



## Review

# Statistical methods for Mendelian randomization in genome-wide association studies: A review



Frederick J. Boehm<sup>a</sup>, Xiang Zhou<sup>a,b,\*</sup>

<sup>a</sup> Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

<sup>b</sup> Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA

## ARTICLE INFO

## Article history:

Received 15 March 2022

Received in revised form 8 May 2022

Accepted 9 May 2022

Available online 14 May 2022

## Keywords:

Genome-wide association study

Mendelian randomization

Genomics

Causal inference

Confounding

Horizontal pleiotropy

## ABSTRACT

Genome-wide association studies have yielded thousands of associations for many common diseases and disease-related complex traits. The identified associations made it possible to identify the causal risk factors underlying diseases and investigate the causal relationships among complex traits through Mendelian randomization. Mendelian randomization is a form of instrumental variable analysis that uses SNP associations from genome-wide association studies as instruments to study and uncover causal relationships between complex traits. By leveraging SNP genotypes as instrumental variables, or proxies, for the exposure complex trait, investigators can tease out causal effects from observational data, provided that necessary assumptions are satisfied. We discuss below the development of Mendelian randomization methods in parallel with the growth of genome-wide association studies. We argue that the recent availability of GWAS summary statistics for diverse complex traits has motivated new Mendelian randomization methods with relaxed causality assumptions and that this area continues to offer opportunities for robust biological discoveries.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Contents

1. Introduction	2339
2. Assumptions of Mendelian randomization	2339
3. Statistical models and methods for MR with one instrument	2340
3.1. Ratio of coefficients method	2340
3.2. Two-stage methods	2340
3.3. Likelihood-based methods	2340
3.4. Semiparametric methods	2341
3.5. Mendelian randomization with multiple independent SNP instruments	2341
4. Statistical models and methods for MR with multiple instruments	2341
4.1. Methods with Individual-level data	2341
4.2. Methods with summary data	2342
4.3. Assumption violations and sensitivity analysis	2343
4.4. Explicit modeling of horizontal pleiotropy	2343
4.5. Correlated horizontal pleiotropy	2346
4.6. Binary, count, and time-to-event outcomes and exposures	2346
5. Recent findings	2347
5.1. Omnigenic MR	2347
5.2. Sample overlap in biobanks	2347
5.3. Bi-directional causal inference with MR to account for reverse causation	2347
5.4. Practice recommendations	2347

\* Corresponding author at: Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA.

E-mail address: [xzhousph@umich.edu](mailto:xzhousph@umich.edu) (X. Zhou).

6. Summary and outlook .....	2348
CRedit authorship contribution statement .....	2348
Declaration of Competing Interest .....	2348
Acknowledgements .....	2348
References .....	2348

## 1. Introduction

Great interest has developed in inferring causal relationships between complex traits, i.e., traits that seemingly are not inherited in a Mendelian fashion, in observational human genetics studies. Discovery of such relationships is crucial to enhancing our understanding of the biology of health and disease. Inferring causal relationship in genetics is often carried out through Mendelian randomization (MR) analysis. MR analysis is a form of instrumental variable analysis in which genetic markers, typically single nucleotide polymorphisms (SNPs), serve as instruments, or proxies, for inferring causal effects of an exposure variable on an outcome variable [1–7].

MR analysis is facilitated by the development of genome-wide association studies (GWAS), which present unique opportunities for discovery of causal relationships via MR [8]. A GWAS interrogates millions of single nucleotide polymorphisms (SNPs) to infer which are associated with the trait [9,10]. In the nearly 17 years since publication of early studies, researchers have reported thousands of novel SNP-trait associations from GWAS [10–13]. Ever larger sample sizes in GWAS have enhanced statistical power to detect associations and have refined our understanding of human health and disease. The association results from many large-scale GWAS are nowadays readily available, often in the form of summary statistics that include the marginal SNP p-values and/or their effect size estimates and standard errors [14]. The identified SNP associations are used as the main input for MR analysis and thus the wide availability of GWAS summary statistics clear the way for effective MR analysis in complex genetics studies.

Effective MR analyses are enabled by many MR methods developed in recent years. Methods for and uses of MR have appeared at a rapid and accelerating pace since the publication of the early GWAS (Fig. 1). The overall trend of the methodology development is in the direction of increasingly sophisticated modeling of hori-

zontal pleiotropy including both independent and correlated horizontal pleiotropy while attempting to maintain scalability and computational efficiency in the presence of multiple correlated SNPs.

Here, we present a comprehensive review on 47 MR methods, primarily in the context of GWAS summary statistics, to help practitioners to choose which MR methods to use in applied data analysis. We discuss the basics form of MR analysis, the causality assumptions in MR, and how recent MR methods are developed to ensure robust results in the presence of assumption violations. We discuss in detail methods advances before briefly summarizing applications and directions for future research. Different from existing MR reviews [15–17], we discuss in detail recent methods developments that enable modeling of horizontal pleiotropy and correlated horizontal pleiotropy and place these developments in the larger context of MR analysis with GWAS data. We present these recent methods developments for an audience of both statisticians and epidemiologists. We hope our review will facilitate the further advance of MR methods and their wide application on GWAS data.

## 2. Assumptions of Mendelian randomization

Mendelian randomization uses genetic markers in the form of SNP genotypes as the proxy (“instrument” or “instrumental variable”) for the “exposure,” a complex trait, and asks whether the “exposure” variable is causal for the “outcome” variable, which is typically a second complex trait. The exposure and outcome variables can be binary, a count, a time to event, or a continuous variable. For brevity and simplicity, we focus on continuous exposures and outcomes before considering other classes of outcome variables in a later section. Most MR studies use the “two-sample” MR design, in which one cohort of subjects has measurements for the exposure, while a second cohort has measurements for the outcome, with both cohorts sharing the same set of SNP instruments [18]. We focus on the two-sample MR design before considering the one-sample design and the partial two-sample design where samples are partially overlapped between the two cohorts.

MR aims to infer the causal effect of the exposure variable on the outcome variable in observational studies. The proper causal interpretation in an MR study requires the SNP instruments to satisfy three causality assumptions. The first assumption states that SNP instruments are associated with the exposure. If the first assumption, about the association between the SNP instruments and the exposure, holds, but this association is weak, then an amplification of biases, such as those due to violations of assumptions 2 and 3, may result [19]. Indeed, bias in causal estimates can increase with decreases in the strength of association between the SNP instrument and exposure [20]. The second assumption requires independence, conditional on the exposure and all measured and unmeasured confounders, between the SNP instruments and the outcome. The second MR assumption, often termed the exclusion restriction assumption, can also be stated as the need for the SNP instruments to affect the outcome only through the exposure. In a causal diagram, this means that the only path from the SNP instruments to the outcome is that containing the exposure. The last of the three assumptions states that the SNP instru-

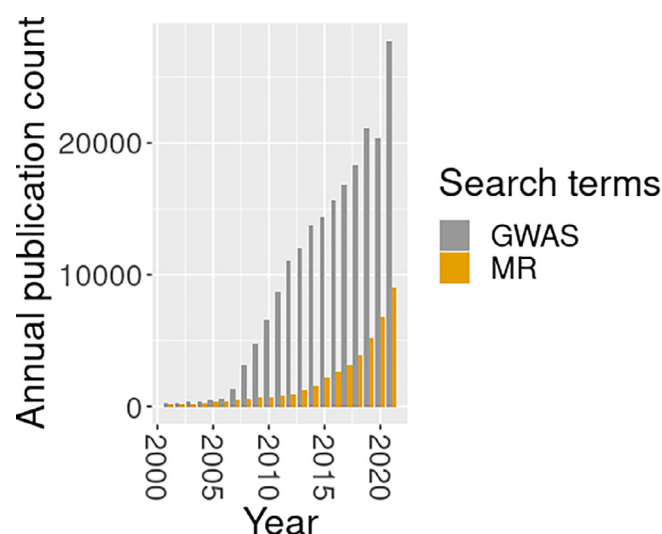


Fig. 1. Upward trend in article counts by year for Google Scholar keyword searches: 1. Mendelian randomization and 2. Genome-wide association study.

ments are independent of all (measured and unmeasured) confounders of the relationship between exposure and outcome. The third assumption ensures that the SNP instruments are independent of all confounders of the exposure–outcome relationship.

When the three MR assumptions hold, one may use instrumental variable statistical methods, with the genetic marker data, the exposure data, and the outcome data, to estimate and to test the causal effect of the exposure on the outcome [7,21,22]. In practice, careful evaluation of the assumptions is needed. Because assumptions 2 and 3, which involve absence of certain types of confounding, can't be verified, practitioners must be cautious when planning MR analyses. Sensitivity analysis is highly recommended as a way to assess robustness of estimates in the presence of possible assumption violations [19]. We will discuss these sensitivity analyses in more detail in a later section and consider their evaluation in MR studies [19].

The causal assumptions of MR also help one to understand the origin of the term “Mendelian randomization.” Recall Mendel's inheritance laws [23], under the assumption that alleles segregate randomly from parent to offspring, the offspring genotypes are unlikely to be associated with confounders of the exposure–outcome relationship. Additionally, reverse causation from the outcome or exposure to the genotypes is unanticipated since germ-line genotypes are fixed at conception and, thus, precede realization of other observed variables. Therefore, using SNP instruments in the instrumental variable analysis is often referred to as Mendelian randomization.

### 3. Statistical models and methods for MR with one instrument

We begin our discussion of Mendelian randomization in GWAS data by considering the simplest case, where there is a single SNP instrument and a single outcome variable. Approaches to MR with one SNP instrument can be classified into four categories: ratio of coefficients method, two-stage methods, likelihood-based methods, and semiparametric methods [15]. We discuss each of these before turning to methods that leverage multiple SNP instruments.

#### 3.1. Ratio of coefficients method

The ratio of coefficients method, also known as the Wald method, estimates the causal effect of the exposure  $X$  on the outcome  $Y$  by using a single SNP instrument [24]. For a continuous outcome, the causal effect estimator, (1).

$$\hat{\gamma}_{\text{Wald}} = \frac{\hat{\beta}_{YZ}}{\hat{\beta}_{XZ}}$$

In Eq. (1),  $\hat{\beta}_{YZ}$  and  $\hat{\beta}_{XZ}$  are the slope estimates from the regressions of the outcome and exposure, respectively, on the SNP instrument. Since the Wald method requires only the regression coefficients, it can be used with summary data. However, the Wald method doesn't accommodate multiple SNP instruments, which limits its direct use in the GWAS setting as will be discussed later.

#### 3.2. Two-stage methods

A two-stage statistical model involves two regression models [15,25]. For a continuous outcome, one may perform “two-stage least squares,” which involves two linear regressions. First, the exposure variable is regressed on the instrument (i.e., the SNP genotypes) (Equation (2)). We denote each subject's exposure variable value as  $x_i$  in Equation (2), while  $\alpha_0$  denotes an intercept term and  $\alpha_1$  is the slope.  $z_i$  denotes the SNP instrument genotype for

subject  $i$ , while  $\epsilon_i$ , the error term, is assumed to be independent among subjects and normally distributed with a shared common variance and mean zero.

