

# Inferring the direction of a causal link and estimating its effect via a Bayesian Mendelian randomization approach

Ioan Gabriel Bucur,  Tom Claassen and Tom Heskes 

Statistical Methods in Medical Research  
2020, Vol. 29(4) 1081–1111

© The Author(s) 2019



Article reuse guidelines:

[sagepub.com/journals-permissions](http://sagepub.com/journals-permissions)

DOI: 10.1177/0962280219851817

[journals.sagepub.com/home/smm](http://journals.sagepub.com/home/smm)



## Abstract

The use of genetic variants as instrumental variables – an approach known as Mendelian randomization – is a popular epidemiological method for estimating the causal effect of an exposure (phenotype, biomarker, risk factor) on a disease or health-related outcome from observational data. Instrumental variables must satisfy strong, often untestable assumptions, which means that finding good genetic instruments among a large list of potential candidates is challenging. This difficulty is compounded by the fact that many genetic variants influence more than one phenotype through different causal pathways, a phenomenon called horizontal pleiotropy. This leads to errors not only in estimating the magnitude of the causal effect but also in inferring the direction of the putative causal link. In this paper, we propose a Bayesian approach called BAYESMR that is a generalization of the Mendelian randomization technique in which we allow for pleiotropic effects and, crucially, for the possibility of reverse causation. The output of the method is a posterior distribution over the target causal effect, which provides an immediate and easily interpretable measure of the uncertainty in the estimation. More importantly, we use Bayesian model averaging to determine how much more likely the inferred direction is relative to the reverse direction.

## Keywords

Causal inference, Mendelian randomization, Bayesian model averaging, instrumental variables, sparsity prior, genetic epidemiology, robust estimation

## 1 Introduction

Finding and estimating causal effects from various measured *exposures* (phenotypes, biomarkers, risk factors) to disease and health-related *outcomes* is a crucial task in epidemiology. If we knew there was a causal link from a modifiable exposure (e.g. the level of low-density lipoprotein cholesterol (LDL-C)) to an outcome of interest (e.g. the incidence of cardiovascular disease), we could intervene on the former to produce a desired effect on the latter. Moreover, if we knew the *magnitude* of the causal effect, we could then predict the efficacy of treating the disease by altering the level of the exposure.

*Randomized controlled trials* (RCTs) are the best means of establishing the causal relation between exposure and outcome.<sup>1</sup> In epidemiological research, RCTs have been used, for example, to identify causal risk factors or to determine the efficacy of a new drug. However, it is often not practical or ethical to perform an RCT, in which case we want to be able to extract causal information from observational epidemiological studies. Directly inferring causation from observational data is problematic since observed correlations (associations) do not imply any particular causal relationship on their own. The correlation between two variables  $X$  and  $Y$  can be explained by the existence of a causal link from  $X$  and  $Y$  (*direct causation*) or from  $Y$  to  $X$  (*reverse causation*), but it can also be explained by an unobserved common cause (*confounder*) influencing the two variables simultaneously. It can even be the result of a combination of causal and confounding effects.

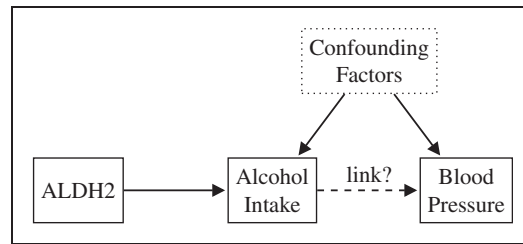
---

Data Science Department, Institute for Computing and Information Sciences, Radboud University, Nijmegen, The Netherlands

### Corresponding author:

Ioan Gabriel Bucur, Faculty of Science, Radboud University, Postbus 9010, 6500 GL Nijmegen, The Netherlands.

Email: [g.bucur@cs.ru.nl](mailto:g.bucur@cs.ru.nl)



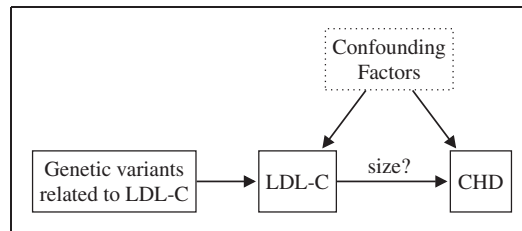
**Figure 1.** Chen et al.<sup>6</sup> used the ALDH2 genetic variant as a proxy for alcohol intake to determine if there is a causal link between the latter and blood pressure. They found that alcohol intake has a marked positive causal effect on blood pressure.

One method for finding and estimating the magnitude of causal effects from observational data that has recently gained traction in the field of epidemiology is *Mendelian randomization* (MR). In MR, germline *genetic variants* are used as proxy variables for environmentally modifiable exposures with the goal of making causal inferences about the outcomes of the modifiable exposure.<sup>2</sup> The basic principle behind MR is that the differences between individuals resulting from genetic variation are not subject to the confounding or reverse causation biases that distort observational findings.<sup>3</sup> The natural “randomization” of *alleles* (variant forms) can thus be likened to the random allocation of treatment performed in RCTs in the sense that groups defined by the value of a relevant genetic variant will experience an on-average difference in exposure, while not differing with respect to confounding factors.<sup>4</sup>

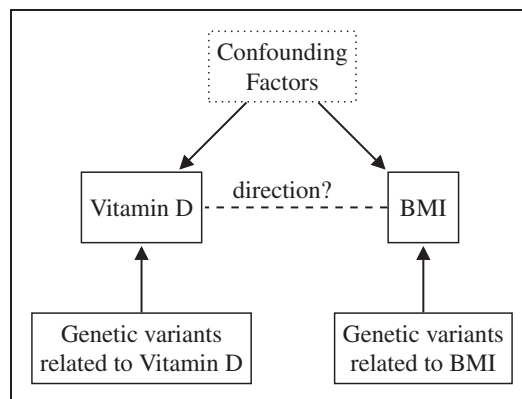
MR has already proved very useful in extracting causal information from observational studies. MR has been used successfully for drug target validation, for predicting side effects of drug targets, for discovering causal risk factors in clinical practice, or for a better understanding of complex molecular traits.<sup>5</sup> In a classic MR study, Chen et al.<sup>6</sup> examined the nature of the association between alcohol intake and blood pressure. In this example, the exposure variable (alcohol intake) is a modifiable behavior, but conducting an RCT would be problematic for both ethical and practical reasons: alcohol consumption is potentially damaging to the subject’s health and measuring the intake is prone to error. Chen et al.<sup>6</sup> employed the genetic variant ALDH2 as a surrogate for measuring alcohol intake (see Figure 1). ALDH2 encodes *aldehyde dehydrogenase*, a major enzyme involved in alcohol metabolism. The (\*2\*2) variant of the ALDH2 gene is associated with an accumulation of *acetaldehyde* in the body after drinking alcohol, and therefore unpleasant symptoms. Because of this, carriers of this particular variant tend to limit their alcohol consumption. Chen et al.<sup>6</sup> found that the individuals possessing the (\*2\*2) variant also have lower blood pressure on average. The authors considered the possibility that the (\*2\*2) *genotype* influences blood pressure via a different causal pathway not mediated by alcohol intake (*horizontal pleiotropy*). However, they found no association between ALDH2 and hypertension in females, for whom drinking levels in any genotype group were similar (very low). If there were pleiotropic effects, an effect on blood pressure would be seen in both sexes. The findings of Chen et al.<sup>6</sup> support the idea that alcohol intake increases blood pressure and the risk of hypertension, to a much greater extent than previously thought. Their results indicate that previous observational epidemiological studies suggesting the cardioprotective effects of moderate alcohol consumption might have been misleading.

In a more recent study, Ference et al.<sup>7</sup> performed an MR meta-analysis for the purpose of estimating the effect of long-term exposure to lower-plasma LDL-C on the risk of *coronary heart disease* (CHD). Unlike in the previous example, the causal link from LDL-C to CHD was already well established at the time of the study, but the magnitude of the (long-term) causal effect had not been reliably quantified (see Figure 2). To estimate the causal effect of LDL-C exposure on the risk of CHD, Ference et al.<sup>7</sup> considered nine *polymorphisms* (genetic variations resulting in the occurrence of several different alleles at a locus) in six different genes. For each polymorphism, they computed a causal effect estimate using MR and then they combined the results to obtain a more precise estimate. Finally, they compared the estimated causal effect of long-term exposure to lower LDL-C with the clinical benefit resulting from the same magnitude of LDL-C reduction during treatment with a statin. Ference et al.<sup>7</sup> found that Mendelian random allocation to lower LDL-C levels was responsible for a 54.5% reduction in CHD risk for each mmol/L lower LDL-C. They concluded that prolonged (lifetime) exposure to lower LDL-C results in a threefold greater reduction in the risk of CHD compared to the clinical benefit of a statin-based treatment producing the same magnitude of LDL-C reduction.

Finally, Vimalaswaran et al.<sup>8</sup> investigated the causality and direction of the association between body mass index (BMI) and 25-hydroxvitamin D (25(OH)D). In this example, it is not clear whether the larger vitamin D



**Figure 2.** Ference et al.<sup>7</sup> used Mendelian randomization for estimating the (positive) causal effect of low-density lipoprotein cholesterol (LDL-C) on the risk of coronary heart disease (CHD). They estimated that for each mmol/L lower LDL-C, the risk of CHD is reduced by 54.5% (95% CI (48.8%, 59.5%)).



**Figure 3.** Vimalleswaran et al.<sup>8</sup> explored the causal direction of the relationship between body mass index (BMI) and 25-hydroxyvitamin D [25(OH)D] via bidirectional Mendelian randomization. They concluded that higher BMI leads to lower vitamin D levels and not the other way around.

storage capacity of obese individuals (vitamin D is stored in fatty tissues) lowers their 25(OH)D levels or whether 25(OH)D levels influence fat accumulation. The latter is an example of *reverse causation*, which occurs when the outcome of interest precedes and leads to changes in the exposure instead of the other way around.<sup>9</sup> Reverse causation is an important source of bias in observational epidemiological studies,<sup>2,10</sup> which can lead to misinterpretation in the observed association with respect to the potential impact of interventions.<sup>11</sup> In their work, Vimalleswaran et al.<sup>8</sup> used a bidirectional MR approach (see Figure 3), which implies performing an MR analysis in both directions, to distinguish between the two causal models. For this approach to work, it must be known on which phenotypic trait each genetic variant has a primary influence.<sup>4</sup> If vitamin D deficiency influences obesity, then the genetic variants primarily associated with lower 25(OH)D levels should also be associated with higher BMI. Conversely, if obesity leads to lower vitamin D levels, then the genetic variants primarily associated with higher BMI should also be related to lower 25(OH)D concentrations. Vimalleswaran et al.<sup>8</sup> concluded that higher BMI (obesity) leads to lower vitamin D levels and not the other way around. Their findings provided evidence for the role of obesity as a causal risk factor for vitamin D deficiency and suggested that interventions meant to reduce BMI are expected to decrease the prevalence of vitamin D deficiency, as later shown in intervention studies.<sup>12,13</sup>

In this paper, we introduce a Bayesian framework (Bayesian approach to Mendelian randomization (BAYESMR)) that extends the MR approach to situations where the direction of the causal effect between the two phenotypes of interest is unknown or uncertain. In our approach, we do not need to know in advance which of the two phenotypes is primarily influenced by the genetic variants. In the experimental section, we look at both directions of the association between smoking and coffee consumption using the same genetic variants in an attempt to determine which direction is more likely. This flexibility enables the researcher to apply the instrumental variable approach to a much broader set of genetic variants. The rest of this paper is organized as follows. In Section 2, we discuss the assumptions behind MR and recent extensions to the approach. Then, we introduce our proposed Bayesian model in Sections 3 and 4. We perform a series of simulations in Section 5 to

illustrate the strengths of our approach. In Section 6, we apply our BAYESMR method to a couple of real-world problems where the standard MR approaches might produce biased results. We conclude by discussing the advantages and disadvantages of our proposed method as well as potential applications in Section 7.

## 2 Instrumental variables and MR

### 2.1 Instrumental variable assumptions

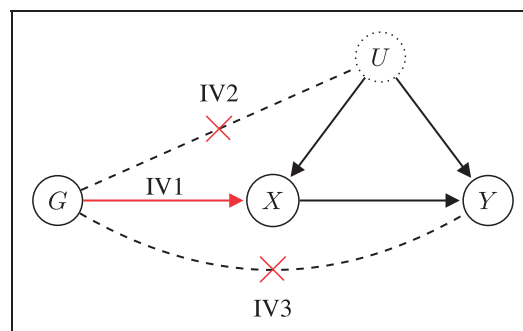
MR is an approach in which genetic variants are used as instrumental variables for estimating the effect of a modifiable phenotype or exposure on an outcome variable of clinical interest. A (valid) *instrumental variable* (IV) or *instrument*<sup>14</sup> is a factor which:

- (IV1) is associated with the exposure;
- (IV2) is independent of all confounders of the exposure–outcome association;
- (IV3) influences the outcome only through its effect on the exposure.

The three core IV assumptions are depicted by the *directed acyclic graph* in Figure 4, where  $G$  is the genetic variant used as an instrumental variable,  $X$  is the variable measuring the exposure,  $Y$  is the variable measuring the outcome of interest, and  $U$  is an unmeasured variable that encapsulates the potential confounding effects affecting the exposure–outcome association. Note that for the graphical model to be acyclic, there must be no feedback loops in the system. Acyclicity is a standard assumption in MR and implies that the causal link of interest is unidirectional.<sup>15</sup> Assumption IV1 is represented in Figure 4 by the edge between  $G$  and  $X$ . Note that the association between the genetic variant and the exposure need not be causal, although this is often assumed in the context of MR. Assumption IV2 is denoted by a lack of paths from  $G$  to  $U$ . Finally, assumption IV3 is denoted by an absence of paths from  $G$  to  $Y$ , except for the one that goes through  $X$ . Genetic variants are particularly suitable as candidate instrumental variables because they are fixed at conception and “randomized” in a population according to Mendel’s laws of inheritance, meaning that their associations with non-genetic measures are generally not susceptible to confounding or reverse causation.<sup>5,10</sup> In graphical terms, this is equivalent to removing all arrows from non-genetic variables ( $X, Y, U$  in Figure 4) to genetic variables ( $G$  in Figure 4).

### 2.2 The MR pipeline

The suitability of using genetic variants as instrumental variables combined with the wide availability of data on genetic associations extracted by means of high-throughput genomic technologies have contributed to the increasing popularity of MR applications.<sup>4,16</sup> The first step in MR is to find genetic variants that are robustly associated with the exposure under study.<sup>17</sup> Ideally, MR is performed using a single variant whose biological effect on the exposure is well understood. In practice, however, such a single variant is not always available or known. Moreover, genetic variants typically only explain a small proportion of the variance in the exposure.<sup>18</sup> When a single reliable genetic instrument is not available, it has become common to use multiple genetic variants in MR



**Figure 4.** Directed acyclic graph showing the instrumental variable assumptions. Causal effects are unidirectional and denoted by arrows. The association entailed by IV1 is highlighted in red, while the crossed-out dashed lines signify the absence of an association (causal or non-causal). The dotted border of  $U$  means that the variable is unobserved.

studies. *Genome-wide association studies* (GWAS) provide a rich source of potential instruments for MR analysis.<sup>19</sup> Increasing the number of exposure-related genetic variants will increase the proportion of variance explained by the instruments, which will lead to improved precision in the causal estimate and to increased power in testing causal hypotheses.<sup>17</sup> What's more, the use of multiple genetic instruments provides opportunities to test for violations of the IV assumptions.<sup>20</sup>

Genetic variants are typically selected based on the size of the genetic effect on the exposure of interest and the specificity of the association.<sup>19</sup> After deciding on a set of relevant exposure-related genetic instruments, the next step is to examine if they are also associated with the disease outcome. If no genotype-outcome association is found, then it becomes likely that there is no causal link from exposure to outcome. Otherwise, by making additional assumptions, it is possible to compute a consistent estimate of the causal relationship's strength.<sup>15</sup> A commonly made assumption is that the causal relation between the exposure  $X$  and the outcome  $Y$  is linear

$$Y = \beta X + \epsilon \quad (1)$$

The strength or magnitude of this relationship is then given by the parameter  $\beta$ , which is the *causal effect* of the exposure on the outcome. In an MR analysis, the estimate of  $\beta$  is derived from the associations with the genetic instruments. For example, when using a single genetic variant as an instrumental variable, a consistent estimate of the causal effect under the linear model assumption (1) is Wald's ratio<sup>2</sup> (the Wald method<sup>17</sup>)

$$\hat{\beta}^{IV} = \frac{r_{Y|G}}{r_{X|G}} \quad (2)$$

where  $r_{Y|G}$  is the coefficient of the regression of the outcome on the genetic variant and  $r_{X|G}$  is the coefficient of the regression of the exposure on the genetic variant. When using multiple genetic instruments, the individual IV estimates can be combined similarly to performing a meta-analysis

$$\hat{\beta}^{IVW} = \frac{\sum_{j=1}^J \hat{\beta}_j^{IV} s_j^{-2}}{\sum_{j=1}^J s_j^{-2}} \quad (3)$$

