

A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization

Jack Bowden,^{a,*†} Fabiola Del Greco M,^b Cosetta Minelli,^c
George Davey Smith,^a Nuala Sheehan^d and John Thompson^d

Mendelian randomization (MR) uses genetic data to probe questions of causality in epidemiological research, by invoking the Instrumental Variable (IV) assumptions. In recent years, it has become commonplace to attempt MR analyses by synthesising summary data estimates of genetic association gleaned from large and independent study populations. This is referred to as two-sample summary data MR. Unfortunately, due to the sheer number of variants that can be easily included into summary data MR analyses, it is increasingly likely that some do not meet the IV assumptions due to pleiotropy. There is a pressing need to develop methods that can both detect and correct for pleiotropy, in order to preserve the validity of the MR approach in this context. In this paper, we aim to clarify how established methods of meta-regression and random effects modelling from mainstream meta-analysis are being adapted to perform this task. Specifically, we focus on two contrasting approaches: the Inverse Variance Weighted (IVW) method which assumes in its simplest form that all genetic variants are valid IVs, and the method of MR-Egger regression that allows all variants to violate the IV assumptions, albeit in a specific way. We investigate the ability of two popular random effects models to provide robustness to pleiotropy under the IVW approach, and propose statistics to quantify the relative goodness-of-fit of the IVW approach over MR-Egger regression. © 2017 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd

Keywords: instrumental variables; Mendelian randomization; meta-analysis; MR-Egger regression; pleiotropy

1. Introduction

The fundamental aim of Epidemiology is to determine the root causes of illness, with the focus of many epidemiological analyses being to examine whether an environmental exposure modifies the severity, or the risk of, disease. Of course, it is well known that causal conclusions cannot strictly be drawn from mere statistical associations between an exposure and outcome, unless all possible confounders of the association are identified, perfectly measured and appropriately adjusted for. Mendelian randomization (MR) [1, 2] offers an alternative way to probe the issue of causality in epidemiological research, by using additional genetic information satisfying the Instrumental Variable (IV) assumptions. Briefly, the genetic data must both predict the exposure and predict the outcome only through the exposure. Figure 1 shows an illustrative causal diagram relating a single nucleotide polymorphism (SNP) G_j to an exposure, X and outcome, Y , in the presence of an unmeasured confounder, U . Here, a single U is used to denote the combined influence of all unmeasured confounders.

To state the IV assumptions more formally: the j th SNP must be: associated with X (IV1); independent of U (IV2) and independent of Y given X and U (IV3). These assumptions are encoded by the arrows in Figure 1. If IV1–IV3 are satisfied for variant G_j , then traditional IV methods can be employed to reliably test for a causal effect using G_j , X and Y alone, without any attempt to adjust for U at all.

^aMRC Integrative Epidemiology Unit, University of Bristol, U.K.

^bCenter for Biomedicine, EURAC research, Bolzano, Italy

^cPopulation Health and Occupational Disease, NHLI, Imperial College, London, U.K.

^dDepartment of Health Sciences, University of Leicester, Leicester, U.K.

*Correspondence to: Jack Bowden, MRC Integrative Epidemiology Unit, Oakfield House, BS8 2BN.

†E-mail: jack.bowden@bristol.ac.uk

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

An array of sophisticated techniques exist for estimating the causal effect with individual participant data. However, the sharing of such data is often impractical and in recent years, it has become much more common to attempt MR analyses using summary data estimates of SNP-exposure and SNP-outcome associations based on large and independent populations [3,4]. This is referred to as two-sample summary data MR [5, 6]. The SNPs in question are usually those identified as ‘hits’ in separate genome wide association studies (GWAS). They are generally picked from distinct genomic regions in order to be mutually independent, or not in linkage disequilibrium (LD). The veracity of this assumption can be further scrutinised by using external data resources that catalogue LD structure across the genome (see for example <http://www.internationalgenome.org/>). In many cases, this restriction still allows hundreds of SNPs to be included in the analysis. If all of the incorporated genetic variables provide independent and unbiased estimates of the same causal effect, then summary data MR can be reliably reduced to a fixed effect meta-analysis of the set of causal estimates obtained using each genetic variant in turn. Many assumptions are necessary for this to be true, however. For example, the SNPs must at the very least be uncorrelated and valid IVs. They must also collectively satisfy additional modelling restrictions with respect to the exposure and outcome, which will be discussed in detail in Section 2. Unfortunately, due to the sheer number of variants that can now be easily included in such MR analyses (often with limited knowledge of their functional role), it is increasingly likely that some fail to meet the fundamental assumption of being a valid IV, due to pleiotropy.

Pleiotropy occurs when a single SNP affects multiple phenotypes related to the outcome [2, 7], and is a potential barrier to obtaining reliable inferences via MR. Crucially, however, only pleiotropy of a certain type poses a threat. For example, in Figure 1 G_j is a valid IV, but when X causally affects Y , G_j will be associated with both X and Y (through X). Indeed, MR exploits this very fact to test for causality. Pleiotropy is problematic when a SNP affects Y through phenotypic pathways other than X , because this could invalidate its use as an IV. This is illustrated by the addition of dashed lines in Figure 2. For example the dashed line linking variant G_j to U would make G_j invalid due to violation of IV2, and the dashed line linking variant G_j to Y would make G_j invalid due to violation of IV3. From now on, we use the term ‘pleiotropy’ to refer to IV violations of this sort.

There is a pressing need to develop methods that can both detect and correct for pleiotropy in order to preserve the validity of the MR approach. Indeed, some progress has already been made towards this goal [8–11]. The aim of this paper is to clarify how established methods of meta-regression and random effects modelling from mainstream meta-analysis have been adapted to perform this task. Specifically, we focus on two contrasting approaches: The Inverse Variance Weighted (IVW) method [6, 12] which assumes in its most basic form that IV1–3 hold (e.g. no pleiotropy) and is equivalent to a fixed effect meta-analysis; and secondly a simple adaptation to the MR context of the method of Egger regression [13] used to adjust for small study bias in meta-analysis (referred to as MR-Egger [9]). In doing so, we attempt to make use

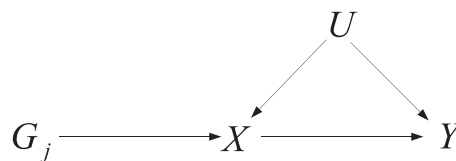


Figure 1. Illustrative diagram showing the hypothesised causal relationship between a genetic variable G_j , environmental exposure X and outcome Y .

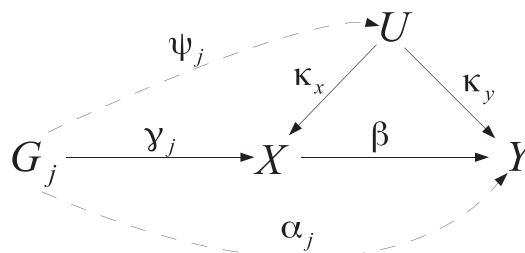


Figure 2. Illustrative diagram showing the causal and parametric relationships between genetic variable G_j , X and Y . Solid lines on their own define G_j as a valid IV, but the addition of dashed lines indicate violations of the IV assumptions.

of, extend and connect the work of Rücker et al. [14] and Henmi and Copas [15] for probing the issue of small study bias in meta-analysis with the work of Del Greco et al [16] and Bowden et al [9, 17] for detecting and accounting for pleiotropy in summary data MR analysis.

In Section 2, we clarify the assumptions necessary for a two-sample summary data MR analysis, by starting from an underlying model for data at the individual level. In Sections 3 and 4, we explore how the IVW and MR-Egger methods can be applied to pleiotropy-affected summary data, by drawing on the meta-analytical framework of Rücker et al [14]. In Section 5, we look at the issue of estimating the causal effect in practice, when some of the crucial assumptions are violated. In Section 6, we illustrate the methods discussed using an application that probes the causal effect of plasma urate levels on cardiovascular disease risk. We conclude with a discussion and point to future research.

2. Modelling assumptions

We start by writing the assumed data generating model for subject i out of N , linking genetic data on L uncorrelated variants $G_{i1}, \dots, G_{iL} \equiv G_i$, to their exposure X_i and outcome Y_i in the presence of an unobserved summary confounder U_i , as represented in Figure 2. X , Y and U are assumed to be continuous and G_{ij} represents the number of alleles of SNP j assigned to subject i (0, 1 or 2) at the point of conception. For simplicity, we ignore intercept terms in the models below:

$$U_i | G_i = \sum_{j=1}^L \psi_j G_{ij} + \epsilon_{ui} \tag{1}$$

$$X_i | U_i, G_i = \sum_{j=1}^L \gamma_j G_{ij} + \kappa_x U_i + \epsilon_{xi} \tag{2}$$

$$Y_i | X_i, U_i, G_i = \beta X_i + \kappa_y U_i + \sum_{j=1}^L \alpha_j G_{ij} + \epsilon_{yi} \tag{3}$$

Here, β represents the causal effect of X on Y that we wish to estimate and γ_j represents the underlying true association between SNP j and the exposure X (i.e. not through U). We assume that the genetic data has been coded to reflect the number of exposure increasing alleles (be they ‘minor’ or ‘major’), so that $\gamma_j \geq 0$. We define IV1 to be satisfied for variant j when γ_j is non-zero. Conversely, when they are non-zero, the parameters ψ_j and α_j summarise all violations of IV2 and IV3 for SNP j , respectively. κ_x and κ_y represent the influence of all confounders on the exposure and outcome, respectively. Only when κ_x and κ_y are both non-zero can confounding take effect, thus biasing a traditional observational analysis so that the association between X and Y is not equal in general to β (hence the need for IV methods). In the main body of this paper, we will interpret all parameters indexing these (and subsequent) models as fixed values in the ‘classical’ sense. Alternative Bayesian formulations are of course possible, a topic we return to in the discussion.