The resulting fitted values for the exposure variable from Eq. (2) are the independent variable in the second linear regression, where the outcome is the dependent variable. The causal effect estimate, then, is the coefficient  $\hat{\beta}_1$  obtained from the second regression analysis (Eq. (3)). Note also that the random errors in Eq. (3) are assumed independent of those in Eq. (2). Like the  $\epsilon_i$  in Eq. (2), the  $\tau_i$  are assumed independent and identically distributed normal random errors.

$$x_i = \alpha_0 + \alpha_1 z_i + \epsilon_i$$

$$y_i = \beta_0 + \beta_1 \hat{x}_i + \tau_i$$

Note that the uncertainty in the fitted values from the first regression is not considered when performing the second regression. For this reason, the variance of the estimator is incorrect in two-stage calculations [15,26]. This and other observations led researchers to develop likelihood-based MR methods.

#### 3.3. Likelihood-based methods

Likelihood-based MR methods, unlike two-stage methods, provide maximum likelihood estimates with their many desirable properties [27]. Limited information maximum likelihood from econometrics is the earliest approach for likelihood-based inference in MR [28,29]. Limited information maximum likelihood with a single SNP instrument is modeled with two equations (Eqs. (4) and (5)), where the random errors follow a bivariate normal distribution [15].

$$x_i = \alpha_0 + \alpha_1 z_{i1} + \epsilon_{xi}$$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_{yi}$$

Limited information maximum likelihood is sometimes called the maximum likelihood counterpart of two-stage least squares, and it yields the same causal estimate as two-stage least squares and the ratio method when used with a single SNP instrument. Additionally, the limited information maximum likelihood framework can accommodate more than one SNP instrument by replacing  $\alpha_1 z_{i1}$  with the sum  $\sum_{k=1}^K \alpha_k z_{ik}$ .

One may also use Bayesian methods to obtain likelihood-based estimators [30]. Kleibergen [31] examined a model that is similar to that from the limited information maximum likelihood framework (Eq. (6)). The Bayesian model differs from the limited information maximum likelihood model in that the causal effect parameter,  $\beta_1$  represents the effect between the true means for the exposure and the outcome. In the limited information maximum likelihood model, the causal effect is that of the measured effect on the measured outcome.

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \xi_i \\ \eta_i \end{pmatrix}, \Sigma \right)$$

$$\xi_i = \alpha_0 + \alpha_1 z_{i1}$$

$$\eta_i = \beta_0 + \beta_1 \xi_i$$

For each subject  $i$ , the exposure and outcome values come from a bivariate normal distribution. The mean of the exposure distribution is assumed linear in the SNP instrument, and the mean of the outcome distribution is linear in the mean exposure [15,32]. Kleibergen [31] demonstrated that the Bayesian model, with weak instruments, outperforms the limited information maximum likelihood model in terms of frequentist coverage levels.

### 3.4. Semiparametric methods

Semiparametric instrumental variable methods, which feature parametric and nonparametric components, typically assume a parametric model connecting the exposure and outcome, but don't impose distributional assumptions on the errors [15]. Compared to fully parametric models, semiparametric models often are more robust to model misspecification [15,33]. We follow Burgess [15] by discussing three semiparametric strategies, generalized method of moments, continuous updating estimator, and G-estimation of structural mean models.

Generalized method of moments can be viewed as a more flexible form of two-stage least squares that handles heteroscedastic errors and nonlinearity in the two regressions [15]. Before specifying the model, we introduce notation.  $E(Y \vee \text{do}(X = x))$  is the conditional expectation of  $Y$  if we forced  $X$  to take value  $x$  for every subject [34]. The generalized method of moments equations, with a single SNP instrument, then, can be written as in Eq. (7).

$$\sum_i (y_i - f(x_i; \beta)) = 0$$

$$\sum_i z_i (y_i - f(x_i; \beta)) = 0$$

The GMM estimate is the value of the vector  $\beta$  that satisfies Eq. (7), where  $f(x_i; \beta) = E(Y \vee \text{do}(X = x))$ . Pearl [35] developed numerical methods for obtaining estimates from Eq. (7).

Another semiparametric estimation method is G-estimation of a structural mean model [15,36,37]. We follow Burgess [15] by defining the potential outcome  $Y(x)$  as the outcome value that we would have observed had we set the exposure value  $X$  to  $x$ . For example,  $Y(0)$  denotes the observed outcome had we set the exposure to zero instead of its observed value  $x$ . The structural mean model for a continuous outcome is displayed in Eq. (8).

$$E(Y(x) - Y(0)|X = x, Z = z) = \beta_1 x$$

The causal effect parameter is  $\beta_1$ . Burgess [15] derives the estimating equations for  $\beta_1$ , after noting that the conditional expectation,  $E(Y(0)|X = x, Z = z)$  is independent of  $Z$  and reasoning that the causal effect is that value of  $\beta_1$  that yields zero covariance between  $Z$  and  $E(Y(0) \vee X = x, Z = z)$  (Eq. (9)).

$$\sum_i (z_{ik} - z_k)(y_i - \beta_1 x_i) = 0$$

Note that  $k$  indexes SNP instruments and ranges from 1 to  $K$  in Eq. (9).

### 3.5. Mendelian randomization with multiple independent SNP instruments

While using a single SNP instrument for MR analysis can be effective, Greenland [38] recognized that many genetic variants individually explain a small proportion of the variation in a trait, and, thus, sufficient statistical power of MR analysis with a single SNP instrument would require sample sizes in the tens of thousands [39,40]. Schatzkin [41] resolved this issue by proposing use of multiple SNP instruments in MR analysis. In particular, Palmer [42] reasoned that a single causal estimate derived from a collection of causal SNPs would have greater precision than an estimate derived from only one SNP [39]. Recognition of complex traits' diverse genetic architectures has thus fueled development of two-sample MR methods with multiple SNP instruments [42–45].

There are two important considerations for MR analysis with multiple SNP instruments. The first consideration is how to choose a subset of the available SNPs to serve as instruments. The solution to this task differs among published methods and majority of MR methods chose a set of independent SNPs as instruments [46,47]. The independent SNPs may be selected through linkage disequilibrium (LD) clumping [48]. The second consideration is how to make

use of GWAS summary statistics for MR analysis, as sharing of GWAS summary statistics was encouraged and became a common practice [49]. [50] recognized that one could calculate the Wald ratio estimate of the causal effect from GWAS summary statistics, and [39] devised a strategy for determining a single causal effect from summary statistics for a collection of independent SNPs. Below, we discuss a number of MR methods that make use of multiple independent SNP instruments.

## 4. Statistical models and methods for MR with multiple instruments

### 4.1. Methods with Individual-level data

One approach to integrating multiple SNP instruments into a MR framework is through calculation of allele scores [51]. Harbord [51] calculate an unweighted score as the number of exposure-increasing alleles in the subject's genotypes. They also calculate a weighted allele score by using the exposure effects as weights. For example, a subject with  $g_k$  copies of exposure-increasing alleles for SNP  $k$  has an unweighted allele score  $z = \sum_{k=1}^K g_k$  and a weighted allele score  $z = \sum_{k=1}^K w_k g_k$ . Harbord [51] found that when weights are obtained from external data or from cross-validation or jackknife approaches applied to the analysis data, the allele score functions as a single instrumental variable and greatly diminishes the bias compared to that of the two-stage least squares estimator [15,52].

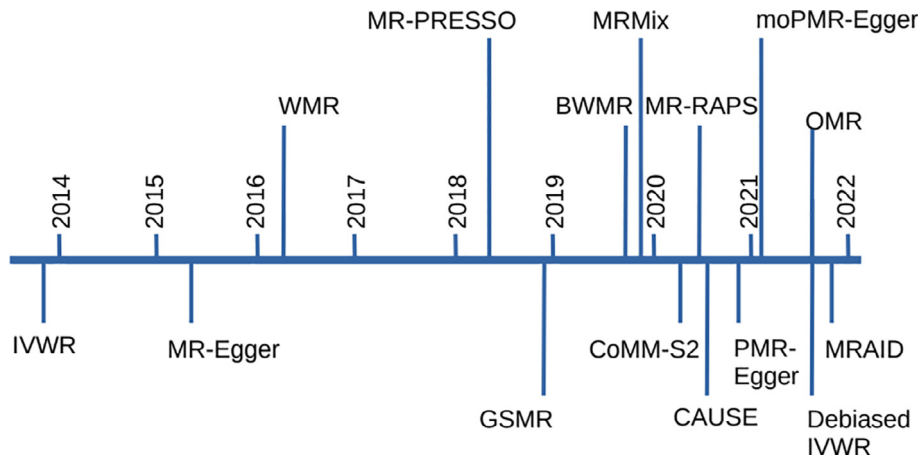
Angrist [53] developed a LASSO-based method, sisVIVE, to identify invalid SNP instruments [54]. sisVIVE's advantage over earlier methods is that it doesn't require the analyst to know which SNP instruments are valid. Instead, it requires that at least 50% of the instruments be valid. sisVIVE outperforms two-stage least squares in many ways and performs similar to oracle two-stage least squares. Simulations and data analysis results reveal that sisVIVE is robust to possibly invalid instruments.

Tibshirani [55], building on the research from [53], implemented an adaptive LASSO-based estimator after recognizing that sisVIVE misclassifies valid SNP instruments as invalid when the invalid SNP instruments have strong effects on the exposure. Consistent selection of invalid SNP instruments, they found, depends on SNP instrument correlations. To address this issue, Tibshirani [55] proposed a median estimator with consistency that doesn't depend on SNP instrument-exposure association strength or the SNP instrument correlation structure. They then applied methods from Windmeijer [56] to achieve a consistent estimator with the same asymptotic distribution as the oracle two-stages least squares. One important limitation of this thread of research is that both Angrist [53] and Tibshirani [55] require individual-level data.

Zou [57], working with individual-level GWAS data, share naive and smoothed constrained instrumental variable methods. Central to their work is the proposal to create a new instrumental variable as a linear combination of genotype data from a collection of SNP instruments. They require that the new instrumental variable be standardized to have norm 1 and be orthogonal to potentially vertically pleiotropic traits. For a collection of SNPs, they choose the  $K$ -vector that meets these criteria and maximizes the correlation with the exposure. To obtain their second estimator, Zou [57] apply a  $l_0$  penalty to constrain the number of SNP instruments that are given nonzero weights in the calculation of the new instrumental variable. They then compare their method with sisVIVE and allele score methods. Zou [57] find that both the smoothed version of their estimator and the allele score method are unbiased in their simulation settings.

Jiang [58] and Spiller [59] present a method, MR-GxE, that exploits gene by environment interactions to detect and to adjust





**Fig. 2.** Two-sample MR with GWAS summary statistics methods publications density increases over time.

for bias due to horizontal pleiotropy. Spiller [59] treat a SNP-covariate interaction as an instrumental variable, and, by so doing, they impose assumptions on that interaction. Together, these assumptions require that MR-GxE assume that horizontal pleiotropy effects are constant across the study cohort. The authors devise a three-step inferential procedure for two-sample MR with GWAS summary statistics. First, they estimate the SNP-exposure and SNP-outcome associations at a range of covariate values. Second, they regress the SNP-outcome estimated associations on the SNP-exposure estimated associations. Third, they treat the resulting slope as the causal effect estimate and the intercept as the mean horizontal pleiotropy effect.

