Methods

Mendelian randomization with Egger pleiotropy correction and weakly informative Bayesian priors

AF Schmidt[1,2,3]* and F Dudbridge[4,5]

[1]Groningen Research Institute of Pharmacy, University of Groningen, Groningen, The Netherlands, [2]Institute of Cardiovascular Science, University College London, London, UK, [3]Department of Cardiology, University Medical Center Utrecht, Utrecht, The Netherlands, [4]Department of Health Sciences, University of Leicester, Leicester, UK and [5]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

*Corresponding author. Institute of Cardiovascular Science, University College London, Gower Street, London WC1E 6BT, UK. E-mail: amand.schmidt@ucl.ac.uk

Abstract

Background: The MR-Egger (MRE) estimator has been proposed to correct for directional pleiotropic effects of genetic instruments in an instrumental variable (IV) analysis. The power of this method is considerably lower than that of conventional estimators, limiting its applicability. Here we propose a novel Bayesian implementation of the MR-Egger estimator (BMRE) and explore the utility of applying weakly informative priors on the intercept term (the pleiotropy estimate) to increase power of the IV (slope) estimate.

Methods: This was a simulation study to compare the performance of different IV estimators. Scenarios differed in the presence of a causal effect, the presence of pleiotropy, the proportion of pleiotropic instruments and degree of 'Instrument Strength Independent of Direct Effect' (InSIDE) assumption violation. Based on empirical plasma urate data, we present an approach to elucidate a prior distribution for the amount of pleiotropy.

Results: A weakly informative prior on the intercept term increased power of the slope estimate while maintaining type 1 error rates close to the nominal value of 0.05. Under the InSIDE assumption, performance was unaffected by the presence or absence of pleiotropy. Violation of the InSIDE assumption biased all estimators, affecting the BMRE more than the MRE method.

Conclusions: Depending on the prior distribution, the BMRE estimator has more power at the cost of an increased susceptibility to InSIDE assumption violations. As such the BMRE method is a compromise between the MRE and conventional IV estimators, and may be an especially useful approach to account for observed pleiotropy.

Key words: Epidemiology methods, Bayesian analysis, Mendelian randomization, instrumental variables, pleiotropy

Key Messages

- Absence of pleiotropy is an essential assumption for instrumental variable analyses using genetic instruments, known as Mendelian randomization. The MR-Egger method corrects for the presence of pleiotropy by introducing a nuisance parameter which captures directional pleiotropy. However, including this nuisance parameter greatly reduces power to detect a causal effect as compared to the traditional inverse variance weighted (IVW) estimator.
- In this paper we propose a novel Bayesian implementation of the MR-Egger, 'BMR-Egger', which increases the power of the causal estimate by introducing a weakly informative prior on the nuisance parameter. Our motivation is that the BMR-Egger can be seen as a compromise between two extreme prior distributions. Specifically, the IVW method corresponds to applying an optimistic informative prior on the intercept with a mean and variance of zero, whereas MR-Egger corresponds to a pessimistic non-informative prior with an infinite variance.
- When the 'Instrument Strength Independent of Direct Effect' (InSIDE) assumption holds, the BMR-Egger has increased power with acceptable type 1 error rates as compared to the MR-Egger. If the InSIDE assumption is violated, all estimators are biased and show inappropriately high rejection rates. In this case, adding prior beliefs increases bias and rejection rates of the BMR-Egger towards that of the IVW estimator.

Introduction

Instrumental variable analyses using genetic instruments, often termed Mendelian randomization (MR) analyses, use genetic exposures as instruments to determine the causal association between an intermediate phenotype, often a biomarker, and a particular outcome such as disease. The estimate of such an MR analysis reflects an unbiased causal estimate of the phenotype effect on the outcome, if (among others) the following assumptions are met.

i. The instruments are associated with the phenotype.
ii. The instruments are independent of observed and unobserved confounders of the phenotype-outcome association.
iii. Conditional on the phenotype and confounders, the instruments are independent of the outcome (i.e. the exclusion restriction assumption).

Given that biomarkers are the (indirect) products of multiple genes, it is often possible to select a set of genetic instruments that meet assumption (i). Furthermore, because genes are randomly allocated at conception,[1] assumption (ii) is often plausible as well. Assumption (iii) states that the genes can only be related to disease due to their effects on the phenotype (i.e. no pleiotropy other than that mediated by the phenotype). Whether this assumption generally holds has been contested.[2] For example, if one is interested in estimating the causal relation between high-density lipoprotein cholesterol (HDL-C) and coronary heart disease (CHD), it is often difficult to find genes that affect HDL-C but not low-density lipoprotein cholesterol (LDL-C) or triglycerides.[3,4] Such a situation may indicate violation of assumption (ii) (when LDL-C is viewed as a confounder of HDL-C and CHD), of assumption (iii) (when a gene effects both

pathways independently) or of both assumptions. In practice, such distinctions are difficult to make and hence robust IV methods are preferred.

Recently Bowden *et al.*[5] proposed a novel method related to the Egger test,[6] 'Mendelian randomization Egger' (MR-Egger/MRE), which corrects for potential violations of assumption (iii) by quantifying the amount of directional pleiotropy. This MR-Egger method assumes that the 'Instrument Strength is Independent of the Direct Effect' (i.e. the InSIDE assumption), which means that across single nucleotide polymorphisms (SNPs), pleiotropic effects are independent of phenotypic effects. The MR-Egger method corrects for pleiotropy by introducing a nuisance parameter which quantifies the average amount of directional pleiotropy. However, including this nuisance parameter greatly reduces precision and power to detect a causal effect. Despite this reduced power, the MRE method has been frequently used in empirical settings.[4,7–9]

In this paper, we propose a novel Bayesian implementation of the MR-Egger method, 'BMR-Egger', which increases power of the causal estimate by introducing a (weakly) informative prior on the nuisance parameter, which is the intercept in a linear regression. From a Bayesian perspective, the standard inverse variance weighted (IVW) estimator and the MRE estimator can be unified by noticing that the IVW method corresponds to putting an optimistic informative prior on the intercept with mean and variance of zero; conversely, the MRE approach can be seen as a pessimistic non-informative prior with infinite variance. Whereas pessimistic and optimistic priors are often used, for example in randomized controlled trials (RCTs) in rare diseases,[10] in genetics considerable data may be available on the magnitude of pleiotropy

and consequently less extreme, more believable priors may be usefully employed.