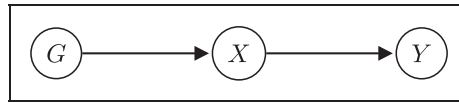
where  $s_j$  is the standard error of  $\hat{\beta}_j^{IV}$  and can be approximated using the delta method.<sup>21</sup>  $\hat{\beta}^{IVW}$  is known as the *inverse-variance weighted* (IVW) estimator and is equivalent to a zero-intercept weighted linear regression of the genetic associations with the outcome on the genetic associations with the exposure.<sup>17</sup>

Unfortunately, when selecting multiple genetic variants for MR based on their association with the exposure, we cannot be certain that all of them are valid instrumental variables. In the worst possible case, none of the selected variants are valid instruments for the exposure–outcome associations. As a result, there is great interest in developing methods robust to the validity of the IV assumptions.<sup>4,5</sup>

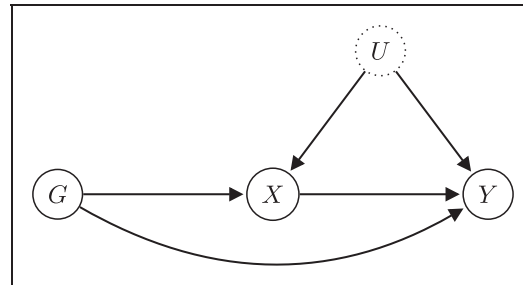
### 2.3 Challenges of MR

The application of MR methods requires genetic variants that are valid instruments for the exposure–outcome association, i.e. genetic variants satisfying the core IV assumptions. If a causal link from exposure to outcome exists, then a genetic variant associated with the exposure will also be associated with the outcome. Under the core IV assumptions, the converse also holds: an association between the genetic variant and the outcome implies the existence of a causal link from the exposure to the outcome. When these assumptions are violated, however, we can no longer be sure that a causal link from exposure to the outcome exists. A discovered genetic association with the disease outcome could be produced via a different causal pathway which is not mediated by the exposure (see Figure 6). In this case, the bias in the causal effect estimate obtained through the MR approach is potentially unbounded.<sup>22</sup> Even worse, if we are dealing with reverse causation for the exposure–outcome association, then the causal link whose effect we are trying to estimate does not exist (see Figure 7).

Unfortunately, only assumption IV1 is directly testable from data. It is typically fulfilled by selecting only those genetic variants that are associated with an exposure of interest. In general, the validity of assumptions IV2 and IV3 cannot be established from the data, since they involve the unobserved confounder  $U$ .<sup>15</sup> However, if the exposure–outcome association is unconfounded (see Figure 5), then we can check the validity of IV2 and IV3 by performing a conditional independence test. In this particular scenario, the genetic variant becomes independent of the disease outcome after conditioning on the exposure variable. This principle of testing the IV assumptions by looking for a conditional independence is used to orient the causal relationship between the two traits in



**Figure 5.** The local causal discovery<sup>23</sup> (LCD) pattern can be discovered from observational data by testing if the genetic variant  $G$  is independent from the outcome  $Y$  when conditioning on the exposure  $X$ .



**Figure 6.** The genetic variant  $G$  is pleiotropic since it affects both phenotypic measures ( $X$  and  $Y$ ). In this illustration,  $G$  exhibits both *vertical pleiotropy* ( $X$  and  $Y$  are influenced by  $G$  via the same causal pathway  $G \rightarrow X \rightarrow Y$ ) and *horizontal pleiotropy* ( $X$  and  $Y$  are influenced by  $G$  via different causal pathways). The former is crucial to the application of MR, while the latter results here in a violation of the IV3 assumption.

mediation-based approaches<sup>24</sup> such as the *likelihood-based causality model selection*,<sup>25</sup> the regression-based *causal inference test*,<sup>26</sup> or the *FINDR*<sup>27</sup> algorithm. In the artificial intelligence literature, this particular model is called the *local causal discovery*<sup>23</sup> (LCD) pattern. The testability of the LCD pattern has been exploited, for example, in the *TRIGGER*<sup>28</sup> algorithm for the purpose of learning the structure of gene regulatory networks.

Assumption IV2 is not completely testable since there can always be unmeasured confounding variables for which we cannot account, but we can at least filter out the genetic variants associated with any potential confounding traits that have been measured.<sup>29</sup> Fortunately, it is not very likely that genetic variants are associated with confounders of the exposure–outcome association.<sup>30</sup> The assumption that is most likely to be violated is IV3, which states that the instrumental variable is only related to the outcome via its effect on the risk factor.<sup>31</sup> Assumption IV3 does not hold if there are alternative causal pathways between the genetic variant and the disease outcome that are not mediated by the exposure (see Figure 6). This often happens when a genetic variant influences more than one phenotypic traits, a property called *pleiotropy*.<sup>32</sup>

As the GWAS sample sizes continue to grow, the number of candidate genetic variants to be used as potential instruments is also increasing. Unfortunately, more and more of the supplied candidates are likely to be invalid instruments. Hemani et al.<sup>33</sup> show that horizontal pleiotropy and reverse causation can easily lead to invalid instrument selection in MR and suggest that the presence of horizontal pleiotropy should be considered the rule rather than the exception. This severely limits the application of MR to cases where previous background knowledge can be used to validate the core assumptions. By relaxing (some of) these assumptions, however, the MR approach can be applied to a much broader range of genetic variants.

### 2.3.1 Pleiotropy

In the genetics of complex diseases it is common for genetic variants to affect multiple disease-related phenotypes.<sup>34,35</sup> If such a *pleiotropic genetic variant* affects the disease outcome through different causal pathways, then IV2 or IV3 or both are potentially violated. This makes pleiotropy problematic for MR studies.<sup>4,15</sup> As a result, many recent extensions of the MR approach have been aimed at producing better causal estimates in the presence of pleiotropic effects.<sup>17,32,36–40</sup> Pleiotropy can be divided into horizontal (biological) pleiotropy and vertical (mediated) pleiotropy. *Horizontal pleiotropy* occurs when the same genetic variant affects multiple trait via different biological pathways, whereas *vertical pleiotropy* occurs when the same genetic variant is associated with multiple traits found on the same biological pathway. We will only focus on the former, since only horizontal pleiotropy can lead to violations of IV3, and refer to it hereafter as “pleiotropy”, for brevity.<sup>32</sup>

Two main approaches have been used for robustifying MR against potentially pleiotropic genetic variants. The first approach is to replace some of the (untestable) IV assumptions, e.g. IV2 or IV3 or both, with a weaker set

of alternative assumptions. The second approach is to assume that the IV assumptions hold for some, but not necessarily for all the genetic variants. In this case, the invalid genetic instruments are allowed to violate the IV assumptions in an arbitrary way.<sup>41</sup>

*MR-Egger regression*<sup>17</sup> is a pleiotropy-robust method that falls into the first category. MR-Egger is an extension to the IVW method introduced in Subsection 2.2, in which an intercept term is added to the weighted linear regression of the genetic associations with the outcome on the genetic associations with the exposure. While IVW is consistent if the pleiotropic effects average to zero across the genetic variants, this extension allows for consistency even when the pleiotropic effects do not average to zero. Both the IVW and MR-Egger methods work under the *instrument strength independent of direct effect* (InSIDE) assumption, which stipulates that the association with the exposure is independent from the direct (pleiotropic) effect on the outcome across different genetic variants.<sup>41</sup> Another robust alternative to the IVW estimator is the *weighted median estimator*<sup>36</sup> (WME), for which we compute a weighted median of the IV estimates instead of a weighted average. The WME is a consistent estimate of the true causal effect as long as at least 50% of the genetic instruments are valid. In a similar vein, Kang et al.<sup>42</sup> have introduced an LASSO<sup>43</sup>-type procedure to identify the valid instruments from within a set of candidate genetic variants. Hartwig et al.<sup>37</sup> have proposed instead to compute the weighted mode of the individual IV estimates to obtain the *mode-based estimate* (MBE). The MBE can be a consistent estimate of the true causal effect even when fewer than 50% of the genetic instruments are valid. However, the application of the method relies on the assumption that the largest number of similar individual IV estimates comes from valid instruments, an assumption termed the *Zero Modal Pleiotropy Assumption* (ZEMPA). Finally, Hemani et al.<sup>33</sup> suggest combining these methods (along with others) in a “mixture of experts” machine learning approach.<sup>44</sup>

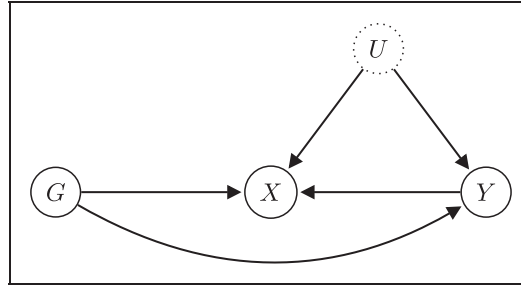
A number of Bayesian modeling approaches for dealing with pleiotropic genetic variants have been proposed, in which all variants used as instruments are allowed to exhibit pleiotropic effects on the outcome. Li<sup>38</sup> introduced empirical Bayes hierarchical models for estimating the causal effect in MR studies where many instruments are invalid, while Berzuini et al.<sup>40</sup> performed a fully Bayesian treatment of the MR problem. In both cases, the authors handled pleiotropy by putting shrinkage priors on the pleiotropic effects, thereby forcing them to take values close to zero unless there is strong posterior evidence for pleiotropy.<sup>45</sup> Thompson et al.,<sup>46</sup> on the other hand, modeled pleiotropy explicitly and used Bayesian model averaging to provide estimates that allow for uncertainty regarding the nature of pleiotropy.

### 2.3.2 Reverse causation

A crucial aspect that is often overlooked in MR studies is the implicit assumption made regarding the direction of the causal link between the exposure and the outcome. While it is sometimes possible to rule out reverse causation, in most situations there is not enough background knowledge available about the underlying causal mechanism. The observed relationship between low levels of LDL cholesterol and risk of cancer is an example where we cannot a priori assume a causal direction. It may be that low levels of LDL cholesterol are causal for cancer, but it is also possible that the presence of the disease has a negative effect on LDL cholesterol.<sup>5</sup> In the case of complex systems such as gene regulatory networks, it is also unclear in which direction the gene regulation occurs.<sup>47</sup>

The existence of alternative causal paths from a genetic variant to the outcome of interest affects not only the estimation of causal effects but also model identification. If there exists a causal pathway from the genetic variant to the outcome that is not mediated by the exposure, then an exposure–outcome causal link is no longer required to explain their association (see Figure 6). In fact, the causal link between the two traits may have the opposite direction (see Figure 7). Ignoring the possibility of reverse causation for the exposure–outcome association is potentially dangerous. If the true causal link is in the expected direction (from  $X$  to  $Y$ , Figure 6), we will be underestimating the uncertainty of our estimate by not considering the alternative model. If the true causal link is in the opposite direction (from  $Y$  to  $X$ , Figure 7), then the MR approach will produce a misleading result, as we are trying to estimate the wrong causal effect. Bidirectional MR<sup>4</sup> is one approach used to distinguish between an “exposure” having a causal effect on an “outcome” and the reverse. The idea is to perform MR in both directions to tease apart the causal relationships  $X \rightarrow Y$  and  $Y \rightarrow X$ .<sup>5</sup> To make use of this approach, it is essential to know if a genetic variant primarily influences the “exposure” or the “outcome”. However, this can be difficult to assess when utilizing variants with little or no understanding of their biological effects.<sup>4</sup>

Agakov et al.<sup>49</sup> proposed a Bayesian framework built on sparse linear methods developed for regression, which is similar in spirit to BAYESMR. In their approach, they first produce a sparse representation of the data by putting a Laplace prior on the linear coefficients and performing MAP inference. They then compute the evidences (marginal likelihoods) using the Laplace approximation for the direct, reverse, and pleiotropic models, with and without latent confounders. This *sparse instrumental variables* approach has been used to infer the direction of causality between



**Figure 7.** Reverse causation refers to the situation when the outcome  $Y$  precedes and causes the exposure  $X$  instead of the other way around.<sup>48</sup> The term “reverse” refers to the fact that the direction of the causal effect is opposite to what was expected based on the study design. We emphasize that although we denote the exposure by  $X$  and the outcome by  $Y$ , these terms do not have any causal meaning, i.e. they do not imply a particular causal ordering.

vitamin D levels (plasma 25-hydroxyvitamin D) and colorectal cancer.<sup>50</sup> More recently, Hemani et al.<sup>24</sup> have suggested using the *Steiger Z-test*<sup>51</sup> of correlated correlations for elucidating the direction of the causal link. Their *MR Steiger*<sup>24</sup> approach is based on the simple principle that in most circumstances the genetic variant will exhibit stronger correlation with the trait located upstream in the causal chain.

## 2.4 Summary

MR is a powerful principle that can help strengthen causal inference by combining observational data with genetic knowledge. However, MR methods require good candidate genetic variants to be used as instruments, which often limits their application to variants with well understood biology. The wide spread of pleiotropic genetic variants, which violate key IV assumptions (IV2 and especially IV3), is a particularly troublesome issue.<sup>4,52</sup> Many methods have been designed to improve causal effect estimation in the presence of pleiotropic effects,<sup>38,40,46</sup> but they do not incorporate the possibility of reverse causation. At the same time, a number of methods have been developed for inferring the correct direction of the causal link between two phenotypes,<sup>24,33</sup> but these do not provide a measure of uncertainty in the model selection. In the following section, we introduce a general method with which we aim to handle all of these issues simultaneously. We first perform Bayesian model averaging to account for the uncertainty in the direction of the causal link and then, for each model, we obtain an informative posterior distribution for the causal effect of interest.

## 3 Bayesian model specification

### 3.1 Setup

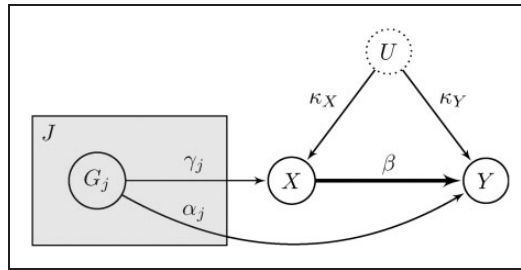
We assume that the data are generated from the model illustrated in the causal diagram from Figure 8, where all relationships between variables are linear with no effect modification. We consider  $J$  independent genetic variables, which we denote by  $G_j$ ,  $j \in \{1, 2, \dots, J\}$ . Their genetic variation is given by the *allele* (variant form) found at a specific locus. The two allele copies inherited at a specific locus form the *genotype* of an individual. Here, we only consider binary *polymorphisms*, i.e. each allele has two possible forms.<sup>2</sup> For convenience, we will refer to these forms as *reference alleles* and *effect alleles*, without assigning any biological meaning to the terms. Each variable  $G_j$  can take the values 0, 1, or 2 corresponding to the number of effect alleles, as opposed to reference alleles, at the locus. Their corresponding *effect allele frequencies* (EAF) are denoted by  $p_j$ ,  $j \in \{1, 2, \dots, J\}$ .

The generative model from Figure 8 is described by the following linear structural equation system

$$\begin{aligned} X &:= \mu_X + \boldsymbol{\gamma}^T \mathbf{G} + \kappa_X U + \epsilon_X \\ Y &:= \mu_Y + \boldsymbol{\alpha}^T \mathbf{G} + \beta X + \kappa_Y U + \epsilon_Y \end{aligned} \quad (4)$$

where  $\mathbf{G}$  is the random vector containing the  $J$  independent variables  $G_j$ ,  $\boldsymbol{\gamma}$  is the corresponding vector of instrument strengths, while the vector  $\boldsymbol{\alpha}$  collects the potential pleiotropic effects of the genetic variants.  $\beta$  represents the putative causal effect of the exposure  $X$  on the outcome  $Y$ , which is what we wish to estimate.  $\kappa_X$  and  $\kappa_Y$  represent the effect of the confounder  $U$  on  $X$  and  $Y$ , respectively. Finally,  $\mu_X$  and  $\mu_Y$  are the average values of the exposure and outcome variables, while  $\epsilon_X$  and  $\epsilon_Y$  represent independent measurement error.





**Figure 8.** Graphical description of our assumed generative model. We denote the exposure variable by  $X$  and the outcome variable by  $Y$ . We are interested in the causal effect from  $X$  to  $Y$ , which is denoted by  $\beta$ . The association between  $X$  and  $Y$  is obfuscated by the unobserved variable  $U$ , which we use to model unmeasured confounding explicitly. The shaded plate indicates replication across the  $J$  independent genetic variants  $G_j, j \in \{1, 2, \dots, J\}$ . Note that the replication also applies to the parameters  $\gamma_j$  and  $\alpha_j$  and their corresponding edges.

