidence interval. On the other hand, although the bias and variance of the pIVW estimator might not be negligible when η (or η_δ) is less than 5, they are still much smaller than those of the dIVW estimator (see Figure 2(A) and (B)).

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

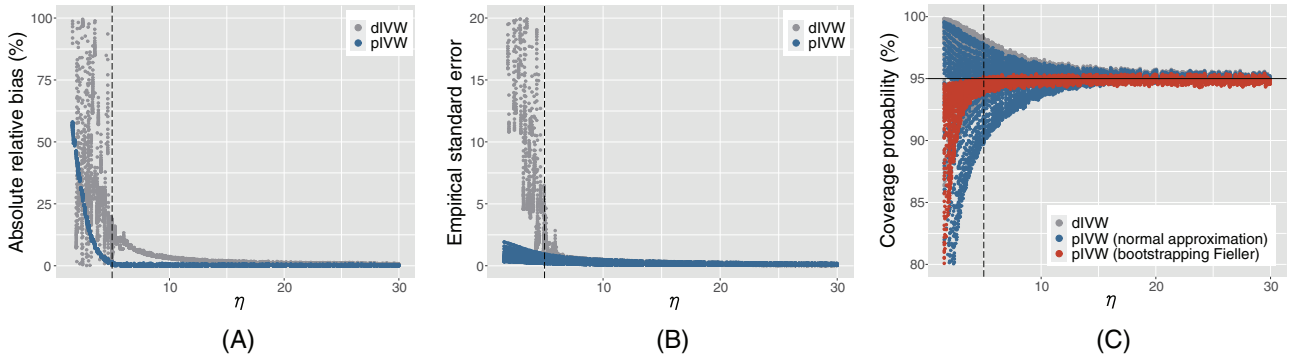

FIGURE 2 The plots of (A) the absolute relative biases (biases divided by β); (B) the empirical standard errors; and (C) the coverage probabilities of the 95% confidence intervals for the dIVW estimator and the pIVW estimator ($\lambda_{opt} = 1$) against the effective sample size η. The dashed line shows $\eta = 5$. The dots represent the simulation results under different settings of parameters based on 10,000 replicates. There is no horizontal pleiotropy ($\tau = 0$) or IV selection. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

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). More detailed data description is provided in Web Table 10. To exclude correlated IVs, we perform the linkage-disequilibrium clumping to remove the correlated genetic variants within 10 Mb pairs and with the linkage disequilibrium $r^2 < 0.001$. The numbers of IVs included into the analysis are from 1768 to 2338 for different datasets (see Web Table 11 for details).

From the results of the pIVW estimator with $\lambda_{opt} = 1$, 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 (see Table 4 and Web Figures 7 and 8). Our findings agree with some recent epidemiological studies (Nakeshbandi et al., 2020; Popkin et al., 2020). Some previous MR studies have also found significant causal effects of BMI on the COVID-19 outcomes (Leong et al., 2021; Ponsford et al., 2020), 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 $\delta = \sqrt{2 \log p} = 3.87$. To our knowledge, there is a lack of MR studies about 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. More results can be found in Web Table 11 and Web Figures 7 and 8.

The other competing MR methods provide very different causal effect estimates when the estimated effective sample size $\hat{\eta}$ or $\hat{\eta}_\delta$ is small (see Table 4; see Web Appendix H for the estimation of $\hat{\eta}$ and $\hat{\eta}_\delta$). For peripheral vascular disease with a very small $\hat{\eta}$ when no IV selection is performed, the IVW estimator has a very small estimate (0.007) that might be biased toward zero, because its denominator is a biased estimator of μ_2 and might overestimate μ_2 in the presence of many weak IVs (Ye et al., 2021). In contrast, the dIVW estimator yields a relatively large estimate (4.413) with an extreme estimated standard error (95.942), which possibly overestimates the causal effect due to the presence of many weak IVs. In this case, the pIVW estimator adjusts the dIVW estimator by the correction factor $r_\lambda = 20.86$, and provides a causal effect estimate being 0.219 with the estimated standard error being 0.429. After we perform IV selection to remove the weak IVs, the estimates from all the methods are in similar magnitudes. For BMI with large $\hat{\eta}$ and $\hat{\eta}_\delta$, there is a smaller discrepancy among these methods, and the pIVW estimator and the dIVW estimator provide similar results in this case.