One reasonable approach may be a prior belief that extreme departures from balanced pleiotropy are unlikely, as strong pleiotropic effects of genetic variants may have been previously identified. This is similarly optimistic to the IVW method; however, instead of (unrealistically) assuming a zero prior variance, we suggest use of weakly informative priors to allow for a degree of pleiotropy. Alternatively, as we will discuss using an empirical example of urate and coronary heart disease (CHD),[7] often considerable (aggregated) data will be available on potential pleiotropic pathways, which can be used to further elucidate a prior distribution to fit the specific data at hand. We note that defining what constitutes a pleiotropic pathway is difficult and will depend on subjective criteria such as statistical significance and the availability of relevant datasets (such as MR-base[11]).

In the following, we introduce notations, and the outcome model, followed by a review of the MR estimators and the proposed novel Bayesian MR estimator. Subsequently we evaluate the discussed methods in a simulation study, and the empirical example noted above.

Methods

Notation, and outcome model

Let us assume there are data available from $j = 1, \ldots, J$ independent single-nucleotide polymorphisms (SNPs) G (an $i = 1, \ldots, n$ subject by J matrix), with α_j representing the (marginal) effect of SNP j on a biomarker X, β_j the (marginal) SNP effect on an outcome Y, and variance of their estimators $\sigma^2_{\alpha_j}$ and $\sigma^2_{\beta_j}$. We note that α_j and β_j may be estimated from the same data (one-sample MR study) or in separate data (two-sample MR study); we focus on the latter.[12] Based on these data we are interested in estimating the causal effect of X on Y, assuming Y is generated by the linear model $Y_i = \sum_j^J \phi_j G_{ij} + \theta X_i + \varepsilon_y$, with θ a scalar, and Y, G and X as defined above. When assumptions (ii–iii) hold, $\phi_j = 0$ and $Y_i = \theta X_i + \varepsilon_y$. Note that the absence of an intercept term in the above equations should be interpreted as meaning the intercept (arbitrarily) equals zero, and should not be misinterpreted as an absence of pleiotropy which is represented by ϕ_j.

MR estimators

When there are multiple instruments available, the causal (IV) effect of X on Y can be estimated using a weighted ordinary least squares (OLS) regression of β_j on α_j while supressing the intercept. Given that β_j and α_j are unknown,

they are estimated from the data, with the estimates collected in the following matrices:

$$A = \begin{bmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_J \end{bmatrix},$$

$$B = \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_J \end{bmatrix},$$

$$\Omega_{jk} = \hat{\sigma}_{\beta_j} \hat{\sigma}_{\beta_k} p_{jk},$$

where Ω is the sample variance-covariance matrix for B, with $p_{j=k} = 1$ and, assuming that SNPs are independent, $p_{j \neq k} = 0$. In the case of correlated SNPs, $p_{j \neq k}$ can be estimated based on the pairwise between SNP correlations[13] and the regression fitted by generalized least squares. The following regression is weighted by the precision of the SNP effect estimates, giving the IVW point estimate and standard error estimates (assuming no pleiotropy, or balanced positive and negative pleiotropic effects under the InSIDE assumption)[14] as:

$$\hat{\theta}_{IVW} = (A^t \Omega^{-1} A)^{-1} A^t \Omega^{-1} B, \tag{1}$$

with the weighted residuals:

$$\hat{\varepsilon}_j = diag(\Omega^{-1})^{\frac{1}{2}} (B - \hat{\theta}_{IVW} A),$$

where $diag(\cdot)$ indicates the diagonal elements. The variance of the error term is then:

$$\hat{\sigma}^2_\varepsilon = \frac{1}{J-k} + \sum_{j=1}^{J} \hat{\varepsilon}^2_j,$$

where k equals the number of parameters ($k = 1$ in this specific case). Finally, the standard error of the slope is:

$$\hat{\sigma}_{\theta_{IVW}} = diag\left(\hat{\sigma}^2_\varepsilon (A^t \Omega^{-1} A)^{-1}\right)^{\frac{1}{2}}. \tag{2}$$

Here, and in the following derivations, the sigma term $\hat{\sigma}^2_\varepsilon$ will only be included if it is larger than 1, resulting in standard errors following a multiplicative random effects model.[15]

The MR-Egger method corrects for (unmeasured) directional pleiotropy by introducing an intercept term which captures the expected effect of an instrument on outcome when it has no effect on the biomarker, and is hence a measure of the average amount of pleiotropy. To implement the MR-Egger we first recode the data as follows:

$$\hat{\beta}_j^* = \hat{\beta}_j sgn(\hat{\alpha}_j),$$

with sgn the sign function,

$$\hat{\alpha}_j^* = |\hat{\alpha}_j|,$$

$$A = \begin{bmatrix} 1 & \hat{\alpha}_1^* \\ \vdots & \vdots \\ 1 & \hat{\alpha}_J^* \end{bmatrix},$$

$$B = \begin{bmatrix} \hat{\beta}_1^* \\ \vdots \\ \hat{\beta}_J^* \end{bmatrix}$$

$$\begin{bmatrix} \hat{\phi}_{MRE} \\ \hat{\theta}_{MRE} \end{bmatrix} = (A^t \Omega^{-1} A)^{-1} A^t \Omega^{-1} B,$$

$$\begin{bmatrix} \hat{\sigma}_{\phi_{MRE}} \\ \hat{\sigma}_{\theta_{MRE}} \end{bmatrix} = diag\left(\hat{\sigma}_\varepsilon^2 (A^t \Omega^{-1} A)^{-1} \right)^{\frac{1}{2}},$$

with $\hat{\sigma}_\varepsilon^2$ derived as before, $\hat{\theta}_{MRE}$ the slope estimate, and $\hat{\phi}_{MRE}$ the intercept estimate.