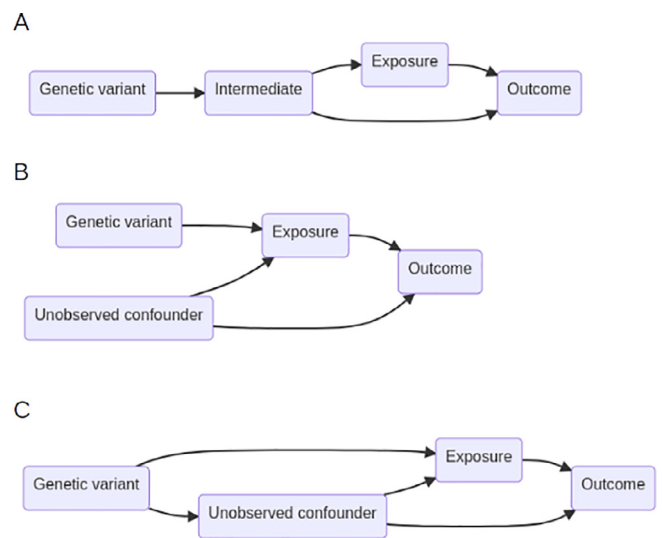
#### 4.2. Methods with summary data

The proliferation of publicly available GWAS data accelerates the development of MR methods with multiple SNP instruments that make use of GWAS summary statistics for model fitting. Our timeline highlights some of the many multi-instrument MR methods with GWAS summary statistics, with an increasing density of methods over time in the last three years (Fig. 2). These methods include the inverse variance weighted MR [39], MR-Egger [43], weighted median estimation [60], Bayesian weighted MR [61], robust adjusted profile score [62], MRMix [63], CAUSE [64], and MRaid [65]. These methods differ in their approaches to three considerations that include instrument selection (from among the available SNPs), modeling and controlling for horizontal pleiotropy, and statistical inference procedure. We summarize some of these MR methods that use independent or correlated SNP instruments below. We will cover the remaining MR methods that model horizontal pleiotropy in a later section.

Smith [39] termed their method “inverse variance weighting” (IVW) because each causal SNP’s effect is weighted by the inverse of the variance of the ratio estimator, and the overall causal effect is the sum of the weighted SNP causal effects. Specifically, Smith [39] combined causal effect ratio estimates from independent SNPs by using Eq. (10).

$$\hat{\beta}_{IVW} = \frac{\sum_{j=1}^M \hat{\theta}_j \text{Var}(\hat{\theta}_j)^{-1}}{\sum_{j=1}^M \text{Var}(\hat{\theta}_j)^{-1}}$$

Spiller [60], noting that consistency with IVW MR is not robust to invalid instrument use, developed a method for consistent MR analysis with weighted median-based estimators, building on



**Fig. 3.** Causal diagrams for MR. A. Scenario where the genetic variant affects an intermediate variable on the pathway to the exposure. Because the intermediate affects the outcome through a pathway that doesn’t involve exposure, this scenario violates the exclusion restriction assumption. B. Scenario where the genetic variant affects an exposure, which in turn affects an outcome, possibly in the presence of unmeasured confounding. C. Correlated horizontal pleiotropy occurs when the genetic variant affects the exposure, which in turn affects the outcome, and the genetic variant affects the unobserved confounder, which in turn affects the exposure and the outcome independently.

research from [66]. Its breakdown point is 50%, meaning that up to 50% of SNP instruments can be invalid while maintaining consistency in estimation [67]. Specifically, let  $\tilde{\beta}_j$  denote the  $j^{\text{th}}$  ordered ratio estimate statistic, from least to greatest. If an odd number, say  $2K + 1$ , of SNP instruments is used, then the simple median estimator is defined as the  $K + 1$  ordered ratio estimate. For an even number, say  $2K$ , number of SNP instruments, choose the mid-point between the two middle-ranking ordered ratio estimates,  $\frac{\tilde{\beta}_K + \tilde{\beta}_{K+1}}{2}$ .

Due to the inefficiency of the simple median estimator, Spiller [60] define the weighted median estimator. To construct this estimator, order the ratio estimates from least to greatest, as with the simple median estimator. The weighted median estimator, then is the median of a distribution defined by having  $\tilde{\beta}_j$  as the  $p_j = 100$  percentile, where  $w_j$  is the weight assigned to the  $j^{\text{th}}$  ordered ratio estimate. In this setting, the simple median estimator is seen to be the weighted median estimator when all weights are equal.

Spiller [60] also studied the penalized weighted median estimator. They defined the penalized weights,  $w^i = w_j' \times \min(1, 20q_j)$ , where  $q_j$  is the p-value resulting from the comparison of  $Q_j = w_j'(\tilde{\beta}_j - \hat{\beta}_{IVW})^2$  to a  $\chi_1^2$  distribution [60,68]. The penalized weighted median estimator leaves most variants unaffected, but downweights those with outlying Wald ratio estimates. While the above MR methods all use independent SNP instruments, using independent instruments may not be ideal when there are multiple causal SNPs in linkage disequilibrium with each other. In this setting, discarding some of the causal SNPs may only capture a small proportion of trait variance explained by the exposure and lead to a power loss in MR [15,39,69,70]. Consequently, many MR methods have been recently developed to model multiple correlated instruments.

Cochran [69] developed methods for using correlated SNPs in IVW MR by inserting a covariance matrix into the standard IVW formulation. However, many MR methods use a set of pre-selected SNPs as instruments. Typically, these SNPs are selected to be statistically independent. This restriction to independent SNPs is needed for valid inference in methods like standard inverse variance weighting MR [39].

#### 4.3. Assumption violations and sensitivity analysis

It is worthwhile to consider ways in which the standard MR assumptions may be violated and strategies for identifying such violations. Let's first consider the exclusion restriction assumption (assumption 2 above). It states that the SNP instrument must affect the outcome only through the exposure (Fig. 3B). The causal diagram in Fig. 3A illustrates a scenario that violates this assumption [19]. Bias in causal effect estimation would result. However, if the intermediate is measured and treated as the exposure, then an unbiased estimate of causal effect is possible, since the intermediate captures the two pathways, intermediate  $\rightarrow$  outcome and intermediate  $\rightarrow$  exposure  $\rightarrow$  outcome, through which the SNP instrument affects the outcome Y [19].

Recognition of the many ways that the MR assumptions may be violated has inspired methods advances that enable assumption relaxations and has motivated use of sensitivity analysis to quantify the impact of possible assumption violations. Sensitivity analysis is recommended and widely used in MR studies because of the inability to verify the three MR assumptions with observational data. The goal of a sensitivity analysis is to gain insight into how the results might differ if the assumptions be violated. Because not all confounding variables are known or measured, assumptions 2 and 3 are not fully verifiable in MR studies. Recognition of this fact has inspired the development of sensitivity analysis tools. While we can't assess whether a SNP instrument is associated with every confounder of the exposure-outcome association, it is possible to examine the associations of the SNP instrument with the measured covariates. While absence of such associations doesn't guarantee satisfaction of the assumption, presence of SNP-covariate associations must be investigated carefully, as they may constitute assumption violations [71].

Burgess [71] considered a collection of sensitivity analysis methods when working with GWAS summary statistics. Burgess [71] presented methods for both assessing the MR assumptions, to the extent possible, and performing robust analyses. For example, they used measured covariates to assess for possible associations with the SNP instruments. While they can't rule out the possibility of unmeasured confounding, they can study the possibility of measured covariates serving as confounders.

To illustrate their approach, Burgess [71] shared a case study in which they examined the causal effect of C-reactive protein

(CRP) levels on coronary artery disease risk with four genetic variants in the CRP gene region and 17 other genetic variants that affect coronary artery disease risk. They used measured covariates to probe for SNP-covariate associations, and followed it up with scatter plots and Cochran's Q test on the causal estimates to inquire about whether the SNP instruments all identify the same causal parameter [72–74]. Additionally, Burgess [71] suggested using a funnel plot, like those in the meta-analysis literature, to visualize possible evidence of directional pleiotropy, where the average pleiotropic effects of the SNP instruments is nonzero [75]. Additionally, Egger regression can be useful in this setting [43,76–78].

Spiller [59] point out that MR-GxE can be used as a sensitivity analysis tool to choose a set of valid SNP instruments from a collection of SNPs in settings where there may be a violation of the constant pleiotropy assumption.

MR-GENIUS is a framework for two-sample MR analysis when individual-level data are available [79]. Bowden [79] focused efforts on relaxing the exclusion restriction assumption and building on existing G-estimation methods to create an estimator that is robust to violations of the exclusion restriction and to additive unmeasured confounding [80]. They observe that their estimator, in some settings, reduces to that of [81], which is widely used in econometrics, but not in the MR literature.

#### 4.4. Explicit modeling of horizontal pleiotropy

Besides sensitivity analysis, several methods have been developed to directly model horizontal pleiotropy to ensure the validity of MR assumptions. Pleiotropy, where a single genetic variant affects multiple traits, has a long history of study in genetics and complex traits [82–84]. In MR setting, horizontal pleiotropy occurs when a SNP instrument affects the outcome through at least one pathway that bypasses the exposure variable [46]. The presence of horizontal pleiotropy constitutes a violation of the standard MR assumptions and can lead to biased causal effect estimates and diminished statistical power. Watanabe [46] documented widespread horizontal pleiotropy in GWAS. We discuss approaches to modeling and accounting for horizontal pleiotropy below.

In the short time since publication of [46], researchers have recognized two types of horizontal pleiotropy [65]. The first occurs via exposure-independent paths. The resulting horizontal pleiotropic effects are independent of the SNP-exposure relationships. The second type of horizontal pleiotropy manifests in the presence of unobserved exposure-outcome confounding. It induces correlation between horizontal pleiotropic effects and SNP-exposure effects. Both types of horizontal pleiotropy violate standard MR modeling assumptions and can bias causal effect estimates and can increase false discoveries [65]. Early MR analyses avoided confounding from horizontal pleiotropy by discarding instrumental SNPs that might be associated with the outcome [46,47]. More recent methods have attempted to model horizontal pleiotropy [64,65,85]. CAUSE [64] and MRMix [63] both use a mixture of normal distributions to control for both types of horizontal pleiotropy. Modeling both types of horizontal pleiotropy is particularly challenging because the MR model likelihood often involves an integral that can't be solved analytically. Because of this issue with the model likelihood, both MRMix and CAUSE use other, non-likelihood-based methods for inference.

Pierce [43] adapted Egger regression, originally developed to detect bias in meta-analyses, to detect bias from horizontal pleiotropy in two-sample MR studies with GWAS summary statistics or individual-level data [76]. Their approach is termed MR-Egger. They formulate their model as

$$X_i = \sum_{j=1}^J \gamma_j G_{ij} + U_i + \epsilon_{Xi}$$

$$Y_i = \sum_{j=1}^J \alpha_j G_{ij} + \beta X_i + U_i + \epsilon_{Yi}$$

where  $U$  designates confounders,  $X$  and  $Y$  denote the exposure and outcome, respectively,  $G$  is a SNP instrument, and  $\epsilon_X$  and  $\epsilon_Y$  are the zero-mean, normally distributed random errors.

