Statistica Sinica Preprint No: SS-2023-0344	
Title	Leveraging Local Distributions in Mendelian Randomization: Uncertain Opinions are Invalid
Manuscript ID	SS-2023-0344
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0344
Complete List of Authors	Ziya Xu and Sai Li
Corresponding Authors	Sai Li
E-mails	saili@ruc.edu.cn
Notice: Accepted author version.	

Leveraging Local Distributions in Mendelian Randomization: Uncertain Opinions are Invalid

Ziya Xu and Sai Li*

Institute of Statistics and Big Data, Renmin University of China

Abstract: Mendelian randomization (MR) considers using genetic variants as instrumental variables (IVs) to infer causal effects in observational studies. However, the validity of causal inference in MR can be compromised when the IVs are potentially invalid. In this work, we propose a new method, MR-Local, to infer the causal effect in the existence of possibly invalid IVs. By leveraging the distribution of ratio estimates around the true causal effect, MR-Local selects the cluster of ratio estimates with the least uncertainty and performs causal inference within it. We establish the asymptotic normality of our estimator in the two-sample summary-data setting under either the plurality rule or the balanced pleiotropy assumption. Extensive simulations and analyses of real datasets demonstrate the reliability of our approach.

Key words and phrases: Causal inference, instrumental variable, Mendelian randomization, pleiotropy.

*Corresponding to: saili@ruc.edu.cn

1. Introduction

The instrumental variable (IV) approach is widely used to infer causal effects in the existence of unmeasured confounders. It relies on the valid IV assumption that instruments only affect the outcome through the exposure of interest. In epidemiology and biological studies, genetic variants are often utilized as IVs to detect causal relationships between phenotypes. Such causal studies are known as Mendelian randomization (MR) and gain popularity in various disciplines (Davey Smith and Ebrahim, 2003). However, the statistical foundations of MR are still evolving due to concerns regarding the potential invalidity of genetic instruments.

The large availability of genome-wide association studies (GWAS) has made genetic variants, particularly single nucleotide polymorphisms (SNPs), a popular choice of IVs. However, the exclusion restriction assumption, a key assumption in conventional IV methods, may not be credible when using genetic instruments. Many genetic variants exhibit pleiotropic effects, meaning that they can affect multiple phenotypes simultaneously (Davey Smith and Hemani, 2014). Hence, a variant can affect the outcome through more than one pathway, violating the exclusion restriction assumption. Moreover, the effects and functions of most SNPs are still largely unknown in terms of many biological traits. Researchers are likely to incorporate some invalid IVs in a causal study, which poses a statistical challenge in dealing with the possible invalidity given a large number of candidate IVs.

1.1 Model set-up

We first specify the potential outcome model with exposure $d_i \in \mathbb{R}$, outcome $y_i \in \mathbb{R}$, and a set of candidate IVs $\boldsymbol{z}_i \in \mathbb{R}^p$, $i = 1, \ldots, n$. We consider the additive linear, constant effects model (Holland, 1988; Small, 2007), which is

$$y_i^{(d,\boldsymbol{z})} - y_i^{(d',\boldsymbol{z}')} = (d - d')\beta + (\boldsymbol{z} - \boldsymbol{z}')^T\boldsymbol{\eta}, \quad \mathbb{E}[y_i^{(0,\mathbf{0})}|\boldsymbol{z}_i] = \boldsymbol{z}_i^T\boldsymbol{\kappa},$$

where $\beta \in \mathbb{R}$ denotes the causal effect of the exposure on the outcome, $\boldsymbol{\eta} \in \mathbb{R}^p$ denotes the direct effect of IVs on the outcome, and $\boldsymbol{\kappa} \in \mathbb{R}^p$ denotes the effect of IVs on the outcome through unmeasured confounders. Let $u_i = y_i^{(0,\mathbf{0})} - \mathbb{E}[y_i^{(0,\mathbf{0})}|\boldsymbol{z}_i]$. It gives that

$$y_i^{(d,\boldsymbol{z})} = d\beta + \boldsymbol{z}^T(\boldsymbol{\kappa} + \boldsymbol{\eta}) + u_i, \ \mathbb{E}[u_i|\boldsymbol{z}_i] = 0.$$

Let $\boldsymbol{\pi} = \boldsymbol{\kappa} + \boldsymbol{\eta} \in \mathbb{R}^p$. The potential outcome model gives the following model for the observed data. For $i = 1, \ldots, n$,

$$y_i = d_i\beta + \boldsymbol{z}_i^T\boldsymbol{\pi} + u_i, \ \mathbb{E}[u_i|\boldsymbol{z}_i] = 0. \tag{1.1}$$

For the exposure d_i, we fit a linear working model as

$$d_i = \boldsymbol{z}_i^T \boldsymbol{\gamma}_D + v_i, \text{ where } \boldsymbol{\gamma}_D = \mathbb{E}^{-1}[\boldsymbol{z}_i \boldsymbol{z}_i^T]\mathbb{E}[\boldsymbol{z}_i d_i] \text{ and } \mathbb{E}[v_i \boldsymbol{z}_i] = 0. \tag{1.2}$$

Plugging (1.2) into (1.1), we arrive at the reduced-form representation:

$$y_i = \boldsymbol{z}_i^T \boldsymbol{\gamma}_Y + \beta v_i + u_i, \tag{1.3}$$

where $\mathbb{E}[\boldsymbol{z}_i(v_i, u_i)] = \mathbf{0}$ and

$$\gamma_{Y,j} = \gamma_{D,j}\beta + \pi_j, \quad j = 1, \ldots, p. \tag{1.4}$$

In Equation (1.4), $\gamma_{Y,j}$ represents the total effect of the j-th IV on the outcome, which can be decomposed into two components: the effect through the causal pathway, $\gamma_{D,j}\beta$, and the invalid (pleiotropic) effect, π_j. We define the set of relevant IVs as $\mathcal{S} = \{1 \leq j \leq p : \gamma_{D,j} \neq 0\}$ and the set of valid IVs as $\mathcal{V} = \{1 \leq j \leq p : \gamma_{D,j} \neq 0, \pi_j = 0\}$. Under conventional IV assumptions, $\mathcal{V}$ is a known (nonempty) set. State-of-the-art methods, such as two-stage least squares (Basmann, 1957) and inverse-variance weighting (Burgess et al., 2013), can be applied to infer the causal effect. In contrast, we consider the presence of potentially invalid IVs, allowing $\mathcal{V}$ and $\mathcal{S}$ to be unknown a priori.

1.2 Existing identification assumptions

Regarding models (1.1) and (1.2), recent research has focused on estimating the causal effect in the presence of invalid IVs. One class of identification assumptions is based on the prior distribution of the invalid effects $\boldsymbol{\pi}$. For example, MR-Egger (Bowden et al., 2015) allows for non-zero $\boldsymbol{\pi}$, provided that the vector $\boldsymbol{\pi}$ has to be uncorrelated with the instrument strength vector $\boldsymbol{\gamma}_D$. This condition is known as the InSIDE assumption. On the other hand, MR-Raps (Zhao et al., 2020), a method based on a profile likelihood, allows for Gaussian invalid effects with zero mean. This condition is referred to as the balanced pleiotropy assumption. An alternative identification condition, proposed by Kolesár et al. (2015), assumes that $\boldsymbol{\pi}$ and $\boldsymbol{\gamma}_D$ are orthogonal, which is equivalent to the InSIDE assumption under the balanced pleiotropy assumption. The above assumptions all relax the standard IV assumptions. However, they are not practically verifiable and may be violated due to the existence of either correlated pleiotropy, where SNPs affect the outcome through confounders (Morrison et al., 2020), or directional pleiotropy, where $\boldsymbol{\pi}$ has a non-zero mean.

Another class of identification assumptions is based on the sparsity of $\boldsymbol{\pi}$. Define $\beta^{[j]} = \gamma_{Y,j}/\gamma_{D,j}$ for $j \in \mathcal{S}$, also known as the Wald ratio based on the j-th IV. The majority rule (Bowden et al., 2016; Kang et al., 2016; Windmeijer et al., 2019) assumes that at least half of π_j's are zero for $j \in \mathcal{S}$, implying that the median of

$\{\beta^{[j]}\}_{j\in\mathcal{S}}$ gives the true causal effect β. A weaker assumption than the majority rule, the plurality rule (Guo et al., 2018; Windmeijer et al., 2021) or the ZEMPA assumption (Hartwig et al., 2017), identifies β as the mode of $\{\beta^{[j]}\}_{j\in\mathcal{S}}$. Specifically, let $C_j = \{k \in \mathcal{S} : \beta^{[k]} = \beta^{[j]}\}$ denote the IVs that have same Wald ratio as the j-th IV. The plurality rule assumes that $C_{j^*} = \mathcal{V}$ for $j^* = \arg\max_{j\in\mathcal{S}} |C_j|$ (Guo et al., 2018). In other words, if we cluster the relevant IVs according to their Wald ratios, then the largest cluster must be formed by valid IVs. Hence, the IVs in the largest cluster can be used to identify the true causal effect. Alternatively, we can define the distribution function of these Wald ratios as $d_\beta(t) = \sum_{j\in\mathcal{S}} \mathbb{1}(\beta^{[j]} = t)/|\mathcal{S}|$ for $t \in \mathbb{R}$, so the plurality rule implies that the causal effect can be identified as the mode of $d_\beta(t)$. The ZEMPA assumption is equivalent to the plurality rule, and in our work, we use the term "plurality rule" to refer to this category of assumptions. Although weaker than the conventional IV assumptions, neither the majority rule nor the plurality rule are verifiable based on data.

