

7

Robust methods using variants from multiple gene regions

In this part of the book, we describe more advanced methods for Mendelian randomization. We begin by considering robust methods, which do not require all genetic variants to be valid instrumental variables (IVs) to give consistent estimates of a causal parameter.

If a particular gene region (or regions) has a specific biological link to the exposure, then we would generally advocate basing the primary analysis on variants from that region. However, particularly for complex risk factors such as body mass index or blood pressure, there is no single gene region that encodes the risk factor, and so a polygenic Mendelian randomization analysis is necessary. If several genetic variants in different gene regions have similar associations with the outcome, then a polygenic analysis may even provide stronger evidence of a causal relationship, as the analysis is not dependent on the validity of the IV assumptions for a single gene region. However, in many cases, not all genetic variants associated with the exposure will tell the same story.

Three important factors when comparing methods are bias, coverage, and efficiency. Bias is the difference between the true value of a parameter and the average value of an estimate when it is calculated multiple times in independent datasets. If a method produces biased estimates, then on average the estimates from the method will be too large or too small compared to the true parameter value. Asymptotic bias means bias in the limiting case as the sample size increases towards infinity. Coverage is the probability that a confidence interval contains the true parameter value. By definition, a 95% confidence interval should contain the true parameter value 95% of the time. However, in practice, this may not be true, due to approximations and distributional assumptions made in constructing the interval. Coverage of a method under the null hypothesis is related to the Type 1 error rate (or false positive rate), in that if the coverage is $(1 - \alpha)$ then the Type 1 error rate is α . Maintaining a nominal Type 1 error rate (for instance, having a Type 1 error rate close to 5% for a test at a nominal 5% significance level) is

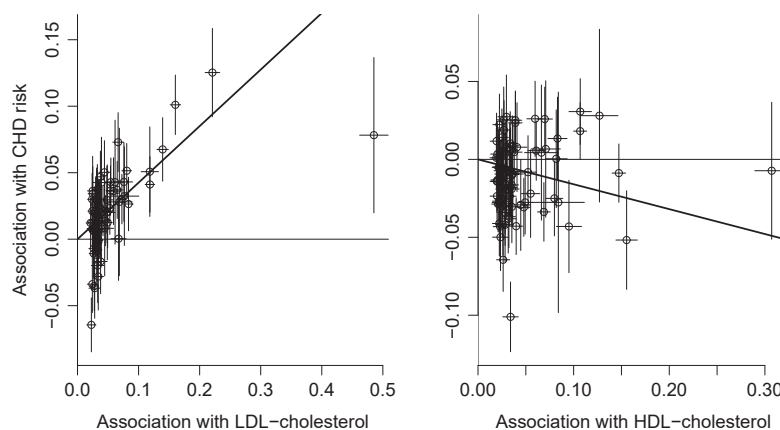
especially important to avoid making false positive causal claims. Efficiency of an estimate relates to the power to detect a true causal effect. A method is more efficient if it produces estimates with smaller standard errors, and thus narrower confidence intervals and greater power.

7.1 Motivating example: LDL- and HDL-cholesterol and coronary heart disease

In Section 6.2.1, we showed that five genetic variants influencing low-density lipoprotein (LDL) cholesterol had associations with coronary heart disease (CHD) risk that were proportional to their associations with LDL-cholesterol. This is consistent with LDL-cholesterol being a causal risk factor for CHD, as has been demonstrated in randomized controlled trials. If rather than selecting variants with a specific biological link to LDL-cholesterol, we consider instead all 75 uncorrelated variants associated with LDL-cholesterol from a previous genome-wide association study (GWAS) [Do et al., 2013], then there is some additional heterogeneity in the genetic associations with the outcome, but a consistent story still emerges (Figure 7.1, left panel): the vast majority of genetic variants that are positively associated with LDL-cholesterol are also positively associated with CHD risk.

For high-density lipoprotein (HDL) cholesterol, the picture is different. We consider 86 uncorrelated genetic variants previously associated with HDL-cholesterol at a genome-wide level of significance, and display their associations with HDL-cholesterol and CHD risk in Figure 7.1 (right panel). We see that there is considerable heterogeneity in the variant-specific causal estimates, with some variants suggesting a positive causal effect of HDL-cholesterol on CHD risk, and others suggesting a negative causal effect. The inverse-variance weighted (IVW) method gives a log odds ratio of -0.160 , corresponding to an odds ratio of 0.85 per 1 standard deviation increase in genetically-predicted HDL-cholesterol, with 95% confidence interval in a random-effects model from 0.76 to 0.95. However, Cochran's Q statistic is 439.7, corresponding to an I^2 statistic of 80.7%, suggesting that not all these genetic variants are valid IVs for HDL-cholesterol. Indeed, while trials of drugs that lower LDL-cholesterol have consistently shown reductions in CHD risk [Cholesterol Treatment Trialists' Collaboration, 2005; Sabatine et al., 2017], CETP inhibitors that raise HDL-cholesterol have generally shown null results in trials [Tardif et al., 2015; Lincoff et al., 2017].