Pierce [43] then write the outcome in terms of the SNP instrument  $j$  and the effects defined in Eq. (11) and define the new parameter  $\Gamma_j = \alpha_j + \beta\gamma_j$  (Eq. (12)).

$$Y_i = (\alpha_j + \beta\gamma_j) G_{ij} + \epsilon_{Yij}$$

$$= \Gamma_j G_{ij} + \epsilon_{Yij}$$

In Eq. (12),  $\alpha_j = 0$  when SNP  $j$  is a valid instrument.

MR-Egger fits the regression model in Eq. (13).

$$\hat{\Gamma}_j = \beta_{0E} + \beta_E \hat{\gamma}_j$$

The notation in Eq. (13) could be a little misleading. The independent and dependent variables are  $\hat{\gamma}_j$  and  $\hat{\Gamma}_j$ , respectively, while the regression parameters are  $\beta_{0E}$  and  $\beta_E$ .

The slope coefficient  $\hat{\beta}_E$  estimates the causal effect of the exposure on the outcome. Under the assumption that the SNP-exposure relationship is independent of the SNP's pleiotropic effects, Egger's test offers a valid test of the null causal hypothesis and consistently estimates the causal effect, even if all SNPs are invalid instruments [43]. In efforts to detect the presence of horizontal pleiotropy in two-sample MR studies with GWAS summary data, [72] developed the between-instrument heterogeneity test. In a meta-analysis of MR Wald ratio estimates, one per SNP instrument, they calculate  $Q$  in Eq. (14).

$$Q = \sum_{k=1}^K w_k \left( \hat{\beta}_{XY}^{(k)} - \mu_F \right)^2$$

In Eq. (14),  $w_k$  is the inverse variance of the Wald estimator, and

$$\mu_F = \frac{\sum_{k=1}^K w_k \hat{\beta}_{XY}^{(k)}}{\sum_{k=1}^K w_k}$$

While the  $Q$  test tends to be conservative in small sample sizes, its power increases with increasing sample size and increasing degree of pleiotropy [72].

Yuan [86], working in the setting of two-sample MR with summary GWAS data, recognized that measurement error in outcome and exposure can give misleading results with the inferred causal arrow in the wrong direction. To work around this issue, they extended MR with a method that infers the causal direction between two traits. Specifically, Yuan [86] adapt Steiger's  $Z$  test for a difference in correlations [87] to assess which of the two traits in an analysis has the stronger correlation with the SNP instrument. That with the stronger correlation is treated as the exposure. Yuan [86] applied their test in a study of DNA methylation and gene expression, where either direction of causality is plausible, and found that many methylation traits cause changes in gene expression.

Steiger [88] developed a weighted mode-based MR causal effect estimation method for two-sample MR analysis with GWAS summary data. They first calculate Wald ratio estimates for every instrumental SNP. They then apply smoothing to the empirical distribution of ratio estimates. The mode of this smoothed distribution is the simple mode-based estimate of the causal effect [88]. The inverse variance-weighted mode-based estimate is obtained by weighting the empirical distribution of ratio estimates by the inverse variance of each estimate. Steiger [88] compared their weighted and unweighted mode-based estimators with weighted and unweighted median-based estimators, MR Egger regression, and IVW. They found that their mode-based estimators demon-

strated less bias and lower type I error rates than other estimators in some simulation settings. The mode-based estimators possessed lower statistical power to detect a causal effect compared to the IVW and weighted median-based methods, but their power exceeded that of MR Egger regression [88]. The mode-based estimators consistently estimated the causal effect when the mode across SNP instruments of the horizontal pleiotropy effects was zero [88]. In this manner, the mode-based estimators demonstrated a greater robustness to horizontal pleiotropy than did the other estimators.

Hartwig [89] approached the challenge of horizontal pleiotropy by developing three new methods for two-sample MR analysis with GWAS summary data: robust regression, penalized weights, and LASSO penalization. The first two can be viewed as modifications of MR-Egger and IVW methods. Together, they offer three strategies for downweighting or excluding variants with heterogeneous causal estimates. Recall that MR-Egger offers consistent causal effect estimates when there is no horizontal pleiotropy or when the horizontal pleiotropy effects adhere to the "Instrument Strength Independent of Direct Effect" (InSIDE) assumption [43]. The InSIDE assumption is satisfied when there is no correlation between the pleiotropic SNP instrument-outcome effects and the SNP instrument-exposure effects. Hartwig [89] uses MM estimation, which is one technique for robust linear regression [90], to formulate an estimator with high breakdown point and high efficiency. The second estimation method, with penalized weights, is much like the penalized median estimator, from above, except that the second argument in the minimum function is, instead of the above. Finally the LASSO-based method fashions an estimator by applying a penalty to the objective function from MR-Egger to get Eq. (15) as the objective function.

$$\sum_j \text{Var}(\hat{\beta}_j)^{-1} \left( \hat{\beta}_{Yj} - \theta_{0j} - \theta_1 \hat{\gamma}_j \right)^2 + \lambda \sum_j |\theta_{0j}|$$

Hartwig [89] studied two procedures for choosing  $\lambda$ : a heterogeneity stopping rule and a cross-validation rule. Taken together, these three methods offer additional contributions to a suite of sensitivity analysis tools that practitioners should consider.

Koller [91] introduced the contamination mixture method for robust and efficient MR analysis with hundreds of SNPs, some of which may be invalid instruments by violating one or more of the three core MR assumptions. Working in the setting of two-sample MR with GWAS data, they begin by considering a collection of candidate SNP instruments, which may contain some invalid SNP instruments. For each candidate SNP instrument, they calculate the Wald ratio estimate of the causal effect. With the collection of Wald ratio estimates, Koller [91] reason that the set of valid SNP instruments have Wald ratio estimates that arise from a normal distribution centered on the true causal parameter with variance equal to that of the Wald ratio estimator. The invalid instruments, however, are assumed to follow a normal distribution centered at zero with a variance greater than that of the Wald ratio estimator. They then work with profile log likelihood over a grid of causal parameter values to make inferences.

Zhao [62] presented MR-RAPS, a method that leverages a robust adjusted profile score to accomplish statistical inference in two-sample MR analyses with GWAS summary data. Dividing pleiotropic effects into systematic and idiosyncratic, they model the systematic pleiotropy with random effects. In so doing, no SNP satisfies the exclusion restriction assumption. Adjusting their profile likelihood estimator from the setting without pleiotropy, [62] obtain an estimator with consistency and asymptotic normality [92]. Idiosyncratic pleiotropy is addressed through robustification of the adjusted profile score [93]. They then demonstrate the properties of their estimator by analyzing simulated and real data [62].

Huber [94], working with a collection of SNP instruments, reasoned that if all three MR assumptions are satisfied, then the collection of Wald ratio estimates should be homogeneous; thus, any departures from homogeneity may signal a violation of one or more assumptions. They reported modified weights, in addition to studying weights derived from first order and second order approximations to the variance. Their modified weights lead to an estimator that is useful for quantifying heterogeneity and detecting outliers.

Bowden [95] developed the MR-TRYX framework for exploiting horizontal pleiotropy to identify alternative causal pathways. They did this by positing that, for a collection of putative SNP instruments, any heterogeneity in the Wald ratio estimates may be due to horizontal pleiotropy. Once they identify a SNP with outlying ratio estimates, they search GWAS results by querying the MR-Base database to identify complex traits (in other studies) that associate with the outlying SNP [96]. They follow up promising trait associations to account for their observations with multiple causal pathways.

Hemani [97] recognized that in some MR studies, the Wald ratio estimates for different SNP instruments may form clusters, and these clusters may represent distinct causal mechanisms and may identify the SNPs that are involved in them. They devised a method to detect these clusters in the collection of Wald ratio estimates and found SNP instruments that point to distinct causal pathways of biological importance [97].

Fang [98] sought a probabilistic model that accounts for both linkage disequilibrium among SNP instruments and horizontal pleiotropy. They termed their method MR-LDP. Using an approach from [99,98] derived an approximate likelihood for their regression models of the exposure and outcome on the SNP instruments. Incorporation of a random effect into their models accounted for the variance in causal effect estimates due to horizontal pleiotropy. They situate inference within an empirical Bayesian framework and present a parameter-expanded variational Bayes expectation–maximization algorithm for estimation [100]. By jointly modeling the distribution of GWAS summary statistics and causal effects, [98] efficiently accommodates multiple SNP instruments in linkage disequilibrium. One limitation of MR-LDP is its reliance on the InSIDE assumption and resulting inability to accommodate correlated horizontal pleiotropy.

Liu [101] present IMRP, a method for causal effect estimation and horizontal pleiotropy detection. They describe a test to distinguish vertical pleiotropy from horizontal pleiotropy. By combining this test with two-sample MR analysis of GWAS summary statistics, they develop an iterative procedure for estimating the causal effect of exposure on outcome and test for presence of horizontal pleiotropy in the SNP instruments. The iterative procedure alternates between 1) updating the estimate of the causal effect and 2) testing for horizontal pleiotropy with the included SNP instruments. At each test for horizontal pleiotropy, some SNP instruments may be discarded if they demonstrate evidence of horizontal pleiotropic effects. The procedure ends when there remain only SNP instruments that don't exhibit horizontal pleiotropy. In the authors' simulations and data analysis sections, their new method performs about as well as GSMR, IVW, and MR-PRESSO in the settings with and without balanced pleiotropy and with and without satisfaction of the InSIDE assumption [101]. Finally, the authors note that IMRP is three orders of magnitude faster than simulations-based MR-PRESSO.

Zhu [102] build on research from Grant [103] to present a method that applies a  $\ell_1$  penalty to coefficients of covariates in two-sample MR with GWAS summary statistics. Importantly, their method doesn't penalize the SNP-exposure association. Instead, it shrinks towards zero those coefficients for covariates that demonstrate little pleiotropy. In so doing, they extend work from [104]

and [105] to get a causal effect estimator that needs no valid SNP instruments and is robust to measured vertical pleiotropy and more statistically efficient than previous MR methods.

Burgess [106] consider two classes of horizontal pleiotropy with their method, MRCIP. They partition horizontal pleiotropic effects into correlated or idiosyncratic horizontal pleiotropy. Idiosyncratic horizontal pleiotropy is that with a large effect size, while correlated horizontal pleiotropy, like its definition elsewhere in our manuscript, refers to correlation between the effects of the SNP instruments on the exposure and the effects of the SNP instruments on the outcome. MRCIP relaxes both the exclusion restriction assumption and the InSIDE assumption by directly modeling the correlated horizontal pleiotropy with random effects. Furthermore, to accommodate the idiosyncratic pleiotropy, the authors downweight SNP instruments that demonstrate strong direct effects on the outcome. They find that their MRCIP provides valid causal inference even when there are no valid SNP instruments (i.e., when all SNP instruments are invalid) and when the InSIDE assumption is violated [106]. Their PRW-EM algorithm adds a reweighting step to the traditional expectation maximization algorithm to estimate the causal effect parameter in MR analysis [107]. Finally, they compare MRCIP to other MR methods with both simulated and real data sets [106], like the developers of MR-RAPS, assume that the horizontal pleiotropy is balanced, i.e., that the mean horizontal pleiotropic effect is zero. One direction for future study is to augment MRCIP to permit unbalanced horizontal pleiotropy.