Next, we describe our proposed Bayesian MR-Egger method [BMRE] using a bivariate normal likelihood and the conjugate prior distribution with hyperparameters for the prior mean and variance of the intercept and slope:

$$\mu_0 = \begin{bmatrix} \mu_\phi \\ \mu_\theta \end{bmatrix}$$

$$\Sigma_0 = \begin{bmatrix} \sigma_{\mu_\phi}^2 & 0 \\ 0 & \sigma_{\mu_\theta}^2 \end{bmatrix}.$$

Then the posterior distribution is bivariate normal with mean μ_N and variance-covariance matrix Σ_N:

$$\lambda_0 = \Sigma_0^{-1},$$
$$\mu_N = (A^t \Omega^{-1} A + \lambda_0)^{-1} (\lambda_0 \mu_0^t + A^t \Omega^{-1} B),$$
$$\Sigma_N = \hat{\sigma}_\varepsilon^2 (A^t \Omega^{-1} A + \lambda_0)^{-1},$$

with $\hat{\sigma}_\varepsilon^2$ derived as before.

To explore the effect of including prior information using weakly informative priors, we performed the simulation study described below. Specifically, we were interested in exploring the advantage of specifying a prior on the intercept $\hat{\phi}_{BMRE}$ to increase precision of the posterior $\hat{\theta}_{BMRE}$ and on the robustness of prior misspecification. In our empirical example, we illustrate how to use empirical data on observed pleiotropy signals to elucidate reasonable priors, decreasing the likelihood of prior misspecification.

Our results will also discuss a further method to allow for pleiotropy, the weighted median (WM) estimator.[16] This estimator assumes that at least 50% of the weights,

$w_j = \frac{\hat{\alpha}_j^2}{\hat{\sigma}_{\beta_j}^2}$, come from valid instruments. If this assumption is true, a consistent estimate of causal effect is the 50th percentile of the empirical distribution function of SNP-specific IV estimates $\frac{\hat{\alpha}_j}{\hat{\beta}_j}$, with the percentile distribution based on $\frac{s_j - \frac{w_j}{2}}{S}$; where $s_j = \sum_{k=1}^{j} w_k$, the cumulative sum up to the jth SNP, and $S = \sum_{k=1}^{J} w_k$.

Data-generating process

Similar to the original publication by Bowden *et al.*,[5] data for $i = 1, \ldots, n$ subjects were simulated, with $n = 1000$ and $J = 20$ SNPs. G_{ij} were sampled from a trinomial distribution with minor allele frequency $p_j = 0.30$ under Hardy–Weinberg equilibrium. An unmeasured confounder was generated based on $U_i = \sum_j^J \omega_{1j} G_{ij} + \epsilon_u;\ \epsilon_u \sim N(0,2)$ and a biomarker $X_i = \sum_j^J \alpha_j G_{ij} + \omega_2 U_i + \epsilon_x;\ \epsilon_x \sim N(0,2)$. Finally, the outcome was generated following $Y_i = \sum_j^J \phi_j G_{ij} + \theta X_i + \omega_3 U_i + \epsilon_y;\ \epsilon_y \sim N(0,2)$. Based on the two-sample MR principle,[12] this algorithm was run twice (with the same parameters) to generate two independent datasets, the first used to derive the genetic effects on the biomarker by fitting the linear model $X_i = \alpha_j G_{ij} + \epsilon_x$, and the second to estimate the genetic effects on the outcome from the linear model $Y_i = \beta_j G_{ij} + \epsilon_y$.

Simulation scenarios

The above defined MR estimators were evaluated in five scenarios (Table 1). In scenario I there was no pleiotropy, hence $\phi_j = 0$, and the confounder was independent of the SNPs, $\omega_{1j} = 0$. In scenario II pleiotropy was generated based on $\phi_j \sim U(0.00, 0.20)$, and in scenario III the InSIDE assumption was violated by setting $\omega_{1j} \sim U(L, U)$, $L = 0, U = 0.50$. In scenario IV the InSIDE assumption was met, $\omega_{1j} = 0$, and pleiotropy was generated based on $\phi_j \sim U(0.00, 0.50)$ with probability $q = \{0.1, 0.2, 0.3, 0.4\}$ and 0 otherwise, resulting in (on average) qJ SNPs violating assumption (iii). In this scenario the average pleiotropy depends on q and ranged between $E(\phi_j)q = \{0.025, \ldots, 0.100\}$. Subsequently, in scenario V $q = 0.4$ with ϕ_j and ω_{1j} generated based on q as in scenario IV. Different types and severities of InSIDE assumption violations were generated by first setting $L = 0$ and $B = \{0.10, 0.30, 0.60, 1.00\}$, and subsequently setting $L = -B$ and $B = \{0.10, 0.30, 0.60, 1.00\}$. All scenarios were repeated under the null- and alternative-hypotheses setting $\theta = \{0.00, 0.05\}$. The BMRE estimator was evaluated using the following hyperparameters: $\mu_\phi = \{0, 0.05, 0.10, 0.15\}$, $\mu_\theta = 0$, with every element of μ_ϕ evaluated with five different variance hyperparameters: $\sigma_{\mu_\phi}^2 = \{10, 10^{-2}, 10^{-2.4}, 10^{-2.7}, 10^{-3}\}$, and $\sigma_{\mu_\theta}^2 = 10$.

Table 1. Simulation scenarios of a multi-SNP Mendelian randomization study, with potential pleiotropic effects (i.e. violation of the exclusion restriction assumption)[a]

Parameters	Scenario I (no pleiotropy)	Scenario II (pleiotropy)	Scenario III (InSIDE violated)	Scenario IV (partial pleiotropy)	Scenario V (partial pleiotropy and InSIDE violated)
Number of subjects n	1000	1000	1000	1000	1000
Number of SNPs J	20	20	20	20	20
Proportion of pleiotropic SNPs q	1.0	1.0	1.0	**{0.1, 0.2, 0.3, 0.4}**	**0.4**
Minor allele frequency p_j	0.30	0.30	0.30	0.30	0.30
Effect of G_j on U_i (ω_{1j})	0.00	0.00	**$Unif(L, U)$**	0.00	**$Unif(L, U)$**
Lower limit of ω_{1j} L			**$L = 0.00$**		**$L = 0$**
					$B = \{0.10, 0.30, 0.60, 1.0\}$, and
Upper limit of ω_{1j} U			**$U = 0.50$**		**$L = -B$**
					$B = \{0.10, 0.30, 0.60, 1.0\}$
Effect of G_j on X_i (α_j)	$Unif(0.5, 4)$	$Unif(0.5, 4)$	$Unif(0.5, 4)$	$Unif(0.5, 4)$	$Unif(0.5, 4)$
Effect of G_j on Y_i (ϕ_j)	0.00	**$Unif(0, 0.2)$**	$Unif(0, 0.2)$	$Unif(0, 0.2)$	$Unif(0, 0.2)$
Effect of U_i on X_i (ω_2)	1	1	1	1	1
Effect of U_i on Y_i (ω_3)	1	1	1	1	1
Effect of X_i on Y_i (θ)	{0.00, 0.05}	{0.00, 0.05}	{0.00, 0.05}	{0.00, 0.05}	{0.00, 0.05}

[a]Changes from the previous scenario (to the left) are presented in bold.