We consider robust methods for Mendelian randomization in three

**FIGURE 7.1**

Genetic associations with LDL-cholesterol (left panel) and HDL-cholesterol (right panel, both standard deviation units) and with CHD risk (log odds ratios). Horizontal and vertical lines represent 95% confidence intervals for the genetic associations. Diagonal lines represent IVW estimates. Adapted from Burgess and Davey Smith [2017].

categories: consensus methods, outlier-robust methods, and modelling methods. Table 7.1 summarizes the methods presented.

As the majority of Mendelian randomization analyses are conducted using summarized data, we focus on methods that can be implemented using summarized data on genetic associations with the exposure ($\hat{\beta}_{X_j}$) and with the outcome ($\hat{\beta}_{Y_j}$) and their standard errors ($se(\hat{\beta}_{X_j})$ and $se(\hat{\beta}_{Y_j})$), or else the variant-specific ratio estimates ($\hat{\theta}_j = \frac{\hat{\beta}_{Y_j}}{\hat{\beta}_{X_j}}$) and their approximate standard errors ($se(\hat{\theta}_j) = \frac{se(\hat{\beta}_{Y_j})}{\hat{\beta}_{X_j}}$), for genetic variants $j = 1, \dots, J$. We assume that all the variants are uncorrelated. While robust methods based on the IVW method can be adapted for correlated variants, it is likely that if one variant in a given gene region is an invalid IV, then other variants in the same gene region will also be invalid IVs. Hence if it is uncertain which variants are valid IVs, it is sensible to prune down to one genetic variant per gene region before proceeding with the analysis, to ensure that no gene region has a strong influence on the analysis. Throughout this chapter, we assume that all associations of the genetic variants with the exposure and

outcome are homogeneous (that is, do not differ between individuals) and linear, and the effect of the exposure on the outcome is homogeneous and linear (Section 5.2.1).

7.2 Consensus methods

The two-stage least squares and inverse-variance weighted estimates can both be expressed as a weighted mean of the ratio estimates for the individual variants (Section 5.2.2). If the ratio estimates are close to zero for all but one of the genetic variants, but non-zero for just one variant, then the weighted mean of the ratio estimates will differ from zero. This implies that if the true causal effect is null, a single pleiotropic genetic variant can lead to rejection of the causal null hypothesis, a false positive finding. The property that a single incorrect datapoint can lead to an arbitrarily large bias in an estimate is referred to as a 0% breakdown point.

A consensus method is one that takes its causal estimate as a summary measure of the distribution of the variant-specific estimates. We consider the median method that has a 50% breakdown point ('majority valid' assumption) and the mode-based estimation method that has a higher breakdown point ('plurality valid' assumption).

7.2.1 Median method

The most straightforward consensus method is the median method [Bowden et al., 2016]. Rather than taking a weighted mean of the ratio estimates as in the IVW method, we take the median of the ratio estimates. Under the linearity and homogeneity assumptions of Section 5.2.1, all genetic variants that are valid instrumental variables estimate the same causal parameter. In a large sample size, this means that the ratio estimates for the valid IVs will all tend towards the same value. Provided that over 50% of the genetic variants are valid IVs, this means that the median of the ratio estimates will tend towards the true causal effect. In a finite sample size, estimates from invalid IVs will still influence the median estimate, but they will have far less influence than for the IVW estimate (Figure 7.2).

In the simple median method, all genetic variants receive equal weight in the analysis. A weighted version of the median method can also be calculated. In the weighted median method, we consider an empirical distribution in which each variant receives a weight corresponding to the inverse of the variance of

| Method | Consistency assumption | Strengths and weaknesses | Reference |
|---------------------------|---------------------------------------|--|---------------------------|
| Inverse-variance weighted | All valid or balanced pleiotropy | Most efficient (greatest statistical power), biased if average pleiotropic effect differs from zero | Burgess et al. [2013] |
| Weighted median | Majority valid | Robust to outliers, sensitive to addition/removal of genetic variants | Bowden et al. [2016] |
| Mode-based estimation | Plurality valid | Robust to outliers, sensitive to bandwidth parameter and addition/removal of genetic variants, often less efficient | Hartwig et al. [2017] |
| MR-PRESSO | Outlier-robust | Removes outliers, efficient with valid IVs, very high false positive rate with several invalid IVs | Verbanck et al. [2018] |
| MR-Robust | Outlier-robust | Downweights outliers, efficient with valid IVs, high false positive rate with several invalid IVs | Rees et al. [2019b] |
| MR-Lasso | Outlier-robust | Removes outliers, efficient with valid IVs, high false positive rate with several invalid IVs | Rees et al. [2019b] |
| MR-Egger | InSIDE | Sensitive to outliers, sensitive to violations of InSIDE assumption, InSIDE assumption often not plausible, often less efficient | Bowden et al. [2015] |
| Contamination mixture | Plurality valid | Robust to outliers, sensitive to variance parameter and addition/removal of genetic variants | Burgess et al. [2020b] |
| MR-Mix | Plurality valid | Robust to outliers, requires large numbers of genetic variants, very high false positive rate in several scenarios | Qi and Chatterjee [2019b] |
| MR-RAPS | Balanced pleiotropy (except outliers) | Downweights outliers, sensitive to violations of balanced pleiotropy assumption | Zhao et al. [2018] |

TABLE 7.1

Comparison of robust methods. This list is not exhaustive, but focuses on methods that can be implemented using summarized genetic associations.