Many other methods have emerged for estimating causal effects in the presence of invalid IVs. For example, Qi and Chatterjee (2019), Hu et al. (2022), and Morrison et al. (2020) propose Gaussian mixture models to account for directional or correlated pleiotropy. However, these methods lack theoretical guarantees, which can make their causal inference results less reliable. Another approach, introduced by Verbanck et al. (2018), is the MR-PRESSO test, designed to detect horizontal pleiotropy and

pleiotropic outlier variants. Additionally, Sun et al. (2022) leverages machine learning methods to model the nuisance parameters including pleiotropic effects. While these new methods relax conventional IV assumptions, their assumptions are hard to verify in a data-dependent way. Thus, it is crucial to develop causal inference methods that rely on even weaker assumptions, enhancing the reliability of causal inference in real-world applications.

1.3 Rationale and our findings

In this work, we propose a method to infer the causal effect β in more general scenarios than the plurality rule and the balanced pleiotropy assumption. Our rationale can be illustrated as follows. Consider that the invalid effects π_j are *i.i.d.* generated from the following model.

$$\pi_j = 0 \text{ if } j \in \mathcal{V} \quad \text{and} \quad \pi_j \sim N(\mu_j, \sigma_\pi^2) \text{ if } j \notin \mathcal{V}. \tag{1.5}$$

Model (1.5) applies to various scenarios with different values of $|\mathcal{V}|$ and μ_j. For instance, $|\mathcal{V}| \geq 0.5p$ indicates the majority rule holding and $\mu_j = 0$ represents the balanced pleiotropy. In practice, we obtain the ratio estimates $\hat{\beta}^{[j]}$, the empirical version of $\beta^{[j]}$, where $\hat{\beta}^{[j]} = \beta^{[j]} + \hat{\epsilon}_j$ contains noise $\hat{\epsilon}_j$, $j = 1, \ldots, p$. Assume for now that $\{\hat{\epsilon}_j\}_{j \leq p}$ are asymptotically normal with variance ξ^2 independently. Then the

ratio estimates $\hat{\beta}^{[j]}$, $j \in \mathcal{S}$, satisfy that

$$\frac{\hat{\beta}^{[j]} - \beta}{\xi} \xrightarrow{D} N(0,1) \text{ if } j \in \mathcal{V}, \quad \frac{\hat{\beta}^{[j]} - (\beta + \mu_j/\gamma_{D,j})}{\sqrt{\xi^2 + \sigma_\pi^2/\gamma_{D,j}^2}} \xrightarrow{D} N(0,1) \text{ if } j \notin \mathcal{V}. \tag{1.6}$$

If $\mu_j/\gamma_{D,j} = \mu_{bias}$ holds for some positive constant μ_{bias} and any $j \notin \mathcal{V}$, then, as shown in (1.6), the density of $\{\hat{\beta}^{[j]}\}_{j \le p}$ has two peaks at β and $\beta + \mu_{bias}$. The peak at β is the center of valid ratio estimates, while the peak at $\beta + \mu_{bias}$ is the center of invalid ratio estimates. Figure 1 provides an illustrative example where $\beta = 0.1$ and $\mu_{bias} = 3$. Methods based on the plurality rule identify β as the mode of ratio estimates, which corresponds to the highest peak in the density plot of $\{\hat{\beta}^{[j]}\}_{j \le p}$. These methods fail to identify β in this example since the highest peak is $\beta + \mu_{bias}$. Nevertheless, we can identify β as the sharpest peak provided that $\sigma_\pi^2 > 0$. This idea

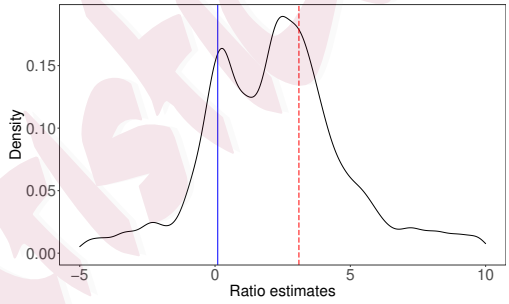

Figure 1: Density plot of the ratio estimates given by 2000 IVs among which 350 are valid. We set $\beta = 0.1$ and $\pi_j = 3\gamma_{D,j}$ for $j \notin \mathcal{V}$. The true causal effect is 0.1 (solid) and the model of causal effect given by invalid IVs is 3.1 (dashed). The data generation process is detailed in the Supplement (Section S4.1).

applies to various scenarios. If the plurality rule holds, the highest peak is also the

sharpest. In the case of balanced pleiotropy, the density of ratio estimates only has a single peak at β since $\mu_j = 0$, which is also the sharpest. This idea can be extended to scenarios involving non-Gaussian invalid effects beyond the model (1.5) provided that the variance of invalid effects is non-zero.

Based on this rationale, we propose a method to identify the causal effect as the sharpest peak of the distribution of ratio estimates. This idea bridges the plurality rule and the balanced pleiotropy and also applies to more general scenarios as described in Section 3.2. We develop a novel method that first calculates the uncertainty of the estimates in the neighborhood of each potential peak and then performs causal inference within the neighborhood with the least uncertainty, corresponding to the sharpest peak. We also establish the proposed estimator's asymptotic normality in any of the three scenarios outlined in Section 3. In extensive experiments, including GWAS-simulated studies and analyses of two-sample GWAS summary data, our approach consistently exhibits robust performance.

1.4 Notations and organization

We introduce some notations. Let a_n and b_n be two sequences of real numbers indexed by n. We define $a_n = O(b_n)$ if there exists a constant $c > 0$ such that $|a_n| \leq cb_n$ for all n, $a_n = o(b_n)$ if $a_n/b_n \to 0$ as $n \to \infty$, $a_n \gg b_n$ if $a_n/b_n \to \infty$ as $n \to \infty$, and $a_n \asymp b_n$ if there exists a constant $c > 0$ such that $c^{-1}b_n \leq |a_n| \leq cb_n$ for all

n. We write $\xrightarrow{\mathbb{P}}$ to denote convergence in probability, and $\xrightarrow{D}$ to denote convergence in distribution. For a random variable X, we denote the expectation of X as $\mathbb{E}(X)$ and the variance of X as $\mathrm{Var}(X)$. If X follows a standard normal distribution, $\Phi(\cdot)$ denotes the cumulative distribution function and $\phi(\cdot)$ denotes the density function. For a set A, we denote the complement of set A as A^c and the number of elements in set A as $|A|$.

In the rest of this paper, we introduce the proposed method in Section 2. Theoretical guarantees are provided in Section 3. Numerical results based on simulated GWAS data are conducted in Section 4. The proposed method is applied to several real studies in Section 5. Section 6 concludes the paper with a discussion and some extensions.

2. Local-distribution based method for MR

In this section, we formalize the idea of using the local distribution to estimate β. We focus on a two-sample MR setting where we obtain $\hat{\boldsymbol{\gamma}}_D$ as an estimate of $\boldsymbol{\gamma}_D$ from one GWAS and $\hat{\boldsymbol{\gamma}}_Y$ as an estimate of $\boldsymbol{\gamma}_Y$ from another independent GWAS. Their standard errors $\sigma_{D,j}$ and $\sigma_{Y,j}$, $j = 1, \ldots, p$ are also available in the corresponding GWAS. While the GWAS statistics are marginal effects, under certain conditions, they are element-wise asymptotically normal (Zhao et al., 2020; Ye et al., 2021) such

that

$$\frac{\hat{\gamma}_{D,j} - \gamma_{D,j}}{\sigma_{D,j}} \xrightarrow{D} N(0,1) \text{ and } \frac{\hat{\gamma}_{Y,j} - \gamma_{Y,j}}{\sigma_{Y,j}} \xrightarrow{D} N(0,1), \ j = 1, \ldots, p,$$

as $(n_D, n_Y) \to \infty$. Formal assumptions on the empirical estimates are given in Condition 1.