In this work, we restrict our attention to the case where the exposure and outcome variables are continuous. We also assume without loss of generality that  $\mu_X = \mu_Y = 0$ . We denote the variance of the intrinsic error terms by  $\text{Var}[\epsilon_X] = \sigma_X^2$  and  $\text{Var}[\epsilon_Y] = \sigma_Y^2$ . Without loss of generality, we can fix the mean of the confounding variable and of the error terms to zero, i.e.  $\mathbb{E}[U] = \mathbb{E}[\epsilon_X] = \mathbb{E}[\epsilon_Y] = 0$ . Also without loss of generality, we can fix the variance of the confounding variable to one ( $\text{Var}[U] = 1$ ) by appropriately rescaling the confounding coefficients  $\kappa_X$  and  $\kappa_Y$ .

### 3.2 Model assumptions

The first two core IV assumptions are implied by our proposed model. The existence of direct edges from each  $G_j$  to the exposure  $X$  implies that IV1 is satisfied for each genetic variant. At the same time, we assume that for any  $j \in \{1, 2, \dots, J\}$  we have  $G_j \perp U$ , where we use the notation  $X \perp Y | S$ <sup>52</sup> to denote that  $X$  is conditionally independent of  $Y$  given  $S$ . These independence statements signify that the genetic variants are not correlated with any confounding variables (IV2). On the other hand, we allow for unrestricted violations of assumption IV3, as long as they do not lead to violation of IV2. We summarize the contribution of the pleiotropic effects in our system with a single directed edge between each genetic variant  $G_j$  and the outcome variable  $Y$  (see Figure 8). The strength of the pleiotropic effect is denoted by  $\alpha_j$ . A genetic variant  $G_j$  is a valid instrumental variable only if its direct effect on the outcome is exactly zero, i.e.  $\alpha_j = 0$ .

In our model, we assume that the instrument strengths  $\gamma_j$  and the pleiotropic effects  $\alpha_j$  are independent a priori. Berzuini et al. have termed this *instrument effects orthogonality*,<sup>40</sup> which is the Bayesian interpretation of the InSIDE assumption.<sup>17</sup> This relaxation of the stronger IV3 assumption is frequently used in pleiotropy-robust MR methods<sup>5,54</sup> and is in line with the principle of independent (causal) mechanisms.<sup>55</sup> Here, we distinguish between the causal mechanism describing the direct association between the genetic variants and  $X$ , which is parametrized by  $\gamma_j$ , and the causal mechanism describing the direct (not through  $X$ ) association between the genetic variants and  $Y$ , which is parametrized by  $\alpha_j$ .

We also assume that the genetic variants are independent, i.e.  $G_i \perp G_j, \forall i, j \in \{1, 2, \dots, J\}, i \neq j$ , although our model could be easily extended to handle correlated genetic variants by introducing additional parameters. This means that we do not allow for the possibility of *linkage disequilibrium*, where one genetic variant can potentially have a causal effect on another related genetic variant. We do not assume, however, that the genetic variants primarily influence one phenotype or the other, as required for instance in the application of standard bidirectional MR. This enables us to use the same set of instruments and the same prior distributions for both the forward and reverse analyses, as the variables  $X$  and  $Y$  become essentially interchangeable.

### 3.3 Likelihood

We assume that the data generating process is described by equation system (4) and that each data sample is a complete set of observed values for the exposure  $X$ , the outcome  $Y$  and the genetic variants  $\mathbf{G} \equiv [G_1, G_2, \dots, G_J]^T$ . The input data set  $\mathbf{D}$  consists of  $N$  i.i.d. samples of  $(\mathbf{G}, X, Y)$ . Our model parameters consist of:

- $(\gamma, \beta, \alpha)$  – the structural coefficients denoting the direct causal effects from  $\mathbf{G}$  to  $X$  (instrument strengths), from  $X$  to  $Y$  (the target causal effect), and from  $\mathbf{G}$  to  $Y$  (pleiotropic effects), respectively;

- $(\sigma_X, \sigma_Y)$  – the scale parameters of the intrinsic error terms  $\epsilon_X$  and  $\epsilon_Y$ ;
- $(\kappa_X, \kappa_Y)$  – the coefficients corresponding to the confounding effect of  $U$ .

We collect all parameters of the model in the vector  $\Theta = (\gamma, \beta, \alpha, \sigma_X, \sigma_Y, \kappa_X, \kappa_Y)$ .

We model each genetic variable  $G_j$  as binomially distributed, where we count the number of effect alleles at the genetic locus, either 0, 1, or 2. If  $p_j$  is the EAF of  $G_j$ , then the binomial distribution has parameters  $n = 2$  and  $p = p_j$ , so  $\text{Var}[G_j] = 2 \cdot p_j \cdot (1 - p_j)$ . In our model, the intrinsic error terms of  $X$  and  $Y$  are independent and follow a Gaussian distribution:  $\epsilon_X \sim \mathcal{N}(0, \sigma_X^2)$ ,  $\epsilon_Y \sim \mathcal{N}(0, \sigma_Y^2)$ . The confounding variable  $U$  follows a standard normal distribution:  $U \sim \mathcal{N}(0, 1)$ . Note that we can set the variance of  $U$  to one without any loss of generality by appropriately rescaling the confounding coefficients  $\kappa_X$  and  $\kappa_Y$ . We obtain a linear-Gaussian model conditional on the value of the genetic variants, in the same vein as Jones et al.,<sup>56</sup> Li,<sup>38</sup> or Berzuini et al.<sup>40</sup>

$$\begin{bmatrix} X \\ Y \end{bmatrix} | \mathbf{G} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma), \quad \text{where} \quad \boldsymbol{\mu} = \begin{bmatrix} \gamma^T \mathbf{G} \\ (\boldsymbol{\alpha} + \gamma\beta)^T \mathbf{G} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_X^2 + \kappa_X^2 & \kappa_X \kappa_Y + \beta(\sigma_X^2 + \kappa_X^2) \\ \kappa_X \kappa_Y + \beta(\sigma_X^2 + \kappa_X^2) & \sigma_Y^2 + \beta^2 \sigma_X^2 + (\kappa_Y + \beta \kappa_X)^2 \end{bmatrix} \quad (5)$$

Given the conditional linear-Gaussian model, the log-likelihood is equal, up to constant terms, to

$$\log \mathcal{L}_{\mathbf{D}} \propto -\frac{N}{2} (\log \det \Sigma + \text{tr}\{\Sigma^{-1} \mathbf{S}\}) \quad (6)$$

where  $\mathbf{S} = \frac{1}{N} \sum_{i=1}^N \left( \begin{bmatrix} X_i - \gamma^T \mathbf{G}_i \\ Y_i - (\boldsymbol{\alpha} + \gamma\beta)^T \mathbf{G}_i \end{bmatrix} \begin{bmatrix} X_i - \gamma^T \mathbf{G}_i & Y_i - (\boldsymbol{\alpha} + \gamma\beta)^T \mathbf{G}_i \end{bmatrix} \right)$ . We use  $\mathbf{G}_i, X_i, Y_i$  to denote the  $i$ th observed value of the genetic vector, exposure variable, and outcome variable, respectively.

Both Berzuini et al.<sup>40</sup> and Li<sup>38</sup> consider the same linear-Gaussian model, but use a slightly different parametrization:  $\Theta' = (\gamma, \beta, \alpha, \kappa_X^2 + \sigma_X^2, \kappa_Y^2 + \sigma_Y^2, \kappa_X \kappa_Y)$ . The difference between our approach and theirs is that we model the confounding explicitly by defining the linear confounding effects  $\kappa_X$  and  $\kappa_Y$ , while they model the confounding implicitly by defining a nonzero covariance parameter for  $(X, Y - \beta X) | \mathbf{G}$ . In other words, using the terminology from Jones et al.,<sup>56</sup> we employ a “full model” parametrization, while Li<sup>38</sup> and Berzuini et al.<sup>40</sup> consider a parametrization equivalent to the “correlated errors model”. In both the full and correlated errors model, the likelihood of the observables  $(X, Y)$  given  $\mathbf{G}$  is bivariate normal.<sup>56</sup> Thompson et al.<sup>46</sup> also assume a linear-Gaussian model, but approximate the discrete distribution over the set of genetic variants with a continuous normal distribution.

### 3.4 Priors

According to Jones et al.,<sup>56</sup> informative priors are crucial to the success of Bayesian MR methods, regardless of whether or not they are identifiable, and the so-called “vague” (uninformative) priors should be avoided whenever possible. Thompson et al.<sup>46</sup> mention two arguments against the use of an uninformative prior in the MR context: (1) that the analysis is already subjective due to the selection of the genetic variants and (2) the data alone are sufficient to produce useful estimates of the causal effect. However, care must be taken when specifying informative priors, as substantial variation in parameter estimates can occur as the prescribed priors are modified.

For most genetic variants, it is unrealistic to expect exactly zero pleiotropy, i.e. that the effect on the outcome variable is completely mediated by the exposure.<sup>57</sup> However, we can differentiate between pleiotropic effects that are weak and strong relative to the other interactions. A weak pleiotropic effect will not introduce significant bias, and therefore still enable us to produce a useful estimate of the causal effect, in spite of the violation of IV3.<sup>58</sup> For this reason, in choosing the prior for our structural parameters, we start from the assumption that the interactions between variables are either “strong” (relevant) or “weak” (irrelevant), in a similar vein to the considerations made by Berzuini et al.<sup>40</sup> This is different from the standard assumption that the pleiotropic effects are either zero or nonzero as used, for example, in the LASSO-based method of Kang et al.<sup>42</sup> A reasonable choice for the prior (of  $\theta \in \Theta$ ) is then a scale mixture of two Gaussian distributions centered at zero

$$\theta \sim w \cdot \mathcal{N}(0, \tau^2) + (1 - w) \cdot \mathcal{N}(0, \lambda \tau^2) \quad (7)$$

where  $\tau > 0$  and  $\lambda \in (0, 1)$ . In the limit of  $\lambda \rightarrow 0$ , we obtain the spike-and-slab prior.<sup>59</sup> This Gaussian mixture prior has a straightforward interpretation, in which the component with higher variance is meant to capture the “strong” interactions, while the component with lower variance is meant to capture the “weak” interactions. It

is informative in the sense that we expect some of the interactions in our data to be negligible, though not exactly zero, and therefore encourage the exclusion of irrelevant nonzero effects.<sup>60</sup> For example, we might expect that some (most) of the genetic variants do not have significant pleiotropic effects, but exhibit some weak unmediated residual interaction.

Our Gaussian mixture prior induces model parsimony by giving a preference for the model with the smallest number of strong parameters, which will then allow us to choose the most parsimonious causal explanation (direct causation, reverse causation, no causation).<sup>61</sup> Moreover, since every interaction is shrunk selectively, in the sense that smaller coefficient estimates are shrunk towards zero more sharply than larger coefficients, a significant gain in efficiency can be achieved.<sup>62</sup> Other shrinkage priors, such as the horseshoe employed by Berzuini et al.,<sup>40</sup> have also been used successfully for Bayesian MR. However, we have chosen the spike-and-slab prior due to ease of implementation and interpretation. In our experiments, we found that nested sampling works better when we specify the structural parameter (marginal) prior density directly instead of using a hierarchical approach. Consequently, we found it is helpful that the spike-and-slab prior density is simply a mixture of Gaussian densities, as opposed to the horseshoe prior density which has no closed-form expression.<sup>63</sup>

### 3.4.1 Rescaling

To make our method independent of the measured variables' scale, we rescale our model parameters as follows. We divide the first equation in equation (4) by  $\sigma_X$ , the scale of the first error term  $\epsilon_X$ , and we divide the second equation by  $\sigma_Y$ , the scale of the second error term  $\epsilon_Y$

$$\begin{aligned}\tilde{X} &:= \tilde{\mu}_X + \tilde{\gamma}^T \tilde{\mathbf{G}} + \tilde{\kappa}_X U + \tilde{\epsilon}_X \\ \tilde{Y} &:= \tilde{\mu}_Y + \tilde{\alpha}^T \tilde{\mathbf{G}} + \tilde{\beta} \tilde{X} + \tilde{\kappa}_Y U + \tilde{\epsilon}_Y\end{aligned}\quad (8)$$

The standardized variables

$$\tilde{X} = \frac{X}{\sigma_X}, \quad \tilde{Y} = \frac{Y}{\sigma_Y}, \quad \tilde{G}_j = \frac{G_j}{\sqrt{2p_j(1-p_j)}}, \quad \forall j \in \{1, 2, \dots, J\},$$

are dimensionless quantities. The parameters of the model are also appropriately rescaled and become dimensionless

$$\tilde{\beta} = \frac{\sigma_X}{\sigma_Y} \beta, \quad \tilde{\gamma}_j = \frac{\sqrt{2p_j(1-p_j)}}{\sigma_X} \gamma_j, \quad \tilde{\alpha}_j = \frac{\sqrt{2p_j(1-p_j)}}{\sigma_Y} \alpha_j.$$

We then put priors on the set of rescaled parameters  $\tilde{\Theta} = (\tilde{\gamma}, \tilde{\beta}, \tilde{\alpha}, \sigma_X, \sigma_Y, \tilde{\kappa}_X, \tilde{\kappa}_Y)$ .

### 3.4.2 Choice of priors

$\tilde{\alpha}$ : A Gaussian mixture prior suitably reflects the uncertainty in the pleiotropy of a genetic variant. If the pleiotropic effect is weak, then the induced bias for computing the causal effect is very small, even though the *exclusion restriction* (IV3) assumption is violated. On the other hand, if the pleiotropic effect is strong, we would like to correct for this bias. Our model incorporates both situations naturally by assigning a Gaussian mixture (spike-and-slab) prior to the (rescaled) pleiotropic interaction. If the interaction is weak (irrelevant), then it will be captured by the “spike”. If the interaction is strong (relevant), then it will be captured by the “slab”.

$\tilde{\gamma}$ : The genetic variants are often preselected using the criterion of association with the exposure of interest, which suggests that the instrument strength is nonzero. A normal prior with large enough variance would be appropriate for this parameter. If we have prior knowledge that the instrument is weak, we can appropriately choose a smaller variance for the prior. It is also possible that variants selected for the association with the exposure are actually more strongly related (upstream) to the outcome. The strength of the genotype-exposure association could then be seen as “weak” or “irrelevant” compared to the genotype-outcome association. With GWAS growing ever larger, the statistical power to detect significant associations that may be influencing the trait downstream of many other pathways increases,<sup>33</sup> in which case a Gaussian mixture prior could also be appropriate.

- $\tilde{\beta}$ : The mixture of Gaussians prior is also suitable for the exposure–outcome causal effect, since it is possible that the dependence  $G \perp\!\!\!\perp Y$  is explained by pleiotropy and not by a causal effect  $X \rightarrow Y$ . Because we do not know a priori if we should expect a significant causal effect, we fix the weights of the mixture to 0.5, which corresponds to an indifference prior.<sup>59</sup>
- $\tilde{\kappa}_X, \tilde{\kappa}_Y$ : We also put a Gaussian mixture prior on the confounding coefficients to express the uncertainty with regard to the presence of confounding.
- $\sigma_X, \sigma_Y$ : For the scale parameters, we choose a weakly informative, but proper half-Gaussian prior with a large standard deviation, as suggested by Gelman.<sup>64</sup>

### 3.4.3 Prior hyperparameters

When we do not know if the (potentially pleiotropic) genetic variants we intend to use for MR are valid or not, it is unclear whether their pleiotropic effects are more likely to be “large” or “small”. To express this uncertainty, we define the hyperparameter  $w_\alpha$ , corresponding to the weight of the Gaussian mixture prior for the pleiotropic effects and assign to it an uninformative uniform hyperprior

$$\begin{aligned}\tilde{\alpha}_j &\sim w_\alpha \cdot \mathcal{N}(0, \tau^2) + (1 - w_\alpha) \cdot \mathcal{N}(0, \lambda \tau^2), \quad \forall j \\ w_\alpha &\sim \mathcal{U}(0, 1)\end{aligned}$$

We can make a similar argument for the strength of the association between the genetic variants and the exposures, i.e. the “instrument strengths”. Analogously, we obtain the hierarchical prior

$$\begin{aligned}\tilde{\gamma}_j &\sim w_\gamma \cdot \mathcal{N}(0, \tau^2) + (1 - w_\gamma) \cdot \mathcal{N}(0, \lambda \tau^2), \quad \forall j \\ w_\gamma &\sim \mathcal{U}(0, 1)\end{aligned}$$

## 4 Bayesian model averaging and inference

### 4.1 Computing the model evidence

The Bayesian view of model comparison involves the use of probabilities to represent uncertainty in the choice of model.<sup>44</sup> We define the prior probability of a model  $p(\mathcal{M})$  to express our preference for different models. We then wish to compute the posterior distribution

$$p(\mathcal{M}|\mathbf{D}) \propto p(\mathbf{D}|\mathcal{M})p(\mathcal{M})$$

which will give us the probability of the model given the data. Bayesian model comparison and selection is a challenging problem, as it involves computing the Bayesian *model evidence*, also called the *marginal likelihood*<sup>44</sup>

$$p(\mathbf{D}|\mathcal{M}) = \int d\Theta p(\mathbf{D}|\mathcal{M}, \Theta)p(\Theta|\mathcal{M}) \quad (9)$$

The model evidence can be viewed as the probability of generating the data set  $\mathbf{D}$  from a model whose parameters are sampled at random from the prior, and therefore expresses the preference shown by the data for different models. Evaluating the model evidence involves marginalizing the likelihood-prior product over all the parameters of the model (see equation (9)). With the exception of a few special cases, this marginalization is analytically intractable and can potentially become very difficult to compute as the number of parameters increases. Fortunately, there are sampling schemes such as *nested sampling*<sup>65</sup> that are designed to accurately compute the model evidence.