The focus of this paper is to consider the case where only summary data estimates of the SNP-exposure and SNP-outcome associations and their standard errors are available for each of the L variants. In order to derive models for these two sets of L summary data estimates, we firstly define G_{-j} as the $(L-1)$ -length vector of genetic variables without G_j . We now re-write models (2) and (3) as models for $X|G_j$ and $Y|G_j$ only, by marginalising over (U, G_{-j}) and (X, U, G_{-j}) , respectively:

$$X_i | G_{ij} = (\gamma_j + \kappa_x \psi_j) G_{ij} + \epsilon'_{xij}$$

where $\epsilon'_{xij} = \sum_{l \neq j} (\gamma_l + \kappa_x \psi_{il}) G_{il} + \kappa_x \epsilon_{ui} + \epsilon_{xi}$ (4)

$$Y_i | G_{ij} = \{ \beta (\gamma_j + \kappa_x \psi_j) + \kappa_y \psi_j + \alpha_j \} G_{ij} + \epsilon'_{yij}$$

where $\epsilon'_{yij} = \beta \epsilon'_{xij} + \sum_{l \neq j} (\kappa_y \psi_l + \alpha_l) G_{il} + \kappa_y \epsilon_{ui} + \epsilon_{yi}$ (5)

Now the coefficients of G_{ij} in (4) and (5) represent the underlying quantities that would be estimated in principle with summary data for SNP j under our assumed model. Note that all other terms have been subsumed into new residual error terms ϵ'_{xij} and ϵ'_{yij} .

2.1. Two-sample Mendelian randomization

It is becoming increasingly common to perform summary data MR analyses using sets of genetic association estimates with the exposure and outcome that have been gleaned from separate samples. The most natural way of justifying this approach is to assume that the SNP-exposure and SNP-outcome samples, whilst being independent, are homogeneous in the sense that models (4) and (5) hold for both. We will refer to them as sample ‘1’ and ‘2’, with N_1 and N_2 subjects, respectively. For clarity, we now write explicit models for the two sets of L SNP-exposure and SNP-outcome association estimates from the two samples. We refer to the j th SNP-exposure association estimate as $\hat{\gamma}_j$ (with variance σ_{Xj}^2) and the j th SNP-outcome association estimate as $\hat{\Gamma}_j$ (with variance σ_{Yj}^2):

$$\text{Sample 1: } \hat{\gamma}_j = \gamma_j + k_x \psi_j + \epsilon_{Xj}, \quad \text{var}(\epsilon_{Xj}) = \sigma_{Xj}^2 \tag{6}$$

$$\text{Sample 2: } \hat{\Gamma}_j = \alpha_j + k_y \psi_j + \beta(\gamma_j + k_x \psi_j) + \epsilon_{Yj} \quad \text{var}(\epsilon_{Yj}) = \sigma_{Yj}^2 \tag{7}$$

From (4) and (5), the magnitude of σ_{Xj}^2 and σ_{Yj}^2 are affected by the allele frequencies of SNPs in G_{-j} and the sample size (N_1 or N_2). Because they relate to homogeneous but independent samples, ϵ_{Xj} and ϵ_{Yj} can be considered as mutually independent. Finally, although they must be estimated from the data, it is common practice to assume that the variances σ_{Xj}^2 and σ_{Yj}^2 are known. We will refer to the various assumptions and simplifications made in Section 2.1 as the ‘Two-Sample Assumptions’ (TSA).

We now state some additional assumptions that are commonly utilised by the IVW or MR-Egger approaches for causal inference (*vis* estimation of β), based on models (6) and (7) and the TSA. All of these assumptions are re-stated for clarity in Table I.

2.2. Additional assumptions

2.2.1. *The NOME assumption.* A pragmatic assumption from the MR literature is to assume σ_{Xj}^2 is small enough to be treated as zero. This would be strictly true if N_1 were infinite, but is frequently reasonable because large study sizes (often over 100 000) are now common and rising year-on-year. This is referred to as the NO Measurement Error (NOME) assumption in Bowden et al [17].

2.2.2. *The InSIDE assumption.* If assumptions IV2 and IV3 are satisfied for variant j ($\alpha_j = \psi_j = 0$), then the total pleiotropic effect of variant j on the outcome is also zero. However, their violation does not necessarily preclude valid causal inference: consistent estimation of the causal effect is still possible if the

Table I. Assumptions used in two-sample MR analyses.	
Assumption	Description
IV assumptions	
IV1	$\gamma_j > 0$: SNP predicts exposure (not through U).
IV2	$\psi_j = 0$: No SNP-confounder association.
IV3	$\alpha_j = 0$: No residual SNP-outcome association after controlling for X and U .
TSA	
TSA1	Models (6) and (7) hold for population 1 and 2.
TSA2	ϵ_{Xj} in (6) and ϵ_{Yj} (7) independent of each other and any other terms.
TSA3	σ_{Xj}^2 and σ_{Yj}^2 in (6) and (7) known.
NOME assumption	
	$\sigma_{Xj}^2 = 0$: Negligible uncertainty in SNP-exposure association.
InSIDE assumption (under IV2)	
General InSIDE	Sample covariance $\text{cov}(\alpha_j, \gamma_j) \rightarrow 0$ as $L \rightarrow \infty$.
Perfect InSIDE	Sample covariance $\text{cov}(\alpha_j, \gamma_j) = 0$ for data at hand.
VIS assumption (under IV2)	
	$\gamma_j \neq \gamma_i$ for some i and j not equal to i Some variation in instrument strengths present.

magnitude of the pleiotropy in equation (7), $\alpha_j + k_x \psi_j$, is independent of the SNP-exposure associations, $\gamma_j + k_x \psi_j$. This was first identified as a crucial assumption for causal inference in the econometrics literature by Kolesar et al [18] and was independently derived for use in MR by Bowden et al [9], who termed it the InSIDE assumption (Instrument Strength Independent of Direct Effect). Because ψ_j is a common factor of both the instrument strength and pleiotropy terms, by far the most natural way to imagine that InSIDE could hold (even in principle) is if IV2 holds ($\psi_j = 0$) and the magnitude of the pleiotropy not via $U(\alpha_j)$ is independent of γ_j across the instruments. We will therefore adopt this convention when referring to InSIDE from now on.

If the InSIDE assumption is satisfied then the sample covariance of the α_j s and γ_j s will tend to zero as the number of variants, L , increases. We will refer to this as ‘general’ InSIDE. However, if the sample covariance of the α_j s and γ_j s is *exactly* zero for the data at hand, we will say that InSIDE is ‘perfectly’ satisfied.

2.2.3. The VIS assumption. If the SNP-exposure association parameter estimates $\hat{\gamma}_1, \dots, \hat{\gamma}_L$ were all identical, then their sample variance would be zero. If this were the case, it would rule out using the $\hat{\gamma}_j$ s as an explanatory variable in any regression model (specifically MR-Egger regression). We therefore define the VIS (Variation in Instrument Strength) assumption as being satisfied when the aforementioned sample variance of the true parameter values $\gamma_1, \dots, \gamma_L$ is non-zero.

2.2.4. NOME, InSIDE and VIS under a weighted analysis. In practice when implementing the IVW and MR-Egger regression approaches, the analysis has traditionally been weighted by the standard error of the SNP-outcome association estimates, σ_{Y_j} , to improve efficiency. For consistency, this is the approach we will subsequently follow. In order to define the NOME, InSIDE and VIS assumptions within the context of this weighted analysis (or any other weighted analysis), it is necessary to divide the relevant parameter by the chosen weight. For example, for VIS to be satisfied in a analysis weighted by σ_{Y_j} , we require some variation between the weighted instrument strength terms γ_j/σ_{Y_j} . Similar transparent modifications would also be required for NOME and InSIDE.

3. Mendelian randomization under violations of IV3

In this section, we discuss the motivation for using the meta-analytical methods of IVW and MR-Egger regression, when all the assumptions outlined in Table I are satisfied except IV3. We will therefore use the following simplified versions of models (6) and (7) as our starting point:

$$\text{Data generating model under: IV1-2, TSA, NOME, InSIDE and VIS} \tag{8}$$

$$\text{SNP-exposure: } \hat{\gamma}_j = \gamma_j \tag{8}$$

$$\text{SNP-outcome: } \hat{\Gamma}_j = \alpha_j + \beta \gamma_j + \epsilon_{Y_j}, \quad \text{var}(\epsilon_{Y_j}) = \sigma_{Y_j}^2 \tag{9}$$

In order to easily discuss violations of IV3, we define the sample mean and variance of the α terms as μ_α and σ_α^2 , respectively. Specifically, we shall investigate four important special cases of model (9) that are defined and differentiated by the chosen values of μ_α and σ_α^2 . Under model (9), the α_j term is the sole source of additional variation in the SNP-outcome association estimates (given the SNP-exposure estimates), apart from the residual error term ϵ_{Y_j} . There may of course be factors inducing additional heterogeneity in the SNP-outcome associations other than pleiotropy, so in practice we can think of α_j as representing the combination of all these sources. For simplicity, we will continue to refer to this as pleiotropy.

We start by assuming that IV3 actually holds for all variants in model (9), which is consistent with $\alpha_j = \mu_\alpha = \sigma_\alpha^2 = 0$. We call this case (a) and it is the natural starting point for an MR analysis. In this case, model (9) can be simplified to the following linear model with no intercept:

$$\frac{\hat{\Gamma}_j}{\sigma_{Y_j}} = \frac{\beta \gamma_j}{\sigma_{Y_j}} + \epsilon_j, \quad \text{var}(\epsilon_j) = 1. \tag{10}$$

Both sides of the equation are divided by σ_{Y_j} in order to make the variance homoscedastic and clarify the mechanics of model fitting under a weighted analysis. An overall estimate can be calculated from model (10) using standard regression theory, which yields the well known meta-analytic formulae:

$$\hat{\beta}_{IVW} = \frac{\sum_{j=1}^L w_j \hat{\beta}_j}{\sum_{j=1}^L w_j}, \quad w_j = 1/\text{var}(\hat{\beta}_j) = \gamma_j^2 / \sigma_{Y_j}^2. \quad (11)$$

In the MR context, this is referred to as the IVW estimate [6, 12]. Formula (11) can alternatively be derived by calculating a weighted average of the set of ratio estimates of causal effect, $\hat{\beta}_j = \hat{\Gamma}_j / \gamma_j$, obtained using each SNP in turn, where the weight given to $\hat{\beta}_j$ is the inverse of its variance, $\sigma_{Y_j}^2 / \gamma_j^2$, as in a fixed effect meta-analysis.