Dempster [108] build on other research using mixture models in the two-sample MR setting with GWAS summary statistics. Like MR-Clust, MR-PATH, as the method from [108] is called, works with the Wald ratio estimates from multiple SNP instruments. Coining the term “mechanistic heterogeneity” to refer to the distribution of Wald ratio estimates due to SNP involvement in distinct causal pathways, [108] specify a  $J$ -component normal mixture model. They use an expectation maximization algorithm to fit the model parameters (the component weights and the parameters for each of the  $J$  normal distributions) and suggest a Bayesian information criterion-based approach to choosing  $J$  [109].

Schwarz [110] propose causal effect test statistics that are robust under weak instrument asymptotics [111]. Specifically, they extend three econometric methods to two-sample MR with GWAS summary statistics [112–114] and study the theoretical properties and the performance in applications. Schwarz [110] extend their methods to demonstrate conditions under which their estimators are equivalent to those in works from [62,55], and [94].

Moreira [115] developed BayesMR, a Bayesian framework for two-sample MR with individual-level data and discuss an approximation for use with GWAS summary statistics. Unlike many MR methods, BayesMR both aims to control for horizontal pleiotropy and to quantitatively assess the possibility of reverse causation, where the putative exposure is actually caused by the putative outcome. For independent SNP instruments, Moreira [115] permit each SNP instrument to have both a direct effect on the exposure and a horizontally pleiotropic direct effect on the outcome. Instrument effects and horizontally pleiotropic effects are assumed independent. Note that this is the Bayesian analog of the InSIDE assumption. After specifying the model, they present a nested sampling scheme [116–118] that simultaneously computes the model evidence for both causal directions (i.e., exposure causes outcome and outcome causes exposure) and acquires samples from the posterior distribution for Bayesian statistical inferences.

Handley [119] present BMRE, a Bayesian implementation of MR Egger. They apply weakly informative priors to the horizontal pleiotropy effect estimator in efforts to increase statistical power to detect a nonzero slope, i.e., the coefficient of the SNP instrument. They develop their method for two-sample MR with individual-



level data. Simulations demonstrate that BMRE outperforms the MR Egger method while maintaining type I error rates in the examined scenarios. The authors emphasize a potential role for BMRE in MR sensitivity analysis.

Schmidt [120] found that multivariable MR, where multiple exposures jointly cause an outcome, consistently estimates the effect of the exposures on the outcome. They examined multivariable MR with both individual-level data and GWAS summary statistics. Schmidt [120] also develop a generalized version of Cochran's Q test for quantifying SNP instrument strength and validity. However, this test requires knowledge of the covariances between the effects of the SNP instruments on the exposures. While these covariances could be estimated from individual-level data, such data are often unavailable in the GWAS setting. Importantly, the authors note that the estimands for multivariable MR and MR differ; multivariable MR estimates the direct causal effects of each exposure on the outcome, while MR estimates the total causal effect of an exposure on the outcome.

MR-link, reported by Sanderson [121], aims to account for unobserved pleiotropy and linkage disequilibrium by using individual-level data on the outcome variable and summary statistics for the exposure. MR-link uses a three-step procedure to estimate causal effects.

Use GCTA-COJO to obtain SNP instruments [122].

Prune all SNPs in linkage disequilibrium with the SNP instruments to obtain a set of “tag” SNPs with correlations less than 0.95.

Solve for the causal effect estimate with ridge regression of the outcome on the matrix resulting from concatenation of the SNP instrument genotypes matrix and the tag SNPs genotypes matrix.

Note that in step 3, ridge regression is preferred to ordinary least squares due to the limited statistical power of ordinary least squares in this setting. Presumably, this is due to collinearity in the matrix of tag SNP genotypes.

#### 4.5. Correlated horizontal pleiotropy

The second type of horizontal pleiotropy, from above, is sometimes called “correlated horizontal pleiotropy.” Correlated horizontal pleiotropy occurs when a SNP affects both the exposure and a variable that confounds the relationship between exposure and outcome (Fig. 3C). CAUSE and MRAID are two methods that aim to model correlated horizontal pleiotropy [64,65]. Their approaches to modeling correlated horizontal pleiotropy differ in important ways that ultimately affect performance. CAUSE allows for a proportion of SNPs to exhibit correlated pleiotropy, which [64] models as an effect on a shared, unobserved factor. The remaining SNPs are independent of this unobserved factor. Every SNP can have a nonzero pleiotropic effect on the outcome, and these pleiotropic effects are uncorrelated with the variant effects on the exposure. The model, then, is written as a mixture of bivariate normal distributions, and inference proceeds by borrowing ideas on adaptive shrinkage from [123]. Finally, [64] compares two models to determine whether the GWAS summary statistics are consistent with a causal effect of the exposure on the outcome. Specifically, they estimate the difference in the expected log pointwise posterior density for the model with causal effect fixed at zero and the model that permits a nonzero causal effect [124].

Morrison [65] found that MRMix is not robust to misspecification of SNP effect sizes and often is biased. CAUSE, they found, yields overly conservative p-values [65]. These observations motivated [65] to develop a new method, MRAID. MRAID accommodates both individual-level data and GWAS summary statistics

[65]. We focus on the approach that uses GWAS summary statistics. Two equations are central to MRAID modeling (Eqs. (16) and (17)).

$$\widehat{\beta}_X = \Sigma_1 \beta + e_x$$

$$\widehat{\beta}_Y = \alpha \Sigma_2 \beta + \Sigma_2 \eta_0 + \Sigma_2 \eta_1 + e_y$$

The estimated marginal effects for the  $p$  SNP instruments on the exposure are denoted by  $\widehat{\beta}_X$ , while  $\widehat{\beta}_Y$  represents the estimated effects of the same SNPs on the outcome. The  $p$  by  $p$  SNP instrument correlation matrices for the exposure and outcome are written as  $\Sigma_1$  and  $\Sigma_2$ , respectively.  $\Sigma_1$  and  $\Sigma_2$  can be estimated with 1000 Genomes Project data, for example, by choosing a subset of 1000 Genomes Project subjects that have similar ancestry [125]. The  $p$ -vector error terms,  $e_x$  and  $e_y$ , follow multivariate normal distributions with mean zero and variances  $\Sigma_1 \frac{\sigma_X^2}{n_1 - 1}$  and  $\Sigma_2 \frac{\sigma_Y^2}{n_2 - 1}$ , respectively. Morrison [65] then constructs a Gibbs sampler to make likelihood-based inferences [126]. To facilitate computations, the investigators make several assumptions about the collection of SNPs:

Relatively small proportion of SNPs have nonzero effects on the exposure

A relatively small proportion of SNPs demonstrate horizontal pleiotropy.

The chosen instrumental SNPs are more likely to display horizontal pleiotropy than are non-instrumental SNPs.

Those SNPs that display horizontal pleiotropy are more likely to demonstrate uncorrelated horizontal pleiotropy than correlated horizontal pleiotropy.

By encoding these assumptions in prior distributions, Morrison [65] enables the inferential procedures to accommodate the observed data in the context of the assumptions on SNPs.

#### 4.6. Binary, count, and time-to-event outcomes and exposures

While we've focused above on continuous outcomes and continuous exposures, some data are more naturally treated as binary, count, or time-to-event variables. MR methods for non-continuous outcomes and exposures is an active area of research. Kleibergen [31] and Geman [127] developed MR strategies for binary exposures and binary outcomes. Kleibergen [31] extended methods from Mendel [24] by fitting two logistic regressions, one for the outcome and one for the exposure. They then calculated the causal effect estimator as the ratio of the two logistic regression coefficient estimators. Geman [127] devised a new MR method for binary outcome and binary exposure by drawing on [128] and treating the exposure and outcome as correlated binary random variables. They implemented an iterative optimization algorithm to simultaneously infer the causal effect parameter and other parameters.

Researchers have also leveraged other generalized linear models for analysis in MR studies [129]. However, this area is relatively unexplored. Nelder et al. [130] treated hospitalizations as a count variable and modeled it with quasi-Poisson methods. Hazewinkel et al. [131] characterized statistical properties, including bias and power, with simulations of count random variables and binary random variables. With growing interest in modeling counts in biomedical data, where molecular phenotyping technologies now acquire RNA molecule counts and protein abundances, there may be opportunities for methods innovations in a generalized linear models framework for MR studies [132–135].

Recently, multiple teams of investigators have analyzed time-to-event traits in MR studies [136–140]. He et al. [140], for example, studied post-diagnosis survival time in women with breast

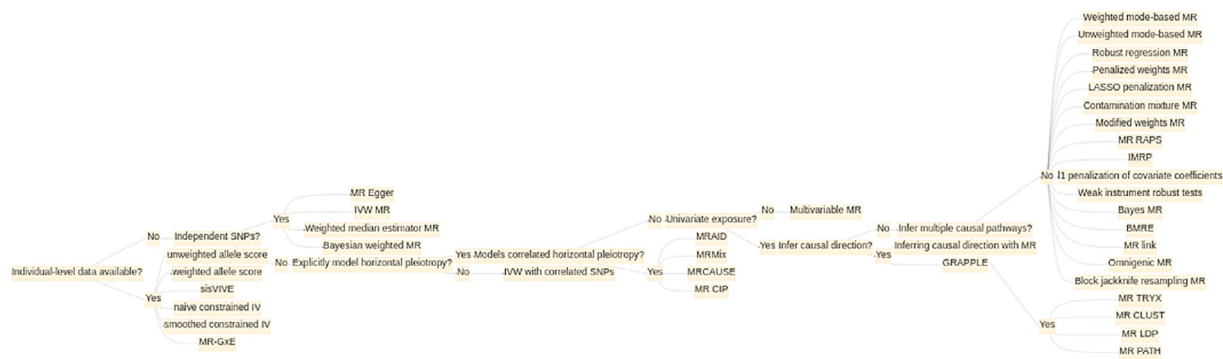


Fig. 4. Decision tree diagram for choosing among multiple-SNP MR methods.

cancer. They treated survival as a time-to-event outcome, which enabled them to model it with a proportional hazards model [141] in two-stage MR with individual-level data. Similarly, Tikkanen [139], in a study of cardiovascular disease, applied proportional hazards regression when examining time to stroke as an outcome variable. In very recent work, Bowden [79] reported methods, MR GENIUS, that are robust to violations of the exclusion restriction assumption (under many assumed MR data generating processes). They considered binary, continuous, and time-to-event outcomes within their framework. We are unaware of MR studies that treat the exposure variable as a time-to-event, but it's conceivable that future research will consider time-to-event traits as exposures in MR analyses.

## 5. Recent findings

### 5.1. Omnigenic MR

Recent research on the omnigenic hypothesis, which posits that every SNP's effect on a trait is nonzero, has informed MR methods development [142]. Boyle [143] recognized the limitations of previous MR methods in the context of the omnigenic hypothesis. In response, Boyle [143] developed a MR method that uses all genome-wide SNPs as instruments. Using GWAS summary statistics as inputs, their method relies on a composite likelihood framework for scalable computation and allows for horizontal pleiotropy. Boyle [143] used extensive simulations, including those with model misspecifications, to characterize their method's statistical power and robustness. Finally, they applied the new method to identify multiple complex traits that affect coronary artery disease and asthma. These causal relationships highlight the important roles of plasma lipids, blood pressure, and the immune system in CAD susceptibility and those of obesity and the immune system in asthma development.