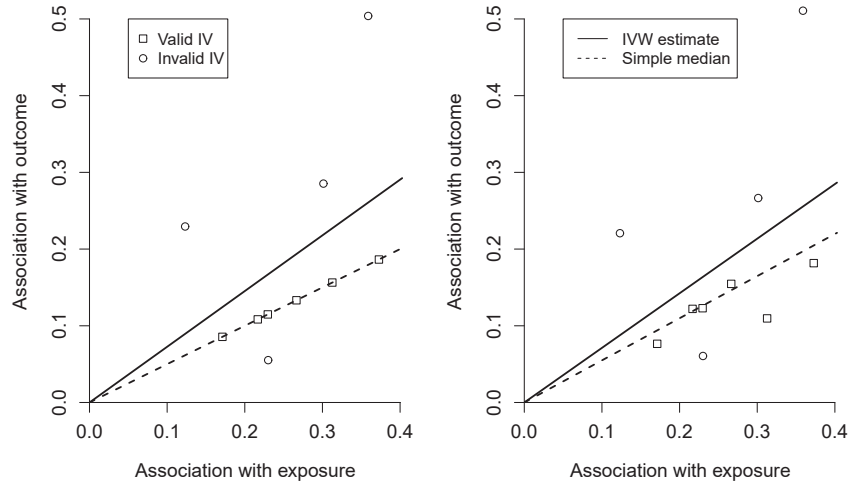


FIGURE 7.2

Illustration of simple median method: synthetic data on genetic associations with exposure and outcome corresponding to infinite sample size (left panel) and finite sample size (right panel) for 6 valid instruments (squares) and 4 invalid instruments (circles). Solid lines represent IVW estimates, dashed lines represent simple median estimates.

the ratio estimate (the same weights as in the IVW method). Estimates are ranked in order, and the weighted median estimate is taken as the estimate at the median (the 50th percentile) of the empirical distribution.

For example, suppose there are five variants, and the variant with the lowest estimate receives 25% of the weight, the next variant 20%, the third variant 15%, the fourth variant 30%, and the variant with the largest estimate 10%. The 50th percentile of the empirical distribution is between the estimates for the second and third variants. As the cumulative weight at the second variant is 45% and the cumulative weight at the third variant is 60%, we extrapolate linearly taking the second estimate plus $\frac{50-45}{60-45} = \frac{5}{15} = \frac{1}{3}$ of the difference between the third and second estimates. This is equivalent to taking two-thirds of the second estimate plus one-third of the third estimate.

Standard errors and confidence intervals for the median method are constructed by parametric bootstrapping of the genetic association estimates, making a normal assumption for the estimate. We refer to the assumption that over 50% of the variants are valid IVs as the ‘majority valid’ assumption.

For the weighted median method, it is required that over 50% of the weight corresponds to valid IVs.

7.2.2 Mode-based estimation method

Suppose that there is an even number of genetic variants and, as the sample size increases, exactly half of them have ratio estimates that tend to one value, and the other half tend to a different value. Under the majority valid assumption, we would not be able to tell which group of variants represents the valid IVs, and which represents the invalid IVs. However, suppose instead that 40% of the variants tend towards one value, 10% towards a second value, 10% towards a third value, 10% towards a fourth value, and so on. We may reasonably state that there is more weight of evidence that the value evidenced by the 40% of variants is the true causal effect. The ‘plurality valid’ assumption states that out of all the different values taken by ratio estimates in large samples (we term these the ratio estimands), the true causal effect is the value taken for the largest number of genetic variants (that is, the modal ratio estimand) [Guo et al., 2018]. This assumption allows the true causal effect to be identified in cases where less than 50% of variants are valid IVs, provided that no larger group of invalid IVs with the same ratio estimand exists. This assumption is also referred to as the Zero Modal Pleiotropy Assumption (ZEMPA) [Hartwig et al., 2017].

This assumption is exploited by the mode-based estimation method [Hartwig et al., 2017]. As no two ratio estimates will be identical in finite samples, it is not possible to take the mode of the ratio estimates directly. In the mode-based estimation method, a normal density is constructed for each genetic variant centered at its ratio estimate. The spread of this density depends on a bandwidth parameter, and (for the weighted version of the mode-based estimation method) the precision of the ratio estimate. A smoothed density function is then obtained by summing these normal densities. The maximum of this distribution is the causal estimate. Again, confidence intervals are constructed by parametric bootstrapping with a normal assumption on the distribution of the causal estimate.