When performing model comparison, it is convenient to work with probability ratios. The *prior odds ratio*  $\frac{p(\mathcal{M})}{p(\mathcal{M}' )}$  expresses the preference we give to model  $\mathcal{M}$  over model  $\mathcal{M}'$  before looking at any of the data. The ratio of evidences for the two competing models,  $\frac{p(\mathbf{D}|\mathcal{M})}{p(\mathbf{D}|\mathcal{M}' )}$ , is called the *Bayes factor*.<sup>44</sup> Combining the prior odds with the Bayes factor, we get the *posterior odds ratio*

$$\frac{p(\mathcal{M}|\mathbf{D})}{p(\mathcal{M}' |\mathbf{D})} = \frac{p(\mathbf{D}|\mathcal{M})}{p(\mathbf{D}|\mathcal{M}' )} \cdot \frac{p(\mathcal{M})}{p(\mathcal{M}' )}$$

In our approach, we compare two competing models corresponding to the two possible causal orderings for the non-genetic variables:  $X \rightarrow Y$  and  $Y \rightarrow X$ . In the first model (Figure 9(a)), which we term  $\mathcal{M}_{X \rightarrow Y}$ , the causal relationship is in the expected direction, i.e. from exposure to outcome. In the alternate model (Figure 9(b)), which we term  $\mathcal{M}_{Y \rightarrow X}$ , the causal relationship is in the opposite direction, which means we are dealing with reverse causation for the exposure–outcome association. MR rules out any other orderings for  $(G_j, X)$  and  $(G_j, Y) \forall j \in \{1, 2, \dots, J\}$ , as the genotype assignment must have temporally preceded the other two variables.

The marginal likelihood for these models is also difficult to compute, with the Gaussian mixture prior on the parameters leading to a multimodal posterior. To solve our problem, we have used the POLYCHORD<sup>66,67</sup> algorithm, which employs a nested sampling scheme, for computing the (log-)evidences of the two models. We then estimate the causal effect in both directions and combine the results by weighting them appropriately with the model evidence. This combined result properly reflects the uncertainty of not knowing the direction of the causal relationship.

If we have background knowledge that can be used to express a prior preference for one of the two models, we can easily incorporate it into our framework. We simply have to change the prior probabilities for each model. For example, if we know that  $X$  precedes  $Y$  temporally, we can eliminate the possibility of reverse causation. In that case,  $p(\mathcal{M}_{X \rightarrow Y}) = 1$ , and hence  $p(\mathcal{M}_{X \rightarrow Y} | \mathbf{D}) = 1$ , so we are only considering the “expected direction” model. This is equivalent to the implicit assumption that the causal link is from exposure to outcome, which is often made in MR studies.

### 4.2 Posterior inference

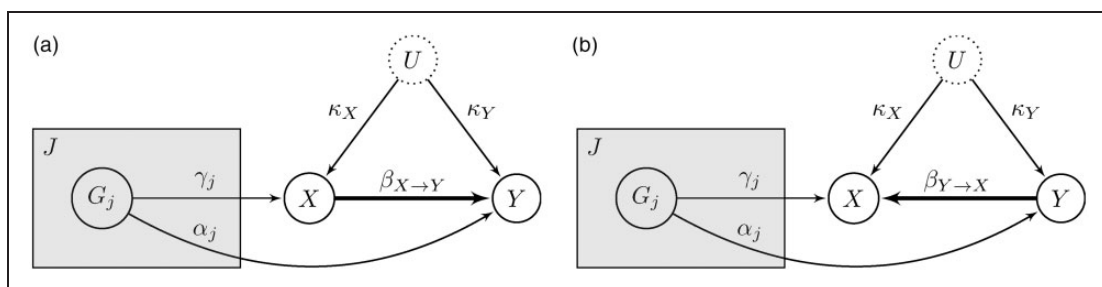
We take as input the complete data set of  $N$  i.i.d. observations  $\mathbf{D}$  or the sample covariance matrix  $\mathbf{S}$ , which is a sufficient statistic for our model (see Subsection 6.1). As discussed in the previous section, we fit the model described in Figure 8 twice by considering both directions for the causal link between  $X$  and  $Y$ , namely  $X \rightarrow Y$  (Figure 9(a)) and  $Y \rightarrow X$  (Figure 9(b)). We first assume a causal link from exposure to outcome and perform Bayesian inference to derive the posterior distribution of  $\beta_{X \rightarrow Y}$ . We then consider the reverse direction of causation and perform Bayesian inference to obtain the posterior distribution of  $\beta_{Y \rightarrow X}$ .

The Bayesian inference process is described in Figure 10 for the expected direction of the causal link. For the reverse direction, we simply switch the positions of  $X$  and  $Y$ . Given our choice of prior distributions for the rescaled parameters (see Subsection 3.4), we draw  $K$  samples from the full posterior distribution

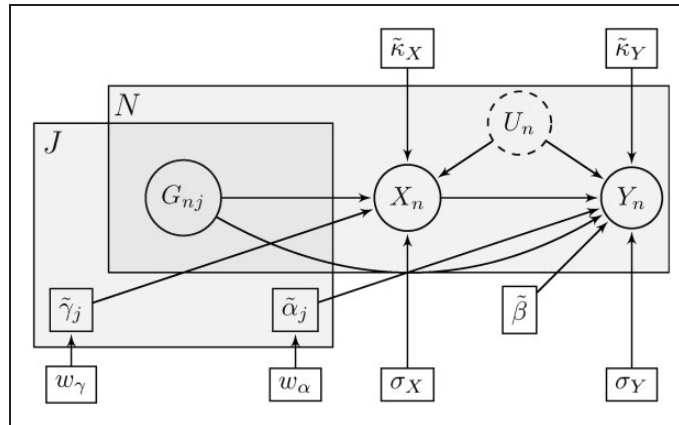
$$p(\tilde{\Theta} | \mathbf{D}) \propto \mathcal{L}_{\mathbf{D}}(\tilde{\Theta})p(\tilde{\Theta})$$

where  $\tilde{\Theta} = (\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}, \sigma_X, \sigma_Y, \tilde{\kappa}_X, \tilde{\kappa}_Y)$ . After sampling, we revert our parameter of interest to the original scale

$$\beta^{(k)} = \frac{\sigma_Y^{(k)}}{\sigma_X^{(k)}} \tilde{\beta}^{(k)}, \quad \forall k \in \{1, 2, \dots, K\}$$



**Figure 9.** Competing models encompassing both possible directions for the causal link between exposure and outcome. (a) ( $\mathcal{M}_{X \rightarrow Y}$ ) Model where the causal relationship is in the expected direction.  $\beta_{X \rightarrow Y}$  is the parameter measuring the (linear) causal effect of the exposure  $X$  on the outcome  $Y$ . The shaded plate indicates replication across the variants  $G_j$ . (b) ( $\mathcal{M}_{Y \rightarrow X}$ ) Alternative model, where the causal relationship is in the reverse direction, from outcome to exposure. Note that  $\beta_{Y \rightarrow X}$  is a different parameter than  $\beta_{X \rightarrow Y}$ . The latter is equal to zero for this model, as there is no causal link from  $X$  to  $Y$ .



**Figure 10.** Compact description of the Bayesian inference process using plate notation. The small rectangles indicate our model parameters, while the circles denote random variables (observed or unobserved). The gray superposed areas signify replication across the  $J$  genetic variants and the  $N$  data points. This is also suggested by the subscripts used for parameters and variables.

Finally, we compute the evidence for both directions, i.e.  $p(\mathbf{D}|\mathcal{M}_{X \rightarrow Y})$  and  $p(\mathbf{D}|\mathcal{M}_{Y \rightarrow X})$ , and combine the posterior distributions over the parameters

$$\begin{aligned}
 p(\beta_{X \rightarrow Y}|\mathbf{D}) &= p(\beta_{X \rightarrow Y}|\mathbf{D}, \mathcal{M}_{X \rightarrow Y}) \cdot p(\mathcal{M}_{X \rightarrow Y}) + \delta_0(\beta_{X \rightarrow Y}) \cdot p(\mathcal{M}_{Y \rightarrow X}) \\
 p(\beta_{Y \rightarrow X}|\mathbf{D}) &= p(\beta_{Y \rightarrow X}|\mathbf{D}, \mathcal{M}_{Y \rightarrow X}) \cdot p(\mathcal{M}_{Y \rightarrow X}) + \delta_0(\beta_{Y \rightarrow X}) \cdot p(\mathcal{M}_{X \rightarrow Y})
 \end{aligned}
 \tag{10}$$

where  $\delta_0(\beta)$  is the Dirac delta function centered at zero.

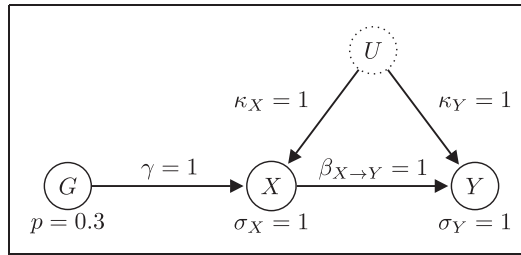
### 4.3 Example: Instrumental variable setting

We first consider a simple generating instrumental variable model with one genetic instrument (see Figure 11) to illustrate the model comparison and averaging procedure. In this example, the instrument strength  $\gamma$ , the causal effect  $\beta$ , and the confounding effect  $\kappa_X \kappa_Y$  are of the same size ( $\gamma = \beta = \kappa_X \kappa_Y = 1$ ). We have purposefully chosen unrealistic structural parameter values for ease of illustration. Note, however, that the parameters can be scaled appropriately to reflect more realistic values. The genetic variant  $G$  has effect allele frequency  $p = 0.3$ . Assuming  $G$  follows a binomial distribution with two allele copies, then  $\text{Var}[G] = 2 \cdot p \cdot (1 - p) = 0.42$ . We also set  $\mu_X = \mu_Y = 0$  and  $\sigma_X = \sigma_Y = 1$ . We generated 10,000 samples from this model, resulting in the following sample covariance and correlation over the observed variables

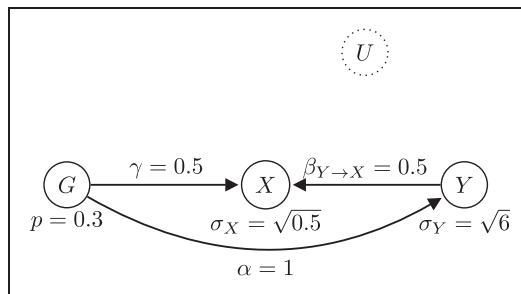
$$\Sigma = \begin{bmatrix} 0.421 & 0.434 & 0.441 \\ 0.434 & 2.447 & 3.439 \\ 0.441 & 3.439 & 6.404 \end{bmatrix}; \quad \mathbf{C} = \begin{bmatrix} 1.000 & 0.427 & 0.268 \\ 0.427 & 1.000 & 0.869 \\ 0.268 & 0.869 & 1.000 \end{bmatrix}$$

We now fit models  $\mathcal{M}_{X \rightarrow Y}$  (Figure 9(a)) and  $\mathcal{M}_{Y \rightarrow X}$  (Figure 9(b)) to the generated data and compute their log-evidence using a mixture Gaussian prior on each of the rescaled structural parameters ( $\tilde{\gamma}, \tilde{\alpha}, \tilde{\beta}, \tilde{\kappa}_X, \tilde{\kappa}_Y$ ), as described in Section 3. We fixed the hyperparameters of the Gaussian mixture prior to  $w = 0.5, \tau = 1, \lambda = 0.01$  for each structural parameter. We obtained a Bayes factor of  $\frac{p(\mathbf{D}|\mathcal{M}_{X \rightarrow Y})}{p(\mathbf{D}|\mathcal{M}_{Y \rightarrow X})} = 1.55$  in favor of the expected direction, which means that the data are 1.55 (95% CI (0.959, 2.505)) times more likely to have been generated from  $\mathcal{M}_{X \rightarrow Y}$  than  $\mathcal{M}_{Y \rightarrow X}$ . If we have no a priori preference for  $\mathcal{M}_{X \rightarrow Y}$  over  $\mathcal{M}_{Y \rightarrow X}$  or vice versa, then the *prior odds* are  $\frac{p(\mathcal{M}_{X \rightarrow Y})}{p(\mathcal{M}_{Y \rightarrow X})} = 1$  and the corresponding *posterior odds* are  $\frac{p(\mathcal{M}_{X \rightarrow Y}|\mathbf{D})}{p(\mathcal{M}_{Y \rightarrow X}|\mathbf{D})} = 1.55$ .

This result is perhaps surprising at first glance, as it implies that there is at least a one in three chance that the causal relationship is in the reverse direction. However, when we allow for pleiotropy, it becomes possible to fit a similarly sparse model where the  $X \rightarrow Y$  edge is reversed. Such a model is illustrated in Figure 12. Note that data generated from either the model in Figure 11 or the one in Figure 12 have the same underlying probability



**Figure 11.** Generating model where the causal relationship is from  $X$  to  $Y$  (the expected direction). This model satisfies all three instrumental variable assumptions.



**Figure 12.** Alternative generating model, statistically equivalent to the one in Figure 11, where the causal relationship is from  $Y$  to  $X$  (the reverse direction).

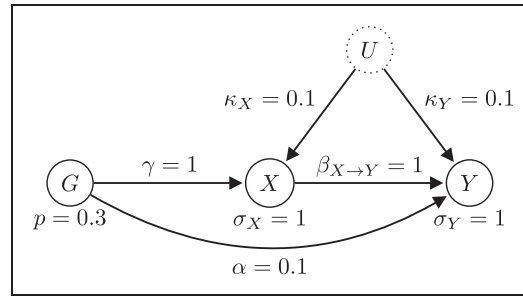
distribution over the observed variables. When we incorporate the assumption that pleiotropic effects are “weak” (close to zero), then the preference for  $\mathcal{M}_{X \rightarrow Y}$  over  $\mathcal{M}_{Y \rightarrow X}$  becomes very strong. We can express this preference in the prior for  $\alpha$ , by setting  $w = 0$  in equation (7), which is equivalent to putting a strongly informative (low variance) Gaussian prior on the size of the pleiotropic effect. Note that this is a weaker assumption than IV3 ( $\alpha = 0$ ), which is equivalent to setting  $w = \lambda = 0$  in the Gaussian mixture prior from equation (7). After appropriately changing the prior of  $\alpha$  to reflect the assumption of “weak” pleiotropy by setting the hyperparameter  $w$  to 0.0 instead of 0.5, the Bayes factor estimate becomes 100.36 (95% CI (59.42, 169.51)), which translates into a probability that the model was  $\mathcal{M}_{Y \rightarrow X}$  given the data of less than 1%.