### 5.2. Sample overlap in biobanks

The issue of overlapping samples in the two-sample MR with GWAS summary data has recently received attention in the published literature. Hemani [97], working with data from the UK Biobank, found that SNP instruments derived from overlapping samples explained a higher proportion of the variance compared to those derived from non-overlapping samples. They argue that block jackknife resampling MR enables causal inference while avoiding bias due to overlapping samples. Wang [144] analyzed 2514 traits from the UK Biobank and evaluated the impact of winner's curse on MR analysis. Winner's curse, they found, amplifies weak instrument bias without inflating the false discovery rate. This finding led the authors to design a pseudoreplication process that reduces bias in MR studies. Sadreev [145] proposed a Bayesian

approach for one-sample MR analysis where SNP instruments are permitted to exhibit pleiotropic effects on the outcome. Working in the setting of one-sample MR with individual-level data, they construct their model with a shrinkage prior, so that many SNP instruments have no horizontal pleiotropic effects. Their method relaxes the exclusion restriction assumption to permit horizontal pleiotropy (for some SNP instruments) and permits correlated SNP instruments, unlike many frequentist methods. Posterior inference proceeds by Markov chain Monte Carlo sampling. They also elaborate their model for a univariate exposure into that for a bivariate exposure to analyze the causal effects of body mass index and serum phenylalanine levels on blood pressure.

### 5.3. Bi-directional causal inference with MR to account for reverse causation

Reverse causation occurs when the outcome at an early time point has a causal effect on the exposure variable. While reverse causation is often assumed to not occur in MR study designs, [146] points out several scenarios where this assumption may not hold [147]. In recognition of this observation, [148] developed a framework, GRAPPLE, for performing two-sample MR with weak and strong instruments. Using GWAS summary statistics as inputs, GRAPPLE extends MR-RAPS and can detect presence of horizontally pleiotropic pathways, infer the causal direction, and perform multivariable MR. Central to the GRAPPLE framework is the observation that a horizontally pleiotropic pathway often gives rise to an additional mode in the profile likelihood. GRAPPLE uses the presence of multiple profile likelihood modes to diagnose horizontal pleiotropy with effects that are grouped by pathway. GRAPPLE facilitates identification of SNP instruments that contribute to the additional profile likelihood modes. A researcher using GRAPPLE may then recognize a confounding factor for each mode. With additional GWAS summary data with traits that reflect the confounding factors, GRAPPLE can fit multivariable MR models when the InSIDE assumption holds for the remaining horizontally pleiotropic effects [148].

### 5.4. Practice recommendations

We now turn attention to the question of how to choose among the MR methods for a given analysis. We focus on the two-sample setting where only GWAS summary statistics are available. The large number of two-sample MR methods with summary statistics may challenge a practitioner seeking to develop an analysis plan. Early considerations include:

- assessing the plausibility of the MR assumptions
- determining which questions to address with sensitivity analysis

assessing causal directionality between exposure and outcome variables

Detailed examples may be found in [149,150], and [151]. In some settings, a practitioner may address these questions by a careful study of the relevant scientific literature. For example, there may be experimental studies that suggest a causal direction for the relationship between exposure and outcome. If both causal directions are plausible, then a bidirectional MR analysis may be needed. If the bidirectional MR analysis (and supporting scientific evidence) is consistent with a single causal direction, then a practitioner may proceed with unidirectional MR methods. Given the pervasive nature of horizontal pleiotropy and its potential impact on MR, we recommend using one or more MR methods explicitly model horizontal pleiotropy. Recent methods, such as CAUSE and MRAID, go a step further by modeling correlated horizontal pleiotropy [64,65].

Other MR methods can be used in sensitivity analysis to assess robustness of findings to possible violations of MR assumptions. Methods like MR-clust have the potential to reveal additional relationships among genes that share a biological pathway [152]. The STROBE-MR statement has additional practice recommendations [153].

## 6. Summary and outlook

Open research questions remain in the MR field. We have highlighted above three recent findings: omnigenic MR, methods for sample overlap in biobanks, and methods for bidirectional causal inference. While we discussed initial findings in these three areas, many opportunities for methods enhancements remain. For example, as biobank sizes grow in number of subjects and number of measured traits, there is increasing demand for MR methods that scale efficiently with large sample sizes. Development of methods for two-sample MR with overlapping samples opens opportunities to study rich biobank data, and we anticipate that this question – how to accommodate overlapping samples – will continue to be an active area of research.

In summary, we have presented a comprehensive review on 47 methods for MR analysis in GWAS (Fig. 4). The overall trend of the methodology development for MR analysis is in the direction of increasingly sophisticated modeling of horizontal pleiotropy including both independent and correlated horizontal pleiotropy while attempting to maintain scalability and computational efficiency in the presence of multiple correlated SNPs. We hope our detailed review would benefit both methodology developers and applied analysts to further advance the development of MR methods and aid in their applications towards large, biobank-scale datasets, thus offering the possibility of discovering even more causal relationships among complex traits [154].

## CRediT authorship contribution statement

**Frederick J. Boehm:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Xiang Zhou:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This study was supported by the National Institutes of Health (NIH) grant R01HG009124.

## References

- [1] Wright PG. *Tariff on animal and vegetable oils*. New York: Macmillan Company; 1928.
- [2] Wright S. Correlation and causation. *J Agric Res* 1921;557–585.
- [3] Wright S. The theory of path coefficients: a reply to Niles's criticism. *Genetics* 1923;8:239–55.
- [4] Stock JH, Trebbi F. Retrospectives: who invented instrumental variable regression? *J Econ Perspect* 2003;17:177–94.
- [5] Katan M. Apolipoprotein e isoforms, serum cholesterol, and cancer. *Lancet* 1986;507–508.
- [6] Katan MB. Commentary: Mendelian randomization, 18 years on. *Int J Epidemiol* 2004;33:10–1.
- [7] Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007;16:309–30.
- [8] Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes MV, White J, et al. Selecting instruments for mendelian randomization in the wake of genome-wide association studies. *Int J Epidemiol* 2016;45:1600–16. <https://doi.org/10.1093/ije/dvw088>.
- [9] The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–78. <https://doi.org/10.1038/nature05911>.
- [10] Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, et al. Complement factor h polymorphism in age-related macular degeneration 2005;308:6.
- [11] Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucl Acids Res* 2014;42:D1001–6. <https://doi.org/10.1093/nar/gkt1229>.
- [12] Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: Biology, function, and translation. *Am J Hum Genet* 2017;101:5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>.
- [13] Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012;90:7–24. <https://doi.org/10.1016/j.ajhg.2011.11.029>.
- [14] Bunieello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. others, The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;47:D1005–12.
- [15] Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for mendelian randomization. *Stat Methods Med Res* 2017;26:2333–55. <https://doi.org/10.1177/0962280215597579>.
- [16] Richmond RC, Smith GD. Mendelian randomization: concepts and scope. *Cold Spring Harbor Perspect Med* 2022;12:a040501.
- [17] Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, et al. Mendelian randomization. *Nat Rev Methods Primers* 2022;2:6. <https://doi.org/10.1038/s43586-021-00092-5>.
- [18] Pierce BL, Burgess S. Efficient design for mendelian randomization studies: Subsample and 2-sample instrumental variable estimators. *Am J Epidemiol* 2013;178:1177–84. <https://doi.org/10.1093/aje/kwt084>.
- [19] VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. *Epidemiology* 2014;25:427–35. <https://doi.org/10.1097/EDE.0000000000000081>.
- [20] Burgess S, Thompson SG. CRP CHD Genetics Collaboration, Avoiding bias from weak instruments in mendelian randomization studies. *Int J Epidemiol* 2011;40:755–64. <https://doi.org/10.1093/ije/dyr036>.
- [21] Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statist Med* 2008;27:1133–63. <https://doi.org/10.1002/sim.3034>.
- [22] Palmer TM, Thompson JR, Tobin MD, Sheehan NA, Burton PR. Adjusting for bias and unmeasured confounding in mendelian randomization studies with binary responses. *Int J Epidemiol* 2008;37:1161–8. <https://doi.org/10.1093/ije/dyn080>.
- [23] Mendel G. Versuche uber pflanzen-hybriden. Vorgelegt in Den Sitzungen 1865.
- [24] Wald A. The fitting of straight lines if both variables are subject to error. *Ann Math Statist* 1940;11:284–300. <https://doi.org/10.1214/aoms/1177731868>.
- [25] Sawa T. The exact sampling distribution of ordinary least squares and two-stage least squares estimators. *J Am Stat Assoc* 1969;64:923–37. <https://doi.org/10.1080/01621459.1969.10501024>.
- [26] Angrist JD, Pischke J-S. Chapter 4: Instrumental variables in action: Sometimes you get what you need. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press; 2008.
- [27] R.A. Fisher, Theory of statistical estimation, in: *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, 1925: pp. 700–725.
- [28] R. Davidson, J.G. MacKinnon, Estimation and inference in econometrics, Oxford New York, 1993.
- [29] Hansen LP, Heaton J, Yaron A. Finite-sample properties of some alternative GMM estimators. *J Bus Econ Stat* 1996;14:262–80.