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

TABLE 4 Estimated causal effects ($\hat{\beta}$) and estimated standard errors (SEs) of three obesity-related exposures (i.e., peripheral vascular disease (PVD), hypertensive disease (HD), and BMI) on the risk of hospitalized COVID-19. The pIVW estimator with the optimal $\lambda_{opt} = 1$

		No IV selection		IV selection $\delta = \sqrt{2 \log p}$	
Exposure	**Method**	$\hat{\eta}$	$\hat{\beta}$ **(SE)**	$\hat{\eta}_\delta$	$\hat{\beta}$ **(SE)**
PVD	IVW	1.68	0.007 (0.013)	3.98	0.233 (0.055)
	MR-Egger		0.022 (0.019)		0.292 (0.071)
	MR-Median		0.032 (0.020)		0.367 (0.098)
	MR-RAPS		0.379 (0.550)		0.419 (0.217)
	dIVW		4.413 (95.942)		0.670 (0.317)
	pIVW		0.219 (0.429)		0.594 (0.245)
HD	IVW	28.85	0.089 (0.033)	12.04	0.135 (0.065)
	MR-Egger		0.068 (0.047)		0.199 (0.091)
	MR-Median		0.109 (0.059)		0.132 (0.099)
	MR-RAPS		0.241 (0.095)		0.161 (0.080)
	dIVW		0.246 (0.093)		0.163 (0.079)
	pIVW		0.244 (0.093)		0.163 (0.078)
BMI	IVW	218.93	0.382 (0.077)	37.36	0.371 (0.100)
	MR-Egger		0.455 (0.105)		0.563 (0.140)
	MR-Median		0.549 (0.141)		0.361 (0.176)
	MR-RAPS		0.466 (0.097)		0.397 (0.106)
	dIVW		0.468 (0.096)		0.397 (0.106)
	pIVW		0.468 (0.096)		0.397 (0.106)

from being close to zero to reduce the bias due to the presence of many weak IVs. Moreover, we allow for the balanced horizontal pleiotropy by adjusting the variance estimation of the proposed pIVW estimator. Both simulation studies and real data analysis demonstrate the improved performance of the proposed pIVW estimator compared to the original IVW estimator and the recent dIVW estimator (Ye et al., 2021).

Our pIVW estimator has multiple advantages. First, our theoretical and numerical results show that the bias of the pIVW estimator with the optimal $\lambda_{opt} = 1$ converges to zero at a faster rate than that of the dIVW estimator as the effective sample size η (or η_δ) increases. Meanwhile, the pIVW estimator with the optimal $\lambda_{opt} = 1$ has smaller variance than the dIVW estimator. 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. The dIVW estimator can also be viewed as a special case of our proposed pIVW estimator, because the dIVW estimator is equivalent to the pIVW estimator with $\lambda = 0$. When $\lambda > 0$, their difference converges to zero as the effective sample size η (or η_δ) increases. 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 might have multiple numerical solutions (Zhao et al., 2019, 2020). In future work, we plan to extend the proposed penalization approach to other MR estimators to handle the weak IV bias, for instance, a penalized MR-Egger estimator (Bowden et al., 2015) to account for the unbalanced horizontal pleiotropy, and to account for the linkage disequilibrium (Wang et al., 2022).

ACKNOWLEDGMENTS

Siqi Xu and Peng Wang contributed equally to this work. The authors thank the Editor, the Associate Editor, and the reviewer for constructive comments that help improve this article. The authors also thank Lin Liu for helpful comments on our earlier draft.