7.2.3 Summary of consensus methods

As these consensus methods take the median or mode of the ratio estimate distribution as the causal estimate, they are naturally robust to outliers, as the median and mode of a distribution are unaffected by the magnitude of extreme values. However, they are still influenced by invalid variants, as these variants contribute to determining the location of the median or mode of

a distribution. These methods can also be sensitive to changes in the ratio estimates for variants that contribute to the median or mode, and to the addition and removal of variants from the analysis. Additionally, the methods may not be as efficient as those that combine the estimates from multiple genetic variants more directly. In particular, the mode-based method has been shown to be less efficient than other methods [Slob and Burgess, 2020].

7.3 Outlier-robust methods

Next, we present three outlier-robust methods. These methods either downweight or remove genetic variants from the analysis that have outlying ratio estimates. They provide consistent estimates under the same assumptions as the IVW method for the set of genetic variants that are not identified as outliers.

7.3.1 MR-PRESSO method

In the MR-Pleiotropy Residual Sum and Outlier (MR-PRESSO) method [Verbanck et al., 2018], the IVW method is implemented by weighted regression using all the genetic variants, and the residual sum of squares (RSS) is calculated from the regression equation. The RSS is a measure of heterogeneity, and is equal to Cochran's Q statistic. Then, the IVW method is performed omitting each genetic variant from the analysis in turn. If the RSS decreases substantially compared to a simulated expected distribution, then that variant is removed from the analysis. This procedure is repeated until no further variants are removed from the analysis. The causal estimate is then obtained by the IVW method using the remaining genetic variants.

7.3.2 MR-Robust method

In MR-Robust, the IVW method is again implemented by weighted linear regression, except that instead of using least squares regression, MM-estimation is used combined with Tukey's biweight loss function [Rees et al., 2019b]. MM-estimation (each 'M' stands for 'maximum likelihood type') provides robustness against influential points and Tukey's loss function provides robustness against outliers. Tukey's loss function is a truncated quadratic function, meaning that there is a limit in the extent to which an outlier contributes to the analysis [Mosteller and Tukey, 1977]. This contrasts

with the quadratic loss function used in least squares regression, which is unbounded, meaning that a single outlier can have an unlimited effect on the IVW estimate.

7.3.3 MR-Lasso method

In MR-Lasso, the IVW regression model is augmented by adding an intercept term for each genetic variant [Rees et al., 2019b]. The IVW estimate is the value of θ that minimizes:

$$\sum_{j=1}^J \text{se}(\hat{\beta}_{Y_j})^{-2} \left(\hat{\beta}_{Y_j} - \theta \hat{\beta}_{X_j} \right)^2. \quad (7.1)$$

In MR-Lasso, we minimize:

$$\sum_{j=1}^J \text{se}(\hat{\beta}_{Y_j})^{-2} \left(\hat{\beta}_{Y_j} - \theta_{0j} - \theta \hat{\beta}_{X_j} \right)^2 + \lambda \sum_{j=1}^J |\theta_{0j}|, \quad (7.2)$$

where λ is a tuning parameter. As the regression equation contains more parameters than there are genetic variants, a lasso penalty term is added for identification [Windmeijer et al., 2018]. The intercept term θ_{0j} represents the direct (pleiotropic) effect of the j th variant on the outcome. It should be zero for a valid IV, but will be non-zero for an invalid IV. The causal estimate is then obtained by the IVW method using the genetic variants that had $\hat{\theta}_{0j} = 0$ in equation (7.2). A heterogeneity criterion is used to determine the value of λ . Increasing λ means that more of the pleiotropy parameters equal zero and so the corresponding variants are included in the analysis; we increase λ step-by-step until one step before there is more heterogeneity in the ratio estimates for variants included in the analysis than expected by chance alone.

7.3.4 Summary of outlier-robust methods

The MR-PRESSO and MR-Lasso methods remove variants from the analysis, whereas MR-Robust downweights variants. These methods will be valuable when there is a small number of genetic variants with heterogeneous ratio estimates, as they will be removed from the analysis or heavily downweighted, and so will not influence the overall estimate. In such a case, these methods are likely to be efficient, as they are based on the IVW method. The methods are likely to be less valuable when there is a larger number of genetic variants that are pleiotropic, or when the average pleiotropic effect of non-outliers is not zero.

One note of caution: if these methods remove a substantial proportion of

the variants from the analysis, they may give a false impression of confidence in the causal estimate due to homogeneity of the ratio estimates amongst the remaining variants. However, it is not reasonable to claim that there is strong evidence for a causal effect after a large number of variants with heterogeneous estimates have been removed from the analysis.

7.4 Modelling methods

Finally, we present four robust methods that attempt to model the distribution of estimates from invalid IVs or make a specific assumption about the way in which the IV assumptions are violated.