Performance metrics

Performance was evaluated using the following metrics: bias defined as $\overline{\theta} - \theta$, with $\overline{\theta}$ equal to the mean of $\hat{\theta}$; the root mean square error $RMSE = \sqrt{bias^2 + ESE^2}$, with ESE equal to the empirical standard error of $\hat{\theta}$: the proportion of rejected null-hypotheses (i.e. depending on whether θ equals 0 this is the type 1 error or power, using an alpha of 0.05).

All simulations were repeated 5000 times, with analyses performed using the statistical package R version 3.1.2 for Unix.[17] The number of replications was chosen to ensure sufficient precision to detect small deviations from the nominal type 1 error rate of 0.05 (the 95% lower and upper bounds were 0.044 and 0.056).

Results

Results of the simulation study

In scenario I all the MR assumptions held, hence all the IVW, WM and MRE estimators were unbiased (Appendix Figures 1–2, available as Supplementary data at *IJE* online). Bias of the BMRE estimates was minimal for the hyperparameters $\mu_0 = \{0.00, 0.05\}$, irrespective of the variance hyperparameter. Type 1 error rates of both the intercept and the slope estimates were generally below 0.05 using the same priors, and the RMSE markedly decreased with smaller values of $\sigma^2_{\mu_\phi}$ (Figure 1). Repeating scenario 1 with the true slope set to 0.05, revealed that power of the BMRE estimator (relative to the MRE) was increased without increasing the intercept type 1 error rate above 0.05, unless $\mu_\phi \geq 0.1$ and then only for small values of $\sigma^2_{\mu_\phi}$ (Figure 2).

Scenario II explored performance in the presence of pleiotropy which biased the IVW estimates, and (because 100% of the SNPs were pleiotropic) the WM. The MRE estimator remained unbiased (Appendix Figures 5–6, available as Supplementary data at *IJE* online), with the BMRE showing a similar pattern of bias as before, with bias depending on the size of $\sigma^2_{\mu_\phi}$ when $\mu_\phi \neq 0.10$. Intercept rejection rates (power) were increased when $\mu_\phi \geq 0.10$ and $\sigma^2_{\mu_\phi} \neq 10^1$; slope rejection rates (type 1 error) were close to nominal for all BMRE using $\sigma^2_{\mu_\phi} \leq 10^{2.4}$ (Appendix Figure 7, available as Supplementary data at *IJE* online). In the same scenario (Appendix Figure 7) the type 1 error rates of the IVW estimator, and (to a lesser extent) the WM estimator, were inflated, at 0.73 and 0.44 respectively. Setting the phenotype effect to 0.05 (Figure 3) showed that power of the slope estimate was improved even when μ_ϕ was misspecified (i.e. not 0.10). Throughout the RMSE of the BMRE, estimators were equal to or lower than for the MRE estimator.