It is precisely assumptions IV2 and IV3 that dictate the intercept in (10) should be constrained to zero. The NOME assumption dictates the simple form for $\text{var}(\hat{\beta}_j)$ because this means that the denominator of $\hat{\beta}_j$ can be treated as a constant (even though in truth it is the random variable $\hat{\gamma}_j$).

3.1. Testing and accounting for balanced pleiotropy

The variance of the IVW estimate under case (a) (the fixed effect model), is given by:

$$\text{Var}(\hat{\beta}_{IVW}) = \frac{1}{\sum_{j=1}^L w_j}. \quad (12)$$

The value of 1 the numerator in (12) is a direct consequence of the fact that the variance of the residual error in model (10) is equal to 1. As suggested by Del Greco et al [16], the data can be scrutinised to assess this assumption by calculating Cochran's Q statistic with respect to the L IV estimates as follows:

$$Q = \sum_{j=1}^L \left(\frac{\hat{\Gamma}_j}{\sigma_{Y_j}} - \frac{\hat{\beta}_{IVW} \gamma_j}{\sigma_{Y_j}} \right)^2 = \sum_{j=1}^L w_j (\hat{\beta}_j - \hat{\beta}_{IVW})^2. \quad (13)$$

If true, Q follows a χ^2 distribution on $L-1$ degrees of freedom. If the L estimates exhibit over-dispersion or heterogeneity (as will often happen) then, under model (9), this must be due to pleiotropy and an extension to the basic model is required. The most natural first extension, we argue, would be to allow for 'balanced' pleiotropy. That is, we relax IV3 to allow non-zero α_j s but assume they have zero mean (e.g. $\mu_\alpha = 0$, $\sigma_\alpha^2 > 0$). We call this case (b) and re-write model (10) as:

$$\frac{\hat{\Gamma}_j}{\sqrt{\sigma_{Y_j}^2 + \sigma_\alpha^2}} = \frac{\beta \gamma_j}{\sqrt{\sigma_{Y_j}^2 + \sigma_\alpha^2}} + \epsilon_j, \quad \text{var}(\epsilon_j) = 1. \quad (14)$$

Model (14) is justified for any L under Perfect InSIDE, because this allows pleiotropy to be treated as independent residual error with respect to the explanatory variable γ_j . It is only strictly valid under General InSIDE as $L \rightarrow \infty$. Fitting model (14) is equivalent to performing a standard additive random effects meta-analysis [19, 20]. That is, we can obtain a new IVW point estimate and variance for the causal effect under model (14) by applying formulae (11) and (12) and replacing $w_j = \gamma_j^2 / \sigma_{Y_j}^2$ with $w_j^* = \gamma_j^2 / (\sigma_{Y_j}^2 + \sigma_\alpha^2)$ using a plug-in estimate for σ_α^2 (for example the DerSimonian and Laird estimate [19]). We will refer to the IVW estimate obtained by fitting model (14) as $\hat{\beta}_{IVW}^{ARE}$. The two IVW estimates $\hat{\beta}_{IVW}$ and $\hat{\beta}_{IVW}^{ARE}$ do differ, because under the latter the variance component σ_α^2 is used to re-weight the contribution of the j th IV estimate. The variance of $\hat{\beta}_{IVW}^{ARE}$ is additionally inflated via w_j^* to account for heterogeneity due to pleiotropy. When a large value of σ_α^2 is estimated from the data, this will have the effect of evening out the weight given to each ratio estimate in the analysis by making them a less direct function of each estimate's precision.

3.1.1. Multiplicative random effects models. The following multiplicative random effects model [21, 22] could instead be applied to the summary data estimates in the presence of observed heterogeneity:

$$\frac{\hat{\Gamma}_j}{\sigma_{Y_j}} = \frac{\beta \gamma_j}{\sigma_{Y_j}} + \frac{\phi^{1/2}}{\sigma_{Y_j}} \epsilon_j, \quad \text{var}(\epsilon_j) = 1. \quad (15)$$

If adopted, model (15) returns the same point estimate for β as under the fixed effect IVW model, $\hat{\beta}_{IVW}$. That is, the variance component σ_α^2 is not allowed to influence its point estimate. However, the addition of

the scale parameter $\phi^{1/2}$ allows the variance of $\hat{\beta}_{IVW}$ to increase when heterogeneity is detected. Specifically, ϕ is estimated from the variance of the residual error using $\hat{\phi}_{IVW} = \frac{Q}{L-1}$ where Q is defined in (13). The variance of the IVW estimate under model (15) is then set to

$$\text{var}(\hat{\beta}_{IVW}) = \frac{\hat{\phi}_{IVW}}{\sum_{j=1}^L w_j},$$

so that it is $\hat{\phi}_{IVW}$ times the variance under the fixed effect model - case (a).

If the additive pleiotropy model (14) is actually correct then the residual error of the multiplicative pleiotropy model (15) is miss-specified, because the constant term $\phi^{1/2}$ should be replaced with $\sqrt{1 + \frac{\sigma_\alpha^2}{\sigma_{Yj}^2}}$, which will usually result in a small loss of efficiency. One reason to prefer the multiplicative model in the general meta-analysis context, is that $\hat{\beta}_{IVW}$ is known to be more robust to small study bias than $\hat{\beta}_{IVW}^{ARE}$ [15]. We will subsequently show that the same principle holds true in the MR context, when ‘directional’ pleiotropy is present in the data, a notion to which we now turn our attention.

3.2. Testing and accounting for directional pleiotropy

‘Directional’ pleiotropy occurs when the mean value of the pleiotropy distribution, μ_α , is non-zero [9]. We will show how this can be addressed by performing MR-Egger regression. We first consider the special case where μ_α is non-zero but $\sigma_\alpha^2 = 0$. This means that $\alpha_j = \mu_\alpha$ for all j in model (7). Although this appears highly unlikely, it is actually what MR-Egger regression (as proposed in [9]) assumes in its simplest form. We call this case (c), which can be written as

$$\frac{\hat{\Gamma}_j}{\sigma_{Yj}} = \frac{\mu_\alpha}{\sigma_{Yj}} + \beta \frac{\gamma_j}{\sigma_{Yj}} + \epsilon_j, \quad \text{var}(\epsilon_j) = 1 \tag{16}$$

MR-Egger regression fits model (16) under case (c) to give an estimate for the intercept μ_α and causal effect β . In order to distinguish them from the IVW estimate, we will refer to the MR-Egger estimate for the causal effect β as $\hat{\beta}_{1E}$ and the mean pleiotropic effect μ_α as $\hat{\beta}_{0E}$, where

$$\begin{pmatrix} \hat{\beta}_{0E} \\ \hat{\beta}_{1E} \end{pmatrix} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\mathbf{\Gamma}}, \quad \text{for } \mathbf{Z} = \begin{pmatrix} \sigma_{Y1}^{-1} & \gamma_1 \sigma_{Y1}^{-1} \\ \sigma_{Y2}^{-1} & \gamma_2 \sigma_{Y2}^{-1} \\ \vdots & \vdots \\ \vdots & \vdots \\ \sigma_{YL}^{-1} & \gamma_L \sigma_{YL}^{-1} \end{pmatrix}, \quad \hat{\mathbf{\Gamma}} = \begin{pmatrix} \hat{\Gamma}_1 \sigma_{Y1}^{-1} \\ \hat{\Gamma}_2 \sigma_{Y2}^{-1} \\ \vdots \\ \hat{\Gamma}_L \sigma_{YL}^{-1} \end{pmatrix}.$$

The variance of the MR-Egger estimate is equal to the lower diagonal element of $(\mathbf{Z}^T \mathbf{Z})^{-1}$ but when $\sigma_{Y1}, \dots, \sigma_{YL}$ are all exactly equal, this reduces to

$$\text{var}(\hat{\beta}_{1E}) = \frac{1}{\sum_{j=1}^L \left(\frac{\hat{\gamma}_j - \bar{\gamma}}{\sigma_{Yj}} \right)^2}. \tag{17}$$

This simple formula indicates that the variance of $\hat{\beta}_{1E}$ is essentially dictated by the variance of the SNP-exposure associations. It also reinforces the importance of the VIS assumption for MR-Egger: its causal estimate is simply undefined when no such variation exists.

Case (c) allows for directional pleiotropy but assumes that the pleiotropic effect is the same across all variants. This means that pleiotropy induces bias but no additional heterogeneity, which implies that the residual error about model (16) equals 1. We can test the plausibility of this assumption via a simple adaptation of the Q' statistic due to Rücker et al [14] to our context:

$$Q' = \sum_{j=1}^L \frac{1}{\sigma_{Yj}^2} \left\{ \hat{\Gamma}_j - (\hat{\beta}_{0E} + \hat{\beta}_{1E} \gamma_j) \right\}^2 = \sum_{j=1}^L w_j \left(\hat{\beta}_j - \frac{\hat{\beta}_{0E}}{\gamma_j} - \hat{\beta}_{1E} \right)^2, \tag{18}$$

where $w_j = \frac{\gamma_j^2}{\sigma_{Yj}^2}$ as before. If case (c) is correct then Q' should follow a χ^2 distribution with $L-2$ degrees of freedom. If a non-zero intercept $\hat{\beta}_{0E}$ is estimated and the Q' statistic is significantly larger than $L-2$, a more reasonable model might be that directional pleiotropy is inducing bias **and** additional heterogeneity into the MR summary data estimates ($\mu_\alpha \neq 0, \sigma_\alpha^2 > 0$). We call this case (d) and re-write model (16) accordingly as an additive random effect model [14]:

$$\frac{\hat{\Gamma}_j}{\sqrt{\sigma_{Yj}^2 + \sigma_\alpha^2}} = \frac{\mu_\alpha}{\sqrt{\sigma_{Yj}^2 + \sigma_\alpha^2}} + \frac{\beta\gamma_j}{\sqrt{\sigma_{Yj}^2 + \sigma_\alpha^2}} + \epsilon_j, \quad \text{var}(\epsilon_j) = 1. \quad (19)$$

As for case (b), model (19) is valid for any L under perfect InSIDE and as $L \rightarrow \infty$ under General InSIDE. Updated estimates $\hat{\beta}_{0E}^{ARE}$ and $\hat{\beta}_{1E}^{ARE}$ could be obtained by changing the definition of the design matrix Z so that the j th row is equal to

$$\left(\frac{1}{\sqrt{\sigma_{Yj}^2 + \sigma_\alpha^2}}, \frac{\gamma_j}{\sqrt{\sigma_{Yj}^2 + \sigma_\alpha^2}} \right)$$

and where σ_α^2 is substituted with a suitable estimate. For the reasons already outlined in the case of the IVW model, we have so far preferred not to use the additive pleiotropy model (19) in practice to fit MR-Egger regression. We have instead opted to preserve the original fixed effect MR-Egger point estimates, and to scale up their variance if heterogeneity is detected, by fitting the multiplicative model:

$$\frac{\hat{\Gamma}_j}{\sigma_{Yj}} = \frac{\mu_\alpha}{\sigma_{Yj}} + \beta \frac{\gamma_j}{\sigma_{Yj}} + \frac{\phi^{1/2}}{\sigma_{Yj}} \epsilon_j, \quad \text{var}(\epsilon_j) = 1. \quad (20)$$

Under model (20), additional heterogeneity can be taken into account by estimating ϕ via $\hat{\phi}_E = Q' / (L-2)$ and scaling up the variance of the MR-Egger estimates to be $\hat{\phi}_E$ times their value under case (c).