7.4.1 Decomposition of genetic associations and the IVW method

Before introducing the remaining robust methods, we consider a parametric decomposition of the genetic associations with the outcome. We write the association of the j th variant as:

$$\beta_{Yj} = \alpha_j + \theta \beta_{Xj} \quad (7.3)$$

where α_j is the pleiotropic (direct) effect of the genetic variant on the outcome, and $\theta \beta_{Xj}$ is the causal (indirect) effect of the genetic variant on the outcome via the exposure, which comprises the genetic association with the exposure (β_{Xj}) multiplied by the causal effect of the exposure on the outcome (θ). Note that these quantities are written without hats, indicating that the decomposition is for the parameters, rather than their estimates.

The ratio estimand (that is, the quantity targeted by the ratio estimate) for the j th variant can therefore be written as:

$$\frac{\beta_{Yj}}{\beta_{Xj}} = \frac{\alpha_j + \theta \beta_{Xj}}{\beta_{Xj}} = \theta + \frac{\alpha_j}{\beta_{Xj}}. \quad (7.4)$$

This is equal to the causal effect θ if and only if $\alpha_j = 0$, which occurs when the pleiotropic effect of the variant is zero. It can be similarly shown that the IVW estimand is equal to θ when a weighted average of the pleiotropic effects of the variants is zero and a weighted correlation between the α_j and the β_{Xj} is zero [Burgess et al., 2016a]. This condition is referred to as ‘balanced pleiotropy’. The IVW method therefore provides a consistent estimate of the causal effect either under the assumption that all variants are valid IVs (‘all valid’) or the assumption of balanced pleiotropy.

The condition that the weighted correlation between the α_j and the β_{X_j} is zero is referred to as the ‘InSIDE’ assumption – Instrument Strength Independent of Direct Effect.

7.4.2 MR-Egger method

The MR-Egger method [Bowden et al., 2015] is performed similarly to the IVW method, except that the regression model contains an intercept term θ_0 :

$$\hat{\beta}_{Y_j} = \theta_0 + \theta \hat{\beta}_{X_j} + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \text{se}(\hat{\beta}_{Y_j})^2). \quad (7.5)$$

The estimate of the slope parameter θ is the MR-Egger estimate. This method differs from the MR-Lasso method, as there is only one intercept term θ_0 , which represents the average pleiotropic effect. The MR-Egger method gives consistent estimates of the causal effect under the InSIDE assumption. In contrast to the IVW method, the average pleiotropic effect does not have to be equal to zero, and (under the InSIDE assumption) will be estimated by the intercept term θ_0 . The average pleiotropic effect not equalling zero is referred to as ‘directional pleiotropy’.

The intercept in MR-Egger also provides a test of the IV assumptions. The intercept will differ from zero when either the average pleiotropic effect is not zero, or the InSIDE assumption is violated. These two conditions (average pleiotropy of zero and InSIDE assumption satisfied) are precisely those required for the IVW estimate to be unbiased (balanced pleiotropy).

An intuitive way of thinking about the MR-Egger method is that it assesses whether there is a dose–response relationship between the genetic associations with the exposure and those with the outcome. Illustrative data are shown in Figure 7.3. For the synthetic data in the left panel, although all five of the genetic variants individually suggest a positive causal effect of the exposure on the outcome, a dose–response relationship in the associations is absent. Genetic variants that have a greater magnitude of association with the exposure do not also have a greater magnitude of association with the outcome. This is contrary to what would be expected if the associations of the genetic variants with the outcome were entirely mediated via the exposure, and hence it is unlikely that all of the genetic variants are valid instrumental variables. While the individual ratio estimates are all positive (as is the IVW estimate), the MR-Egger regression model (dashed line) tells a different story. The intercept term from MR-Egger differs from zero, and the causal estimate from MR-Egger is compatible with the null. This suggests that the set of genetic variants suffers from directional pleiotropy and, once this pleiotropy is accounted for, there is no residual evidence for a causal effect.

A similar conclusion applies to the example of plasma urate and CHD risk

in Figure 7.3 (right panel). Estimates (odds ratio per 1 standard deviation increase in genetically-predicted plasma urate with 95% confidence interval) are 1.11 (1.03, 1.20) for the IVW method, and 1.00 (0.90, 1.10) for the MR-Egger method [Burgess and Thompson, 2017].