We first introduce the key device, the so-called local distribution, for the proposed method. Consider a statistic $\hat{z}_j(b)$ as a function of b, which is

$$\hat{z}_j(b) = \frac{\hat{\gamma}_{Y,j} - b\hat{\gamma}_{D,j}}{\sqrt{\sigma^2_{Y,j} + b^2\sigma^2_{D,j}}}, \ j = 1, \ldots, p. \tag{2.7}$$

In fact, $\hat{z}_j(b)$ is standardized $\hat{\beta}^{[j]} - b$ for $\hat{\beta}^{[j]} = \hat{\gamma}_{Y,j}/\hat{\gamma}_{D,j}$. For $\delta_n > 0$, we define the local distribution of $\hat{z}_j(b)$ as

$$\mathcal{F}(b, \delta_n) = \mathbb{P}(\hat{z}_j(b) \leq t \mid |\hat{\gamma}_{Y,j} - b\hat{\gamma}_{D,j}| \leq \delta_n). \tag{2.8}$$

We will utilize the special property of $\mathcal{F}(\beta, \delta_n)$ to identify β. We see that $\hat{z}_j(\beta) \xrightarrow{D} N(0,1)$ for $j \in \mathcal{V}$. On the other hand, $\{\hat{\beta}^{[j]}\}_{j \in \mathcal{V}}$ are expected to be closer to β than $\{\hat{\beta}^{[j]}\}_{j \in \mathcal{V}^c}$. Thus, for a small enough δ_n, the cluster $\{j : |\hat{\gamma}_{Y,j} - \beta\hat{\gamma}_{D,j}| \leq \delta_n\}$ only contains valid IVs with high probability. In this case, $\mathcal{F}(\beta, \delta_n)$ is approximately a

truncated standard normal distribution:

$$\mathcal{F}(\beta, \delta_n) = \mathbb{P}\left(z \le t \mid |z| \le \frac{\delta_n}{\sqrt{\sigma_{Y,j}^2 + \beta^2 \sigma_{D,j}^2}} \right) + o(1), \tag{2.9}$$

where z is a standard normal variable. We can identify β from a set of candidate values $\{b_j\}_{j \le p}$ based on (2.9). Various methods can be used to measure the difference between $\mathcal{F}(b_j, \delta_n)$ and the distribution in (2.9) such as the Kullback–Leibler divergence, Kolmogorov–Smirnov (KS) statistics, and moment matching. In the next subsection, we use the second moment of $\mathcal{F}(\beta, \delta_n)$ to estimate β. We discuss other possible methods in Section 6.

2.1 Proposed algorithm

Based on the idea formulated previously, we develop a three-step method to estimate the causal effect.

In the first step, we propose a statistic $\widehat{Q}(b)$ to test whether the local distribution (2.9) holds around b. We define a cluster of IVs that have ratio estimates around b as

$$\widehat{\mathcal{C}}(b) = \left\{ 1 \le j \le p : |\hat{\gamma}_{Y,j} - b\hat{\gamma}_{D,j}| \le \tau_0 \sqrt{\sigma_{Y,j}^2 + b^2 \sigma_{D,j}^2} \right\}. \tag{2.10}$$

The tuning parameter τ_0 in (2.10) controls the "bandwidth". Its form is studied both

theoretically in Section 3 and numerically in Section 4. We see that the distribution of $\hat{z}_j(b)$ for $j \in \widehat{\mathcal{C}}(b)$ is $\mathcal{F}(b, \tau_0\sqrt{\sigma_{Y,j}^2 + b^2\sigma_{D,j}^2})$ as defined in (2.8). Consequently, the distribution of $\hat{z}_j(\beta)$ for $j \in \widehat{\mathcal{C}}(\beta)$ is asymptotically a standard normal distribution truncated at $\pm\tau_0$ as given by (2.9), provided that $\widehat{\mathcal{C}}(\beta)$ only contains valid IVs. Based on this result, we construct the test statistic $\widehat{Q}(b)$ as the standardized second moment of $\{\hat{z}_j(b)\}_{j\in\widehat{\mathcal{C}}(b)}$. Formally, define

$$\widehat{Q}(b) = \frac{1}{|\widehat{\mathcal{C}}(b)|}\sum_{j\in\widehat{\mathcal{C}}(b)}\frac{\hat{z}_j^2(b)}{g(\tau_0)} \text{ and } g(\tau_0) = 1 - \frac{2\tau_0\phi(\tau_0)}{\Phi(\tau_0) - \Phi(-\tau_0)}, \tag{2.11}$$

where $g(\tau_0)$ is the theoretical variance of a standard normal variable truncated at $\pm\tau_0$. We use the statistic $\widehat{Q}(b)$ to indicate the uncertainty, or equivalently, the sharpness locally around b.

In the second step, we perform an uncertainty test based on $\widehat{Q}(b)$. Specifically, we know that $\mathbb{E}[\widehat{Q}(\beta)|\widehat{\mathcal{C}}(\beta) = \mathcal{V}] = 1$. Hence, we search for b whose $\widehat{Q}(b)$ is close to 1. A candidate set for searching can be constructed as follows. Suppose that β falls in the interval $[-C_\beta, C_\beta]$ for some positive constant C_β. We define the candidate values $\{b_j\}_{j\le p}$, ranging from $-C_\beta$ to C_β with a step of $2C_\beta/p$. We see that parameter C_β controls the range and step size of b_j. In practice, we can specify C_β as a fixed constant or as a quantile of the ratio estimates as we further studied in Section S4.3 of the Supplement. Given the set $\{b_j\}_{j\le p}$, we define the set of values that pass the

uncertainty test as

$$\widehat{\mathcal{B}} = \left\{ b_j : |\widehat{Q}(b_j) - 1| \leq \sigma_Q \sqrt{\frac{\log p}{|\widehat{\mathcal{C}}(b_j)|}} + \frac{1}{\log p}, |\widehat{\mathcal{C}}(b_j)| \geq \sqrt{p}, 1 \leq j \leq p \right\}, \quad (2.12)$$

where $\sigma_Q^2 = g^{-2}(\tau_0)\left(g(\tau_0)\left(3+\tau_0^2\right) - \tau_0^2\right) - 1$. In (2.12), the threshold for $\widehat{Q}(b_j)$ is determined by the limiting distribution of $\widehat{Q}(b_j)$ when $b_j = \beta$ and also accounts for the possible bias of the GWAS statistics. Furthermore, we exclude the clusters with a cluster size smaller than $\sqrt{p}$, aligning with our assumption that the number of valid IVs is at least $\sqrt{p}$. Within the set $\widehat{\mathcal{B}}$, we find $\hat{b} = \arg\max_{b \in \widehat{\mathcal{B}}} |\widehat{\mathcal{C}}(b)|$, which corresponds to the highest peak or the mode.

In the last step, we adopt the debiased inverse-variance weighting (dIVW) estimator (Ye et al., 2021) with the IVs in $\widehat{\mathcal{C}}(\hat{b})$ as in (2.13) due to its robustness to weak IVs. We also construct the confidence interval based on the limiting distribution of the final estimator.

If $\widehat{\mathcal{B}}$ is empty, it indicates that all the clusters have large uncertainty, which can occur in the case of balanced pleiotropy. In this scenario, we employ the dIVW estimator with all candidate IVs to infer β and adjust its variance.

We refer to our method as "MR-Local" and present the formal algorithm in Algorithm 1.

Remark 1. The format of (2.11) is closely related to Cochran's Q-statistic (Cochran,

1954), which can be used to quantify heterogeneity within a vector of statistics. Bowden et al. (2019) estimate β by a weighted average of $\hat{\beta}^{[j]}$, $j = 1, \ldots, p$, where the weights are selected to minimize the overall Cochran's Q-statistic. Here we use (2.11) to detect the set of valid IVs instead of direct estimation.

Algorithm 1: MR-Local: Local distribution-based method for MR

Input: $\{\hat{\gamma}_{Y,j},\ \sigma_{Y,j}\}_{j=1}^{p}$, $\{\hat{\gamma}_{D,j}, \sigma_{D,j}\}_{j=1}^{p}$, the range of true causal effect $[-C_\beta, C_\beta]$, and a tuning parameter τ_0.

Step 1. Compute the test statistics indicating the uncertainty.

for $j = 1, \ldots, p$ **do**

let $b_j = -C_\beta + 2C_\beta j/p$. Construct clusters $\widehat{\mathcal{C}}(b_j)$ defined in (2.10). Based on each $\widehat{\mathcal{C}}(b_j)$, compute the test statistic $\widehat{Q}(b_j)$ defined in (2.11).

end

Step 2. Conduct the uncertainty test. Calculate the set $\widehat{\mathcal{B}}$ defined in (2.12). If $\widehat{\mathcal{B}}$ is nonempty, $\hat{b} = \arg\max_{b \in \widehat{\mathcal{B}}} |\widehat{\mathcal{C}}(b)|$. If $\widehat{\mathcal{B}}$ is empty, define $\widehat{\mathcal{C}}(\hat{b}) = \{1, \cdots, p\}$.

Step 3. Estimate and infer the causal effect β. Apply the dIVW estimator in the cluster $\widehat{\mathcal{C}}(\hat{b})$ to obtain

$$\hat{\beta}_{\text{dIVW}} = \frac{\sum_{j \in \widehat{\mathcal{C}}(\hat{b})} \sigma_{Y,j}^{-2} \hat{\gamma}_{D,j} \hat{\gamma}_{Y,j}}{\sum_{j \in \widehat{\mathcal{C}}(\hat{b})} \sigma_{Y,j}^{-2} \left(\hat{\gamma}_{D,j}^2 - \sigma_{D,j}^2 \right)}. \tag{2.13}$$

The $(1-\alpha) \times 100\%$ two-sided confidence interval for β is $[\hat{\beta}_{\text{dIVW}} - z_{\alpha/2}\hat{\sigma}_\beta, \hat{\beta}_{\text{dIVW}} + z_{\alpha/2}\hat{\sigma}_\beta]$, where $\hat{\sigma}_\beta$ is defined in (3.16) when $\widehat{\mathcal{B}}$ is nonempty, is defined in (3.17) when $\widehat{\mathcal{B}}$ is empty.