The InSIDE assumption was violated in scenario III which biased all estimators, with the more informative BMRE faring similarly to the IVW or WM estimators (Appendix Figures 9–10, available as Supplementary data at *IJE* online). Whereas the type 1 error rates of the IVW and WM estimators were close to 100%, the BMRE rejection rates depended on $\sigma^2_{\mu_\phi}$ and often less than the IVW or WM methods (Figure 4). The MRE estimator had only
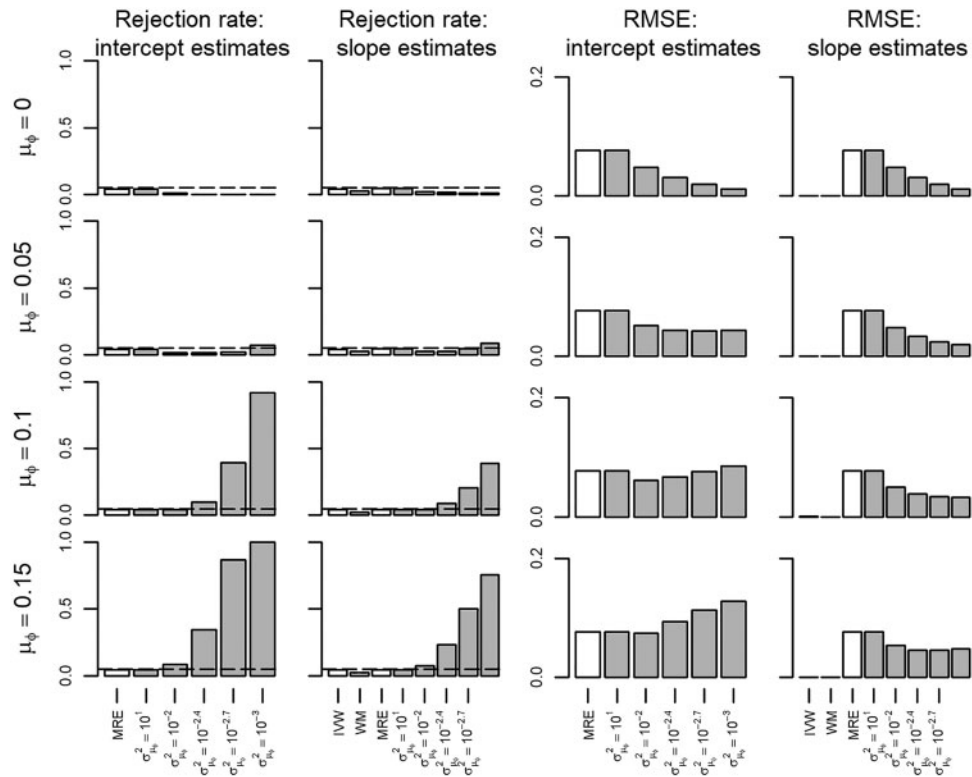
Figure 1. Rejection rate and root mean squared error of a Mendelian randomization study (scenario I) with the true slope of 0 and no unbalanced pleiotropy. IVW, inverse variance weighted; WM, weighted median; MRE, MR-Egger; μ_ϕ indicates the prior mean, and $\sigma^2_{\mu_\phi}$ the prior variance of a Bayesian MRE. The underlying numerical values are presented in Appendix 3, available as Supplementary data at *IJE* online.
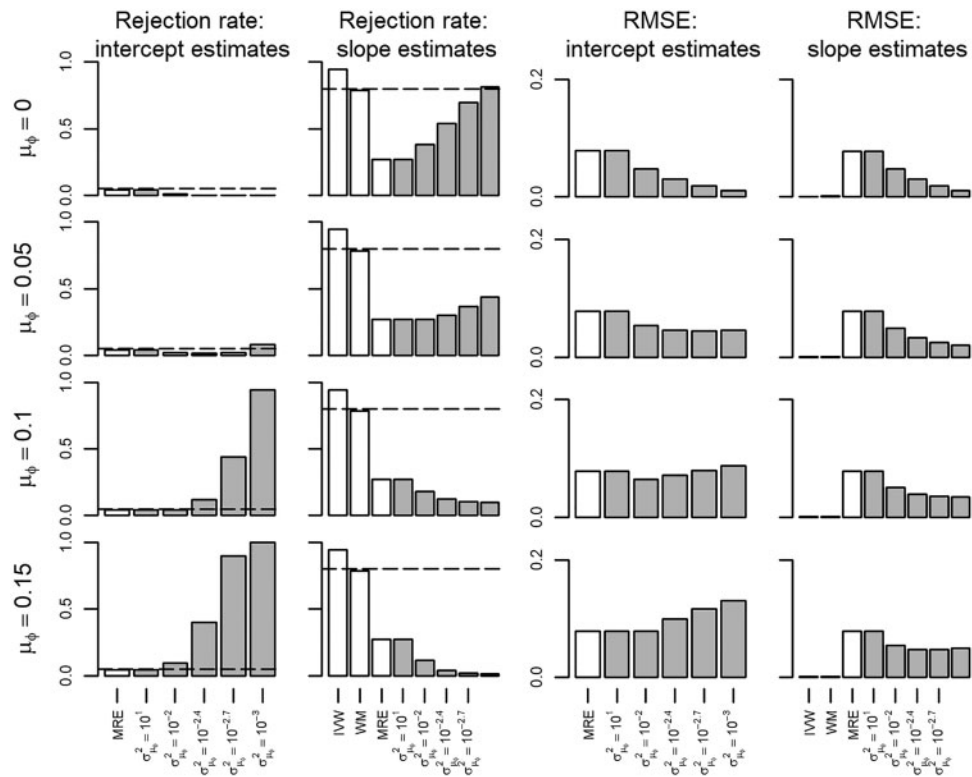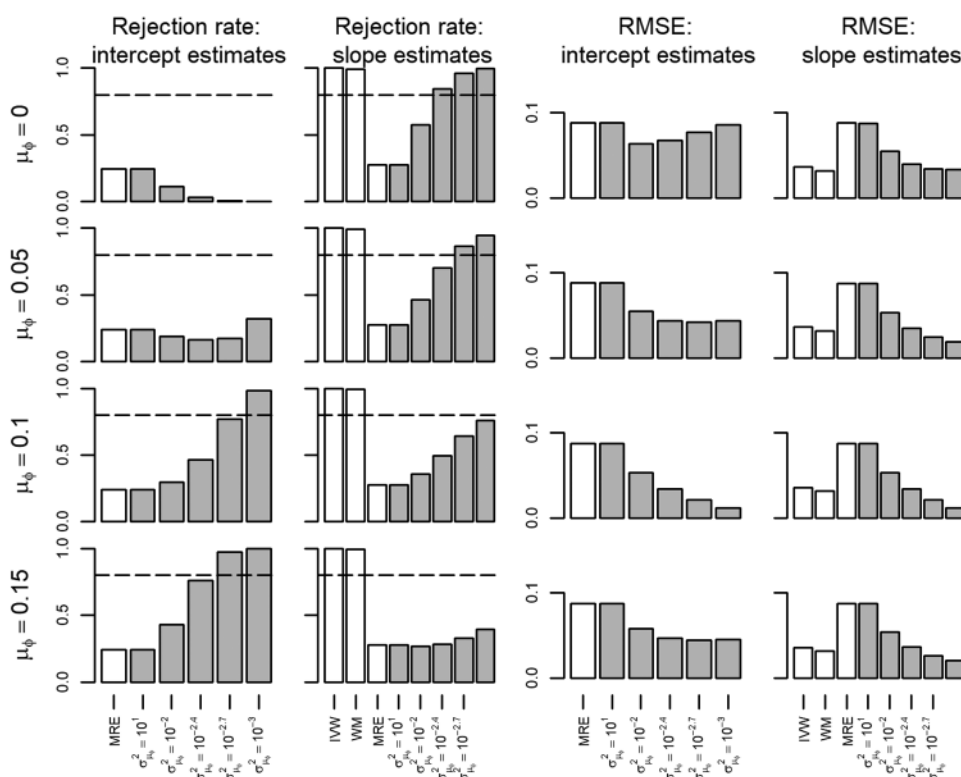
Figure 2. Rejection rate and root mean squared error of a Mendelian randomization study (scenario I) with the true slope of 0.05 and no unbalanced pleiotropy. IVW, inverse variance weighted; WM, weighted median; MRE, MR-Egger; μ_ϕ indicates the prior mean, and $\sigma^2_{\mu_\phi}$ the prior variance of a Bayesian MRE. The underlying numerical values are presented in Appendix 3, available as Supplementary data at *IJE* online.