#### 4.4 Example: Near-conditional independence (near-LCD)

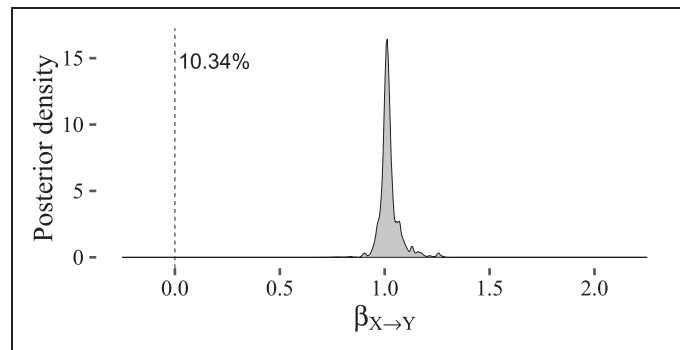
We now look at an example where the confounding and pleiotropic effects are weak relative to the instrument strength and to the exposure–outcome causal effect. The generating model is depicted in Figure 13. We call this example near-LCD, because if we made the confounding and pleiotropic effects infinitely weak, i.e.  $\alpha = \kappa = 0$ , we would arrive at the LCD pattern from Figure 5. Again, we generated 10,000 samples from the model, resulting in the following sample covariance and correlation matrices over the observed variables

$$\Sigma = \begin{bmatrix} 0.425 & 0.430 & 0.481 \\ 0.430 & 1.477 & 1.538 \\ 0.481 & 1.538 & 2.645 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 1.000 & 0.543 & 0.454 \\ 0.543 & 1.000 & 0.778 \\ 0.454 & 0.778 & 1.000 \end{bmatrix}$$

For the prior hyperparameters  $w = 0.5, \tau = 1, \lambda = 0.01$ , we obtained a Bayes factor in favor of the expected direction of:  $\frac{p(\mathcal{M}_{X \rightarrow Y}|\mathbf{D})}{p(\mathcal{M}_{Y \rightarrow X}|\mathbf{D})} = 5.90$ . The confidence interval for the Bayes factor value (95% CI (3.68, 9.47)) suggests that the preference for the expected direction is significant. This example shows that under mild violations of the IV3 assumption and given weak confounding effects, we can reliably indicate the correct causal direction using our method. Moreover, BAYESMR appropriately computes the degree of uncertainty in the model selection, which increases for stronger confounding and/or pleiotropic effects. Finally, our approach results in reasonable



**Figure 13.** Generating model where the causal relationship is from  $X$  to  $Y$  ( $\beta_{X \rightarrow Y} = 1$ ) and in which there is weak confounding ( $\kappa_X = \kappa_Y = 0.1$ ) and pleiotropy ( $\alpha = 0.1$ ). There is a weak dependence  $G \perp\!\!\!\perp Y | X$  that results in a mild violation of the IV3 assumption.



**Figure 14.** Estimated posterior density of the causal effect  $\beta_{X \rightarrow Y}$  for the data generated from the model in Figure 13. The dashed vertical line at zero indicates the estimated probability of the reverse direction (Figure 9(b)), in which case  $\beta_{X \rightarrow Y} = 0$  since there is no causal link from  $X$  to  $Y$ .

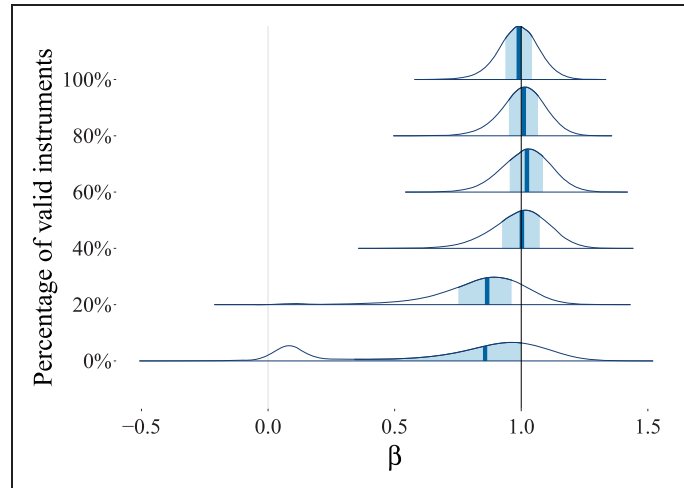
posterior samples, which we show in Figure 14. We can see that most of the mass of the posterior is concentrated around the true value  $\beta_{X \rightarrow Y} = 1$ . The density has two components, given by the two possible causal directions. The dashed vertical line at zero indicates the estimated probability of the reverse direction, for which  $\beta_{X \rightarrow Y} = 0$  since there is no causal link from  $X$  to  $Y$ .

## 5 Simulation studies

### 5.1 Estimation robustness

We imagined a typical MR scenario where there are a number ( $J=25$ ) of genetic variants that may be used as instruments, but it is not known how many of them satisfy the IV assumptions. We analyzed the robustness of the causal effect estimate produced by our method to the presence of pleiotropy. We first treated the case where all variants were valid instruments. We then added pleiotropic effects to the genetic variants, such that only 80%, 60%, 40%, 20%, and finally 0% of them were valid instruments. We generated data according to the Bayesian model specification in Section 3 (see Figure 8). The EAF of each genetic variant were randomly simulated from a uniform distribution. We simulated the instrument strengths ( $\gamma$ ) from a normal distribution with standard deviation  $\sigma_\gamma = 0.1$ . The average absolute strength of the instruments is the mean of the corresponding half-normal distribution:  $\mathbb{E}[|\gamma|] = \frac{2\sigma_\gamma}{\sqrt{2\pi}} \approx 0.08$ . We simulated the intrinsic noise of the exposure and outcome from a half-normal distribution with standard deviation equal to one, which means that  $\mathbb{E}[\sigma_X] = \mathbb{E}[\sigma_Y] = \frac{2}{\sqrt{2\pi}} \approx 0.8$ . The causal effect from  $X$  to  $Y$  was set at  $\beta = 1$ . The pleiotropic effects introduced were simulated from a normal distribution with mean zero and standard deviation  $\sigma_\alpha = 0.1$ , which means that they have a similar magnitude to that of the instrument strengths. This type of pleiotropy is referred to as *balanced* in the literature.<sup>68</sup>





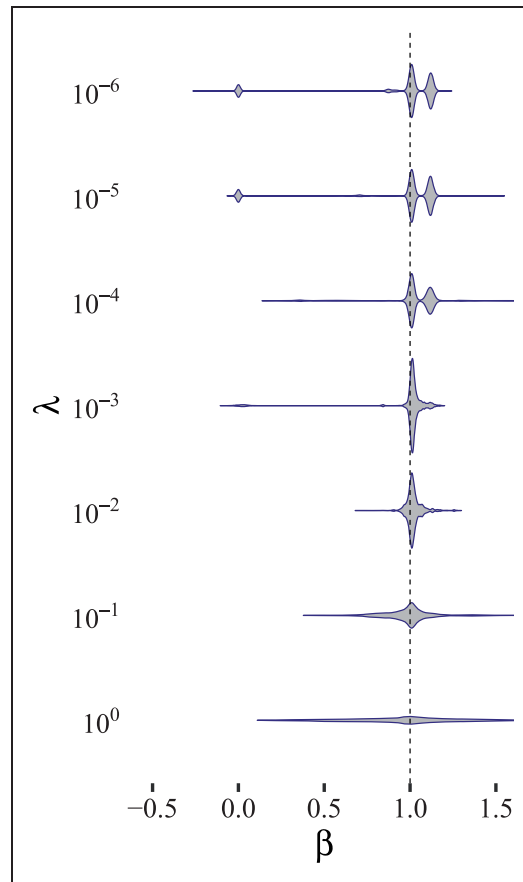
**Figure 15.** The effect of introducing pleiotropic effects on the posterior estimates is that the distribution moves away from the true value  $\beta = 1$ . The shaded area in each posterior distribution corresponds to the 50% posterior uncertainty (credible) interval, with the posterior median in the center depicted with a vertical line. In the worst case scenario, where no genetic variant is a valid instrument, we observe the appearance of a second mode of the distribution, which is close to zero. This mode corresponds to the model explanation of the data where there is a “weak” causal effect from exposure to outcome. We notice, however, that the posterior distribution progression is gradual, thereby showcasing the robustness of BAYESMR to the presence of pleiotropy. When only 40% of the genetic variants were valid instruments, the posterior distribution remained robustly centered around  $\beta = 1$ . Even when none of the genetic variants satisfied the IV assumptions, a significant proportion of the probability mass could be found around the true value.

For the inference, we set the prior hyperparameters to  $\tau = 1$ ,  $\lambda = 0.01$ . The weights indicating the proportion of relevant genetic effects are learned from the data ( $w_\gamma$  and  $w_\alpha$ ), while for the other parameters having a Gaussian mixture prior, the weight hyperparameter is set to  $w = 0.5$ . In Figure 15, we show how the causal effect estimate is affected as fewer and fewer of the genetic variants are valid instruments. We notice that even when only 40% of the genetic variants satisfied the IV assumptions, the posterior distribution remained robustly centered around the true value  $\beta = 1$ . When none of the genetic variants were valid instrumental variables, this was properly reflected in the uncertainty of the posterior distribution. In that particular case, the posterior became bimodal, suggesting two explanations for the observed associations, one where there is a strong causal link between exposure and outcome (the peak around one) and one when the associations are due to a combination of pleiotropic genetic effects and confounding (the peak close to zero).

## 5.2 Sensitivity to prior hyperparameters

In this simulation, we analyzed how the posterior computed by BAYESMR is influenced by the choice of hyperparameters in the Gaussian mixture prior. For this purpose, we again considered the near-LCD scenario introduced in Subsection 4.4 (see Figure 13). We fixed  $\tau = 1$  and varied the parameter  $\lambda$  in equally sized steps on the logarithmic scale. In this configuration, the higher variance (“slab”) component of the Gaussian mixture prior has fixed width, while the lower variance (“spike”) component’s width depends on  $\lambda$ . The mixture weights were fixed at 0.5. For each of the structural parameters on which we put a Gaussian mixture parameter, namely  $\gamma$ ,  $\alpha$ ,  $\beta$ ,  $\kappa_X$ ,  $\kappa_Y$ , we used the same hyperparameter configuration. The posterior distributions obtained for each value of  $\lambda$  are shown in Figure 16. We notice that for every different spike width, the posterior centers around the correct value  $\beta = 1$ . For narrower spikes, the posterior becomes multimodal, indicating other possible sparse solutions. For example, the mode centered around  $\beta = 1.1$  corresponds to the “weak pleiotropy” solution, where  $\alpha \approx 0$ . The mode centered at zero, on the other hand, indicates that there is a very small chance that there is no causal effect from  $X$  to  $Y$ .

The obtained posterior is sensitive to the hyperparameter choice, as reflected by the different mode heights in the multimodal posterior. This is due to the fact that the prior incorporates our belief regarding the magnitude of “weak” and “strong” effects. For instance, the appearance of the second mode around 1.1 for smaller values of  $\lambda$  can be explained by the change in the prior belief that the pleiotropic effect, which is of magnitude  $\alpha = 0.1$ , is



**Figure 16.** Posterior distribution of  $\beta$  for different hyperparameter settings (only  $\lambda$  is varied) in the near-LCD scenario (Figure 13). The true value for  $\beta$  is depicted by a dashed vertical line.

relevant or irrelevant. For  $\lambda = 10^{-2}$ , an effect of magnitude 0.1 has a prior odds of coming from the “spike” (irrelevant interaction) instead of the “slab” (relevant interaction) roughly equal to 6.1. This means that for  $\lambda = 10^{-2}$ , it is easy to “fit” a parameter of that magnitude into the “spike” component, where it is assigned a relatively high prior probability which is similar to the prior probability other small values, including zero. In that case, a solution close to the ground truth, where the confounding and pleiotropy parameters are fit to irrelevant values, will have a high posterior probability. This is reflected by the posterior mode for  $\beta$  around 1 in Figure 16. In contrast, when  $\lambda = 10^{-4}$ , then the prior odds for an effect of magnitude 0.1 to come from the “spike”, instead of the “slab”, is of order  $10^{-20}$ . This means that an effect of magnitude 0.1 is a priori extremely likely to be considered relevant. Because of the sparsifying Gaussian mixture prior, it is however much more advantageous to fit an irrelevant horizontal pleiotropy parameter instead of a relevant one, even though the ground truth is  $\alpha = 0.1$ . This means that the statistically equivalent (resulting in the same covariance matrix) solution where  $\alpha = 0$  and  $\beta = 1.1$  now has a high enough posterior probability, which leads to the appearance of the mode around  $\beta = 1.1$ .

Even though, as we have shown, the posterior distribution is driven by the choice of prior hyperparameters, our approach also enables the researcher to easily incorporate useful background knowledge, such as previously established weak or strong interactions, into the system. The advantage lies in the intuitive interpretation of the hyperparameters for the Gaussian mixture prior (see Subsection 3.4).

### 5.3 Model averaging robustness

In this experiment, we considered a bidirectional MR setting (see Subsection 2.3.2), where we cannot exclude the presence of pleiotropic effects. For simplicity, we considered a single genetic variant  $G_X$  known to be associated with  $X$  and a single genetic variant  $G_Y$  associated with  $Y$ . The true causal direction in the generating model was

**Table 1.** The results of performing a simple bidirectional Mendelian randomization analysis for various degrees of pleiotropy.

$\delta$	$\hat{\beta}_{X \rightarrow Y}^{IV}$	$\hat{\sigma}_{X \rightarrow Y}^{IV}$	$\hat{\beta}_{Y \rightarrow X}^{IV}$	$\hat{\sigma}_{Y \rightarrow X}^{IV}$	$\hat{p}(\mathcal{M}_{X \rightarrow Y}   \mathbf{D})$
-0.5	0.517	0.0346	-0.8715	0.0415	0.167
-0.4	0.615	0.0348	-0.5853	0.0352	0.211
-0.3	0.712	0.0350	-0.3746	0.0305	0.516
-0.2	0.810	0.0353	-0.2130	0.0270	0.632
-0.1	0.908	0.0356	-0.0851	0.0242	0.894
0.0	1.006	0.0359	0.0187	0.0219	0.907
0.1	1.105	0.0362	0.1045	0.0200	0.774
0.2	1.204	0.0366	0.1767	0.0184	0.864
0.3	1.303	0.0370	0.2383	0.0170	0.803
0.4	1.402	0.0374	0.2914	0.0159	0.750
0.5	1.501	0.0378	0.3377	0.0149	0.774

Note: In the first step, we used  $G_X$  as an instrument for  $X$  to estimate the causal effect  $\beta_{X \rightarrow Y}$  (the correct direction). We then used  $G_Y$  as an instrument for  $Y$  to estimate the causal effect  $\beta_{Y \rightarrow X}$  (the wrong direction). We compared these estimates against our posterior probability estimate of  $\mathcal{M}_{X \rightarrow Y}$  given the data, in which analysis we used both instruments concomitantly.

from  $X$  to  $Y$ . This generating model is described by the set of equations

$$\begin{aligned} X &:= \gamma_X G_X + \alpha_X G_Y + \kappa_X U + \epsilon_X \\ Y &:= \alpha_Y G_X + \gamma_Y G_Y + \beta X + \kappa_Y U + \epsilon_Y \end{aligned} \quad (11)$$

where  $\gamma_X$  and  $\gamma_Y$  are the instrument strengths (for  $X$  and for  $Y$ , respectively),  $\alpha_X$  and  $\alpha_Y$  are the pleiotropic (secondary) effects,  $\kappa_X$  and  $\kappa_Y$  are the confounding coefficients, and  $\beta$  is the causal effect of interest.

In our simulation, the genetic variants were strongly associated with their respective phenotypes:  $\gamma_X = \gamma_Y = 1$ . The causal effect parameter was set to  $\beta_{X \rightarrow Y} = 1$ . We varied the pleiotropic effects for each simulation by setting  $\alpha_X = \alpha_Y = \delta$ , with  $\delta \in \{-0.5, -0.4, -0.3, \dots, 0.5\}$ . At the same time, we considered strong confounding:  $\kappa_X = \kappa_Y = 1$ , while the error terms were normally distributed with dispersion given by  $\sigma_X = \sigma_Y = 1$ . We generated 10,000 samples from the above structural equation model with these parameter configurations.

We used BAYESMR to determine how much more likely the  $X \rightarrow Y$  link is than  $Y \rightarrow X$  and to produce a robust causal estimate by combining the estimates for both causal directions. The results of a bidirectional MR analysis, which amounts to computing the Wald ratio from equation (2) for both causal directions, are shown in Table 1 for each value of  $\delta$ . When  $\delta = 0$ , the conditions are ideal for the application of bidirectional MR, since both genetic variants constitute valid instruments for their respective phenotypes. In this case, bidirectional MR would yield an estimate of  $\hat{\beta}^{IV} = 1.006$  (95% CI (0.934, 1.078)) for the  $X \rightarrow Y$  causal direction and  $\hat{\beta}^{IV} = 0.018$  (95% CI (-0.025, 0.062)) for the  $Y \rightarrow X$  causal direction. Since the causal effect estimate is very close to zero for the link  $Y \rightarrow X$ , the inferred direction would be  $X \rightarrow Y$ , for which the causal effect estimate is close to the true value.