We provide further remarks on Algorithm 1. First, unlike the methods proposed by Guo et al. (2018) and Windmeijer et al. (2019), Algorithm 1 does not require

screening for strong IVs. This advantage arises from the combination of two strategies that address weak IV bias: the format of $\hat{z}_j(b)$ in (2.7) and the dIVW estimator. Specifically, the limiting distribution of $\hat{z}_j(b)$ is well-defined when $\hat{\gamma}_{D,j}$ approaches zero, in contrast to the distribution of $\hat{\beta}^{[j]}$. Second, the majority rule (Bowden et al., 2016) and the plurality rule (Guo et al., 2018) can be viewed as a vanilla version of Algorithm 1 with $\widehat{\mathcal{B}} = \{b_j\}_{j \le p}$. These approaches estimate the causal effect solely based on the largest cluster of IVs. In contrast, MR-Local leverages additional information from the distribution of valid ratio estimates as captured by (2.9). Third, besides balanced pleiotropy, an empty set $\widehat{\mathcal{B}}$ in (2.12) can arise in various scenarios, including cases where all instruments exhibit directional pleiotropy. To determine the cause, one can compare the dIVW estimate in (2.13) with the MR-Raps estimate (Zhao et al., 2020) when $\widehat{\mathcal{B}}$ is empty. If the estimates are similar, it suggests that balanced pleiotropy may be the cause.

3. Theoretical justifications for Algorithm 1

In this section, we provide theoretical guarantees for Algorithm 1 in each of the following scenarios: (i) the plurality rule, (ii) the balanced pleiotropy, and (iii) a directional pleiotropy case where the plurality rule can be violated. We focus on a two-sample MR setting where $\hat{\boldsymbol{\gamma}}_D$ and $\hat{\boldsymbol{\gamma}}_Y$ are generated from two independent GWAS with sample sizes n_D and n_Y, respectively. Detailed assumptions are given

as follows.

Condition 1 (Sub-Gaussian noises)**.** Assume that for $j = 1, \cdots, p$,

$$\hat{\gamma}_{D,j} - \gamma_{D,j} = \frac{1}{n_D}\sum_{i=1}^{n_D}\delta_{i,j}^{(D)} + \sigma_{D,j} rem_{D,j} \quad \text{and} \quad \hat{\gamma}_{Y,j} - \gamma_{Y,j} = \frac{1}{n_Y}\sum_{i=1}^{n_Y}\delta_{i,j}^{(Y)} + \sigma_{Y,j} rem_{Y,j},$$

where $\{\delta_{i,j}^{(D)}\}_{i\le n_D, j\le p}$ and $\{\delta_{i,j}^{(Y)}\}_{i\le n_Y, j\le p}$ are mutually independent sub-Gaussian variables with zero means and bounded sub-Gaussian norms. Suppose that $\sigma_{D,j}^2 = \sum_{i=1}^{n_D}\mathrm{Var}[\delta_{i,j}^{(D)}]/n_D^2$, $\sigma_{Y,j}^2 = \sum_{i=1}^{n_Y}\mathrm{Var}[\delta_{i,j}^{(Y)}]/n_Y^2$, $\max_{j\le p}|\gamma_{D,j}| = O(p/\sqrt{n_Y \log^2 p})$, and the reminder terms $\{rem_{D,j}\}_{j\le p}$ and $\{rem_{Y,j}\}_{j\le p}$ satisfy that $\max_{j\le p}\{|rem_{D,j}|, |rem_{Y,j}|\} = o(1/\sqrt{\log p})$. Moreover, $\min\{n_D, n_Y\} \gg \log^4 p$ and $n_Y/n_D = O(p/\log p)$.

Condition 1 assumes that the errors in $\hat{\gamma}_{D,j}$ and $\hat{\gamma}_{Y,j}$ are independent sub-Gaussian variables, allowing for potential bias to account for model misspecification errors and measurement errors. The independence between $\{\hat{\gamma}_{D,j}\}_{j\le p}$ and $\{\hat{\gamma}_{Y,j}\}_{j\le p}$ is ensured through the two-sample MR structure. The independence within $\{\hat{\gamma}_{D,j}\}_{j\le p}$ and $\{\hat{\gamma}_{Y,j}\}_{j\le p}$ is assumed for simplicity as in Zhao et al. (2020) and Ye et al. (2021). In practice, linkage disequilibrium (LD) clumping can be employed to select uncorrelated SNPs. This noise assumption is less restrictive compared to the normality and unbiasedness assumptions made in Zhao et al. (2020) and Ye et al. (2021). We assume that the variances $\sigma_{D,j}^2$ and $\sigma_{Y,j}^2$ are known and decrease with the two GWAS sample sizes n_D and n_Y, respectively. Accurate estimates of these variances can be

obtained in practice (Bowden et al., 2015; Zhao et al., 2020). The upper bound on the maximum IV strength is typically necessary since $\sum_{j\le p}\gamma_{D,j}^2 = O(1)$ if the exposure's variance exists (Zhao et al., 2020). The assumptions regarding n_D, n_Y, and p are mild. Specifically, the assumption on n_Y/n_D is weaker than the assumption made in Zhao et al. (2020) and Ye et al. (2021), where both papers assume that $n_D \asymp n_Y$.

3.1 Conclusions under the plurality rule and the balanced pleiotropy assumption

In this section, we consider two common IV assumptions: the plurality rule and the balanced pleiotropy assumption. We establish the asymptotic normality of our proposed estimator under either assumption.

First, we consider the case where the plurality rule holds.

Condition 2 (Plurality rule)**.** (a) Assume that $|\beta| \le C_\beta$. It holds that $\sup_{\mathcal{C}(b)\neq\mathcal{V}} |\mathcal{C}(b)| < |\mathcal{V}|$, where

$$\mathcal{C}(b) = \left\{ j : |\gamma_{Y,j} - b\gamma_{D,j}| \le 2\tau_0 \sqrt{\sigma_{Y,j}^2 + b^2\sigma_{D,j}^2} \right\}. \tag{3.14}$$

Moreover, $\min_{j\in\mathcal{V}^c} |\pi_j| / \sqrt{\sigma_{Y,j}^2 + C_\beta^2\sigma_{D,j}^2} > 2\tau_0$ and $|\mathcal{V}| \ge \sqrt{p}$.

(b) Denote the average IV strength in a non-empty set $\mathcal{C}$ as $\kappa_{\mathcal{C}}$, where

$$\kappa_{\mathcal{C}} = \frac{1}{|\mathcal{C}|} \sum_{j \in \mathcal{C}} \frac{\gamma_{D,j}^2}{\sigma_{D,j}^2}. \tag{3.15}$$

Suppose that $\max_{j \le p}\{|rem_{D,j}|, |rem_{Y,j}|\} = o(1/\sqrt{|\mathcal{V}|})$, the average valid IV strength $\kappa_{\mathcal{V}} \gg n_D n_Y^{-1} |\mathcal{V}|^{-1} + |\mathcal{V}|^{-1/2}$, and $\max_{j \in \mathcal{V}} \gamma_{D,j}^2 / (\sum_{j \in \mathcal{V}} \gamma_{D,j}^2) = o(1)$.

Part (a) is a finite sample version of the plurality rule. As the sample sizes grow to infinity, $\mathcal{C}(\beta^{[j]})$ defined in (3.14) converges to C_j in Guo et al. (2018) and the condition on $|\pi_j|$ converges to $\min_{j \in \mathcal{V}^c} |\pi_j| > 0$. The former two statements combined with $\sup_{\mathcal{C}(b) \neq \mathcal{V}} |\mathcal{C}(b)| < |\mathcal{V}|$ is equivalent to the definition of the plurality rule in Guo et al. (2018). The condition that $|\mathcal{V}| \ge \sqrt{p}$ ensures the convergence of the empirical valid IV set, which is usually met when the plurality rule holds. We do not make any assumptions regarding the minimum valid IV strength that allows for the existence of weak IVs, in contrast to Guo et al. (2018) and Windmeijer et al. (2021).

In part (b), we assume that the average valid IV strength $\kappa_{\mathcal{V}}$ is not too small to derive the asymptotic normality. To verify Lindeberg's condition, we assume that the maximum valid IV strength is small compared to the sum of valid IV strength. If $n_D \asymp n_Y$ and $|\mathcal{V}| \asymp p$, our condition on $\kappa_{\mathcal{V}}$ simplifies to $\kappa_{\mathcal{V}} \gg p^{-1/2}$, which recovers the assumptions made in Ye et al. (2021).