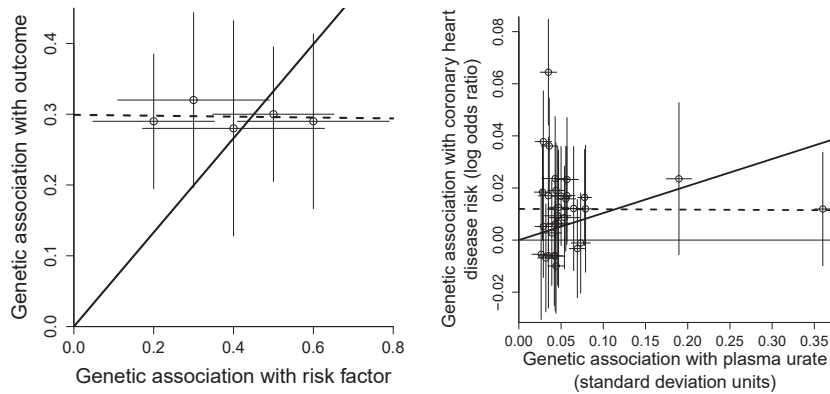


FIGURE 7.3

Graph showing genetic associations for synthetic (left panel) and real data (right panel) examples in which inverse-variance weighted estimate (solid line) and MR-Egger estimate (dashed line) differ substantially. Horizontal and vertical lines are 95% confidence intervals for the genetic associations. In both cases, the inverse-variance weighted estimate is positive, whereas the MR-Egger causal estimate is null with intercept term differing from zero. Taken from Burgess and Thompson [2017].

As the MR-Egger regression model is not invariant to changes in the signs of the genetic association estimates (which would result from switching the reference and effect alleles), we first re-orientate the genetic associations before performing the MR-Egger method by fixing all genetic associations with the exposure to be positive, and correspondingly changing the signs of the genetic associations with the outcome if necessary.

7.4.3 Difficulties with the MR-Egger method

Readers who are familiar with robust methods for Mendelian randomization may be surprised that MR-Egger is not the first robust method presented in this chapter. While the method can be useful, there are several potential issues with its use.

First, while the re-orientation of genetic associations is necessary to ensure

that estimates from the MR-Egger method do not depend on an arbitrary choice of effect alleles, the dependence of the method on a specific orientation of the genetic associations is not ideal. Changing the orientation of a variant affects the definition of the pleiotropic effect α_j . There are therefore different versions of the InSIDE assumption for each of the potential choices of orientation of the variants, and it is only possible for these all to be satisfied if all the pleiotropic effects are zero.

Secondly, the precision of the MR-Egger estimate is not dependent on the proportion of variance in the exposure explained by the genetic variants (as for the IVW method), but on the variance between the genetic associations with the exposure. If these associations are all similar (as in Figure 7.4, left panel), then the MR-Egger estimate will have wide confidence intervals. The precision of the MR-Egger estimate will always be less than that of the IVW estimate, and the difference in precision can be substantial.

Thirdly, the MR-Egger estimate can be strongly influenced by individual variants. In Figure 7.4 (left and right panels), we see how the addition of a single genetic variant can reverse the sign of the MR-Egger estimate, and lead to rejection of the MR-Egger intercept test. The influence on the IVW estimate is less severe.

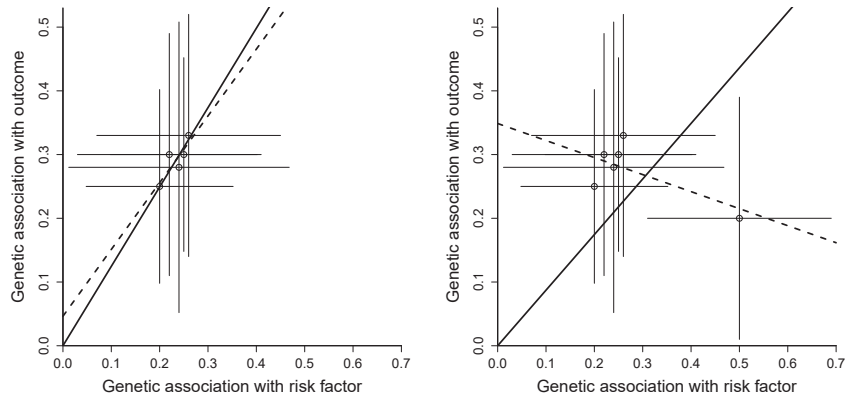


FIGURE 7.4

Synthetic data on genetic associations of five genetic variants (left panel) and with the addition of one extra variant (right panel). Left panel: IVW estimate (solid line) and MR-Egger estimate (dashed line) are similar. Right panel: IVW estimate (solid line) and MR-Egger estimate (dashed line) are markedly different, and the influential genetic variant changes the sign of the MR-Egger estimate. Taken from Burgess and Thompson [2017].

Finally, the InSIDE assumption is generally implausible when variants are pleiotropic [Burgess and Thompson, 2017]. If the pleiotropic effects of genetic variants act via confounders, then the InSIDE assumption will typically be violated even if they influence the outcome via different confounders. If variants influence the outcome via the same confounder, then violation will be more severe still. Even if the pleiotropic effects are not via confounders, for any particular set of genetic variants the correlation between pleiotropic effects and genetic associations with the exposure is likely to differ from zero, meaning that the InSIDE assumption is violated and the MR-Egger estimate is biased.

In favour of MR-Egger, the InSIDE assumption is very different to the majority or plurality valid and outlier-robust assumptions. Robust methods that make the majority or plurality assumption, and so rely on several genetic variants being valid IVs, are likely to give similar answers. In contrast, the MR-Egger method allows all genetic variants to be invalid IVs. However, instead, it relies on the variants satisfying the InSIDE assumption, an untestable and perhaps still implausibly strong assumption.