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are openly available at the following links: the coronavirus disease 2019 outcomes https://www.covid19hg.org/results/r5/; the selection data of body mass index ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/AkiyamaM_28892062_GCST004904; the exposure data of body mass index http://www.nealelab.is/uk-biobank/ ; and the selection data and the exposure data of four obesity-related diseases https://cnsgenomics.com/content/data.

ORCID

Siqi Xu https://orcid.org/0000-0002-6352-2645
Zhonghua Liu https://orcid.org/0000-0003-3048-9823

REFERENCES

Abbott, L., Bryant, S., Churchhouse, C., Ganna, A., Howrigan, D., Palmer, D., et al. (2018) Round 2 GWAS results of thousands of phenotypes in the UK biobank. Available from: http://www.nealelab.is/uk-biobank/ [Accessed 14th November 2018].

Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., et al. (2017) Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nature Genetics*, 49(10), 1458.

Bowden, J., Del Greco M, F., Minelli, C., Smith, G.D., Sheehan, N. & Thompson, J. (2017) A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Statistics in Medicine*, 36(11), 1783–1802.

Bowden, J., Smith, G.D. & Burgess, S. (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology*, 44(2), 512–525.

Bowden, J., Smith, G.D., Haycock, P.C. & Burgess, S. (2016) Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology*, 40(4), 304–314.

Boyle, E.A., Li, Y.I. & Pritchard, J.K. (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell*, 169(7), 1177–1186.

Brockwell, S.E. & Gordon, I.R. (2001) A comparison of statistical methods for meta-analysis. *Statistics in Medicine*, 20(6), 825–840.

Burgess, S., Butterworth, A. & Thompson, S.G. (2013) Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology*, 37(7), 658–665.

Burgess, S. & Thompson, S.G. (2011) Bias in causal estimates from Mendelian randomization studies with weak instruments. *Statistics in Medicine*, 30(11), 1312–1323.

Burgess, S., Thompson, S.G. & CRP CHD Genetics Collaboration, (2011) Avoiding bias from weak instruments in Mendelian randomization studies. *International Journal of Epidemiology*, 40(3), 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(6), 715–718.

COVID-19 Host Genetics Initiative, (2021) Mapping the human genetic architecture of COVID-19. *Nature*, 600, 472–477.

Davies, N.M., Hinke Kessler Scholder, S., Farbmacher, H., Burgess, S., Windmeijer, F. & Smith, G.D. (2015) The many weak instruments problem and Mendelian randomization. *Statistics in Medicine*, 34(3), 454–468.

Didelez, V. & Sheehan, N. (2007) Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research*, 16(4), 309–330.

Fieller, E.C. (1954) Some problems in interval estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 16(2), 175–185.

Hanff, T.C., Mohareb, A.M., Giri, J., Cohen, J.B. & Chirinos, J.A. (2020) Thrombosis in COVID-19. *American Journal of Hematology*, 95(12), 1578–1589.

Hemani, G., Bowden, J. & Smith, G.D. (2018) Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human Molecular Genetics*, 27(R2), R195–R208.

Hwang, J.G. & Hwang, J.T. (1995) Fieller's problems and resampling techniques. *Statistica Sinica*, 5(1), 161–171.

Jordan, R.E., Adab, P. & 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(3), 908.

Lawlor, D.A., Harbord, R.M., Sterne, J.A., Timpson, N. & Smith, G.D. (2008) Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, 27(8), 1133–1163.

Leong, A., Cole, J.B., Brenner, L.N., Meigs, J.B., Florez, J.C. & Mercader, J.M. (2021) Cardiometabolic risk factors for COVID-19 susceptibility and severity: A Mendelian randomization analysis. *PLoS Medicine*, 18(3), e1003553.