Theorem 1 (Asymptotic normality under the plurality rule)**.** *Assume Conditions*

3.1 Conclusions under the plurality rule and the balanced pleiotropy assumption

1 and 2(a). Let $\tau_0 = c_0\sqrt{\log p}$ *for some constant* $c_0 > \sqrt{2}$ *depending on the sub-Gaussian norms in Condition 1. Then, as* $(p, n_D, n_Y) \to \infty$,

$$\mathbb{P}(\widehat{\mathcal{C}}(\hat{b}) = \mathcal{V}) \to 1.$$

Further, assume Condition 2(b). Then, the dIVW estimator defined in (2.13) satisfies that

$$\frac{\hat{\beta}_{\text{dIVW}} - \beta}{\hat{\sigma}_\beta} \xrightarrow{D} N(0,1),$$

where

$$\hat{\sigma}_\beta^2 = \frac{\sum_{j\in\widehat{\mathcal{C}}(\hat{b})} \sigma_{Y,j}^{-4}\left[\hat{\gamma}_{D,j}^2 + \hat{\beta}_{\text{dIVW}}^2\sigma_{D,j}^2\left(\hat{\gamma}_{D,j}^2 + \sigma_{D,j}^2\right)\right]}{\left(\sum_{j\in\widehat{\mathcal{C}}(\hat{b})} \sigma_{Y,j}^{-2}\left(\hat{\gamma}_{D,j}^2 - \sigma_{D,j}^2\right)\right)^2}. \tag{3.16}$$

Under conditions of Theorem 1, we prove that the selected cluster $\widehat{\mathcal{C}}(\hat{b})$ is the set of valid IVs with high probability and the dIVW estimator is asymptotically normal. Given the first conclusion of Theorem 1, the problem is simplified to infer the causal effect in the valid IV scenario. The standard IVW estimator assumes the average IV strength to be larger than p, which is unrealistic in practice. In contrast, we employ the dIVW estimator proposed by Ye et al. (2021), which ensures asymptotic normality when the average valid IV strength $\kappa_{\mathcal{V}} \gg n_D n_Y^{-1}|\mathcal{V}|^{-1} + |\mathcal{V}|^{-1/2}$, given our

sub-Gaussian noises assumption.

Next, we consider the case where the balanced pleiotropy assumption holds.

Condition 3 (Balanced pleiotropy)**.** Assume that $|\beta| \leq C_\beta$ and

$$\pi_j \sim_{i.i.d.} N(0, \sigma_\pi^2) \text{ for } j \leq p,$$

where $\sigma_\pi \geq c_1 \max_{j\leq p} \sqrt{\sigma_{Y,j}^2 + C_\beta^2 \sigma_{D,j}^2}$ for some constant $c_1 > 0$.

Suppose that the average IV strength $\kappa = \kappa_{\{1,\cdots,p\}} \gg n_D n_Y^{-1} p^{-1} + p^{-1/2}(1 + \max_{j\leq p} \sigma_\pi \sigma_{D,j}^{-1})$ and $\max_{j\leq p} \gamma_{D,j}^2 / (\sum_{j\leq p} \gamma_{D,j}^2) = o(1)$.

Condition 3 assumes the balanced pleiotropy as studied in Zhao et al. (2020) and Ye et al. (2021). The condition on the average IV strength κ is used for deriving the asymptotic normality. If $n_D \asymp n_Y$ and $\max_{j\leq p} \sigma_\pi \sigma_{D,j}^{-1} = O(1)$ as assumed in Ye et al. (2021), our condition on κ simplifies to $\kappa \gg p^{-1/2}$, which recovers the assumption made in Ye et al. (2021). The condition on σ_π ensures that the invalid IV estimates exhibit large uncertainty, preventing them from passing the uncertainty test as in (2.12).

Theorem 2 (Asymptotic normality under the balanced pleiotropy)**.** *Assume Conditions 1 and 3. Let* $\tau_0 = c_0 \sqrt{\log p}$ *for some constant* $c_0 > \sqrt{2}$ *depending on the*

sub-Gaussian norms in Condition 1. Then, as $(p, n_D, n_Y) \to \infty$,

$$\frac{\hat{\beta}_{\text{dIVW}} - \beta}{\hat{\sigma}_\beta} \xrightarrow{D} N(0,1),$$

where

$$\hat{\sigma}_\beta^2 = \frac{\sum_{j=1}^p \sigma_{Y,j}^{-4} \left[(\sigma_{Y,j}^2 + \hat{\sigma}_\pi^2)\hat{\gamma}_{D,j}^2 + \hat{\beta}_{\text{dIVW}}^2 \sigma_{D,j}^2 \left(\hat{\gamma}_{D,j}^2 + \sigma_{D,j}^2 \right) \right]}{\left(\sum_{j=1}^p \sigma_{Y,j}^{-2} \left(\hat{\gamma}_{D,j}^2 - \sigma_{D,j}^2 \right) \right)^2} \tag{3.17}$$

and

$$\hat{\sigma}_\pi^2 = \frac{\sum_{j=1}^p \left[(\hat{\gamma}_{Y,j} - \hat{\beta}_{\text{dIVW}}\hat{\gamma}_{D,j})^2 - \sigma_{Y,j}^2 - \hat{\beta}_{\text{dIVW}}^2 \sigma_{D,j}^2 \right] \sigma_{Y,j}^{-2}}{\sum_{j=1}^p \sigma_{Y,j}^{-2}}. \tag{3.18}$$

Theorem 2 justifies the asymptotic normality of our proposed estimator in the balanced pleiotropy case. The variance estimator in (3.17) involves σ_π^2 due to the presence of invalid effects, which is no smaller than the variance estimator in (3.16). We estimate σ_π^2 by (3.18) following the approach in Ye et al. (2021).

Remark 2. Theorem 2 requires that σ_π is no smaller than the order of $\min\{n_D, n_Y\}^{-1/2}$, while Ye et al. (2021) and Zhao et al. (2020) prefer a smaller σ_π that is not greater than the order of $\min\{n_D, n_Y\}^{-1/2}$. Notably, our proposed estimator still enjoys asymptotic normality when $\sigma_\pi = o(1/\sqrt{n_Y \log p})$ since $\widehat{\mathcal{C}}(\beta)$ could pass the uncer-

tainty test as in (2.12). We provide the formal proof in the Supplement (Proposition 1).

According to Theorems 1 and 2, our proposed causal estimate is consistent and the confidence interval constructed in Algorithm 1 achieves the nominal coverage asymptotically. Our method can handle two scenarios: when the plurality rule holds and when there exists balanced pleiotropy. To the best of our knowledge, no existing approach in the literature can handle both scenarios. Therefore, MR-local provides a robust estimation of the causal effect as it is suitable for more general cases.

3.2 Further results under another directional pleiotropy assumption

Besides the two scenarios considered in Section 3.1, we further explore the situation that only a small proportion of valid IVs exists and the pleiotropic effects are not balanced. In this case, methods based on the plurality rule or the balanced pleiotropy assumption can fail. Next, we show that our method still works under mild conditions.

Condition 4 (A directional pleiotropy assumption)**.** Assume that $|\beta| \leq C_\beta$. The invalid effects satisfy that

$$\pi_j \sim_{i.i.d.} N(\mu_j, \sigma_\pi^2) \text{ for } j \in \mathcal{V}^c,$$

where $\min_{j\in\mathcal{V}^c}|\mu_j|/\sqrt{\sigma_\pi^2+\sigma_{Y,j}^2+C_\beta^2\sigma_{D,j}^2}\geq 2\tau_0$ and $\sigma_\pi\geq c_1\max_{j\leq p}\sqrt{\sigma_{Y,j}^2+C_\beta^2\sigma_{D,j}^2}$ for some constant $c_1>1$. Moreover, $\max_{j\in\mathcal{V}^c}|\gamma_{D,j}|/\min_{j\in\mathcal{V}}|\gamma_{D,j}|=o(\sqrt{\log p})$ and $|\mathcal{V}|\geq\sqrt{p}$.

In Condition 4, we assume that the invalid effects π_j follow a Gaussian distribution with mean μ_j and variance σ_π^2 for $j\in\mathcal{V}^c$. The lower bound on $|\mu_j|$ serves as a separation condition to ensure that Wald ratios based on invalid IVs and those based on valid IVs are separated. The lower bound on σ_π is similar to the lower bound in Condition 3. Condition 4 covers scenarios where both the balanced pleiotropy assumption and the plurality rule fail. We provide examples and further explanations in the Supplement (Proposition 2).

Theorem 3 (Asymptotic normality under directional pleiotropy)**.** *Assume Conditions 1 and 4. Let $\tau_0=c_0\sqrt{\log p}$ for some constant $c_0>\sqrt{2}$ depending on the sub-Gaussian norms in Condition 1. Then, as $(p,n_D,n_Y)\to\infty$,*

$$\mathbb{P}(\widehat{\mathcal{C}}(\hat{b})=\mathcal{V})\to 1.$$

Further, assume Condition 2(b). Then the dIVW estimator defined in (2.13) satisfies that

$$\frac{\hat{\beta}_{\text{dIVW}}-\beta}{\hat{\sigma}_\beta}\xrightarrow{D}N(0,1),$$

where $\hat{\sigma}_\beta$ is defined in (3.16).