7.4.4 Contamination mixture method

The contamination mixture method assumes that only some of the genetic variants are valid IVs, and provides consistent estimates under the plurality valid assumption [Burgess et al., 2020b]. To perform the method, a likelihood function is constructed from the ratio estimates. If a variant is a valid instrument, then its ratio estimate $\hat{\theta}_j$ is assumed to be normally distributed about the true causal effect θ with variance $\text{se}(\hat{\theta}_j)^2$. If a variant is not a valid instrument, then its ratio estimate is assumed to be normally distributed about zero with variance $\psi^2 + \text{se}(\hat{\theta}_j)^2$, where ψ^2 represents the variance of the estimands from invalid IVs. This parameter is specified by the analyst. We then maximize the likelihood over different values of the causal effect θ and different configurations of valid and invalid IVs. Maximization is performed by first constructing a profile likelihood as a function of θ , and then maximizing this function with respect to θ . The value of θ that maximizes the profile likelihood is the causal estimate. Confidence intervals are constructed using the likelihood function, and do not rely on bootstrapping or normality of the parameter estimate. The confidence interval is typically not symmetric, and is not even guaranteed to be a single range of values. A confidence interval comprising multiple disjoint ranges occurs when there is uncertainty which of two or more groups of variants supporting different causal effects has more weight of evidence.

7.4.5 MR-Mix method

The MR-Mix method [Qi and Chatterjee, 2019b] is similar to the contamination mixture method, except that rather than dividing the genetic variants into valid and invalid IVs, the method divides variants into four categories: 1) variants that directly influence the exposure only (valid IVs), and 2) variants that influence the exposure and outcome, 3) variants that influence the outcome only, and 4) variants that influence neither the exposure nor outcome. This allows for more flexibility in modelling genetic variants, although potentially leads to more uncertainty in assigning genetic variants to categories.

7.4.6 MR-RAPS method

The MR-Robust Adjusted Profile Score (RAPS) [Zhao et al., 2018] method models the pleiotropic effects of genetic variants directly using a random-effects distribution. The pleiotropic effects are assumed to be normally distributed about zero with unknown variance, independently of the genetic associations with the exposure. Estimates are obtained using a profile likelihood function for the causal effect and the variance of the pleiotropic effects distribution. To provide further robustness to outliers, either Tukey's biweight loss function or Huber's loss function [Mosteller and Tukey, 1977] can be used.

7.4.7 Summary of modelling methods

Modelling methods are likely to be valuable when the modelling assumptions are correct, but not when the assumptions are incorrect. For example, the MR-Egger method requires the InSIDE assumption to be satisfied to give a consistent estimate. The MR-RAPS method is likely to perform well when pleiotropic effects are truly normally distributed about zero, but less well when they are not. The MR-Mix method is likely to require large numbers of genetic variants in order to correctly classify variants into the different categories. The contamination mixture method is less likely to be affected by modelling assumptions as it does not make such strict assumptions, but it can be sensitive to specification of the variance parameter.

7.5 Other methods and comparison

As stated at the beginning of the chapter, we have focused on robust methods that can be implemented using summarized data. Another method that satisfies this criterion fits a similar model to the MR-Lasso method with one pleiotropic parameter per variant, except instead of using lasso penalization to identify the pleiotropic effects of the different variants, it takes a Bayesian approach and imposes a prior distribution on the pleiotropic effects [Berzuini et al., 2020]. Other methods that require individual-level data include: the MR GENIUS method (G-Estimation under No Interaction with Unmeasured Selection), which exploits orthogonality conditions to allow for arbitrary additive pleiotropy functions [Tchetgen Tchetgen et al., 2017]; a constrained optimization approach that includes data on genetic associations with potential confounders, attempting to maximize the association of a linear combination of genetic variants with the exposure while minimizing associations with the confounders [Jiang et al., 2019]; and a confidence interval method that aims to identify groups of variants having similar causal estimates [Windmeijer et al., 2019].

Comparison of different robust methods is difficult. While methods can be compared as to their theoretical properties, it is often unclear how relevant these properties are in practice. The value of comparisons based on simulated data is unclear, as each method typically performs well when its assumptions are satisfied, and poorly when they are not. Comparisons based on real data are difficult to interpret, as there are few examples of exposure–outcome pairs where we are confident to state that the exposure is a causal risk factor for the outcome. Two comparisons of robust methods are Slob and Burgess [2020] and Qi and Chatterjee [2019a]: these investigations prioritized the contamination mixture method and the MR-Mix method as having the best estimation properties. However, these methods were each developed by the authors of the respective comparisons. We discuss recommendations for practice in Section 10.6: in brief, we recommend comparing results from a range of methods that make different assumptions. For example, investigators could perform the weighted median, MR-Egger, and contamination mixture methods. Alternatively, the contamination mixture method could be replaced by the mode-based estimation method or the MR-Mix method.