Figure 3. Rejection rate and root mean squared error of a Mendelian randomization study (scenario II) with the true slope of 0.05 and unbalanced pleiotropy. IVW, inverse variance weighted; WM, weighted median; MRE, MR-Egger; μ_ϕ indicates the prior mean, and $\sigma^2_{\mu_\phi}$ the prior variance of a Bayesian MRE. The underlying numerical values are presented in Appendix 3, available as Supplementary data at *IJE* online.

slightly inflated type 1 error rates close to 0.05. The bias and inflated type 1 error rate of the BMRE persisted even when the intercept prior mean was correctly specified at 0.10 (Figure 4). In these settings, the BMRE estimator was generally more powerful than the MRE approach, which is of limited value given the observed bias and inflated type 1 error rates (Appendix Figure 12, available as Supplementary data at *IJE* online).

The performance of these estimators was further explored in scenario IV by varying the proportion of pleiotropic SNPs. The BMRE results focused on the previously optimally performing combinations of hyperparameters: $\mu_\phi = \{0, 0.05, 0.10\}$, $\sigma^2_{\mu_\phi} = \{10^{-2}, 10^{-2.4}\}$. Note that in this and the next scenario, the average pleiotropy depends on the proportion of pleiotropic SNPs, which ranged between 0.025 (for 10% invalid SNPs) and 0.100 (for 40% invalid SNPs), resulting in differing levels of BMRE misspecification. Figure 5 shows the MRE to be the only unbiased estimator in this scenario. Type 1 error rates were inflated for the IVW and WM methods, with power of the BMRE approach typically surpassing that of the MRE estimator. Next in scenario V, we explored the impact of different degrees of InSIDE assumption violation, revealing a similar amount of bias for all estimators (Figure 6). Type 1 error rates and power were general highest for the IVW,

(closely) followed by WM, the BMRE and MRE methods. As before, the MRE had the largest RMSE throughout, with smaller values for the BMRE, IVW and the WM estimators.

Prior elucidation using empirical data

To illustrate the proposed BMRE method and provide an example of how to elucidate a sensible prior distribution, we consider the study by White *et al*.[7] This study explored the relation between urate and CHD using 31 SNPs collected from 166 486 individuals, 9784 of whom had CHD. White and colleagues used both the IVW and the MRE methods, which showed conflicting results: odds ratio (OR) 1.18 (95% confidence interval (CI) 1.03; 1.34) for the IVW estimate compared with OR 1.05 (95% CI 0.87; 1.27) for the MRE estimate; both re-calculated here using a pleiotropy robust multiplicative random effects model. Aside from the difference in point estimate, the MRE estimate is considerably more variable (standard error (se) of 0.096, compared with an IVW se of 0.066), resulting in wide confidence interval bounds. Interestingly the MRE pleiotropy (intercept) OR estimate of 1.008 (95% CI 0.998; 1.018) is precise, seemingly indicating that amount of directional pleiotropy is minimal, thus questioning the
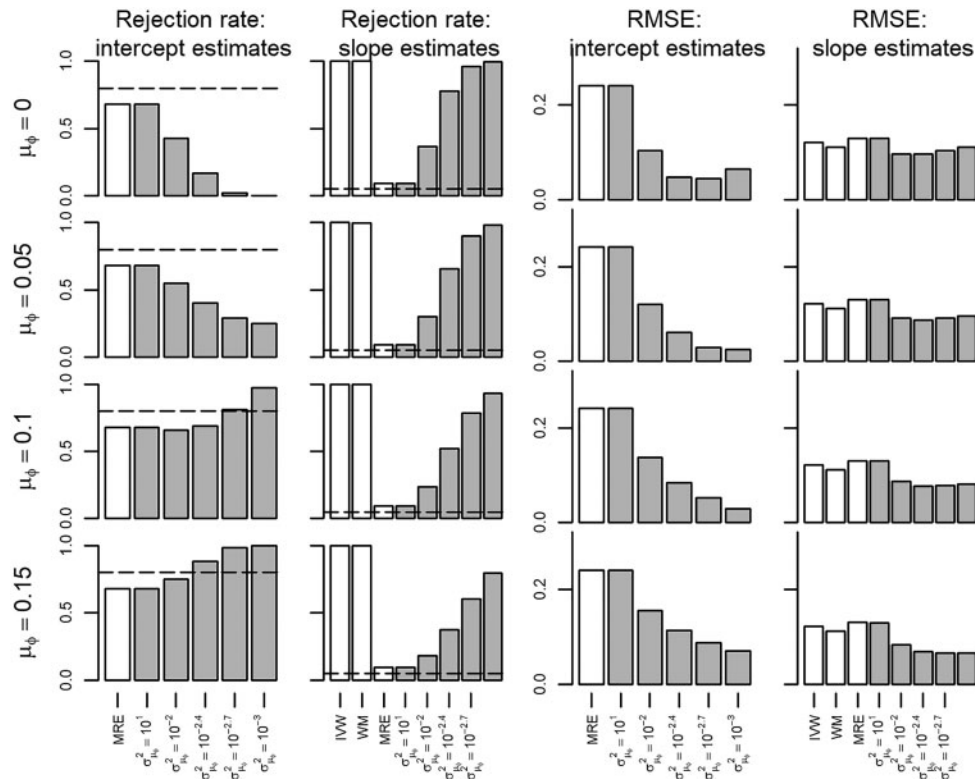
Figure 4. Rejection rate and root mean squared error of a Mendelian randomization study (scenario III) with the true slope of 0.05, and InSIDE assumption violated. IVW, inverse variance weighted; WM, weighted median; MRE, MR-Egger; μ_ϕ indicates the prior mean, and $\sigma^2_{\mu_\phi}$ the prior variance of a Bayesian MRE. The underlying numerical values are presented in Appendix 3, available as Supplementary data at *IJE* online.

necessity of a MRE pleiotropy correction. In the following, we will explore the utility of the BMRE to increase precision of the slope estimate and further explore the necessity of the pleiotropy correction.

White and colleagues not only collected data on CHD and urate, but also on many potential pleiotropic pathways (Appendix Figure 13, available as Supplementary data at *IJE* online) allowing a thorough exploration of the magnitude and direction of observed pleiotropy. We note that four SNPs (rs1260326, rs3741414, rs1178977, rs653178) show clear pleiotropic signals (based on a genome-wide significant *P*-value). Given the number of candidate SNPs, it would be sensible to exclude these SNPs; however, to illustrate the utility of the BMRE we will include these SNPs. Inclusion of pleiotropic SNPs may also occur in practice, for example, when the number of candidate SNPs is modest. Additionally, there is no a priori reason to assume pleiotropy is limited to genome-wide significant signals, hence exclusion of these four SNPs will not necessarily remove all (or even most) of the pleiotropy.