In the presence of pleiotropic effects ( $\delta \neq 0$ ), the bidirectional MR analysis is no longer suitable, since it is possible to arrive at a nonzero causal effect in both directions. In Table 1, we obtain a nonzero causal effect for both  $X \rightarrow Y$  and  $Y \rightarrow X$  anytime  $\delta \neq 0$ . Furthermore, it is possible that the causal effect estimated in the wrong direction ( $Y \rightarrow X$ ) is stronger in magnitude than the causal effect estimated in the correct direction ( $X \rightarrow Y$ ). This shows that bidirectional MR is not a reliable means of inferring the causal direction when dealing with pleiotropic genetic variants. Our approach, on the other hand was robust to the presence of weak pleiotropic effects. In the last column of Table 1, we observe a significant preference for the correct direction when the pleiotropic effects are small.

## 5.4 Sensitivity to nonlinearity and nonnormality

Our assumed Bayesian model, as described in Section 3, entails a linear relationship between the exposure  $X$  and the outcome  $Y$ , as well as a conditional Gaussian distribution given the instrument values,  $p(X, Y | \mathbf{G})$ . In this section, we evaluate the robustness of BAYESMR in situations where the parametric assumptions of linearity and normality are violated. We revisit the near-LCD scenario from Subsection 4.4, where our approach is able to recover the correct direction of the causal relationship with high probability and to produce a reasonable estimate for the causal effect, and use it as the starting point for our sensitivity analysis.

We first explore the situation in which the relationship between exposure and outcome is nonlinear. To make the results comparable to the linear case, we have simulated data where the linear term in the structural equation of  $Y$  from equation (1) is replaced by  $A \tanh(\frac{X}{A})$ , where the parameter  $A$  controls the degree of nonlinearity. In the limit  $A \rightarrow \infty$ ,  $\tanh(\frac{X}{A}) \rightarrow \frac{X}{A}$ , so we recover linearity. Furthermore, the coefficient of the linear tangent approximation at  $X = 0$  has the same value (one) as the structural coefficient  $\beta$  in the near-LCD scenario. We keep the other model parameters as in the example in Subsection 4.4. The results in Table 2 show that despite a nonlinear  $X - Y$  dependence, we are still able to detect the causal effect and its sign. Moreover, the method returns a strong preference for the correct direction for all values of  $A$ . The causal effect estimate  $\hat{\beta}_{X \rightarrow Y}$  remains robust against small to medium deviations from the linear case. It is important to note that when the linearity assumption does not hold, then the estimate returned by our method cannot be interpreted as the effect of an increase in the exposure for an individual, but represents an average causal effect across the population.<sup>69</sup> In case the exposure–outcome relationship is non-monotonic, for instance if it is U-shaped, then it is no longer guaranteed that the method can infer any causal effect. One possible solution then would be to perform a piecewise MR analysis, as suggested by Burgess et al.<sup>69</sup> The idea is to stratify on the “IV-free” exposure distribution so that the localized exposure–outcome relation is approximately linear and then run BAYESMR on each stratum.

In the second simulation, we analyze the effects of including non-Gaussian noise in the generating model. To make the results comparable to the Gaussian case, we have simulated data where the noise term  $\epsilon_Y$  in the structural equation of  $Y$  from equation (1) is distributed according to Student’s  $t$ -distribution. Here, the  $t$ -distribution’s number of degrees of freedom  $\nu$  controls the degree of normality. In the limit  $\nu \rightarrow \infty$ , the  $t$ -distribution becomes a standard normal distribution, so we recover normality. We keep the other model parameters as in the example in Subsection 4.4. The results in Table 3 show that the causal effect estimate for

**Table 2.** The results of running BAYESMR on data generated from the model in Figure 14 for various degrees of nonlinearity, as parametrized by  $A$  (lower  $A$  corresponds to stronger nonlinearity).

A	$\hat{\beta}_{X \rightarrow Y}$					$\hat{p}(\mathcal{M}_{X \rightarrow Y}   \mathbf{D})$
	Q1	Median	Mean	Q3		
0.5	0.260	0.289	0.276	0.309		(0.744, 0.880)
1	0.494	0.511	0.500	0.530		(0.817, 0.919)
2	0.739	0.758	0.737	0.786		(0.821, 0.921)
4	0.906	0.921	0.927	0.940		(0.780, 0.900)
8	0.970	0.985	0.992	1.003		(0.815, 0.918)
16	0.990	1.006	1.011	1.025		(0.776, 0.899)
32	0.991	1.009	0.996	1.028		(0.827, 0.924)

Note: We report the 95% CI for the estimated probability that the direction of the causal link is  $X \rightarrow Y$ , as opposed to  $Y \rightarrow X$ . For the causal effect estimate  $\hat{\beta}_{X \rightarrow Y}$ , we report summary measures, including the first (Q1) and third quartiles (Q3).

**Table 3.** The results of running BAYESMR on data generated from the model in Figure 14 for various degrees of nonnormality, as parametrized by the number of degrees of freedom of the  $t$ -distributed noise (lower  $\nu$  corresponds to stronger nonnormality).

$\nu$	$\hat{\beta}_{X \rightarrow Y}$					$\hat{p}(\mathcal{M}_{X \rightarrow Y}   \mathbf{D})$
	Q1	Median	Mean	Q3		
1	-1.541	0.001	0.294	1.998		(0.635, 0.823)
2	0.904	1.010	0.915	1.066		(0.743, 0.876)
4	0.990	1.019	0.990	1.044		(0.879, 0.948)
8	1.013	1.032	1.021	1.053		(0.809, 0.914)
16	0.998	1.015	1.017	1.033		(0.828, 0.924)
32	0.983	0.999	0.980	1.017		(0.852, 0.936)
64	1.007	1.024	1.025	1.043		(0.800, 0.913)

Note: We report the 95% CI for the estimated probability that the direction of the causal link is  $X \rightarrow Y$ , as opposed to  $Y \rightarrow X$ . For the causal effect estimate  $\hat{\beta}_{X \rightarrow Y}$ , we report summary measures, including the first (Q1) and third quartiles (Q3).

$\beta_{X \rightarrow Y}$  remains robust for a large range of values for  $\nu$ , corresponding to small to medium deviations from the linear case. Moreover, the method returns a strong preference for the correct direction of the causal link for all values of  $\nu$ . When  $\nu=1$ , which corresponds to a  $t$ -distribution with undefined mean and variance, our approach can no longer detect the causal link from  $X$  to  $Y$ , as the posterior distribution of the causal effect estimate becomes too broad. One potential solution for dealing with strong violations of the normality assumption would then be to adopt a non-Gaussian structural equation model, in the vein of Shimizu and Bollen.<sup>70</sup>

## 6 Real-world applications

In this section, we will showcase the potential of BAYESMR by applying it to a number of real-world problems. We start, however, by explaining how our method can be applied when only summary data are available. In the first application, we focus on estimating the causal effect of birth weight on adult fasting glucose levels. In the second application, we analyze the effect of BMI on the risk of Parkinson's disease (PD). In the third and final application, we use our approach to examine the direction of the causal link between coffee consumption and cigarette smoking.

### 6.1 Using summary data

BAYESMR requires as input the first- and second-order moments of the observed data vector  $\mathbf{Z} = (\mathbf{G}, X, Y)$ . These moments are needed in the expression of the likelihood function and constitute the sufficient statistics with respect to our model (Section 3). We can easily compute estimates for these moments when individual-level data is available. However, we often only have access to summary statistics in the form of regression (beta) coefficients and their standard errors. In this subsection, we show how to derive approximations for the first- and second-order moments from the summary data in order to obtain the required sufficient statistics for our method. We require (at least) the following summary statistics:

- $\hat{p}_j$ : the effect allele frequency (EAF) of  $G_j$
- $m$ : the number of allele copies (very often equal to two, since humans are diploid organisms)
- $\hat{\gamma}_j, \hat{\sigma}_{\hat{\gamma}_j}, N_{\hat{\gamma}_j}$ : measures of the gene-exposure association, including the coefficient obtained by regressing  $X$  on  $G_j$ , its standard error and the sample size
- $\hat{\Gamma}_j, \hat{\sigma}_{\hat{\Gamma}_j}, N_{\hat{\Gamma}_j}$ : measures of the gene-outcome association, including the coefficient obtained by regressing  $Y$  on  $G_j$ , its standard error and the sample size
- $\hat{\beta}$ : the coefficient obtained by regressing  $X$  and  $Y$  (observational exposure–outcome association).

Summary data on gene-exposure and gene-outcome associations from GWAS has become increasingly available, so we can typically get estimates for  $\hat{\gamma}_j, \hat{\Gamma}_j$  together with the associated standard errors and sample sizes. The effect allele frequency  $\hat{p}_j$  is usually also reported. In addition, we require a measure of the association between the exposure and the outcome ( $\hat{\beta}$ ) to derive an estimate of  $\text{Cov}[X, Y]$ . This estimate can be obtained from observational studies for determining potential risk factors for the outcome.

To derive the expected values and variances for the genetic variants, we employ the binomial distribution assumption by plugging in the EAF as the estimated success probability and using the appropriate formulas. We can assume without loss of generality that the mean parameters  $\mu_X$  and  $\mu_Y$  in equation (4) are zero, as their location does not influence the regression slope. For estimating the second-order moments, we then employ the following well-known approximations from simple linear regression

$$\begin{aligned}\hat{\gamma}_j &\approx \frac{\text{Cov}[G_j, X]}{\text{Var}[G_j]} \\ \hat{\Gamma}_j &\approx \frac{\text{Cov}[G_j, Y]}{\text{Var}[G_j]} \\ \hat{\beta} &\approx \frac{\text{Cov}[X, Y]}{\text{Var}[X]} \\ \hat{\sigma}_{\hat{\gamma}_j}^2 &\approx \frac{1}{N_{\hat{\gamma}_j}} \left( \frac{\text{Var}[X]}{\text{Var}[G_j]} - \hat{\gamma}_j^2 \right)\end{aligned}$$

$$\hat{\sigma}_{\hat{\Gamma}_j}^2 \approx \frac{1}{N_\Gamma} \left( \frac{\text{Var}[Y]}{\text{Var}[G_j]} - \hat{\Gamma}^2 \right)$$

Note that these approximations also apply in a multivariate setting when the regressors are independent. We use these approximations to finally derive the following estimates for the moments from summary statistics

$$\begin{aligned} \mathbb{E}[G_j] &\approx m \cdot \hat{p}_j \quad (= \mathbb{E}[\widehat{G}_j]) \\ \mathbb{E}[X] &\approx \sum_j \mathbb{E}[\widehat{G}_j] \cdot \hat{\gamma}_j \\ \mathbb{E}[Y] &\approx \sum_j \mathbb{E}[\widehat{G}_j] \cdot \hat{\Gamma}_j \\ \text{Var}[G_j] &\approx m \cdot \hat{p}_j \cdot (1 - \hat{p}_j) \quad (= \text{Var}[\widehat{G}_j]) \\ \text{Cov}[G_j, X] &\approx \text{Var}[\widehat{G}_j] \cdot \hat{\gamma}_j \\ \text{Cov}[G_j, Y] &\approx \text{Var}[\widehat{G}_j] \cdot \hat{\Gamma}_j \\ \text{Var}[X] &\approx \text{Var}[\widehat{G}_j] \cdot (\hat{\gamma}_j^2 + N_{\hat{\gamma}_j} \cdot \hat{\sigma}_{\hat{\gamma}_j}^2) \quad (= \text{Var}[\widehat{X}]) \\ \text{Var}[Y] &\approx \text{Var}[\widehat{G}_j] \cdot (\hat{\Gamma}_j^2 + N_{\hat{\Gamma}_j} \cdot \hat{\sigma}_{\hat{\Gamma}_j}^2) \\ \text{Cov}[X, Y] &\approx \text{Var}[\widehat{X}] \cdot \hat{\beta} \end{aligned} \tag{12}$$

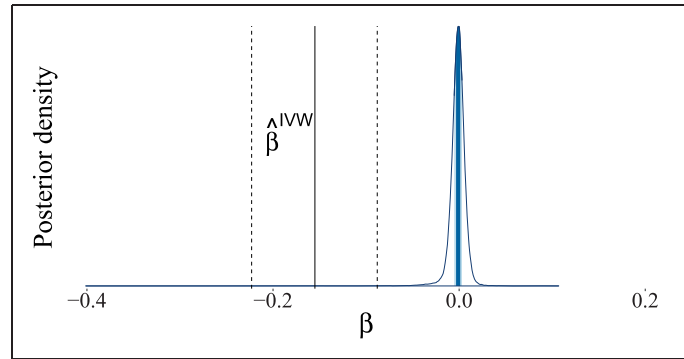
When we have information on multiple genetic variants, we obtain multiple estimates of  $\text{Var}[X]$  and  $\text{Var}[Y]$  in equation (12), in which case we use the average over the estimates. Our approach also requires specifying a sample size. Since the summary statistics are likely to be computed from different samples, we conservatively choose the minimum of their sizes as input to BAYESMR in order not to overestimate the precision of the data. If the sample size for the exposure–outcome association measure is also available, we take it into consideration when calculating the minimum of the sample sizes.

## 6.2 Effect of birth weight on fasting glucose

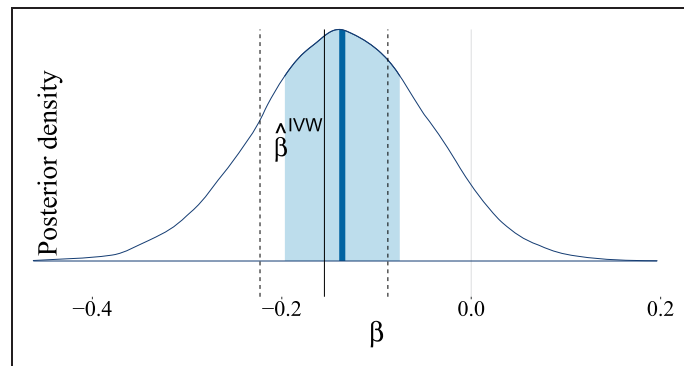
Del Greco et al.<sup>71</sup> performed an IVW meta-analysis of MR Wald estimates (see equations (2), (3)) to analyze the relationship between low birth weight and adult fasting glucose levels. The authors chose seven genetic variants associated with birth weight to use as instruments in an MR analysis. The meta-analysis of the seven MR estimates suggested a significant protective effect of  $-0.155$  mmol/L (95% CI  $(-0.233, -0.088)$ ) reduction in adult fasting glucose level per standard deviation increase (484 g) in birth weight. In their analysis, Del Greco et al.<sup>71</sup> investigated the presence of pleiotropy by means of the between-instrument heterogeneity  $Q$  test. They reported significant evidence of heterogeneity across instruments ( $p=0.03$ ), which they believe suggests that IV3 might be violated for some of the genetic variants. The heterogeneity was primarily due to the variant rs9883024 in gene ADCY5. However, even after removing said variant from the analysis, they obtained a significant negative effect estimate of  $-0.098$  (95% CI  $(-0.168, -0.027)$ ).

We ran our model using the summary statistics for the genotype-exposure and genotype-outcome associations provided in Table IV of Del Greco et al.<sup>71</sup>. The regression coefficient for the observational exposure–outcome association is taken from Daly et al.,<sup>72</sup> who reported a decrease of 0.01 mmol/L in fasting glucose levels per kilogram increase in birth weight in a retrospective cohort study on a multiethnic sample of 855 New Zealand adolescents. The regression coefficients they computed were adjusted for age, sex, and ethnicity. One potential limitation of using this data is that there may be biases induced by the different adjustments made in the study.<sup>31</sup> Another potential issue is the fact that the observational study sample is taken from an underlying population which is different in terms of ethnicity and age.<sup>73</sup> Nevertheless, the associations found in the observational study are consistent with those found in subsequent studies.<sup>74</sup>

Our estimated posterior in Figure 17 indicates that no causal effect is necessary to explain the observed data. This conclusion differs significantly from that of Del Greco et al.<sup>71</sup> because we are making less stringent assumptions regarding the parameter strengths. However, we can arrive at their conclusion if we adapt



**Figure 17.** Estimated posterior distribution for the causal effect of birth weight on adult fasting glucose levels. The light shaded area in the posterior represents the interquartile range, while the dark shaded line indicates the median. For the Gaussian mixture prior in equation (7), we have taken  $\tau^2 = 1$  and  $\lambda = 10^{-4}$ . The IVW estimate reported in Del Greco et al.<sup>70</sup> and its confidence bounds are shown for comparison.