We prove that under Conditions 1 and 4, the selected cluster $\widehat{\mathcal{C}}(\hat{b})$ is the set of valid IVs with high probability. If the strengths of valid IVs satisfy certain conditions as assumed in Condition 2(b), our estimator is asymptotically normal. Theorem 3 demonstrates the reliability of MR-Local beyond the plurality rule and balanced pleiotropy assumptions. A special case of Theorem 3 is where the plurality rule holds and the invalid IVs satisfy the directional pleiotropy assumption, as detailed in Corollary 1 of the Supplement. In contrast to MR-Egger (Bowden et al., 2015), our method does not rely on the InSIDE assumption and provides a viable alternative for handling directional pleiotropy.

Moreover, MR-Local can achieve consistent causal inference under different combinations of the three previously stated assumptions: the plurality rule, the balanced pleiotropy, and the directional pleiotropy assumption. We provide detailed theoretical guarantees in Section S3.1 in the Supplement.

4. Simulated GWAS experiments

In this section, we conduct numerical experiments based on simulated GWAS summary statistics. The code for all the methods in comparison is available at `https://github.com/saili0103/MR-Local`.

4.1 Randomly generated effect sizes

In the first experiment, we simulate the summary statistics with $n_D = n_Y = 10^5$ and $p = 2000$. Let $h_D = 0.1$ represents the heritability of trait D. We generate $\gamma_{D,j} \sim N(0, h_D/p)$ independently for $j = 1, \ldots, p$. Therefore, $\sum_{j=1}^{p} \gamma_{D,j}^2 \approx h_D$. By the reduced form, we set $\gamma_{Y,j} = \gamma_{D,j}\beta + \pi_j$, where β and π_j vary in different settings. We generate the standard error of $\gamma_{D,j}$, $\sigma_{D,j} \sim U[0.8, 1]/\sqrt{n_D}$ and the standard error of $\gamma_{Y,j}$, $\sigma_{Y,j} \sim U[0.8, 1]/\sqrt{n_Y}$ independently. The corresponding IV strength $\kappa \approx 6.25$ in this set-up.

(a) $\mathcal{V}$ is a random subset of $\{1, \cdots, p\}$ with $|\mathcal{V}| = 0.5p$. Set $\pi_j = 0$ for $j \in \mathcal{V}$ and $\pi_j \sim N(0, 0.05/p) + 2.5\gamma_{D,j}$ for each $j \notin \mathcal{V}$.

(b) $\mathcal{V}$ is a random subset of $\{1, \cdots, p\}$ with $|\mathcal{V}| = 0.5p$. Set $\pi_j = 0$ for $j \in \mathcal{V}$ and $\pi_j \sim N(0, 0.5/p) + 2.5\gamma_{D,j}$ for each $j \notin \mathcal{V}$.

(c) $\pi_j \sim_{i.i.d.} N(0, 0.1/p)$ for $j = 1, \ldots, p$.

(d) $\pi_j \sim_{i.i.d.} N(0, 0.05/p)$ for $j = 1, \ldots, p$.

(e) $\mathcal{V}$ is a random subset of $\{1, \cdots, p\}$ with $|\mathcal{V}| = 0.28p$. Set $\pi_j = 0$ for $j \in \mathcal{V}$ and $\pi_j \sim_{i.i.d.} N(0, 0.05/p) + 3\gamma_{D,j}$ for each $j \notin \mathcal{V}$.

In (a) and (b), the plurality rule holds. The valid ratio estimates form the highest peak in both settings while the invalid ratio estimates are more spread out in (b).

In (c) and (d), the pleiotropic effects are balanced with different σ_π^2. In (e), valid IVs are relatively few and the invalid ratio estimates form the highest peak. In (a), (b), and (e), a fixed effect $2.5\gamma_{D,j}$ or $3\gamma_{D,j}$ is added to each π_j, corresponding to the separation condition in Condition 4. The invalid effects constructed in (a), (b), and (e) correspond to the scenario of correlated pleiotropy (Morrison et al., 2020), since π_j is correlated with $\gamma_{D,j}$, $j = 1, \ldots, p$.

In each experiment, we generate $\hat{\gamma}_{D,j} \sim N(\gamma_{D,j}, \sigma_{D,j}^2)$ and $\hat{\gamma}_{Y,j} \sim N(\gamma_{Y,j}, \sigma_{Y,j}^2)$ independently. Our proposal is compared with three other MR methods which are robust to pleiotropy under certain assumptions: MR-MBE (Hartwig et al., 2017), MR-Raps (Zhao et al., 2020), and two-stage hard thresholding (TSHT, Guo et al. (2018)). For the robust performance of MR-MBE and TSHT, we first screen out weak IVs and only use the IVs such that $|\hat{\gamma}_{D,j}|/\sigma_{D,j} \geq \sqrt{2 \log p}$. Each setting is replicated based on 2000 Monte Carlo experiments.

The proposed MR-Local method incorporates two empirical adjustments for robust inference. First, we apply mild IV strength screening using Algorithm 1 with the criterion $|\hat{\gamma}_{D,j}|/\sigma_{D,j} \geq \tau_0$, where τ_0 is the tuning parameter in (2.10). We report results for different choices of τ_0 in the Supplement (Section S4.3). Second, if $\widehat{\mathcal{B}} \neq \emptyset$, the dIVW estimator $\hat{\beta}_{\text{dIVW}}$ is subject to post-selection effects. Hence, we estimate its standard deviation, $\hat{\sigma}_\beta$, via bootstrap. Additionally, we introduce another variant of MR-Local called MR-Local+, which incorporates the uncertainty

test statistic $\widehat{Q}(b)$ along with a skewness test statistic $\widehat{K}(b)$. The motivation is to refine the selected $\widehat{\mathcal{B}}$ by leveraging the fact that when $b = \beta$, the skewness of $\hat{z}_j(b)$ for $j \in \widehat{\mathcal{C}}(b)$ should be close to zero. Further details of MR-Local+ can be found in the Supplement (Section S4.2). In Figure 2, we present the estimation errors of the

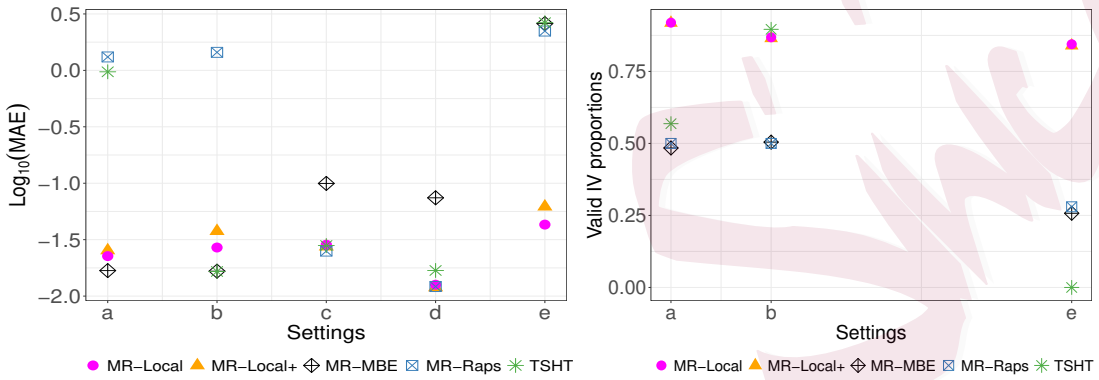

Figure 2: The logarithm of mean absolute errors (MAE, left panel) and the proportion of valid IVs used to estimate the causal effect (right panel) based on MR-Local ($\tau_0 = 1.6$), MR-Local+ ($\tau_0 = 1.6$), MR-Median, and MR-Raps in five settings.

five methods in settings (a)-(e). Our proposed methods, MR-Local and MR-Local+, have reliable performances across all settings. In contrast, the other methods exhibit significant estimation errors in at least one setting. Specifically, MR-Raps performs poorly when the Gaussian assumption of pleiotropic effects is violated. MR-MBE becomes less accurate in settings (c)-(e) when the proportion of invalid IVs is small. TSHT, designed for the plurality rule, exhibits large errors in settings (a) and (e). In the right panel of the figure, we show the proportion of valid IVs in $\widehat{\mathcal{C}}(\hat{b})$ for our methods. For the other methods, we report the proportion of valid IVs based on

their screening steps. Since MR-Raps lacks a screening step and MR-MBE only filters out weak IVs, their proportions of valid IVs are relatively low. TSHT employs a voting step to select valid IVs, which is successful in setting (b) but not in settings (a) and (e), explaining its performance in the estimation errors shown in the left panel. In contrast, our proposal demonstrates a relatively high proportion of valid IVs in $\widehat{\mathcal{C}}(\hat{b})$, highlighting the effectiveness of our screening approach based on the uncertainty test. This, in turn, explains its reliable performance in estimation.