Nakeshbandi, M., Maini, R., Daniel, P., Rosengarten, S., Parmar, P., Wilson, C., et al. (2020) The impact of obesity on COVID-19 complications: a retrospective cohort study. *International Journal of Obesity*, 44(9), 1832–1837.

Park, J.-H., Wacholder, S., Gail, M.H., Peters, U., Jacobs, K.B., Chanock, S.J. & Chatterjee, N. (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature Genetics*, 42(7), 570–575.

Pascarella, G., Strumia, A., Piliego, C., Bruno, F., Del Buono, R., Costa, F., et al. (2020) COVID-19 diagnosis and management: a comprehensive review. *Journal of Internal Medicine*, 288(2), 192–206.

Piegorsch, W.W. & Casella, G. (1985) The existence of the first negative moment. *The American Statistician*, 39(1), 60–62.

Ponsford, M.J., Gkatzionis, A., Walker, V.M., Grant, A.J., Wootton, R.E., Moore, L.S., et al. (2020) Cardiometabolic traits, sepsis, and severe COVID-19: a Mendelian randomization investigation. *Circulation*, 142(18), 1791–1793.

Popkin, B.M., Du, S., Green, W.D., Beck, M.A., Algaith, T., Herbst, C.H., et al. (2020) Individuals with obesity and COVID-19: a global perspective on the epidemiology and biological relationships. *Obesity Reviews*, 21(11), e13128.

Press, S.J. (1969) The t-ratio distribution. *Journal of the American Statistical Association*, 64(325), 242–252.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., et al. (2007) Plink: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559–575.

Sheehan, N.A., Didelez, V., Burton, P.R. & Tobin, M.D. (2008) Mendelian randomisation and causal inference in observational epidemiology. *PLoS Medicine*, 5(8), e177.

Shi, H., Kichaev, G. & Pasaniuc, B. (2016) Contrasting the genetic architecture of 30 complex traits from summary association data. *The American Journal of Human Genetics*, 99(1), 139–153.

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

Smith, G.D. & Ebrahim, S. (2004) Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology*, 33(1), 30–42.

Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. (2018) Detection of widespread horizontal pleiotropy in causal relationships inferred

from Mendelian randomization between complex traits and diseases. *Nature Genetics*, 50(5), 693–698.

Wang, A., Liu, W. & Liu, Z. (2022) A two-sample robust Bayesian Mendelian randomization method accounting for linkage disequilibrium and idiosyncratic pleiotropy with applications to the COVID-19 outcomes. *Genetic Epidemiology*, 46(3-4), 159–169.

Ye, T., Shao, J. & Kang, H. (2021) Debiased inverse-variance weighted estimator in two-sample summary-data Mendelian randomization. *Annals of Statistics*, 49(4), 2079–2100.

Zhao, Q., Chen, Y., Wang, J. & Small, D.S. (2019) Powerful three-sample genome-wide design and robust statistical inference in summary-data Mendelian randomization. *International Journal of Epidemiology*, 48(5), 1478–1492.

Zhao, Q., Wang, J., Hemani, G., Bowden, J. & Small, D.S. (2020) Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Annals of Statistics*, 48(3), 1742–1769.

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

Zheng, Z., Peng, F., Xu, B., Zhao, J., Liu, H., Peng, J., et al. (2020) Risk factors of critical & mortal COVID-19 cases: a systematic literature review and meta-analysis. *Journal of Infection*, 81, e16–e25.

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

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 2–5 are available with this paper at the Biometrics website on Wiley Online Library. The R package for the pIVW method including source code and example data is publicly available at https://github.com/siqixu/mr.pivw. This package is also available at the Biometrics website on Wiley Online Library.

How to cite this article: Xu, S., Wang, P., Fung, W.K. & Liu, Z. (2022) A novel penalized inverse-variance weighted estimator for Mendelian randomization with applications to COVID-19 outcomes. *Biometrics*, 1–12. https://doi.org/10.1111/biom.13732