3.3. When will the IVW and MR-Egger methods be unbiased?

Table II attempts to clarify the additional assumptions that are required for unbiased estimation of β by the IVW and MR-Egger methods under cases (a) to (d). We use the word ‘additional’ to distinguish them from baseline assumptions (which we define as TSA, NOME and IV2) as well as those that are implied by each case (such as IV3 in case (a)). If a specific case automatically invalidates a method, so that unbiased estimation is not possible in general, we mark it with an ‘x’.

Table II. Requirements for unbiased estimation of β by the IVW and MR-Egger regression approaches for special cases (a) to (d).

Case (model) baseline & implied assumptions	Pleiotropy distribution		Unbiased estimation of β possible under additional assumptions?	
	μ_α	σ_α^2	IVW	MR-Egger
(a): No pleiotropy (Fixed effect IVW) TSA,NOME,IV2 IV3, Perfect InSIDE	0	0	IV1	VIS
(b): Balanced pleiotropy (Random effect IVW) TSA,NOME,IV2	0	> 0	IV1, General InSIDE	VIS, General InSIDE
(c): Directional pleiotropy (Fixed effect MR-Egger) TSA,NOME,IV2 Perfect InSIDE	$\neq 0$	0	x	VIS
(d): Directional pleiotropy (Random effect MR-Egger) TSA,NOME, IV2	$\neq 0$	> 0	x	VIS, General InSIDE

Under case (a) (no pleiotropy), assumptions IV2 and IV3 hold which means that Perfect InSIDE is satisfied because $\text{cov}(0, \gamma_j) \equiv 0$. The IVW method therefore only requires IV1 to hold and MR-Egger only requires VIS to hold in order to be unbiased.

Under case (b) (balanced pleiotropy), both methods are consistent (asymptotically unbiased as $L \rightarrow \infty$) under the General InSIDE assumption because heterogeneous pleiotropy contributes to the residual error. Under cases (c) and (d) (directional pleiotropy), the IVW estimate cannot consistently estimate β in general because it does not adjust for a non-zero mean pleiotropic effect. Under case (c), Perfect InSIDE is satisfied because $\text{cov}(\mu_\alpha, \gamma_j) = 0$. The only additional assumption required by MR-Egger for unbiased estimation is VIS. Under case (d), MR-Egger requires both VIS and General InSIDE to be consistent.

4. Can we navigate between models using Rücker's framework?

Moving from the underlying model in case (a) to those in (b), (c) and finally (d) represents a natural way to progressively relax assumption IV3 using heterogeneity statistics when performing MR. Following the approach of Rücker et al in the general meta-analysis context, we would start by assuming all genetic variants are valid - case (a) - but move to case (b) if Cochran's Q is sufficiently extreme with respect to a χ^2_{L-1} distribution, in order to allow for balanced pleiotropy. We would then assess whether directional pleiotropy is present by fitting model (c) and calculate the heterogeneity about this model via Rücker's Q' . If the difference $Q - Q'$ is sufficiently extreme with respect to a χ^2_1 distribution, we would infer that directional pleiotropy is an important factor, and adopt model (c). Finally, we would assess whether Q' is sufficiently extreme with respect to a χ^2_{L-2} distribution. If so, we would opt for model (d) to account for the fact that pleiotropy is inducing heterogeneity and bias.

Unfortunately, this framework does not seamlessly translate to the MR setting, because IVW model (10) is not strictly nested within MR-Egger model (16). This means that $Q - Q'$ is only approximately χ^2_1 distributed. Q' may in fact be larger than Q , making $Q - Q'$ negative. The equivalent IVW and Egger regression models assumed by Rücker et al, are nested so that the equivalent $Q - Q'$ difference has the required distribution.

Therefore, we can only informally use a $Q - Q'$ to decide whether MR-Egger regression represents a better fit than IVW. We therefore suggest the user considers moving from IVW model's (a) and (b) to MR-Egger model's (c) and (d) when $Q - Q'$ is large and positive, and the MR-Egger intercept parameter estimate is sufficiently far from zero and precise.

4.1. An illustrative example

Figure 3 (left) shows an illustrative scatter plot of SNP-outcome and SNP-exposure associations consistent with case (a) (solid black dots, no pleiotropy) and case (b) (hollow black dots, balanced pleiotropy). Under both cases (a) and (b) the points yield a regression line perfectly through the origin. This means that the IVW and MR-Egger causal estimates would be identical ($\hat{\beta}_{IVW} = \hat{\beta}_{1E}$). Under case (a), $\hat{\phi}_{IVW}$ and $\hat{\phi}_E$ would be approximately equal to each other and to 1, and under case (b) $\hat{\phi}_{IVW}$ and $\hat{\phi}_E$ would be approximately equal to each other with a value greater than 1 (2 in this instance). For the solid black data, Cochran's Q would provide no evidence to support a move from the basic model (a) and the IVW estimate. For the hollow black data, Cochran's Q would suggest sticking with the IVW estimate, but to allow

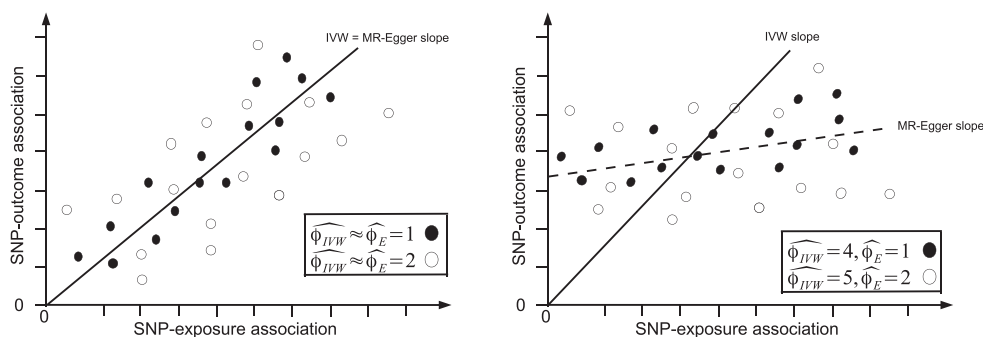


Figure 3. Scatter plot of SNP-outcome versus SNP-exposure estimates for four fictional MR analyses. Left: under cases (a) (solid black dots) and (b) (hollow black dots). Right: under cases (c) (solid black dots) and (d) (hollow black dots).

for over-dispersion due to balanced pleiotropy by moving to model (b). Because the MR-Egger estimate is identical to the IVW estimate, this means that $\hat{\beta}_{0E} = 0$ and $Q - Q' = 0$. Therefore, no evidence exists to support moving to model (c) and the MR-Egger estimate.

Figure 3 (right) shows an illustrative scatter-plot of SNP-outcome and SNP-exposure associations consistent with case (c) (solid black dots, homogeneous directional pleiotropy) and case (d) (hollow black dots, heterogeneous directional pleiotropy). Under both cases (c) and (d) the points are consistent with a regression line with the same non-zero intercept. Under case (c), $\hat{\phi}_E$ is equal to 1 whereas $\hat{\phi}_{IVW}$ is much larger (4 in this instance). Under case (d) $\hat{\phi}_E$ is equal to 2 whereas $\hat{\phi}_{IVW}$ is again larger (5 in this instance). Now, the solid black data would support a move from model (a) (and the IVW estimate) to model (c) (and the MR-Egger estimate) because Q , $\hat{\beta}_{0E}$ and $Q - Q'$ are 'large', but Q' on its own is not. The hollow black dots would additionally yield a large enough Q' in order to opt for model (d), thereby suggesting inference should be based on the MR-Egger estimate accounting for heterogeneity and bias due to pleiotropy.

As in the general meta-analysis context, heterogeneity statistics such as Q and Q' are only strictly valid under the assumption that the precision of SNP-exposure and SNP-outcome estimates are known, despite being estimated from the data. Likewise, their power to 'detect' heterogeneity will be small when the number of genetic variants, L , is small. This could mean that, even if model (d) were true, it may often not be possible to infer this by rejecting models (a), (b) and (c) at the historically popular 5% significance level. Furthermore, choosing a model (and by extension a causal estimate) that depends on the value of a statistical test (i.e 'testimation' [23, 24]) could also induce unwanted bias. Additional research is required to fully understand the operating characteristics of the schema suggested here in realistic data settings, especially the use of heterogeneity statistic estimators, before they can be confidently used in practice.