Setup	β	MR-Local			MR-Local+			MR-MBE		TSHT		MR-Raps	
		cov.	s.d.	$\emptyset$	cov.	s.d.	$\emptyset$	cov.	s.d.	cov.	s.d.	cov.	s.d.
a	0	0.995	0.06	0.00	0.975	0.04	0.01	0.970	0.02	0.147	0.04	0.000	0.04
	0.1	0.996	0.07	0.00	0.973	0.04	0.00	0.945	0.02	0.164	0.04	0.000	0.04
b	0	0.963	0.03	0.01	0.994	0.05	0.00	0.961	0.03	0.658	0.02	0.000	0.04
	0.1	0.960	0.03	0.01	0.988	0.06	0.00	0.978	0.03	0.846	0.02	0.000	0.04
c	0	0.967	0.03	1.00	0.966	0.03	1.00	0.839	0.09	0.351	0.01	0.982	0.03
	0.1	0.963	0.03	1.00	0.961	0.03	1.00	0.891	0.09	0.499	0.01	0.978	0.03
d	0	0.994	0.02	1.00	0.992	0.02	1.00	0.949	0.08	0.607	0.01	0.994	0.02
	0.1	0.993	0.02	1.00	0.991	0.02	1.00	0.956	0.08	0.713	0.01	0.993	0.02
e	0	0.991	0.10	0.00	0.986	0.07	0.01	0.000	0.12	0.000	0.01	0.000	0.04
	0.1	0.986	0.12	0.00	0.974	0.07	0.01	0.000	0.12	0.000	0.01	0.000	0.04

Table 1: Average coverage probabilities (cov.) and average standard deviation (s.d.) for the 95% confidence intervals computed via proposed MR-Local ($\tau_0 = 1.6$), MR-Local+ ($\tau_0 = 1.6$), MR-MBE, TSHT, and MR-Raps in settings (a) to (e). The column "$\emptyset$" indicates the chance of occuring $\{\widehat{\mathcal{B}} = \emptyset\}$.

In Table 1, we present the inference results for each method. We observe that

both MR-Local and MR-Local+ demonstrate coverage probabilities close to the nominal level. Notably, MR-Local+ exhibits higher efficiency with smaller estimated standard errors. As expected, these two methods estimate $\widehat{\mathcal{B}} = \emptyset$ in settings (c) and (d), successfully detecting the presence of balanced pleiotropy. The inference results of MR-Raps are sensitive to the assumption of Gaussian pleiotropy. MR-MBE shows reliable coverage probabilities in settings (a), (b), and (d), but its confidence intervals tend to be longer compared to other methods, resulting in lower efficiency. TSHT exhibits relatively low coverage probabilities, and its intervals are narrow. This is due to the absence of adjustment for post-selection effects in its standard error estimates, whereas the bootstrap method employed in MR-Local, MR-Local+, and MR-MBE partially accounts for selection uncertainty.

4.2 Simulations based on BMI and SBP GWAS

In this subsection, we conduct simulations using data from a GWAS on body mass index (BMI) and another GWAS on systolic blood pressure (SBP). The data is available in the R package "mr.raps" (Zhao, 2018). Since the original dataset only contains 160 SNPs, we replicate the observed effect vectors and standard error vectors 10 times each to create a new dataset with 1600 SNPs. Specifically, we set $\boldsymbol{\gamma}_D$ to be the observed effects from the BMI GWAS after replication, and the standard errors $\{\sigma_{D,j}\}_{j\le p}$ and $\{\sigma_{Y,j}\}_{j\le p}$, are set to be observed standard errors in the BMI dataset.

The pleiotropic effects π_j are generated according to each of the following settings. Let $h_D^* = \sum_{j=1}^p \gamma_{D,j}^2$.

(a) $\mathcal{V}$ is a random subset of $\{1, \cdots, p\}$ with $|\mathcal{V}| = 0.6p$. Set $\pi_j = 0$ for $j \in \mathcal{V}$ and $\pi_j \sim N(0, 0.5h_D^*/p) + 2.5\gamma_{D,j}$ for each $j \notin \mathcal{V}$.

(b) $\mathcal{V}$ is a random subset of $\{1, \cdots, p\}$ with $|\mathcal{V}| = 0.4p$. Set $\pi_j = 0$ for $j \in \mathcal{V}$ and π_j to be the j-th effect on SBP obtained from the data for each $j \notin \mathcal{V}$.

(c) $\pi_j \sim_{i.i.d.} N(0, h_D^*/p)$ for $j = 1, \ldots, p$.

(d) $\mathcal{C}$ is a random subset of $\{1, \cdots, p\}$ with $|\mathcal{C}| = 0.4p$. Set $\pi_j \sim_{i.i.d.} N(0, h_D^*/p)$ for $j \in \mathcal{C}$ and $\pi_j \sim_{i.i.d.} N(0, 4h_D^*/p)$ for $j \notin \mathcal{C}$.

(e) $\mathcal{V}$ is a random subset of $\{1, \cdots, p\}$ with $|\mathcal{V}| = 0.25p$. Set $\pi_j = 0$ for $j \in \mathcal{V}$ and $\pi_j \sim_{i.i.d.} N(0,1)|\gamma_{D,j}| + 5\gamma_{D,j}$ for each $j \notin \mathcal{V}$.

In (a) and (b), the plurality rule holds. The pleiotropic effects in (b) are set to be the observed effects, which are different from random Gaussian effects. In (c) and (d), the pleiotropic effects are balanced. In (d), π_j are generated from a Gaussian mixture model centered at zero. Setting (e) corresponds to the situation considered in Section 3.2.

From Figure 3, we observe that MR-Local and MR-Local+ perform robustly across various settings. MR-MBE has the smallest estimation errors in settings (a)

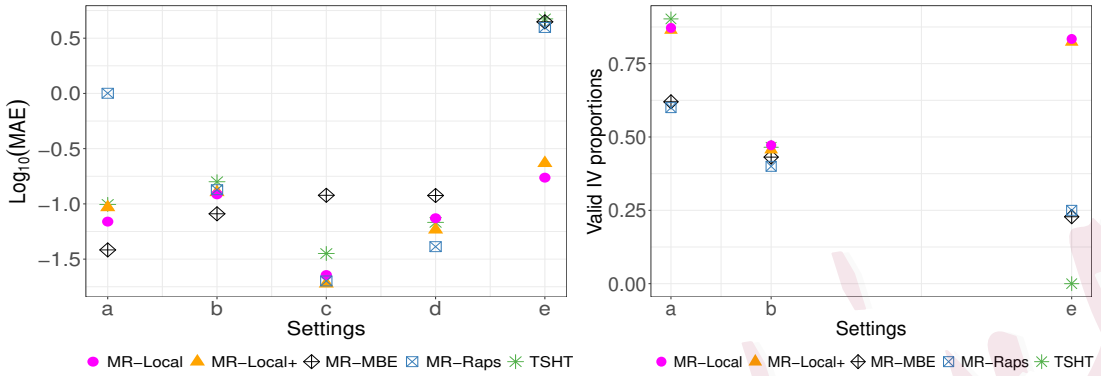

Figure 3: The logarithm of mean absolute errors (MAE, left panel) and the proportion of valid IVs used to estimate the causal effect (right panel) based on MR-Local ($\tau_0 = 1.6$), MR-Local+ ($\tau_0 = 1.6$), MR-Median, and MR-Raps in five settings.

Setup	β	MR-Local			MR-Local+			MR-MBE		TSHT		MR-Raps	
		cov.	s.d.	$\emptyset$	cov.	s.d.	$\emptyset$	cov.	s.d.	cov.	s.d.	cov.	s.d.
a	-0.2	0.969	0.09	0.01	0.882	0.07	0.03	0.974	0.06	0.104	0.02	0.000	0.04
	0	0.964	0.10	0.01	0.906	0.08	0.02	0.988	0.06	0.180	0.02	0.000	0.04
b	-0.2	0.425	0.05	0.44	0.295	0.04	0.55	0.959	0.11	0.000	0.02	0.017	0.04
	0	0.411	0.05	0.44	0.314	0.04	0.54	0.972	0.10	0.000	0.02	0.018	0.04
c	-0.2	0.977	0.03	0.98	0.987	0.03	0.99	0.960	0.12	0.622	0.02	0.980	0.03
	0	0.976	0.04	0.97	0.987	0.03	0.98	0.964	0.12	0.796	0.02	0.984	0.03
d	-0.2	0.942	0.05	0.95	0.960	0.05	0.97	0.810	0.14	0.088	0.02	0.966	0.05
	0	0.963	0.05	0.97	0.951	0.05	0.96	0.856	0.14	0.174	0.02	0.946	0.05
e	-0.2	0.967	0.21	0.01	0.926	0.18	0.03	0.000	0.23	0.000	0.03	0.000	0.08
	0	0.955	0.22	0.00	0.934	0.20	0.03	0.000	0.24	0.000	0.03	0.000	0.08

Table 2: Average coverage probabilities (cov.) and average standard deviation (s.d.) for the 95% confidence intervals computed via proposed MR-Local ($\tau_0 = 1.6$), MR-Local+ ($\tau_0 = 1.6$), MR-MBE, TSHT, and MR-Raps in settings (a) to (e). The column "$\emptyset$" indicates the chance of occuring $\{\widehat{\mathcal{B}} = \emptyset\}$.

and (b) but shows relatively large errors in other settings. MR-Raps has the smallest errors in settings (c) and (d) but becomes less accurate when the pleiotropic effects are not Gaussian. TSHT shows significant errors in setting (e), where the plurality rule fails. In Table 2, MR-Local achieves coverage probabilities close to the nominal levels, except for setting (b) where some invalid effects are close to zero, posing a challenge in distinguishing valid IVs from invalid ones. Specifically, MR-Local estimates $\widehat{\mathcal{B}} = \emptyset$ approximately 50% of the time in setting (b), indicating that no valid peak is identified in half of the cases. In setting (d), where the pleiotropic effects follow a Gaussian mixture distribution, MR-Raps exhibits greater robustness than our proposal.