- [30] Kleibergen F, Zivot E. Bayesian and classical approaches to instrumental variable regression. *J Economet* 2003;114:29–72. [https://doi.org/10.1016/S0304-4076\(02\)00219-1](https://doi.org/10.1016/S0304-4076(02)00219-1).
- [31] Burgess S, Thompson SG. Improving bias and coverage in instrumental variable analysis with weak instruments for continuous and binary outcomes. *Statist Med* 2012;31:1582–600. <https://doi.org/10.1002/sim.4498>.
- [32] Jones EM, Thompson JR, Didelez V, Sheehan NA. On the choice of parameterisation and priors for the Bayesian analyses of Mendelian randomisation studies: Bayesian analyses of Mendelian randomisation studies. *Statist. Med.* 2012;31:1483–501. <https://doi.org/10.1002/sim.4499>.
- [33] Clarke PS, Windmeijer F. Instrumental variable estimators for binary outcomes. *J Am Stat Assoc* 2012;107:1638–52. <https://doi.org/10.1080/01621459.2012.734171>.
- [34] Pearl J. *Causality*. Cambridge University Press; 2009.
- [35] Palmer TM, Sterne JAC, Harbord RM, Lawlor DA, Sheehan NA, Meng S, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *Am J Epidemiol* 2011;173:1392–403. <https://doi.org/10.1093/aje/kwq026>.
- [36] Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Modell* 1986;7:1393–512. [https://doi.org/10.1016/0270-0255\(86\)90088-6](https://doi.org/10.1016/0270-0255(86)90088-6).
- [37] Greenland S, Lanes S, Jara M. Estimating effects from randomized trials with discontinuations: The need for intent-to-treat design and g-estimation. *Clin Trials* 2008;5:5–13. <https://doi.org/10.1177/1740774507087703>.
- [38] Smith GD. Mendelian randomization: Prospects, potentials, and limitations. *Int J Epidemiol* 2004;33:30–42. <https://doi.org/10.1093/ije/dyh132>.
- [39] Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;37:658–65. <https://doi.org/10.1002/gepi.21758>.
- [40] Schatzkin A, Abnet CC, Cross AJ, Gunter M, Pfeiffer R, Gail M, et al. Mendelian randomization: How it can—and cannot—help confirm causal relations between nutrition and cancer. *Cancer Prevent Res* 2009;2:104–13.
- [41] Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res* 2012;21:223–42. <https://doi.org/10.1177/0962280210394459>.
- [42] Pierce BL, Ahsan H, VanderWeele TJ. Power and instrument strength requirements for mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* 2011;40:740–52. <https://doi.org/10.1093/ije/dyq151>.
- [43] Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through egger regression. *Int J Epidemiol* 2015;44:512–25. <https://doi.org/10.1093/ije/dyv080>.
- [44] Hormozdiani F, Gazal S, van de Geijn B, Finucane HK, Ju C-J-T, Loh P-R, et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nat Genet* 2018;50:1041–7. <https://doi.org/10.1038/s41588-018-0148-2>.
- [45] Watanabe K, Stringer S, Frei O, Umičević Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* 2019;51:1339–48. <https://doi.org/10.1038/s41588-019-0481-0>.
- [46] Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet* 2018;50:693–8. <https://doi.org/10.1038/s41588-018-0099-7>.
- [47] Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 2018;9:224. <https://doi.org/10.1038/s41467-017-02317-2>.
- [48] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
- [49] Asking for more. *Nat Genet* 2012;44:733.
- [50] Harbord RM, Didelez V, Palmer TM, Meng S, Sterne JAC, Sheehan NA. Severity of bias of a simple estimator of the causal odds ratio in mendelian randomization studies. *Statist Med* 2013;32:1246–58. <https://doi.org/10.1002/sim.5659>.
- [51] Burgess S, Thompson SG. Use of allele scores as instrumental variables for mendelian randomization. *Int J Epidemiol* 2013;42:1134–44. <https://doi.org/10.1093/ije/dyt093>.
- [52] Angrist JD, Imbens GW, Krueger AB. Jackknife instrumental variables estimation. *J Appl Econ* 1999;14:57–67.
- [53] Kang H, Zhang A, Cai TT, Small DS. Instrumental variables estimation with some invalid instruments and its application to mendelian randomization. *J Am Stat Assoc* 2016;111:132–44. <https://doi.org/10.1080/01621459.2014.994705>.
- [54] Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc: Ser B (Methodol)* 1996;58:267–88.
- [55] Windmeijer F, Farbmacher H, Davies N, Davey Smith G. On the use of the lasso for instrumental variables estimation with some invalid instruments. *J Am Stat Assoc* 2019;114:1339–50. <https://doi.org/10.1080/01621459.2018.1498346>.
- [56] Zou H. The adaptive lasso and its oracle properties. *J Am Stat Assoc* 2006;101:1418–29.
- [57] Jiang L, Ouakacha K, Didelez V, Ciampi A, Rosa-Neto P, Benedet AL, et al. and for the Alzheimer's Disease Neuroimaging Initiative, Constrained instruments and their application to mendelian randomization with pleiotropy. *Genet Epidemiol* 2019;43:373–401. <https://doi.org/10.1002/gepi.22184>.
- [58] Spiller W, Slichter D, Bowden J, G. Davey Smith, Detecting and correcting for bias in mendelian randomization analyses using gene-by-environment interactions. *Int J Epidemiol* 2018. <https://doi.org/10.1093/ije/dvy204>.
- [59] Spiller W, Hartwig FP, Sanderson E, Smith GD, Bowden J. Interaction-based mendelian randomization with measured and unmeasured gene-by-covariate interactions. *Epidemiology* 2020. <https://doi.org/10.1101/2020.07.27.20162909>.
- [60] Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016;40:304–14. <https://doi.org/10.1002/gepi.21965>.
- [61] Zhao J, Ming J, Hu X, Chen G, Liu J, Yang C. Bayesian weighted mendelian randomization for causal inference based on summary statistics. *Bioinformatics* 2019;btz749. <https://doi.org/10.1093/bioinformatics/btz749>.
- [62] Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score. *Ann Statist* 2020;48. <https://doi.org/10.1214/19-AOS1866>.
- [63] Qi G, Chatterjee N. Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. *Nat Commun* 2019;10:1941. <https://doi.org/10.1038/s41467-019-09432-2>.
- [64] Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat Genet* 2020;52:740–7. <https://doi.org/10.1038/s41588-020-0631-4>.
- [65] Yuan Z, Liu L, Guo P, Yan R, Xue F, Zhou X. Likelihood-based mendelian randomization analysis with automated instrument selection and horizontal pleiotropic modeling 2022;8:1–15. <https://doi.org/10.1126/sciadv.abl5744>.
- [66] Han C. Detecting invalid instruments using L1-GMM. *Econ Lett* 2008;101:285–7.
- [67] Hampel FR. A general qualitative definition of robustness. *Ann Math Stat* 1971;42:1887–96.
- [68] Cochran WG. The comparison of percentages in matched samples. *Biometrika* 1950;37:256–66.
- [69] Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in mendelian randomization: Comparison of allele score and summarized data methods. *Statist Med* 2016;35:1880–906. <https://doi.org/10.1002/sim.6835>.
- [70] Burgess S, Thompson SG. Bias in causal estimates from mendelian randomization studies with weak instruments. *Statist Med* 2011;30:1312–23. <https://doi.org/10.1002/sim.4197>.
- [71] Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity analyses for robust causal inference from mendelian randomization analyses with multiple genetic variants. *Epidemiology* 2017;28:30–42. <https://doi.org/10.1097/EDE.0000000000000559>.
- [72] Greco FD, Minelli MC, Sheehan NA, Thompson JR. Detecting pleiotropy in mendelian randomisation studies with summary data and a continuous outcome: Detecting pleiotropy in mendelian randomisation studies with summary data and a continuous outcome. *Statist Med* 2015;34:2926–40. <https://doi.org/10.1002/sim.6522>.
- [73] Small DS. Sensitivity analysis for instrumental variables regression with overidentifying restrictions. *J Am Stat Assoc* 2007;102:1049–58. <https://doi.org/10.1198/016214507000000608>.
- [74] Higgins JPT. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–60. <https://doi.org/10.1136/bmj.327.7414.557>.
- [75] Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002–d. <https://doi.org/10.1136/bmj.d4002>.
- [76] Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629–34. <https://doi.org/10.1136/bmj.315.7109.629>.
- [77] Burgess S, Thompson SG. Interpreting findings from mendelian randomization using the MR-egger method. *Eur J Epidemiol* 2017;32:377–89. <https://doi.org/10.1007/s10654-017-0255-x>.
- [78] Bowden J, Del Greco F, M, C. Minelli, G. Davey Smith, N. Sheehan, J. Thompson. A framework for the investigation of pleiotropy in two-sample summary data mendelian randomization: A framework for two-sample summary data MR. *Statist Med* 2017;36:1783–802. <https://doi.org/10.1002/sim.7221>.
- [79] Tchetgen Tchetgen E, Sun B, Walter S. The GENIUS approach to robust mendelian randomization inference. *Statist Sci* 2021;36. <https://doi.org/10.1214/20-STS802>.
- [80] Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Commun Stat-Theory Methods* 1994;23:2379–412.
- [81] Lewbel A. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *J Busin Econ Stat* 2012;30:67–80.
- [82] Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: Challenges and strategies. *Nat Rev Genet* 2013;14:483–95. <https://doi.org/10.1038/nrg3461>.
- [83] Visscher PM, Yang J. A plethora of pleiotropy across complex traits. *Nat Genet* 2016;48:707–8. <https://doi.org/10.1038/ng.3604>.