4.1.1. A measure of relative fit. In Figure 3 (left), we see strong evidence that the IVW estimate is a reliable measure of causal effect. Because it is also guaranteed to be more precise, there is no reason whatsoever to prefer the MR-Egger estimate. Conversely, in Figure 3 (right), we see strong evidence that the IV assumptions have been violated in a meaningful way and the MR-Egger approach has more appeal. We therefore seek a statistic that will favour the IVW approach *a priori* unless MR-Egger regression is a demonstratively better fit to the data. A simple statistic, Q_R encapsulates this reasoning, and is again closely connected to the work of Rücker et al [14]. It simply measures the relative value of the residual heterogeneity under both the IVW and MR-Egger methods:

$$Q_R = \left(\frac{Q'}{Q} \right) \tag{21}$$

Q_R will generally lie in the unit interval, but because the MR-Egger and IVW models are non-nested, in rare cases the inequality $0 \leq Q' \leq Q$ will not hold. In Figure 3 (left), we see that (under case (a) or (b)) Q_R is approximately equal to 1, which tells us that the IVW method is to be preferred over MR-Egger. In Figure 3 (right), we see that Q_R is equal to 1/4 in case (c) and 2/5 in case (d), indicating that inferences from the MR-Egger approach should be given due consideration. In the limit, if the IVW approach was an increasingly bad fit to the data relative to MR-Egger, then Q_R would tend to zero and MR-Egger should be the preferred choice.

5. Sensitivity analysis

5.1. Estimation under violations of IV2 and IV3

In Section 3, we assumed that IV2 held to demonstrate how established meta-analytic methods can, in theory, be transparently adapted to the MR setting. We now explore, from a theoretical standpoint, how IV2 and IV3 violation distort the causal estimand identified by the IVW and MR-Egger approaches. To do this, we return to and assume models (8) and (9) hold, set $\kappa_x = \kappa_y = 1$, but allow the ψ_j terms to be non-zero. When this is the case, the true estimand for the causal effect identified by variant j using the ratio method is:

$$\beta_j = \beta + \frac{\alpha_j + \psi_j}{\gamma_j + \psi_j} \tag{22}$$

We now make an additional simplifying assumption that the SNP-outcome standard error is constant across variants ($\sigma_{\gamma_j} = \sigma_\gamma$). When this is true, the fixed effect IVW estimand is equal to

$$\beta_{IVW} = \beta + \frac{\sum_{j=1}^L (\gamma_j + \psi_j)(\alpha_j + \psi_j)}{\sum_{j=1}^L (\gamma_j + \psi_j)^2} = \beta + \frac{\text{cov}(\gamma_j + \psi_j, \alpha_j + \psi_j) + E[\gamma_j + \psi_j]E[\alpha_j + \psi_j]}{\text{var}(\gamma_j + \psi_j) + E[\gamma_j + \psi_j]^2}, \quad (23)$$

and the fixed effect MR-Egger estimand is equal to

$$\beta_{1E} = \beta + \frac{\text{cov}(\gamma_j + \psi_j, \alpha_j + \psi_j)}{\text{var}(\gamma_j + \psi_j)}, \quad (24)$$

where $E[\cdot]$, $\text{cov}(\cdot)$ and $\text{var}(\cdot)$ refer to sample expectations, covariances and variances, respectively. If IV2 were satisfied ($\psi_j = 0$ for all j), then the MR-Egger estimand tends to β as L grows large under General InSIDE because the numerator of the bias term in (24) will tend to zero. If, in addition, the sample mean of the α_j s were zero (General InSIDE plus balanced pleiotropy) then the IVW estimand would also tend to β as L grew large because the numerator of the bias term in (23) would tend to zero.

In order to investigate the bias of the IVW and MR-Egger estimators under increasing IV2 violation, we explore the value of their underlying estimands, by firstly generating three independent fixed vectors of $L = 50$ α , γ and ψ parameters with means 1/2, 2 and 1/2, respectively. For clarity, care was taken to ensure that the sample covariance of all 50 α and γ parameters was very close to zero, so that Perfect InSIDE is satisfied across all variants when the ψ_j terms are 0 (IV2 satisfied). The full parameter vectors are given in Table A1 in the Appendix. In addition, Figure A.1 in the Appendix plots the bias term of the MR-Egger causal estimand in equation (24) under IV2 when the included set of instruments is increased sequentially from 2 to 50. This is shown to illustrate the point that, even when α_j and γ_j are generated by independent processes (i.e. under General InSIDE), a sufficiently large number of variants is needed to ensure that their scaled sample covariance (i.e. the MR-Egger bias term) settles at 0. In Figure 4 (left), we plot the vector γ versus the vector α to show the instrument strength and pleiotropic effect parameters are uncorrelated under IV2 and Perfect InSIDE. We label the axes with the addition of ' $0 \times \psi$ ' to stress why this is the case.

In Figure 4 (right), we plot $\gamma + 3\psi$ versus $\alpha + 3\psi$ to show the instrument strength and direct pleiotropic effect parameters when IV2 and InSIDE are violated. In this case, non-zero ψ parameters induce a strong positive correlation between the combined instrument strength and pleiotropy terms. We now define the sensitivity parameter

$$\rho_A = \text{cor}(\gamma + A\psi, \alpha + A\psi).$$

That is, ρ_A equals the correlation between the instrument strength and direct effect due to pleiotropy when ψ_j is set to $A \times \psi_j$. In Figure 4 (left), $\rho_0 = 0$ and in Figure 4 (right), $\rho_3 = 0.62$.

In Figure 5 (left), we plot the value of β_{IVW} and β_{1E} in equations (23) and (24) as a function of A and ρ_A for $\beta = 0$. Their values can therefore be interpreted as a bias. Because the mean pleiotropic effect is non-zero, β_{IVW} is biased even under InSIDE, whereas β_{1E} is not. However, we see that β_{1E} is more strongly

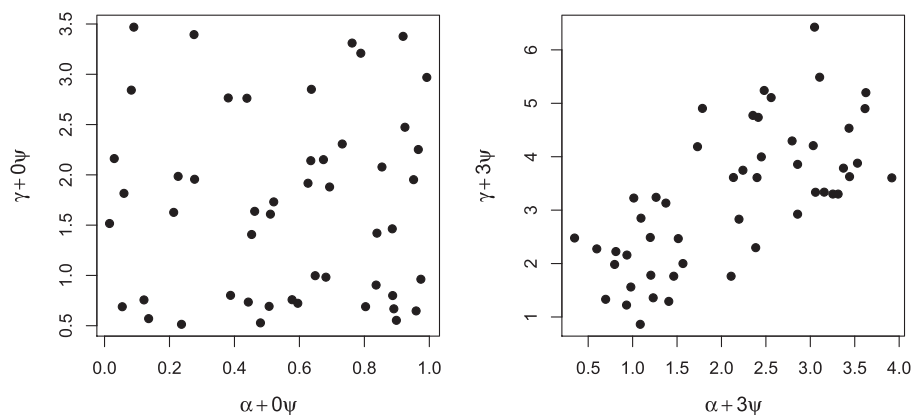


Figure 4. Left: Scatter plot of $\gamma + 0 \times \psi$ versus $\alpha + 0 \times \psi$ parameters, where α and γ are generated to satisfy Perfect InSIDE. Right: Scatter plot of $\gamma + 3 \times \psi$ versus $\alpha + 3 \times \psi$.

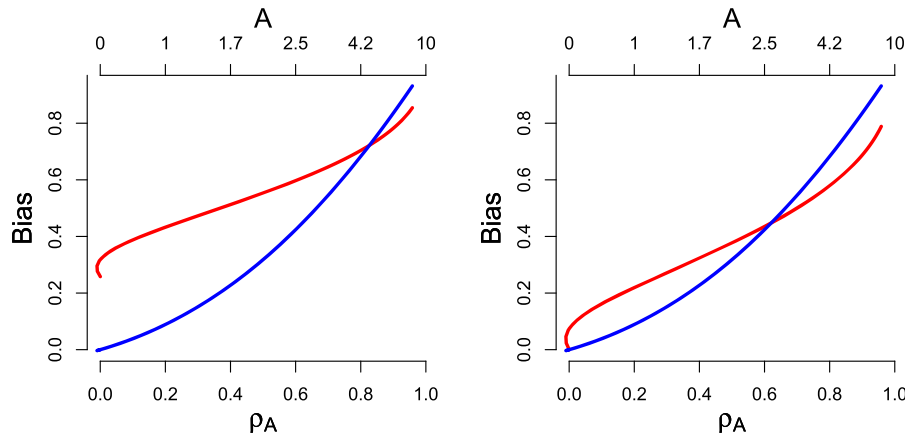


Figure 5. Theoretical bias in the MR-Egger estimand β_{1E} (blue) and IVW estimand β_{IVW} (red) as a function of InSIDE violation via ρ_A for directional pleiotropy (left) and balanced pleiotropy (right). [Colour figure can be viewed at wileyonlinelibrary.com]

affected by violations of InSIDE than β_{IVW} . When $A \geq 4.2$ and $\rho_A \geq 0.8$, β_{1E} is actually further away from the truth than β_{IVW} . In Figure 5 (right), we plot the value of the IVW and MR-Egger estimands as in Figure 5 (left) when α_j is replaced with $\alpha_j - \bar{\alpha}$, in order to make the pleiotropy balanced at $A=0$. Consequently, when $\rho_0 = 0$, β_{IVW} is unbiased. Furthermore, the IVW estimand is less biased than the β_{1E} for $A \geq 2.5$ and $\rho_A \geq 0.65$.

5.2. Fixed versus additive random effects for IVW

In Section 3, we cautioned against using the additive random effects IVW estimate $\hat{\beta}_{IVW}^{ARE}$ as opposed to the fixed effect estimate $\hat{\beta}_{IVW}$ when directional pleiotropy is present. In order to explain the reason for our caution, we plot the value of the additive random effects IVW estimand, β_{IVW}^{ARE} versus its fixed effect counterpart (at $\beta=0$) using our fixed parameter constellation under increasing violations of the InSIDE assumption as before.

We calculated β_{IVW}^{ARE} using equation (11) by plugging in the estimand for β_j from equation (22) and using random effects weights $w_j^* = (\gamma_j + A\psi_j)^2 / (\sigma_{\gamma_j}^2 + \sigma_{\alpha}^2)$. In order to specify these weights, we set σ_{γ_j} equal to a constant value of 0.3, and then used the DerSimonian and Laird method to estimate σ_{α}^2 for each given value of A . Figure 6 shows the results for the two previously discussed scenarios of directional pleiotropy using (α, γ, ψ) and balanced α pleiotropy using $(\alpha - \bar{\alpha}, \gamma, \psi)$. In order to aid interpretation, the x -axis includes scales for ρ_A and, additionally, the amount of heterogeneity as measured by Higgins' I^2 statistic [25]: $(Q - (L - 1))/Q$.