7.6 Example: LDL- and HDL-cholesterol and coronary heart disease reprised

We return to the examples of LDL- and HDL-cholesterol and risk of CHD using genome-wide significant variants introduced at the beginning of the chapter. We proceed to perform the IVW method and each of the robust methods presented in this chapter in turn. Default options were used for each method, including random-effects models and weights for the mode-based estimation method. We recall that clinical trials suggest that LDL-cholesterol is a causal risk factor for CHD, but suggest that HDL-cholesterol may not be.

Results are given in Table 7.2. For LDL-cholesterol, estimates from all methods are positive and the 95% confidence intervals all exclude the null. For HDL-cholesterol, estimates are more variable. Estimates from the weighted median, mode-based estimation, and MR-Egger methods are attenuated towards the null, and the 95% confidence intervals include the null. The MR-Egger intercept excludes the null, suggesting violation of the InSIDE assumption and/or directional pleiotropy. The confidence interval from the contamination mixture method consists of two disjoint ranges, suggesting that there are two groups of variants supporting different effect estimates (see Burgess et al. [2020b] for a discussion of this). Estimates from the outlier-robust methods are all similar, and the confidence intervals from these methods exclude the null.

While we would discourage comparing methods based on their results for a single dataset, for HDL-cholesterol we see disagreement between the estimates from different methods, indicating that there is less confidence in a causal finding for HDL-cholesterol compared with for LDL-cholesterol.

7.7 Computer implementation

Several robust methods are implemented for R in the *MendelianRandomization* package available from the Comprehensive R Archive Network (CRAN) [Yavorska and Burgess, 2017]. The median method is implemented as:

```
mr_median(mr_input(bx, bxse, by, byse), weighting="simple")
mr_median(mr_input(bx, bxse, by, byse), weighting="weighted")
```

The mode-based method is implemented as:

```
mr_mbe(mr_input(bx, bxse, by, byse))
```

| Method | LDL-cholesterol Estimate (95% CI) | HDL-cholesterol Estimate (95% CI) |
|-------------------------|--|---|
| IVW | 1.530 (1.402, 1.669) | 0.852 (0.764, 0.951) |
| Simple median | 1.525 (1.389, 1.675) | 0.767 (0.679, 0.867) |
| Weighted median | 1.587 (1.454, 1.732) | 0.953 (0.867, 1.048) |
| Mode-based estimation | 1.699 (1.509, 1.912) | 0.993 (0.903, 1.092) |
| MR-PRESSO | 1.530 (1.402, 1.669) | 0.852 (0.764, 0.951) |
| MR-Robust | 1.610 (1.448, 1.791) | 0.864 (0.764, 0.976) |
| MR-Lasso | 1.615 (1.528, 1.707) | 0.861 (0.803, 0.923) |
| MR-Egger (intercept) | 1.605 (1.391, 1.852) −0.003 (−0.011, 0.005) | 1.102 (0.930, 1.306) −0.015 (−0.023, −0.007) |
| Contamination mixture | 1.679 (1.597, 1.783) | 0.671 (0.595, 0.772 and 0.888, 0.953) |
| MR-Mix | 1.716 (1.587, 1.856) | 0.549 (0.409, 0.736) |
| MR-RAPS | 1.570 (1.457, 1.692) | 0.879 (0.797, 0.970) |

TABLE 7.2

Comparison of estimates from different robust methods for the effect of LDL- and HDL-cholesterol on CHD risk. Estimates represent odds ratios (95% confidence intervals) per 1 standard deviation increase in genetically-predicted values of the lipid fraction, except for the MR-Egger intercept, which is the untransformed regression coefficient.

The MR-Robust method is implemented as:

```
mr_ivw(mr_input(bx, bxse, by, byse), robust=TRUE)
```

The MR-Lasso method is implemented as:

```
mr_lasso(mr_input(bx, bxse, by, byse))
```

The MR-Egger method is implemented as:

```
mr_egger(mr_input(bx, bxse, by, byse))
```

The contamination mixture method is implemented as:

```
mr_conmix(mr_input(bx, bxse, by, byse))
```

The MR-PRESSO method is implemented using the *MR-PRESSO* package, available from <https://github.com/rondolab/MR-PRESSO>. The MR-Mix method is implemented using the *MRMix* package, available from <https://github.com/gqi/MRMix>. The MR-RAPS method is implemented using the *mr.raps* package, available from the CRAN.

The median, mode-based, and MR-Egger methods are implemented for Stata in the *mrrobust* package [Spiller et al., 2019].

7.8 Summary

In this chapter, we have presented a wide range of robust methods for Mendelian randomization. Findings from a polygenic Mendelian randomization investigation are more reliable when several methods that make different assumptions give similar answers. However, each of these methods still relies on untestable assumptions, and appropriate caution in their interpretation is needed, particularly when results from methods differ.