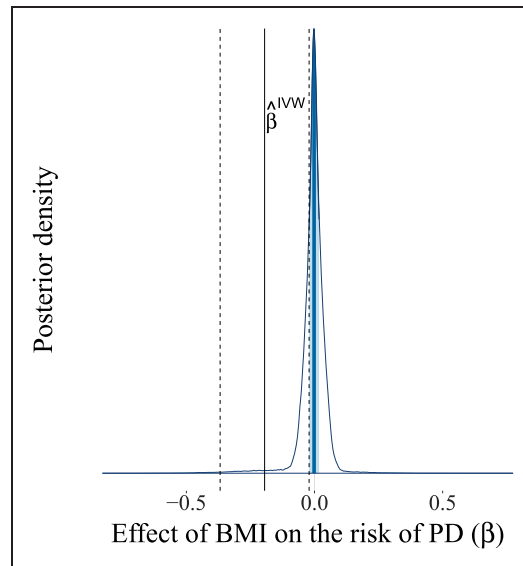


**Figure 18.** Estimated posterior distribution for the causal effect of birth weight on adult fasting glucose after adapting the prior knowledge to fit the classic instrumental variable setting. The light shaded area in the posterior represents the interquartile range, while the dark shaded line indicates the median. For the Gaussian mixture prior in equation (7), we have taken  $\tau^2 = 1$  and  $\lambda = 10^{-4}$ , respectively. The IVW estimate reported in Del Greco et al.<sup>70</sup> and its confidence bounds are shown for comparison.

our prior knowledge by incorporating stronger prior assumptions. To obtain the posterior distribution in Figure 18, we assumed that the pleiotropic effects are negligible (by setting  $w_{\alpha} = 0$ , equivalent to a weaker form of IV3) and that the instrument strengths are relevant (by setting  $w_{\gamma} = 1$ , equivalent to a weaker form of IV1). Additionally, we set  $w = 0$  in the Gaussian mixture prior (equation (7)) for  $\tilde{\kappa}_X$ ,  $\tilde{\kappa}_Y$ , and  $\tilde{\beta}$ , so as to not penalize (potentially) relevant confounding and causal effects too strongly. Since this new set of assumptions is much closer to the typical assumptions made in an IV analysis, we were able to recover the IVW estimate. Our results show, however, that if we penalize every strong interaction between variables equally (see Subsection 3.4), then the posterior estimate of  $\beta$  indicates a negligible causal effect of birth weight on adult fasting glucose.

### 6.3 Effect of BMI on the risk of PD

There have been conflicting findings regarding the association between BMI and PD in observational studies. Both positive and negative associations between higher BMI and the risk of PD have been reported.<sup>75</sup> Wang et al.<sup>76</sup> performed a meta-analysis of 10 cohort studies and their obtained pooled risk ratio (RR) for the association of a 5 kg/m<sup>2</sup> higher BMI with the risk of PD was 1.00 (95% CI (0.89, 1.12)), which suggests that BMI is not associated with the risk of PD. In an MR analysis, Noyce et al.<sup>75</sup> estimated the causal effect of BMI on the risk of PD by combining the ratio estimates corresponding to 77 selected single-nucleotide polymorphisms (SNPs) via an IVW scheme. The results of their analysis suggest a putative protective causal effect of BMI. More specifically, their MR



**Figure 19.** Estimated causal effect of BMI on the risk of PD expressed as the difference in log-odds of PD per 5 kg/m<sup>2</sup> increase in BMI. The light shaded area in the posterior represents the interquartile range, while the dark shaded line indicates the median. The IVW estimate  $\hat{\beta}^{IVW}$  derived from equation (3) along with its 95% confidence bounds are shown for comparison.

analysis yielded an IVW log-odds ratio (log-OR) of  $-0.195$  (95% CI  $(-0.368, -0.021)$ ), meaning that a lifetime exposure of 5 kg/m<sup>2</sup> higher BMI was associated with a lower risk of PD.

We also took the log-OR as the continuous outcome variable. We used the summary data from Noyce et al.<sup>75</sup> for the genetic associations with BMI and the log-OR of PD. We used the RR reported by Wang et al.<sup>76</sup> as an approximation for the OR, since the incidence of PD is very low, resulting in an estimated log-OR of zero for the observational association of a 5 kg/m<sup>2</sup> higher BMI with the risk of PD. We fitted our model assuming the direction of causality from BMI to PD and we obtained a causal effect estimate centered around zero, with 96.3% of the probability mass in the interval  $(-0.1, 0.1)$  (see Figure 19). This result casts doubt on the existence of a causal link between BMI and the risk of PD. When looking at the scatter plot comparing the genetic associations of the 77 variants with the outcomes and their associations with the exposures (Figure 20), we observe the existence of two outliers, corresponding to the red triangles. These two variants show a low association with the outcome relative to the others given their strength as instruments. It is possible that the unusually large negative association with the risk of PD is due to unobserved pleiotropic effects. This claim is supported by our estimates of the pleiotropic effect  $\alpha$  for these two genetic variants, which are shown in Figure 21.

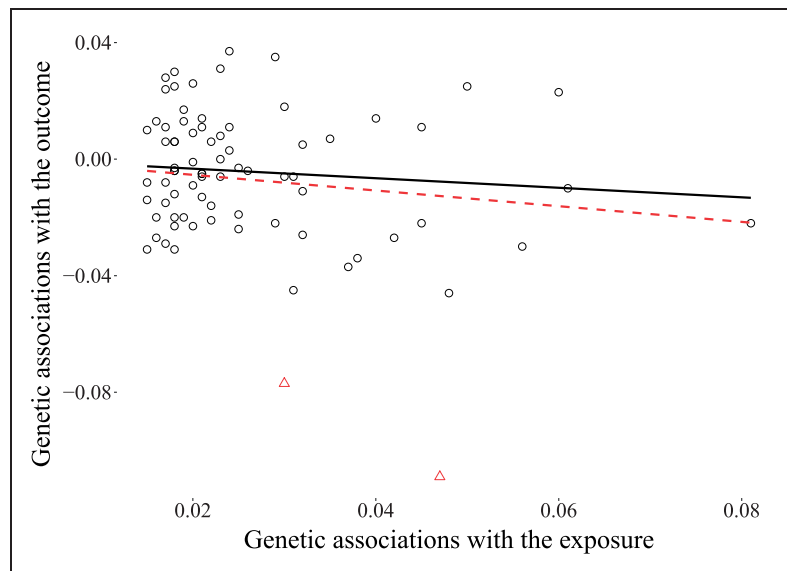
To further substantiate our claim that these variants represent outliers for the IVW and MR-Egger analysis, we used the radial regression approach of Bowden et al.<sup>77</sup> By running the radial IVW and MR-Egger regressions using first-order weights and specifying a statistical significance threshold of 0.01, we discovered the same two outliers observed visually in the scatter plot. Alternatively, we used the MR *pleiotropy residual sum and outlier* (MR-PRESSO) test<sup>57</sup> to identify potential pleiotropic outliers. With MR-PRESSO, we detected the two genetic variants highlighted in Figure 20 as outliers for a significance threshold of 0.1. The outlier-corrected causal estimate produced by MR-PRESSO remained “barely” significant at the 0.05 level:  $\hat{\beta} = -0.1669$  (95% CI  $(-0.3335, -0.0003)$ ).

To sum up, the results of our investigation suggest that the protective causal effect of BMI on the risk of PD discovered by Noyce et al.<sup>74</sup> could also be due to some of the genetic variants exhibiting horizontal pleiotropy. We suspect two variants in particular to be pleiotropic outliers in the MR analysis, as suggested by our method and other approaches.<sup>57,77</sup>

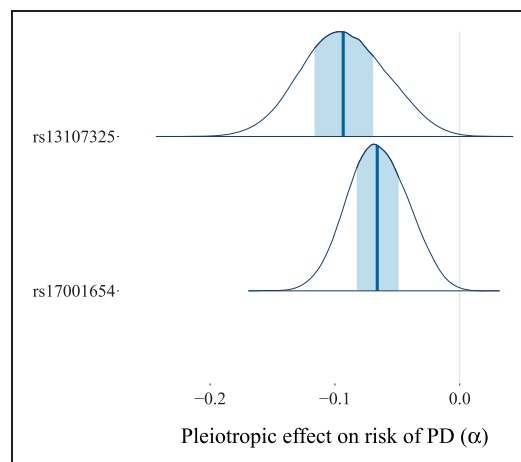
#### 6.4 Does coffee consumption influence smoking?

In this experiment, we investigated the association between coffee consumption and cigarette smoking. It is unclear whether this association is causal. An MR study performed by Treur et al.<sup>78</sup> provided no evidence for causal effects of smoking on caffeine or vice versa. Bjørngaard et al.,<sup>79</sup> on the other hand, found that higher cigarette



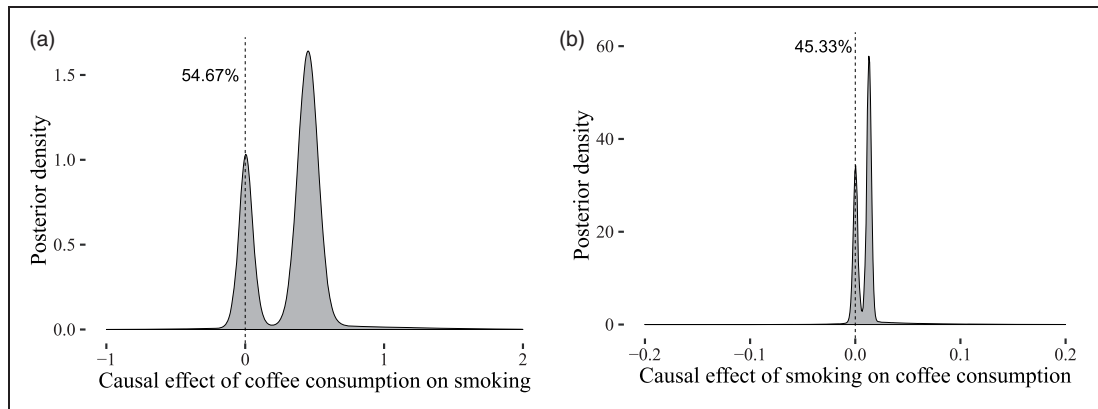


**Figure 20.** Scatter plot of the genetic associations with BMI (horizontal axis) and PD risk (vertical axis) for 77 genetic variants. The two outliers (triangles) show a relatively strong association with the outcome given their association with the exposure. The regression line including the outliers is dashed, while the regression line obtained without the outliers is continuous.



**Figure 21.** Estimated pleiotropic effects for the two genetic variants suspected of being pleiotropic outliers (rs17001654 and rs13107325). The light shaded area in the posterior represents the interquartile range, while the dark shaded line indicates the median. Most of the posterior mass is distributed away from zero, thereby supporting the suspicion that these two variants exhibit horizontal pleiotropy.

consumption causally increases coffee intake. If a causal link between coffee consumption and cigarette smoking exists, its direction is also unclear. Verweij et al.<sup>54</sup> employed a two-sample bidirectional MR study to investigate, among other things, the causal association between the use of nicotine and caffeine, but found little evidence of a causal relationship in either direction. In another study, Ware et al.<sup>80</sup> assessed the impact of coffee consumption on the heaviness of smoking, but also obtained inconclusive results: one of their two-sample MR analyses indicated that heavier consumption of caffeine might lead to reduced heaviness of smoking, while in other MR analyses they found no evidence of a causal relationship between coffee consumption and heaviness of smoking. Ware et al.<sup>80</sup> concluded it is unlikely that coffee consumption has a major causal impact on cigarette smoking and suggested the possibility of reverse causation, i.e. smoking impacting coffee consumption, or confounding as alternative explanations for the observed association.



**Figure 22.** Comparison of the causal effect estimates between  $X$  (coffee consumption in cups per day) and  $Y$  (heaviness of smoking in cigarettes per day) for the two possible causal directions. The estimated evidence for the two models is  $p(\mathcal{M}|X \rightarrow Y) = 54.67\%$  and  $p(\mathcal{M}|Y \rightarrow X) = 45.33\%$ , respectively. In the left figure, we see the estimate of  $\beta_{X \rightarrow Y}$ , which is the causal effect of coffee consumption on heaviness of smoking, under the assumption that the causal link  $X \rightarrow Y$  exists. In the right figure, we see the estimate of  $\beta_{Y \rightarrow X}$ , which is the causal effect of heaviness of smoking on coffee consumption, under the assumption that the causal link  $Y \rightarrow X$  exists. (a) Posterior distribution of the putative causal effect of coffee consumption on smoking. In the case of reverse causation (causal link from smoking to coffee consumption), this effect is zero, as indicated by the vertical dashed line. The estimate next to the line (54.67%) is the evidence for the reverse model. (b) Posterior distribution of the putative causal effect of smoking on coffee consumption. In the case of reverse causation (causal link from coffee consumption to smoking), this effect is zero, as indicated by the vertical dashed line. The estimate next to the line (45.33%) is the evidence for the reverse model.

We used the summary statistics reported by Ware et al.<sup>80</sup> to explore this association. The exposure variable is coffee consumption measured in cups per day, while the outcome variable is smoking measured in cigarettes per day. The summary measurements for the gene-exposure association were taken from the European replication sample ( $N \leq 30,062$ ) of the coffee and caffeine genetics consortium genome-wide association study meta-analysis.<sup>81</sup> Eight independent coffee-related variants meeting the threshold for genome-wide significance in the trans-ethnic GWAS meta-analysis were considered. The associations of coffee-related SNPs with the outcome were obtained from the UK Biobank ( $N = 8072$ ). Ware et al.<sup>80</sup> also computed an observational association of 0.45 additional cigarettes per day for each consumed cup of coffee among the 8072 current daily smokers who reported consuming coffee.

In our approach, we considered two candidate models corresponding to the two possible causal directions. In the first model (Figure 6), we assumed a causal link from coffee consumption to smoking, while in the second model (Figure 7), we assumed the reverse causal link. We computed the evidence for the two models and then, assuming a priori that either of the two models is equally likely, we obtained a posterior probability of 42.6% for the first model and 57.4% for the second model. The posterior over the causal effect is shown for both directions in Figure 22(a) and (b). According to our results, there is significant uncertainty in the direction of the causal effect. This suggests that the observed association may in fact be due to some confounding factor. Still, that does not exclude the possibility of a causal effect between coffee consumption and heaviness of smoking, but more data are required to substantiate this claim.

## 7 Discussion

In this paper, we have proposed a BAYESMR for finding the probable direction of a causal link between two phenotypes and for estimating the magnitude of its causal effect. The novelty in our method consists in inferring the direction of causality, estimating the causal effects and promoting model sparsity in a single-step approach. This is achieved by using a nested sampling scheme<sup>65–67</sup> to compute the model evidence for both causal directions and to sample from the parameter posterior distribution simultaneously.

Many traditional MR methods can be seen as limiting cases of our more general approach. For example, the “no pleiotropy” assumption<sup>68</sup> made when using genetic variants as instrumental variables can be incorporated by setting the prior proportion of the slab component for the pleiotropic effect to zero ( $w_\alpha = 0$ ) and then taking  $\lambda \rightarrow 0$ . Berzuini et al.,<sup>40</sup> for example, have proposed a similar Bayesian approach, but they assume a particular

causal direction. This assumption can be incorporated into our model by setting the prior probability of reverse causation to be zero, as shown in Section 4. Our Bayesian solution subsumes standard MR (Figure 4) and the LCD pattern (Figure 5), is robust to the presence of pleiotropic effects (Figure 6), and incorporates the possibility of reverse causation in the exposure–outcome association (Figure 7). The resulting output appropriately reflects the uncertainty of having to consider two possible causal directions due to the existence of alternative causal pathways.

In the era of high-throughput genomics, it is becoming increasingly likely to select invalid instruments as GWAS sample sizes continue to grow.<sup>33</sup> Our approach allows for the selection of a much broader range of genetic variants, since it does not require any of the instruments to be valid. Furthermore, we can select even those variants for which it is not clear whether they primarily influence the exposure or the outcome. In this sense, we envision running BAYESMR on a large number of potential genetic candidates collected from GWAS, which can be related to either the exposure or the outcome. Another advantage of our method is that it does not rely on individual-level data. This is becoming increasingly important in the era of large GWA studies,<sup>16</sup> where summary statistics regarding genetic associations obtained from large independent samples are made publicly available. Our approach, however, is currently not designed to handle potentially correlated genetic variants, which means it cannot use variants that are in linkage disequilibrium with each other as input. Another concern is the computational scalability of the Bayesian inference to a large number of instrumental variables. In the future, we plan to explore more scalable nested sampling approaches, for instance the one proposed by Buchner,<sup>82</sup> and to extend BAYESMR to account for correlation among genetic variants.

A key aspect of our approach lies in the structural assumption that interactions between variables are either “weak” (irrelevant) or “strong” (relevant).<sup>61</sup> This assumption is different from the traditional zero–nonzero dichotomy and, in our opinion, more realistic than assuming, for example, that a genetic effect on an outcome variable is “completely mediated” by another variable. The Gaussian mixture prior is a natural choice for expressing this assumption and has the intuitive interpretation of capturing the small, irrelevant effects in the “spike” (lower variance) component and the large, relevant effects in the “slab” (higher variance) component. The mixture Gaussian prior has already been used, for example, by Li<sup>38</sup> to induce sparsity in the estimation of pleiotropic effects. Here, we extend the usage of this prior to the other structural parameters, which allows for a more general view of MR analyses. For example, putting a Gaussian mixture prior on the strength of instruments will enable us to make an automatic selection of the relevant instruments in a batch of preselected genetic variants, where the proportion of relevant instruments can be determined by learning the shared hyperparameter  $w_\gamma$ .