Under the directional pleiotropy scenario (Figure 6, top-left), both β_{IVW}^{ARE} and β_{IVW} are biased even when the InSIDE assumption is satisfied ($\rho_A = 0$). The directional pleiotropy manifests itself as heterogeneity about the fixed effect IVW estimate (as measured by Cochran's Q statistic), and at $\rho_A=0$, $I^2=0.45$. As ρ increases β_{IVW} is more robust than β_{IVW}^{ARE} , staying closer to the true value of zero for $\rho_A \leq 0.8$. This occurs because the observed heterogeneity is caused, disproportionately, by β_j terms emanating from variants with a weaker SNP-exposure association, because their bias terms in equation (22) will (in general) be the largest.

When it detects heterogeneity, the additive random effects estimate compounds the problem by re-weighting the ratio estimates more evenly across the board, thereby giving more weight to the weaker (and more biased) ratio estimates. The fixed effect estimate, and by extension the multiplicative random effects estimate, naturally gives ratio estimates from variants with weaker SNP-exposure associations little weight in the analysis, regardless of whether heterogeneity is detected or not.

Under the balanced pleiotropy example (Figure 6, top-right), both β_{IVW} and β_{IVW}^{ARE} are unbiased at $\rho_A=0$. At $\rho_A=0$, I^2 is very close to zero also, indicating that almost all of the heterogeneity observed in the directional pleiotropy case discussed earlier at $\rho_A=0$ was due to bias (i.e. non-zero μ_{α}). As ρ_A increases, β_{IVW} and β_{IVW}^{ARE} remain close together. This is because there is no longer a directional trend in the β_j s as a function of the j th SNP-exposure association. In order to stress this point, Figure 6 (bottom) shows funnel plots of the inverse standard errors versus their causal estimands β_j at $\rho_A = 0$ under directional (bottom-

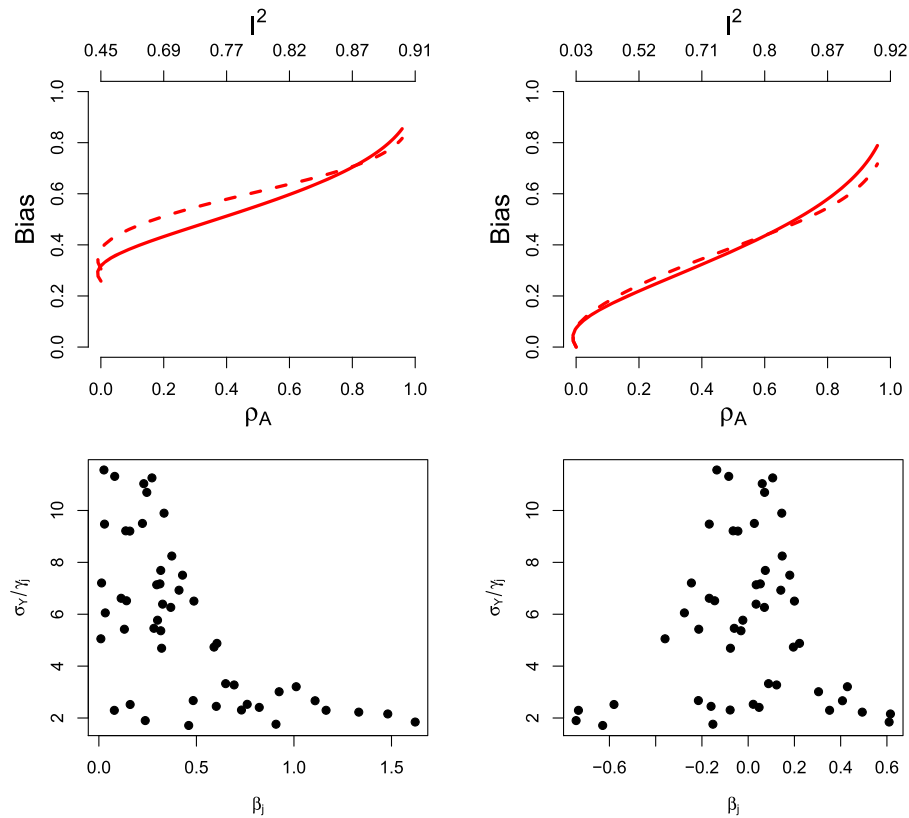


Figure 6. Top: Theoretical bias of the fixed effect estimand β_{IVW} (red solid line) and additive random effect estimand β_{IVW}^{ARE} (red dashed line) as a function of InSIDE violation in the case of directional pleiotropy (left) and balanced pleiotropy (right). Bottom: Funnel plots of the inverse standard errors versus their causal estimands β_j at $\rho=0$ and $\sigma_Y = 0.3$ for directional pleiotropy (left) and balanced pleiotropy (right). [Colour figure can be viewed at wileyonlinelibrary.com]

left) and balanced (bottom-right) pleiotropy. Inverse standard errors used in the y-axis are calculated under NOME. The funnel plot is used as a graphical tool to aid detection of small study bias in meta-analysis [26] and has been transferred to the MR setting by Bowden et al [9]. Only in the directional pleiotropy case will asymmetry be observed in the funnel plot. That is, imprecise causal effect estimates are not seen to ‘funnel’ in towards the more precise causal estimates equally from either side, but appear to be skewed in one particular direction. Therefore, funnel plot asymmetry is a sure sign that a large difference will exist between the fixed and additive random effects IVW estimates, as well as between the IVW and MR-Egger estimates.

5.3. Estimation under violations of NOME

So far, in this paper, we have assumed that the NOME assumption is satisfied. In truth, however, NOME will never hold exactly because the variance of the SNP-exposure association, $\sigma_{X_j}^2$ will always be positive, so that $\hat{\gamma}_j \neq \gamma_j$. In this context, we can think of IVW and MR-Egger approaches as fitting regression models using an explanatory variable that is affected by measurement error. This is known to induce an attenuation in the subsequent effect estimates towards zero due to ‘regression dilution’ [27, 28]. The implications of NOME violation on the two approaches is discussed at length in separate work by Bowden et al [17], but are summarised below.

For clarity, suppose that the pleiotropy distribution and additional assumptions defining either case (a) or (b) holds (Table II) so that both the IVW and MR-Egger methods yield unbiased estimates for β under NOME (i.e. when $\sigma_{X_j}^2 = 0$ for all j). When NOME is in fact violated for variant j ($\sigma_{X_j}^2 > 0$), the expected attenuation in its individual ratio estimate, $\hat{\beta}_j$, can be approximated by:

$$E[\hat{\beta}_j] \approx \beta \left(\frac{F_j - 1}{F_j} \right),$$

where $F_j = \hat{\gamma}_j^2 / \sigma_{X_j}^2$. Because it is a weighted average of ratio estimates, the expected attenuation in the IVW estimate, $\hat{\beta}_{IVW}$, can therefore be approximated by substituting F_j in the above formula with $\bar{F} = \sum_{j=1}^L w_j F_j / \sum_{j=1}^L w_j$, where w_j are as defined in (11). The attenuation in the IVW estimate can be alleviated by increasing the strength of each instrument through F_j so that IV1 is more strongly satisfied.

In contrast, the extent of attenuation in the MR-Egger estimate, $\hat{\beta}_{1E}$, due to NOME violation is not governed by the same F statistic. It can instead be accurately gauged using a modification of the I^2 statistic applied directly to the entire set of SNP-exposure associations (and referred to as I_{GX}^2 in [17]) as below:

$$E[\hat{\beta}_{1E}] \approx \beta I_{GX}^2, \quad \text{where} \quad I_{GX}^2 = (Q_{GX} - (L - 1)) / Q_{GX} \quad \text{for} \quad Q_{GX} = \sum_{j=1}^L \frac{(\hat{\gamma}_j / \sigma_{Y_j} - \bar{\hat{\gamma}})^2}{\sigma_{X_j}^2 / \sigma_{Y_j}^2},$$

where $\bar{\hat{\gamma}}$ is the mean of the weighted $\hat{\gamma}_j$ s. The I_{GX}^2 statistic measures the spread of variation in the estimated instrument strengths (the $\hat{\gamma}_j$ s) relative to their average uncertainty. It lies between 0 and 1 but is generally reported as a percentage. An I_{GX}^2 close to 100% ensures that the attenuation will be minimal, and means that the VIS assumption is strongly satisfied.

In most circumstances, the extent of attenuation is likely to be far worse for MR-Egger than for the IVW estimate. Bowden et al [17] employ the method of Simulation Extrapolation (SIMEX) [29, 30] to adjust the MR-Egger estimate for this dilution. Briefly, this involves generating pseudo-data sets based on the original summary data estimates, but under increasingly strong violations of the NOME assumption, to obtain a series of increasingly diluted parameter estimates. A statistical model is then fitted to this series in order to extrapolate back to the estimate that would have been obtained if NOME had been satisfied. In the next section, we show the SIMEX approach applied in practice to both the IVW and MR-Egger estimates.

6. Assessing the causal role of plasma urate on CHD risk

Although traditional epidemiological studies have suggested that urate levels are associated with a range of cardiometabolic diseases, it is not certain whether the relationship is causal. Recently, Mendelian randomization has been applied to assess the possibility of a causal link [31, 32]. Specifically, White and colleagues [32] conducted a two-sample summary data MR analysis to assess the causal role of plasma urate levels in influencing cardiovascular disease risk. Summary data estimates of the association of 31 uncorrelated genetic variants with plasma urate concentration were obtained from the genome wide association studies catalogue <http://www.genome.gov/gwastudies/> with p -values less than 5×10^{-7} . Estimates for the association of the 31 variants with cardiovascular disease risk were obtained from a separate study population using CARDIoGRAM [3], currently the largest genetic association study measuring cardiovascular outcomes. Here, we repeat and extend the analysis of White et al's for the IVW and MR-Egger approaches, using the methods described thus far. Full results are shown in Table III.