To elucidate and model the likely (known and unknown) pleiotropic effects, we plot the SNP associations with the different phenotypes (Appendix Figure 13), which shows a symmetrical (balanced) pattern centred on a null effect, with most of the estimates between $\pm$ 0.05.

Although reassuring, this does not preclude the possibility of unobserved pleiotropy via different pathways. Based on the observed pleiotropy effects (Appendix Figure 13), we set the mean prior hyperparameter to $\mu_\phi = 0.00$ and considered the following prior variance hyperparameters: $\sigma^2_{\mu_\phi} = \{10^{-6}, 10^{-5.8}, 10^{-5.6}, 10^{-5.4}, 10^{-5.2}\}$. These values of the prior variance parameters were chosen to initially approximate the IVW estimator, incrementally including more uncertainty and thereby allowing for additional pleiotropy. Second, in an alternative approach we use the empirical data to also elucidate the prior variance hyperparameter by selecting a prior variance $\sigma^2_{\mu_\phi} = 6.508 \times 10^{-4} \sim 10^{-3.2}$, putting approximately 95% of the prior distribution $\pm$ 0.05 (the range containing most of the observed pleiotropy signals).

Results of the first approach are shown in Table 2, with the BMRE method showing larger slope estimates (OR ranges from 1.17 to 1.13), than the attenuated MRE OR estimate of 1.05 and the WM 1.12 (95% CI 0.99; 1.27). The BMRE credible intervals included the neutral value of 1 at a prior variance of $10^{-5.6}$; under this prior the intercept odds has 95% probability of lying in (0.997,1.003) suggesting that the balanced pleiotropy assumption has a relevant impact on our IV estimates. Similarly when using the empirically elucidated variance hyperparameter of
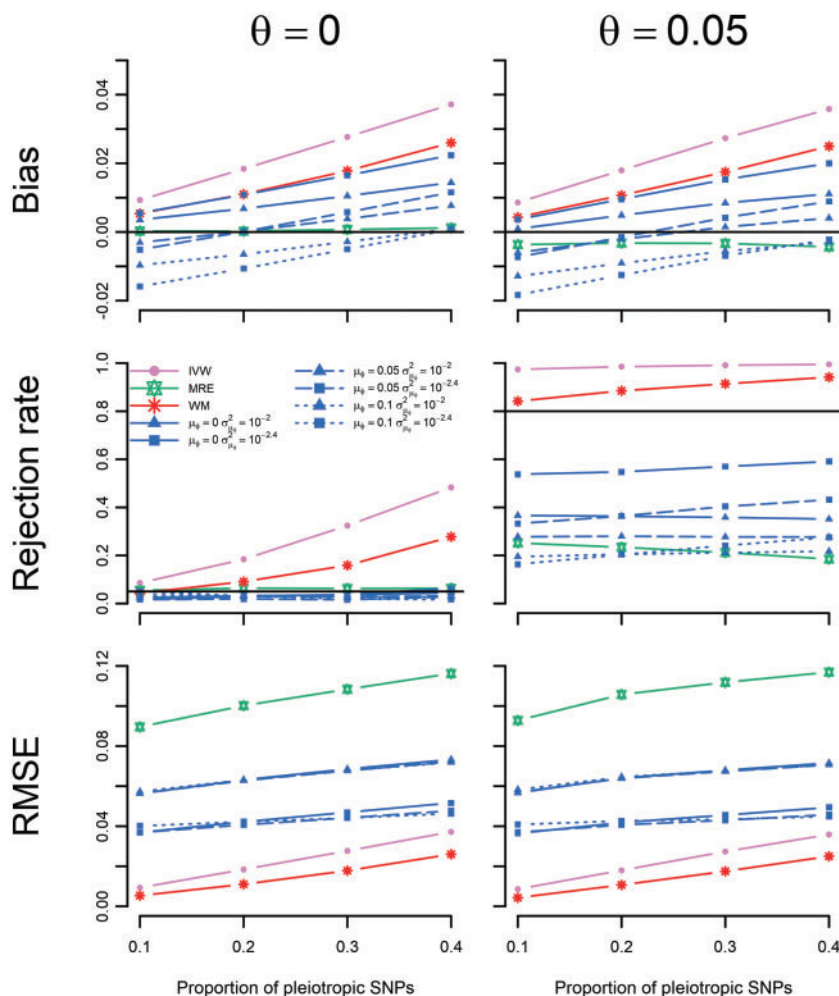
Figure 5. Simulation results of scenario IV: the causal effect estimated in Mendelian randomization study with different proportions of pleiotropic SNPs. IVW, inverse variance weighted; WM, weighted median; MRE, MR-Egger; μ_ϕ indicates the prior mean, and $\sigma^2_{\mu_\phi}$ the prior variance of a Bayesian MRE. The underlying numerical values are presented in Appendix 3, available as Supplementary data at *IJE* online.

6.508×10^{-4}, the BMRE slope estimate becomes OR 1.05 (95% CI 0.87; 1.27) which is identical (to 2 dp) to the MRE estimate. Using the BMRE method we can thus confidently say that despite the empirical data showing balanced pleiotropy, and the tight confidence interval around the MRE intercept estimate, there is relevant directional pleiotropy and the pleiotropy-corrected estimates should be preferred over the IVW estimate.

Discussion

In this paper, we introduce a novel Bayesian implementation of the MR-Egger (BMRE) method for instrumental variable analyses, robust to violation of the exclusion restriction assumption due to pleiotropy. We show that under the InSIDE assumption, the BMRE estimator with weakly informative priors on the intercept increases power to detect a causal effect, while retaining acceptable type 1

error rates. Additionally, the root mean square error of the BMRE estimator was lower than that of the traditional MRE method and, in the presence of pleiotropy, lower than the IVW estimator. Using the empirical example of urate and CHD, we present an approach to evaluate and elucidate sensible prior parameters for the presence of pleiotropy.