Our chosen prior is informative in the sense that it informs which effects we consider a priori to be relevant. Because of this, care must be taken when setting the prior hyperparameters. While this informativeness may be seen as a weakness of the Bayesian approach, it also empowers the research to input sensible assumptions regarding the expected magnitude of effects in an intuitive fashion. If one has a prior idea regarding which effect sizes are deemed relevant and which irrelevant, then one can appropriately tune the variances of the “spike” and the “slab” to reflect this belief. When it is not clear how to distinguish between relevant and irrelevant effects, one can treat all detectable effects as relevant by letting  $\lambda \rightarrow 0$  in a first attempt. At the same time, one can choose  $\tau$  large enough to give support to effect sizes that are substantively different from zero, but not so large that unrealistic values are supported.<sup>60</sup> Furthermore, if one has a prior belief regarding the relevance of a particular parameter, this can be expressed in the prior by appropriately setting the  $w$  parameter, which determines the mixture component proportion. Other sparsifying priors have been proposed for use in MR Bayesian methods, such as the *horseshoe prior*<sup>40</sup> or the *Laplace prior*.<sup>49</sup> It might be interesting to compare the inferences obtained by using these different priors and to assess how easy it is to incorporate prior biological knowledge for each of them.

One important limitation of the approach proposed here lies in the strong parametric assumptions of linearity and Gaussianity.<sup>15</sup> These standard assumptions are commonly made for simplicity and computational convenience in similar works such as those by Thompson et al.,<sup>46</sup> Li,<sup>38</sup> or Berzuini et al.,<sup>40</sup> but when they do not hold, the estimands can potentially be far off the mark and their interpretation is rendered incorrect. We have shown that BAYESMR is robust against mild violations of linearity and Gaussianity, even though great care must be taken when interpreting the results. In future work, it would be interesting to test our method as well as other established methods against a broader range of violations. The current method is also not directly applicable to discrete phenotypic variables. However, as we have shown in Section 6, we can incorporate log-OR of binary values in our model by applying a logit transformation. With our approach, we allow for violations of the exclusion restriction assumption (IV3) and we even allow for violations of IV1 (genetic variant is not robustly

associated with the exposure), although this is typically not considered an issue. However, we still rely on the genetic variants being independent from any unmeasured confounding variables (IV2).

For future work, we plan to make our approach even more general by relaxing the IV2 assumption, and therefore taking into account the possibility that the genetic variables might be associated with unmeasured confounders. This would also mean that the InSIDE assumption, which is required for applying MR-Egger<sup>17</sup> or the Bayesian method proposed by Berzuini et al.,<sup>40</sup> does not hold. Another immediate extension to our work is handling potential measurement error for the exposure and outcome variables. Traditional methods such as MR-Egger rely on having no measurement error in the exposure (the so-called NOME<sup>83</sup> assumption), which is difficult to achieve in practical applications and may lead to erroneous results. Finally, we are also interested in analyzing the links between variables that exert a mutual causal influence on each other. For this purpose, we intend to extend BAYESMR to handle cyclic causal models.

The code implementing the proposed method will be made publicly available by the authors at <https://github.com/igbucur/BayesMR>.

### Acknowledgements

We would like to thank the anonymous reviewers for their thoughtful comments and efforts towards improving our manuscript.


### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received financial support for the research, authorship, and/or publication of this article: This research has been partially financed by the Netherlands Organisation for Scientific Research (NWO), under project 617.001.451.

### ORCID iDs

Ioan Gabriel Bucur  <https://orcid.org/0000-0002-7852-052X>

Tom Heskes  <https://orcid.org/0000-0002-3398-5235>

### References

1. Fletcher RH, Fletcher SW and Fletcher GS. *Clinical epidemiology: the essentials*. Philadelphia: Lippincott Williams & Wilkins, 2012.
2. Lawlor DA, Harbord RM, Sterne JAC, et al. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008; **27**: 1133–1163.
3. Davey Smith G, Paternoster L and Relton C. When will Mendelian randomization become relevant for clinical practice and public health? *JAMA* 2017; **317**: 589–591.
4. Davey Smith G and Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014; **23**: R89–R98.
5. Zheng J, Baird D, Borges MC, et al. Recent developments in Mendelian randomization studies. *Curr Epidemiol Rep* 2017; **4**: 330–345.
6. Chen L, Smith GD, Harbord RM, et al. Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. *PLOS Med* 2008; **5**: e52.
7. Ference BA, Yoo W, Alesh I, et al. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. *J Am Coll Cardiol* 2012; **60**: 2631–2639.
8. Vimalaswaran KS, Berry DJ, Lu C, et al. Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. *PLOS Med* 2013; **10**: e1001383.
9. Rothman KJ, Greenland S and Lash TL. *Modern epidemiology*. Philadelphia: Lippincott Williams & Wilkins, 2008.
10. Davey Smith G and Ebrahim S. ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32**: 1–22.

11. Wade KH, Richmond RC and Smith GD. Physical activity and longevity: how to move closer to causal inference. *Br J Sports Med* 2018; **52**: 890–891.
12. Ibero-Baraibar I, Navas-Carretero S, Abete I, et al. Increases in plasma 25(OH)D levels are related to improvements in body composition and blood pressure in middle-aged subjects after a weight loss intervention: Longitudinal study. *Clin Nutr* 2015; **34**: 1010–1017.
13. Gangloff A, Bergeron J, Pelletier-Beaumont E, et al. Effect of adipose tissue volume loss on circulating 25-hydroxyvitamin D levels: results from a 1-year lifestyle intervention in viscerally obese men. *Int J Obes* 2015; **39**: 1638–1643.
14. Bowden RJ and Turkington DA. *Instrumental variables*. Cambridge: Cambridge University Press, 1990.
15. Didelez V and Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007; **16**: 309–330.
16. Visscher PM, Wray NR, Zhang Q, et al. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* 2017; **101**: 5–22.
17. Bowden J, Davey Smith G and Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015; **44**: 512–525.
18. Brion MJA, Shakhbazov K and Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol* 2013; **42**: 1497–1501.
19. Swerdlow DI, Kuchenbaecker KB, Shah S, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int J Epidemiol* 2016; **45**: 1600–1616.
20. Palmer TM, Lawlor DA, Harbord RM, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res* 2012; **21**: 223–242.
21. Burgess S, Butterworth A and Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013; **37**: 658–665.
22. Cornia N and Mooij JM. Type-II errors of independence tests can lead to arbitrarily large errors in estimated causal effects: an illustrative example. In: *Proceedings of the UAI 2014 conference on causal inference: learning and prediction*. Volume 1274. Germany: CEUR-WS.org, pp.35–42.
23. Cooper GF. A simple constraint-based algorithm for efficiently mining observational databases for causal relationships. *Data Min Knowl Discov* 1997; **1**: 203–224.
24. Hemani G, Tilling K and Smith GD. Orienting the causal relationship between imprecisely measured traits using genetic instruments. *bioRxiv* 2017; p.117101.
25. Schadt EE, Lamb J, Yang X, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* 2005; **37**: 710–717.
26. Millstein J, Zhang B, Zhu J, et al. Disentangling molecular relationships with a causal inference test. *BMC Genet* 2009; **10**: 23.
27. Wang L and Michael T. Efficient and accurate causal inference with hidden confounders from genome-transcriptome variation data. *PLoS Comput Biol* 2017; **13**: e1005703.
28. Chen LS, Emmert-Streib F and Storey JD. Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biol* 2007; **8**: R219.
29. Yin P and Voight BF. MeRP: a high-throughput pipeline for Mendelian randomization analysis. *Bioinformatics* 2015; **31**: 957–959.
30. Davey Smith G, Lawlor DA, Harbord R, et al. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLOS Med* 2007; **4**: e352.
31. Lawlor DA. Commentary: two-sample Mendelian randomization: opportunities and challenges. *Int J Epidemiol* 2016; **45**: 908–915.
32. van Kippersluis H and Rietveld CA. Pleiotropy-robust Mendelian randomization. *Int J Epidemiol* 2017; **47**: 1279–1288.
33. Hemani G, Bowden J, Haycock PC, et al. Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. *bioRxiv* 2017; p.173682.
34. Sivakumaran S, Agakov F, Theodoratou E, et al. Abundant pleiotropy in human complex diseases and traits. *Am J Hum Genet* 2011; **89**: 607–618.
35. Solovieff N, Cotsapas C, Lee PH, et al. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet* 2013; **14**: 483–495.
36. Bowden J, Davey Smith G, Haycock PC, et al. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016; **40**: 304–314.
37. Hartwig FP, Davey Smith G and Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017; **46**: 1985–1998.
38. Li S. Mendelian randomization when many instruments are invalid: hierarchical empirical Bayes estimation. 2017.
39. Burgess S, Zuber V, Gkatzionis A, et al. Improving on a modal-based estimation method: model averaging for consistent and efficient estimation in Mendelian randomization when a plurality of candidate instruments are valid. *bioRxiv* 2017; p.175372.
40. Berzuini C, Guo H, Burgess S, et al. A Bayesian approach to Mendelian randomization with multiple pleiotropic variants. *Biostatistics* 2018, <https://doi.org/10.1093/biostatistics/kxy027>.

41. Burgess S and Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* 2017; **32**: 377–389.
42. Kang H, Zhang A, Cai TT, et al. Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *J Am Stat Assoc* 2016; **111**: 132–144.
43. Tibshirani R. Regression shrinkage and selection via the Lasso. *J Royal Stat Soc. Series B* 1996; **58**: 267–288.
44. Bishop C. *Pattern recognition and machine learning*. Information Science and Statistics. New York: Springer-Verlag, 2006.
45. Burgess S, Foley CN and Zuber V. Inferring causal relationships between risk factors and outcomes from genome-wide association study data. *Annu Rev Genomics Hum Genet* 2018; **19**: 303–327.
46. Thompson JR, Minelli C, Bowden J, et al. Mendelian randomization incorporating uncertainty about pleiotropy. *Stat Med* 2017; **36**: 4627–4645.
47. Aten JE, Fuller TF, Lusic AJ, et al. Using genetic markers to orient the edges in quantitative trait networks: the NEO software. *BMC Syst Biol* 2008; **2**: 34.
48. Flegal KM, Graubard BI, Williamson DF, et al. Reverse causation and illness-related weight loss in observational studies of body weight and mortality. *Am J Epidemiol* 2011; **173**: 1–9.
49. Agakov FV, McKeigue P, Krohn J, et al. Sparse instrumental variables (SPIV) for genome-wide studies. In: Lafferty JD, Williams CKI, Shawe-Taylor J, et al. (eds) *Advances in neural information processing systems*. Volume 23. USA: Curran Associates, Inc., 2010, pp.28–36.
50. Zgaga L, Agakov F, Theodoratou E, et al. Model selection approach suggests causal association between 25-hydroxyvitamin D and colorectal cancer. *PLoS One* 2013; **8**: e63475.
51. Steiger JH. Tests for comparing elements of a correlation matrix. *Psychol Bull* 1980; **87**: 245–251.
52. Hemani G, Bowden J and Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet* 2018; **27**: 195–208.
53. Dawid AP. Conditional independence in statistical theory. *J R Stat Soc Ser B Stat Methodol* 1979; **41**: 1–31.
54. Verweij KJH, Treur JL and Vink JM. Investigating causal associations between use of nicotine, alcohol, caffeine and cannabis: a two-sample bidirectional Mendelian randomization study. *Addiction* 2018; **113**: 1333–1338.
55. Peters J, Janzing D and Schölkopf B. *Elements of causal inference: foundations and learning algorithms*. Cambridge: MIT Press, 2017.
56. Jones EM, Thompson JR, Didelez V, et al. On the choice of parameterisation and priors for the Bayesian analyses of Mendelian randomisation studies. *Stat Med* 2012; **31**: 1483–1501.
57. Verbanck M, Chen CY, Neale B, et al. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018; **50**: 693–698.
58. Zhu Z, Zheng Z, Zhang F, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 2018; **9**: 224.
59. Ishwaran H and Rao JS. Spike and slab variable selection: frequentist and Bayesian strategies. *Ann Stat* 2005; **33**: 730–773.
60. George EI and McCulloch RE. Variable selection via Gibbs sampling. *J Am Stat Assoc* 1993; **88**: 881–889.
61. Bucur IG, Claassen T and Heskes T. Robust causal estimation in the large-sample limit without strict faithfulness. In: Singh A and Zhu J (eds) *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*. Fort Lauderdale, FL, USA: PMLR, 2017, pp.1523–1531.
62. Zhao Q, Chen Y, Wang J, et al. Powerful genome-wide design and robust statistical inference in two-sample summary-data Mendelian randomization. 2018, <https://ui.adsabs.harvard.edu/abs/2018arXiv180407371Z>.
63. Carvalho CM, Polson NG and Scott JG. Handling sparsity via the horseshoe. In: van Dyk D, Welling M (eds) *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*. Florida USA: PMLR, 2009, pp.73–80.
64. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal* 2006; **1**: 515–534.
65. Skilling J. Nested sampling for general Bayesian computation. *Bayesian Anal* 2006; **1**: 833–859.
66. Handley WJ, Hobson MP and Lasenby AN. Polychord: nested sampling for cosmology. *Mon Not R Astron Soc Lett* 2015; **450**: L61–L65.
67. Handley WJ, Hobson MP and Lasenby AN. Polychord: next-generation nested sampling. *Mon Not R Astron Soc* 2015; **453**: 4384–4398.
68. Bowden J, Del Greco FM, Minelli C, et al. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* 2017; **36**: 1783–1802.
69. Burgess S, Davies NM and Thompson SG. Instrumental variable analysis with a nonlinear exposure–outcome relationship. *Epidemiology* 2014; **25**: 877–885.
70. Shimizu S and Bollen K. Bayesian estimation of causal direction in acyclic structural equation models with individual-specific confounder variables and non-Gaussian distributions. *J Mach Learn Res* 2014; **15**: 2629–2652.
71. Del Greco FM, Minelli C, Sheehan NA, et al. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat Med* 2015; **34**: 2926–2940.
72. Daly B, Scragg R, Schaff D, et al. Low birth weight and cardiovascular risk factors in Auckland adolescents: a retrospective cohort study, [http://www.nzma.org.nz/\\_\\_data/assets/pdf\\_file/0003/17931/Vol-118-No-1220-12-August-2005.pdf](http://www.nzma.org.nz/__data/assets/pdf_file/0003/17931/Vol-118-No-1220-12-August-2005.pdf).

73. Thompson JR, Minelli C and Del Greco FM. Mendelian randomization using public data from genetic consortia. *Int J Biostat* 2016; **12**. <https://doi.org/10.1515/ijb-2015-0074>
74. Norris SA, Osmond C, Gigante D, et al. Size at birth, weight gain in infancy and childhood, and adult diabetes risk in five low- or middle-income country birth cohorts. *Diabetes Care* 2011; **35**: 72–79.
75. Noyce AJ, Kia DA, Hemani G, et al. Estimating the causal influence of body mass index on risk of parkinson disease: a Mendelian randomisation study. *PLOS Med* 2017; **14**: e1002314.
76. Wang YL, Wang YT, Li JF, et al. Body mass index and risk of Parkinson’s disease: a dose-response meta-analysis of prospective studies. *PLOS One* 2015; **10**: e0131778.
77. Bowden J, Spiller W, Del-Greco F, et al. Improving the visualisation, interpretation and analysis of two-sample summary data Mendelian randomization via the radial plot and radial regression. *bioRxiv* 2017; p.200378.
78. Treur JL, Taylor AE, Ware JJ, et al. Smoking and caffeine consumption: a genetic analysis of their association. *Addict Biol* 2016; **22**: 1090–1102.
79. Bjørngaard JH, Nordestgaard AT, Taylor AE, et al. Heavier smoking increases coffee consumption: findings from a Mendelian randomization analysis. *Int J Epidemiol* 2017; **46**: 1958–1967.
80. Ware JJ, Tanner JA, Taylor AE, et al. Does coffee consumption impact on heaviness of smoking? *Addiction* 2017; **112**: 1842–1853.
81. Cornelis MC, Byrne EM, Esko T, et al. Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption. *Mol Psychiatry* 2015; **20**: 647–656.
82. Buchner J. Collaborative nested sampling: big data vs. complex physical models. 2017, <https://ui.adsabs.harvard.edu/abs/2017arXiv170704476B>.
83. Bowden J, Del Greco F, Minelli C, et al. Improving the accuracy of two-sample summary data Mendelian randomization: moving beyond the NOME assumption. *bioRxiv* 2018; p.159442.