- [84] Stearns FW. One hundred years of pleiotropy: A retrospective. *Genetics* 2010;186:767–73. <https://doi.org/10.1534/genetics.110.122549>.
- [85] Yuan Z, Zhu H, Zeng P, Yang S, Sun S, Yang C, et al. Testing and controlling for horizontal pleiotropy with probabilistic mendelian randomization in transcriptome-wide association studies. *Nat Commun* 2020;11:3861. <https://doi.org/10.1038/s41467-020-17668-6>.
- [86] Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* 2017;13:e1007081.
- [87] Steiger JH. Tests for comparing elements of a correlation matrix. *Psychol Bull* 1980;87:245.
- [88] F.P. Hartwig, G. Davey Smith, J. Bowden, Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption, *International J Epidemiol* 46 (2017) 1985–1998. [10.1093/ije/dyx102](https://doi.org/10.1093/ije/dyx102).
- [89] Rees JMB, Wood AM, Dudbridge F, Burgess S. Robust methods in mendelian randomization via penalization of heterogeneous causal estimates. *PLoS ONE* 2019;14:e0222362.
- [90] Koller M, Stahl WA. Sharpening wald-type inference in robust regression for small samples. *Comput Stat Data Anal* 2011;55:2504–15.
- [91] Burgess S, Foley CN, Allara E, Staley JR, Howson JMM. A robust and efficient method for mendelian randomization with hundreds of genetic variants. *Nat Commun* 2020;11:376. <https://doi.org/10.1038/s41467-019-14156-4>.
- [92] McCullagh P, Tibshirani R. A simple method for the adjustment of profile likelihoods. *J Roy Stat Soc: Ser B (Methodol)* 1990;52:325–44. <https://doi.org/10.1111/j.2517-6161.1990.tb01790.x>.
- [93] Huber PJ. Robust estimation of a location parameter. *Ann Math Stat* 1964;35. <https://doi.org/10.1214/aoms/1177703732>.
- [94] J. Bowden, F. Del Greco M, C. Minelli, Q. Zhao, D.A. Lawlor, N.A. Sheehan, J. Thompson, G. Davey Smith, Improving the accuracy of two-sample summary-data mendelian randomization: Moving beyond the NOME assumption, *Int J Epidemiol* 48 (2019) 728–742. [10.1093/ije/dyy258](https://doi.org/10.1093/ije/dyy258).
- [95] Y. Cho, P.C. Haycock, E. Sanderson, T.R. Gaunt, J. Zheng, A.P. Morris, G. Davey Smith, G. Hemani, Exploiting horizontal pleiotropy to search for causal pathways within a mendelian randomization framework, *Nat Commun* 11 (2020) 1010. [10.1038/s41467-020-14452-4](https://doi.org/10.1038/s41467-020-14452-4).
- [96] Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. *eLife* 2018;7:e34408.
- [97] S. Fang, G. Hemani, T.G. Richardson, T.R. Gaunt, G.D. Smith, Evaluating and implementing block jackknife resampling mendelian randomization to mitigate bias induced by overlapping samples, *medRxiv*. (2021).
- [98] Cheng Q, Yang Y, Shi X, Yeung K-F, Yang C, Peng H, et al. A two-sample mendelian randomization for GWAS summary statistics accounting for linkage disequilibrium and horizontal pleiotropy. *NAR Genom Bioinform* 2020;2:lqaa028. <https://doi.org/10.1093/nargab/lqaa028>.
- [99] Zhu X, Stephens M. Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. *Ann Appl Stat* 2017;11. <https://doi.org/10.1214/17-AOAS1046>.
- [100] Liu C, Rubin DB, Wu YN. Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika* 1998;85:755–70.
- [101] Zhu X, Li X, Xu R, Wang T. An iterative approach to detect pleiotropy and perform mendelian randomization analysis using GWAS summary statistics. *Bioinformatics* 2021;37:1390–400. <https://doi.org/10.1093/bioinformatics/btaa985>.
- [102] Grant AJ, Burgess S. An efficient and robust approach to mendelian randomization with measured pleiotropic effects in a high-dimensional setting. *Biostatistics* 2020;xxaa045. <https://doi.org/10.1093/biostatistics/kxaa045>.
- [103] Lin W, Feng R, Li H. Regularization methods for high-dimensional instrumental variables regression with an application to genetical genomics. *J Am Stat Assoc* 2015;110:270–88.
- [104] Burgess S, Thompson SG. Multivariable mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015;181:251–60. <https://doi.org/10.1093/aje/kwv283>.
- [105] S. Burgess, F. Dudbridge, S.G. Thompson, Re: “Multivariable mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects, *Am J Epidemiol* 181 (2015) 290–291. 10.
- [106] Xu S, Fung WK, Liu Z. MRCIP: A robust mendelian randomization method accounting for correlated and idiosyncratic pleiotropy. *Briefings Bioinf* 2021;22:bbab019. <https://doi.org/10.1093/bib/bbab019>.
- [107] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc: Ser B (Methodol)* 1977;39:1–22.
- [108] D. long, Q. Zhao, Y. Chen, A latent mixture model for heterogeneous causal mechanisms in mendelian randomization, *arXiv:2007.06476 [Stat]*. (2020). <http://arxiv.org/abs/2007.06476> (accessed March 1, 2022).
- [109] Schwarz G. Estimating the dimension of a model. *Ann Statis* 1978;461–464.
- [110] Wang S, Kang H. Weak-instrument robust tests in two-sample summary-data mendelian randomization. *Biometrics* 2021;biom.13524. <https://doi.org/10.1111/biom.13524>.
- [111] Staiger DO, Stock JH. Instrumental variables regression with weak instruments. (1994).
- [112] Anderson TW, Rubin H. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann Math Stat* 1949;20:46–63.
- [113] Kleibergen F. Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* 2002;70:1781–803. <https://doi.org/10.1111/1468-0262.00353>.
- [114] Moreira MJ. A conditional likelihood ratio test for structural models. *Econometrica* 2003;71:1027–48. <https://doi.org/10.1111/1468-0262.00438>.
- [115] Bucur IG, Claassen T, Heskes T. Inferring the direction of a causal link and estimating its effect via a bayesian mendelian randomization approach. *Stat Methods Med Res* 2020;29:1081–111. <https://doi.org/10.1177/0962280219851817>.
- [116] Skilling J. Nested sampling for general bayesian computation. *Bayesian Anal* 2006;1:833–59.
- [117] Handley W, Hobson M, Lasenby A. POLYCHORD: Next-generation nested sampling. *Month Notices R Astron Soc* 453 (2015) 4384–4398.
- [118] Handley W, Hobson M, Lasenby A. PolyChord: Nested sampling for cosmology, monthly notices of the royal astronomical society. *Letters* 2015;450:L61–5.
- [119] Schmidt AF, Dudbridge F. Mendelian randomization with egger pleiotropy correction and weakly informative bayesian priors. *Int J Epidemiol* 2018;47:1217–28. <https://doi.org/10.1093/ije/dyx254>.
- [120] Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable mendelian randomization in the single-sample and two-sample summary data settings. *Int J Epidemiol* 2019;48:713–27. <https://doi.org/10.1093/ije/dvy262>.
- [121] A. van der Graaf, A. Claringbould, A. Rimbert, BIOS Consortium, B.T. Heijmans, P.A.C.T. Hoen, et al., Mendelian randomization while jointly modeling cis genetics identifies causal relationships between gene expression and lipids, *Nat Commun* 11 (2020) 4930. [10.1038/s41467-020-18716-x](https://doi.org/10.1038/s41467-020-18716-x).
- [122] Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, J. Yang, T. Ferreira, A.P. Morris, S.E. Medland, P.A.F. Madden, A.C. Heath, et al., Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits, *Nat Genet* 44 (2012) 369–375. [10.1038/ng.2213](https://doi.org/10.1038/ng.2213).
- [123] Stephens M. False discovery rates: A new deal, *Biostat* (2016) kxw041. [10.1093/biostatistics/kxw041](https://doi.org/10.1093/biostatistics/kxw041).
- [124] Vehtari A, Gelman A, Gabry J. Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 2017;27:1413–32. <https://doi.org/10.1007/s11222-016-9696-4>.
- [125] The 1000 Genomes Project Consortium, L. Clarke, X. Zheng-Bradley, R. Smith, E. Kulesha, C. Xiao, I. Toneva, et al., The 1000 genomes project: Data management and community access, *Nat Methods* 9 (2012) 459–462. [10.1038/nmeth.1974](https://doi.org/10.1038/nmeth.1974).
- [126] Geman S, Geman D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* 1984;721–741.
- [127] Allman PH, Aban I, Long DM, Bridges SL, Srinivasasainagendra V, MacKenzie T, et al. A novel mendelian randomization method with binary risk factor and outcome. *Genet Epidemiol* 2021;45:549–60. <https://doi.org/10.1002/gepi.22387>.
- [128] Gauvreau K, Pagano M. The analysis of correlated binary outcomes using multivariate logistic regression. *Biom J* 1997;39:309–25. <https://doi.org/10.1002/bimj.4710390306>.
- [129] Nelder JA, Wedderburn RW. Generalized linear models. *J R Stat Soc: Ser A (General)* 1972;135:370–84.
- [130] Hazewinkel A-D, Richmond RC, Wade KH, Dixon P. Mendelian randomization analysis of the causal impact of body mass index and waist-hip ratio on rates of hospital admission. *Econ Hum Biol* 2022;44. <https://doi.org/10.1016/j.ehb.2021.101088>.
- [131] Allman PH, Aban I, Long DM, Patki A, MacKenzie T, Irvin MR, et al. Mendelian randomization in the multivariate general linear model framework. *Genet Epidemiol* 2022;46:17–31. <https://doi.org/10.1002/gepi.22435>.
- [132] Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* 2018;50:1–14. <https://doi.org/10.1038/s12276-018-0071-8>.
- [133] Ozsolak F, Milos PM. RNA sequencing: Advances, challenges and opportunities. *Nat Rev Genet* 2011;12:87–98. <https://doi.org/10.1038/nrg2934>.
- [134] Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science* 2006;312:212–7. <https://doi.org/10.1126/science.1124619>.
- [135] Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: A critical review. *Anal Bioanal Chem* 2007;389:1017–31. <https://doi.org/10.1007/s00216-007-1486-6>.
- [136] Burgess S. Commentary: Consistency and collapsibility. *Epidemiology* 2015;26:411–3. <https://doi.org/10.1097/EDE.0000000000000282>.
- [137] Arvanitis M, Qi G, Bhatt DL, Post WS, Chatterjee N, Battle A, et al. Linear and nonlinear mendelian randomization analyses of the association between diastolic blood pressure and cardiovascular events: The j-curve revisited. *Circulation* 2021;143:895–906. <https://doi.org/10.1161/CIRCULATIONAHA.120.049819>.
- [138] Tikkanen E, Gustafsson S, Knowles JW, Perez M, Burgess S, Ingelsson E. Body composition and atrial fibrillation: A mendelian randomization study. *Eur Heart J* 2019;40:1277–82. <https://doi.org/10.1093/eurheartj/ehz003>.
- [139] He L, Culminskaya I, Loika Y, Arbeeve KG, Bagley O, Duan M, et al. Causal effects of cardiovascular risk factors on onset of major age-related diseases: A time-to-event mendelian randomization study. *Exp Gerontol* 2018;107:74–86. <https://doi.org/10.1016/j.exger.2017.09.019>.
- [140] Guo Q, Burgess S, Turman C, Bolla MK, Wang Q, Lush M, et al. Body mass index and breast cancer survival: A mendelian randomization analysis. *Int J Epidemiol* 2017;46:1814–22. <https://doi.org/10.1093/ije/dyx131>.

- [141] Cox DR. Regression models and life-tables. *J Roy Stat Soc: Ser B (Methodol)* 1972;34:187–202. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>.
- [142] Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: From polygenic to omnigenic. *Cell* 2017;169:1177–86. <https://doi.org/10.1016/j.cell.2017.05.038>.
- [143] L. Wang, B. Gao, Y. Fan, F. Xue, X. Zhou, Mendelian randomization under the omnigenic architecture, (2021) 15.
- [144] I.I. Sadreev, B.L. Elsworth, R.E. Mitchell, L. Paternoster, E. Sanderson, N.M. Davies, et al., Navigating sample overlap, winner's curse and weak instrument bias in mendelian randomization studies using the UK biobank, medRxiv (2021).
- [145] Berzuini C, Guo H, Burgess S, Bernardinelli L. A bayesian approach to mendelian randomization with multiple pleiotropic variants. *Biostatistics* 2020;21:86–101. <https://doi.org/10.1093/biostatistics/kxy027>.
- [146] Burgess S, Swanson SA, Labrecque JA. Are mendelian randomization investigations immune from bias due to reverse causation? *Eur J Epidemiol* 2021;36:253–7. <https://doi.org/10.1007/s10654-021-00726-8>.
- [147] Davey Smith G, Ebrahim S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease?\*. *Int J Epidemiol* 2003;32:1–22. <https://doi.org/10.1093/ije/dyg070>.
- [148] Wang J, Zhao Q, Bowden J, Hemani G, Davey Smith G, Small DS, et al. Causal inference for heritable phenotypic risk factors using heterogeneous genetic instruments. *PLoS Genet* 2021;17:e1009575.
- [149] Zeng P, Zhou X. Causal association between birth weight and adult diseases: Evidence from a Mendelian randomization analysis. *Front Genet* 2019;10:618. <https://doi.org/10.3389/fgene.2019.00618>.
- [150] Zeng P, Wang T, Zheng J, Zhou X. Causal association of type 2 diabetes with amyotrophic lateral sclerosis: New evidence from mendelian randomization using GWAS summary statistics. *BMC Med* 2019;17:225. <https://doi.org/10.1186/s12916-019-1448-9>.
- [151] Zeng P, Zhou X. Causal effects of blood lipids on amyotrophic lateral sclerosis: A mendelian randomization study. *Hum Mol Genet* 2019;28:688–97. <https://doi.org/10.1093/hmg/ddy384>.
- [152] Foley CN, Mason AM, Kirk PDW, Burgess S. MR-clust: Clustering of genetic variants in mendelian randomization with similar causal estimates. *Bioinformatics* 2021;37:531–41. <https://doi.org/10.1093/bioinformatics/btaa778>.
- [153] Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomization: The STROBE-MR statement. *JAMA* 2021;326:1614. <https://doi.org/10.1001/jama.2021.18236>.
- [154] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203–9. <https://doi.org/10.1038/s41586-018-0579-z>.