5. Real studies based on two-sample GWAS summary data

In this section, we estimate the causal relationships based on the GWAS summary statistics from the OpenGWAS database (Elsworth et al., 2020). We infer the causal effect of body mass index (BMI) on four other commonly studied traits: height (HGT), systolic blood pressure (SBP), type-2 diabetes (T2D), and coronary artery disease (CAD). To confirm the reliability of our proposal, we also estimate the causal effect of BMI on BMI, where the true value is one if certain assumptions are satisfied by two datasets. Specifically, we use the summary statistics for BMI from the GIANT consortium as the exposure and the summary statistics for BMI from the MRC-

IEU consortium as the outcome. Detailed information on the studies and the pre-screening steps are given in the Supplement (Section S5.1).

Outcome	MR-Local	MR-Local+	MR-MBE	TSHT	MR-Raps
BMI	1.033(0.018)	1.033(0.017)	1.002(0.035)	1.000(0.006)	1.010(0.007)
HGT	-0.055(0.012)*	-0.055(0.012)*	-0.041(0.030)	-0.033(0.004)	-0.054(0.011)
SBP	0.125(0.010)*	0.125(0.010)*	0.126(0.048)	0.130(0.006)	0.120(0.010)
T2D	0.009(0.002)	0.009(0.001)	0.009(0.003)	0.007(0.000)	0.008(0.001)
CAD	0.379(0.093)	0.379(0.055)	0.333(0.172)	0.389(0.026)	0.368(0.029)

Table 3: Causal inference of BMI on five traits: BMI, HGT, SBP, T2D, and CAD. Each column reports the estimated causal effects (estimated standard deviations). The results with "*" corresponds to $\widehat{\mathcal{B}} = \emptyset$.

The results are reported in Table 3. The estimates for the causal effect of BMI on BMI are close to one with 95% confidence intervals covering the value one. Although the true causal effect of BMI on BMI should be one, the actual causal effect based on these two populations may not be exactly one since the two datasets in use may be subject to different batch effects. For the causal effect of BMI on HGT, MR-Local and MR-Local+ yield estimates similar to MR-Raps, whereas the estimates based on the plurality rule show slight differences. MR-Local and MR-Local+ algorithms output $\widehat{\mathcal{B}} = \emptyset$, suggesting that the results based on the balanced pleiotropy assumption are likely more reliable. Regarding the causal effect estimates of BMI on CAD, all methods produce similar results, but MR-MBE has a significantly larger estimated standard deviation, resulting in a less efficient estimate. We also perform sensitivity

analyses regarding the strength of IVs, the number of IVs, and the tuning parameter τ_0 in Section S5.2 of the Supplement.

6. Discussion

This work introduces MR-Local, a method that utilizes the local distribution to remove invalid IVs and then conduct causal inference. Our proposed causal estimate enjoys consistency and asymptotic normality under mild conditions and has reliable performance across various numerical scenarios. MR-Local is applicable under both the plurality rule and the balanced pleiotropy assumption. Hence, MR-Local is more robust than methods relying on these two assumptions in complex real-world situations.

We discuss other statistics beyond the uncertainty measure defined in (2.11). While the variance entails second-order information on the local distribution, alternative ways can include higher-order information. On the other hand, A popular nonparametric statistic can be the KS-type statistic, which measures the distance between the empirical distribution and the true distribution. Formally,

$$\widehat{\mathrm{KS}}(b) = \sup_{t\in[-\tau_0,\tau_0]} \left| \frac{1}{|\widehat{\mathcal{C}}(b)|} \sum_{k\in\widehat{\mathcal{C}}(b)} \mathbb{1}(\hat{z}_k(b) \le t) - \frac{\Phi(t) - \Phi(-\tau_0)}{\Phi(\tau_0) - \Phi(-\tau_0)} \right|. \tag{6.19}$$

If $\widehat{\mathrm{KS}}(b)$ is large, the IVs in the j-th clique are unlikely to follow the truncated

standard normal distribution as in (2.9). Hence, one can also screen out a candidate value b_j if $\widehat{\mathrm{KS}}(b_j)$ is beyond a proper threshold based on its limiting distribution. Future works involve relaxing the assumption on the distribution of the invalid effects and generalizing our findings to multivariate MR scope (Burgess and Thompson, 2015; Sanderson et al., 2019), which remains an open question.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (grant no. 12201630).

Supplementary Materials

The supplementary material includes the following sections: S1, technical lemmas; S2, proofs for the theorems; S3, supplementary propositions; S4, further information on simulations; S5, further results on real studies.

References

Basmann, R. L. (1957). A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica: Journal of the Econometric Society 25*(1), 77–83.

Bowden, J., G. Davey Smith, and S. Burgess (2015). Mendelian randomization with invalid instruments:

effect estimation and bias detection through egger regression. *International Journal of Epidemiology 44*(2), 512–525.

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

Bowden, J., F. Del Greco M, C. Minelli, Q. Zhao, D. A. Lawlor, N. A. Sheehan, J. Thompson, and G. Davey Smith (2019). Improving the accuracy of two-sample summary-data mendelian randomization: moving beyond the nome assumption. *International Journal of Epidemiology 48*(3), 728–742.

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

Burgess, S. and S. G. Thompson (2015). Multivariable mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *American Journal of Epidemiology 181*(4), 251–260.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics 10*(1), 101–129.

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

Davey Smith, G. and G. Hemani (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics 23*(R1), R89–R98.

Elsworth, B., M. Lyon, T. Alexander, Y. Liu, P. Matthews, J. Hallett, P. Bates, T. Palmer, V. Haberland, G. D. Smith, et al. (2020). The mrc ieu opengwas data infrastructure. *BioRxiv*.

Guo, Z., H. Kang, T. T. Cai, and D. S. Small (2018). Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *Journal of the Royal Statistical Society Series B: Statistical Methodology 80*(4), 793–815.

Hartwig, F. P., G. Davey Smith, and J. Bowden (2017). Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption. *International Journal of Epidemiology 46*(6), 1985–1998.

Holland, P. W. (1988). Causal inference, path analysis and recursive structural equations models. *ETS Research Report Series 1988*(1), i–50.

Hu, X., J. Zhao, Z. Lin, Y. Wang, H. Peng, H. Zhao, X. Wan, and C. Yang (2022). Mendelian randomization for causal inference accounting for pleiotropy and sample structure using genome-wide summary statistics. *Proceedings of the National Academy of Sciences 119*(28), e2106858119.

Kang, H., A. Zhang, T. T. Cai, and D. S. Small (2016). Instrumental variables estimation with some invalid instruments and its application to mendelian randomization. *Journal of the American Statistical Association 111*(513), 132–144.

Kolesár, M., R. Chetty, J. Friedman, E. Glaeser, and G. W. Imbens (2015). Identification and inference with many invalid instruments. *Journal of Business & Economic Statistics 33*(4), 474–484.

Morrison, J., N. Knoblauch, J. H. Marcus, M. Stephens, and X. He (2020). Mendelian randomization

accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature Genetics 52*(7), 740–747.

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

Sanderson, E., G. Davey Smith, F. Windmeijer, and J. Bowden (2019). An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *International Journal of Epidemiology 48*(3), 713–727.

Small, D. S. (2007). Sensitivity analysis for instrumental variables regression with overidentifying restrictions. *Journal of the American Statistical Association 102*(479), 1049–1058.

Sun, B., Y. Cui, and E. T. Tchetgen (2022). Selective machine learning of the average treatment effect with an invalid instrumental variable. *The Journal of Machine Learning Research 23*(1), 9249–9288.

Verbanck, M., C.-Y. Chen, B. Neale, and R. Do (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nature Genetics 50*(5), 693–698.

Windmeijer, F., H. Farbmacher, N. Davies, and G. Davey Smith (2019). On the use of the lasso for instrumental variables estimation with some invalid instruments. *Journal of the American Statistical Association 114*(527), 1339–1350.

Windmeijer, F., X. Liang, F. P. Hartwig, and J. Bowden (2021). The confidence interval method for selecting valid instrumental variables. *Journal of the Royal Statistical Society Series B: Statistical*

Methodology 83(4), 752–776.

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

Zhao, Q. (2018). *mr.raps: Two Sample Mendelian Randomization using Robust Adjusted Profile Score*. R package version 0.2.

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

Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China.

Email: xuziya0798@ruc.edu.cn

Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China.

Email: saili@ruc.edu.cn