Figure 7 (left) shows a scatter plot of the SNP-exposure and SNP-outcome association estimates, where the genetic data has been recoded in order to make the SNP-exposure associations positive. The SNP-outcome association estimates represent log-odds ratios of CHD for a 1 standard deviation increase in plasma urate. Figure 7 (right) shows the corresponding funnel plot, which can be interpreted as discussed in Section 5. The vertical (and horizontal) lines show the causal effect estimates (and 95% confidence intervals) inferred by IVW and MR-Egger regression.

We start our MR analysis by assuming that all variants are valid IVs and that the NOME, IV2, IV3 and InSIDE assumptions hold, which justifies the use of the fixed effect IVW model (10). The NOME assumption seems reasonable for these data under the IVW method, due a large mean F statistic across the genetic variants of 247 (with the weakest being 22 and the strongest being 4886). The IVW estimate $\hat{\beta}_{IVW}$ suggests that a one unit increase in plasma urate increases the log-odds ratio of CHD by 0.163. However, there is evidence of heterogeneity among the 31 ratio estimates, Cochran's Q statistic on 30 degrees of freedom is 63.9, and the scale factor estimate $\hat{\phi}_{IVW} = 2.13$. We therefore relax assumption IV3 to allow each variant to have a pleiotropic effect (assuming it is 'balanced' under InSIDE) and perform random effects analysis. Applying the multiplicative random effects model (15) and so scaling up the variance of the fixed effect IVW estimate by $\hat{\phi}_{IVW}$, the p -value for $\hat{\beta}_{IVW}$ from a resulting t-test is equal to 0.02. Applying the additive random effects model (14), we estimate the pleiotropic variance parameter σ_α^2 to be 0.093, which yields an adjusted point estimate for $\hat{\beta}_{IVW}^{ARE}$ of 0.222 (an increase of 35%).

Table III. IVW and MR-Egger regression analyses of the urate data.					
Model					
Parameter	Est	S.E.	<i>t</i> -value	<i>p</i> -value	Heterogeneity statistics
IVW approach under model (14)					
β_{IVW}^{ARE}	0.222	0.093	2.43	0.021	$I^2 = 53\%$, $\sigma_\alpha^2 = 0.093$
IVW approach under model (15)					
β_{IVW}	0.163	0.067	2.45	0.020	$Q = 63.9$ ($p=3\times 10^{-4}$), $\hat{\phi}_{IVW} = 2.13$
IVW approach+SIMEX under model (15)					
β_{IVW}	0.164	0.067	2.47	0.020	-
MR-Egger regression under model (20)					
β_{0E}	0.008	0.005	1.61	0.11	-
β_{1E}	0.048	0.096	0.50	0.62	$Q' = 58.6$ ($p=9\times 10^{-4}$), $\hat{\phi}_E = 2.02$
MR-Egger regression+SIMEX under model (20)					
β_{0E}	0.008	0.005	1.59	0.12	-
β_{1E}	0.05	0.097	0.51	0.61	-
Summary stats: $I_{GX}^2 = 99\%$, $\bar{F} = 247$, $Q - Q' = 5.25$, $Q_R = 0.917$					

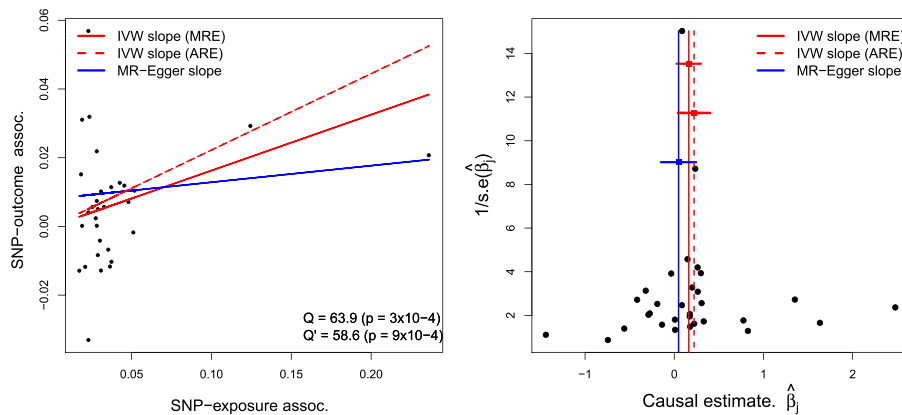


Figure 7. Left: scatter plot of the summary data estimates for the lipids data, with the IVW and MR-Egger slope estimates. Right: corresponding funnel plot of the same data. ARE = additive random effects estimate $\hat{\beta}_{IVW}^{ARE}$, MRE = fixed effect/multiplicative random effects estimate $\hat{\beta}_{IVW}$.

Because both the fixed and additive random effects estimates should be targeting the same quantity under balanced pleiotropy plus InSIDE, this large difference suggests that the pleiotropy could have a significant directional component.

Given the large and significant Q statistic, we now apply MR-Egger regression to these data to probe if there is indeed evidence for directional pleiotropy. We firstly assume that the pleiotropy is directional but constant and fit the fixed effect MR-Egger model (16) under the NOME assumption. The NOME assumption seems reasonable for these data under the MR-Egger method, because the I_{GX}^2 statistic is equal to 99.2% and so a less than 1% dilution in its causal estimate is expected.

Under this model, MR-Egger estimates a non-zero mean pleiotropic effect of 0.008 and, because this is in the same direction, a reduced causal effect estimate, $\hat{\beta}_{1E}$, of 0.048. Calculating the residual heterogeneity about the MR-Egger fit using Rucker's Q' statistic yields $Q' = 58.6$ with a p-value of 9×10^{-4} . We next allow for directional and heterogenous pleiotropy under the InSIDE assumption by fitting multiplicative random effects model (20), and scale up the variance of the fixed effect MR-Egger estimates by a factor of $\hat{\phi}_E = 2.02$. Under this analysis, the p-value for the mean pleiotropic effect (or intercept) parameter estimate $\hat{\beta}_{0E}$ is 0.12 and the p-value for the causal effect estimate $\hat{\beta}_{1E}$ is 0.62. The ratio statistic Q_R is 0.917 indicating that the MR-Egger model explains approximately 8% more of the variation in the SNP-outcome association estimates than the IVW approach and the difference $Q - Q'$ is also large. Although the NOME assumption appears to be sensible for these data, for completeness, we use the

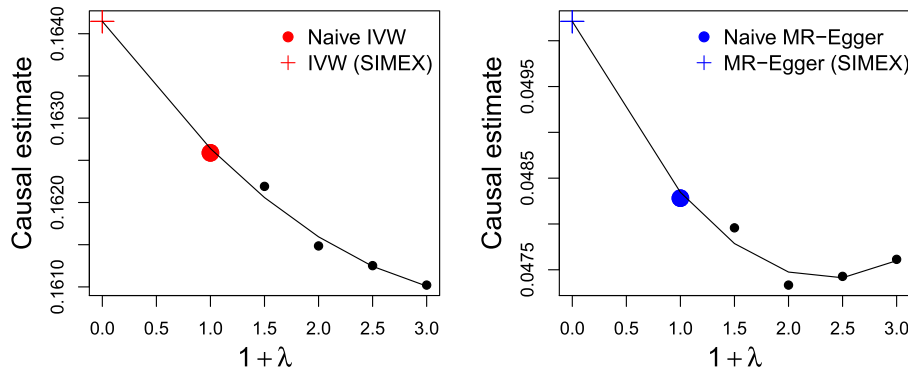


Figure 8. SIMEX adjustment applied to the IVW (left) and MR-Egger (right) causal effect estimate. [Colour figure can be viewed at wileyonlinelibrary.com]

SIMEX method to adjust for NOME violation in the IVW and MR-Egger fits (under multiplicative random effects models only). R code to perform this method can be found in [17]. The results are shown in Figure 8 and Table III.

In summary, careful analysis of these data suggest that the large positive causal association of urate concentration with CHD risk is potentially unreliable due to the presence of directional pleiotropy manifesting itself as heterogeneity among the causal effect estimates. Whilst the application of an additive random effects model accounting for balanced pleiotropy (via the IVW estimate) increases the causal effect estimate further still, MR-Egger regression suggests that the pleiotropy has a positive directional element, and consequently adjusts the causal estimate down towards zero. For these data, there is good evidence that MR-Egger provides a better fit than IVW.

7. Discussion

In this paper, we have attempted to explain the many and varied assumptions that are necessary to apply standard meta-regression and random effects methodology from mainstream meta-analysis to summary data MR. We have also tried to explore the consequences for each method when some of the key assumptions (e.g. IV2, IV3, InSIDE and NOME) are violated, and clarified cases where the violation of one assumption (e.g. IV2 and VIS) acts to promote violation of another (e.g. InSIDE and NOME). We hope that those applying the IVW and MR-Egger regression approaches can use this work to gain a more informed understanding of their strengths and limitations. One such limitation of MR-Egger regression is that it is known to be far less efficient than the IVW method. For this reason, we also gave particular emphasis to examining the properties of the more established IVW estimate under both additive and multiplicative random effects models. We therefore view it as an important finding that the additive approach is less robust to directional pleiotropy, echoing the results of Henmi and Copas [15] for mainstream meta-analyses affected by small study bias. This point was indeed highlighted in the analysis of the urate data affected in Section 6, whereby the IVW estimate actually increased under an additive random effects model in response to positive directional pleiotropy, rather than decreasing towards zero.

In practice, we would encourage users to report the full set of analyses described here, using the various heterogeneity and instrument strength statistics introduced as a guide to the relative importance that could be placed on each method for the data at hand.

7.1. Limitations

A major limitation of this work is that we have assumed throughout that the two-sample assumptions hold, in particular that the SNP-exposure and SNP-outcome association estimates gleaned in separate populations provide information on a common (i.e. identical) set of parameters. This will not be true, for example, in the presence of gene-environment interactions if the distribution of the environmental factor differs in the two samples. As further work, it is vital to understand the consequences for inference if this assumption is violated. In particular, it would be interesting to see whether the approaches discussed would remain valid under weaker assumptions. For example, if the parameters indexing the two population models were not identical, but were instead generated from a common distribution. A Bayesian implementation would then seem natural.

Another major limitation of this work is that we assumed additive linear models with no interaction terms for the SNP-outcome association estimates. As shown by Didelez and Sheehan [33], such a model is required for consistent estimation of the causal effect. In practice (and in our real data example) MR analyses are generally performed with respect to a binary outcome yielding log-odds ratios for the SNP-outcome associations. In this case, when a causal effect between the exposure and outcome exists, ratio estimates of causal effect gleaned from individual variants will be attenuated towards the null by varying amounts due to the non-collapsibility of the odds-ratio [34]. This is enough to invalidate our models in its own right. Furthermore, if the data were collected using case-control sampling, as is also common, then log-odds ratio estimates are also susceptible to ascertainment bias [35]. It will be important to develop summary data methodology to properly account for these two issues as well.

A further limitation of our proposed framework for investigation of pleiotropy, is that it is purely based on statistical evidence, and ignores any *a priori* biological knowledge of possible pleiotropy for individual variants. As more is learned about the association between individual variants and multiple health outcomes, analyses that ignore this additional information will appear increasingly uninformed as well as inefficient.

Finally, we return to the InSIDE assumption. When explaining this assumption and exploring the impact of its violation on the IVW and MR-Egger approaches, we assumed for simplicity that IV2 held. However, it is still possible in theory for InSIDE to hold even when IV2 is violated, albeit in fairly contrived circumstances. For example, suppose that IV2 and IV3 are violated for all variants but the pleiotropy via one route is perfectly negatively correlated with pleiotropy via the other, so that $\alpha_j = -\psi_j$. Alternatively, suppose that IV3 is satisfied for all variants ($\alpha_j = 0$) but the magnitude of violation of IV2 is the same across variants ($\psi_j = \psi$). In both cases, the covariance between the instrument strength and direct effect is zero because the pleiotropy is constant, and Perfect InSIDE holds. Whether these facts could somehow be exploited to improve the performance of MR-Egger regression is a subject for further investigation.

Appendix

Table A1. Model parameters used for the theoretical calculations in Section 5.

Variant	α	γ	ψ
1	0.89	1.62	0.11
2	0.92	3.86	0.27
3	0.51	1.79	0.97
4	0.64	2.42	0.54
5	0.52	1.94	0.63
6	0.44	3.14	0.66
7	0.69	2.11	0.03
8	0.76	3.78	0.60
9	0.68	1.06	0.26
10	0.64	3.24	0.13
11	0.08	3.23	0.80
12	0.38	3.14	0.91
13	0.84	0.97	0.67
14	0.09	3.96	0.99
15	0.67	2.43	0.49
16	0.39	0.85	0.99
17	0.28	2.20	0.11
18	0.02	1.69	0.93
19	0.97	1.04	0.78
20	0.44	0.77	0.16
21	0.48	0.53	0.92
22	0.89	0.85	0.50
23	0.46	1.83	0.86
24	0.06	2.04	0.34

Table A1. Continued.			
Variant	α	γ	ψ
25	0.80	0.72	0.20
26	0.84	1.57	0.05
27	0.24	0.52	0.94
28	0.97	2.54	0.88
29	0.79	3.66	0.52
30	0.12	0.80	0.19
31	0.14	0.58	0.48
32	0.63	2.15	0.19
33	0.92	2.80	0.51
34	0.60	0.76	0.21
35	0.03	2.44	0.11
36	0.28	3.88	0.50
37	0.85	2.34	0.03
38	0.21	1.82	0.20
39	0.45	1.56	0.35
40	0.65	1.08	0.96
41	0.96	0.67	0.99
42	0.99	3.38	0.09
43	0.95	2.19	0.64
44	0.05	0.72	0.71
45	0.51	0.73	0.98
46	0.89	0.70	0.07
47	0.58	0.80	0.86
48	0.90	0.56	0.40
49	0.73	2.61	0.96
50	0.23	2.23	0.38

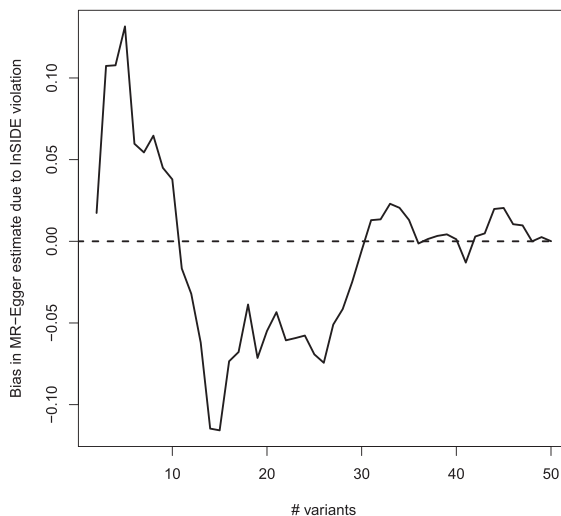


Figure A.1. Theoretical bias of the MR-Egger estimand using the parameter values of Table A1, as the number of included SNPs is increased sequentially from 1:2 to 1:50. Note: the parameters of Table A1 are generated independently so that General InSIDE is satisfied under IV2 for any subset of the 50 SNPs, but that Perfect InSIDE is satisfied when all 50 are considered.

Acknowledgements

We thank Gibran Hemani for his help in accessing the urate data. Jack Bowden is supported by an MRC Methodology Research Fellowship (grant MR/N501906/1). George Davey Smith is supported by the MRC Integrative Epidemiology Unit at the University of Bristol (grant code MCUU12013/1).

References

1. Davey Smith G, Ebrahim S. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease *International Journal of Epidemiology* 2003; **32**:1–22.
2. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics* 2014; **23**:89–98.
3. CARDIoGRAMplusC4D. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genetics* 2003; **45**:25–33.
4. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nature Genetics* 2013; **45**:1274–1283.
5. Pierce B, Burgess S. Efficient design for Mendelian randomization studies: subsample and two-sample instrumental variable estimators. *American Journal of Epidemiology* 2013; **178**:1177–1184.
6. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology* 2013; **37**:658–665.
7. Stearns FW. One hundred years of pleiotropy: a retrospective. *Genetics* 2010; **186**:767–773.
8. Kang H, Zhang A, Cai T, Small D. Instrumental variables estimation with some invalid instruments, and its application to Mendelian randomisation. *Journal of the American Statistical Association* 2016; **111**:132–144.
9. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* 2015; **44**:512–525.
10. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology* 2016; **40**:304–314.
11. Windmeijer F, Farbmacher H, Davies N, Davey Smith G, White IR. Selecting (in) valid instruments for instrumental variables estimation. *Technical report*, University of Bristol, 2015.
12. Johnson T. Efficient calculation for multi-SNP genetic risk scores. *Technical report*, The Comprehensive R Archive Network, 2013. Available at <http://cran.r-project.org/web/packages/gtx/vignettes/ashg2012.pdf> [last accessed 2014/11/19].
13. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 1997; **315**:629–634.
14. Rücker G, Schwarzer G, Carpenter J, Binder H, Schumacher M. Treatment-effect estimates adjusted for small study effects via a limit meta-analysis. *Biostatistics* 2011; **12**:122–142.
15. Henmi M, Copas JB. Confidence intervals for random effects meta-analysis and robustness to publication bias. *Statistics in Medicine* 2010; **29**:2969–2983.
16. Del Greco FM, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Statistics in Medicine* 2015; **34**:2926–2940.
17. Bowden J, Del Greco FM, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for mendelian randomization analyses using MR-Egger regression: the role of the I^2 statistic. *International Journal of Epidemiology* 2016. DOI:10.1093/ije/dyw220.
18. Kolesar M, Chetty R, Friedman F, Glaeser E, Imbens G. Identification and inference with many invalid instruments. *NBER Working Paper 17519*, National Bureau of Economic Research, 2014.
19. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
20. Higgins JPT, Thompson SG, Deeks J, Altman D. Measuring inconsistency in meta-analyses. *British Medical Journal* 2003; **327**:557–560.
21. Thompson SG, Sharp S. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**:2693–2708.
22. Baker R, Jackson D. Meta-analysis inside and outside particle physics: two traditions that should converge *Research Synthesis Methods* 2013; **4**:109–124.
23. Rahman M, Gokhale DV. Testimation in regression parameter estimation. *Biometrical Journal* 1996; **38**:809–817.
24. Abramovich F, Grinshtein V, Pensky M. On optimality of Bayesian testimation in the normal means problem. *The Annals of Statistics* 2007; **35**:2261–2286.
25. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analyses. *Statistics in Medicine* 2002; **21**:1539–1558.
26. Sterne J, Egger M. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology* 2001; **54**:1046–1055.
27. MacMahon S, Peto R, Collins R, Godwin J, Cutler J, Sorlie P, Abbott R, Neaton J, Dyer A, Stamler J. Blood pressure, stroke, and coronary heart disease. Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *The Lancet* 1990; **335**:765–774.
28. Frost C, Thompson SG. Correcting for regression dilution bias: comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2000; **163**:173–189.
29. Cook JR, Stefanski LA. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* 1994; **89**:1314–1328.
30. Hardin JW, Schmiediche H, Carroll RJ. The simulation extrapolation method for fitting generalized linear models with additive measurement error. *Stata Journal* 2003; **3**:373–385.
31. Keenan T, Zhao W, Rasheed A, Ho WK, Malik R, Felix JF, Young R, Shah N, Samuel M, Sheikh N, Mucksavage ML. Causal Assessment of Serum Urate Levels in Cardiometabolic Diseases Through a Mendelian Randomization Study. *Journal of the American College of Cardiology* 2016; **67**:407–416.

32. White J, Sofat R, Hemani G, Shah T, Engmann J, Dale C, Shah S, Kruger FA, Giambartolomei C, Swerdlow DI, Palmer T. Plasma urate concentration and risk of coronary heart disease: a Mendelian randomisation analysis. *The Lancet Diabetes & Endocrinology* 2016; **4**:327–336.
33. Didelez V, Sheehan NA. Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research* 2007; **16**:309–330.
34. Vansteelandt S, Bowden J, Babanezhad M, Goetghebeur E. On instrumental variables estimation of causal odds ratios. *Statistical Science* 2011; **26**:403–422.
35. Bowden J, Vansteelandt S. Mendelian randomization of case-control studies using structural mean models. *Statistics in Medicine* 2011; **30**:678–694.