When the InSIDE assumption was violated, all estimators were biased and showed inappropriately high rejection rates. In this case, adding prior belief increased bias and rejection rates of the BMRE towards those of the IVW estimator. Comparing the BMRE with the WM method showed that (depending on the prior) the BMRE approach had lower type 1 error rates and was more robust to different degrees of InSIDE assumption violation. Furthermore, if 100% of the SNPs were pleiotropic, the BMRE approach generally was less biased, with type 1 error rates closer to nominal than the WM estimator. In the presence of InSIDE
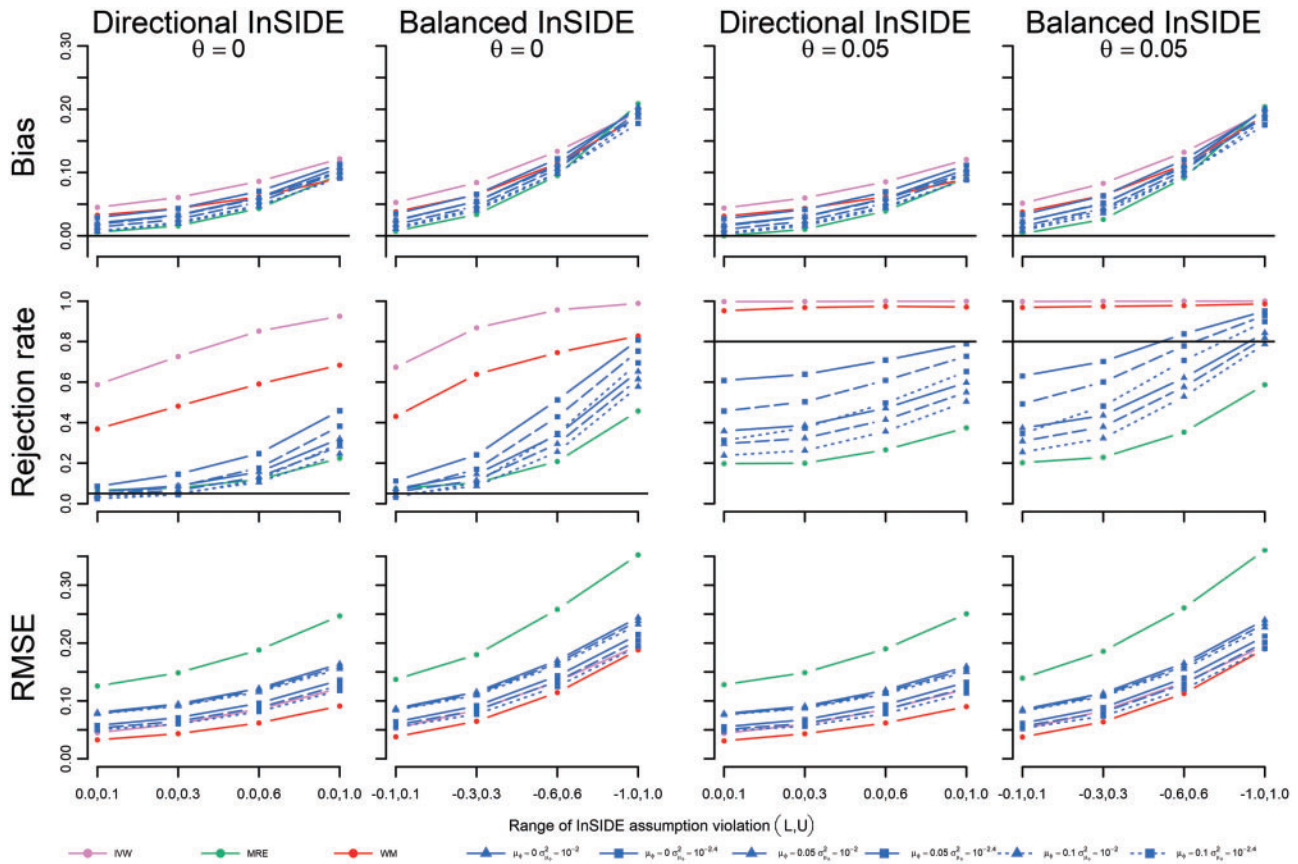
Figure 6. Simulation results of scenario V: the causal effect estimated in a Mendelian randomization study with 40% pleiotropic SNPs and different degrees of InSIDE assumption violation; left panel: no causal effect; right panel: causal effect of 0.05. IVW, inverse variance weighted; WM, weighted median; MRE, MR-Egger; μ_ϕ indicates the prior mean, and $\sigma^2_{\mu_\phi}$ the prior variance of a Bayesian MRE. The underlying numerical values are presented in Appendix 3, available as Supplementary data at *IJE* online.

Table 2. Results of a Mendelian randomization study on the effect of plasma urate on CHD with different IV estimators

	Estimates	
	Intercept OR (95% CI)	Slope OR (95% CI)
IVW		1.18 (1.03; 1.34)
MRE	1.008 (0.998; 1.018)	1.05 (0.87; 1.27)
BMRE		
$\sigma^2_\phi = 10^{-6}$	1.001 (0.998; 1.003)	1.17 (1.02; 1.34)
$\sigma^2_\phi = 10^{-5.8}$	1.001 (0.998; 1.004)	1.16 (1.01; 1.34)
$\sigma^2_\phi = 10^{-5.6}$	1.001 (0.997; 1.006)	1.15 (1.00; 1.33)
$\sigma^2_\phi = 10^{-5.4}$	1.002 (0.997; 1.007)	1.14 (0.99; 1.33)
$\sigma^2_\phi = 10^{-5.2}$	1.003 (0.997; 1.009)	1.13 (0.97; 1.32)
WM		1.12 (0.99; 1.27)

Results presented as odds ratio per 1 SD increase in urate with 95% confidence (or credibility) interval (CI) in brackets. The intercept measures the amount of pleiotropy, the slope estimates the effect of plasma urate on CHD. IVW, inverse variance weighted method; MRE, MR-Egger method; BMRE, Bayesian MR-egger method; WM, weighted median method. $\mu_\phi = 0$, the slope mean and variance priors were 0 and 10 throughout, respectively.

assumption violation, the MRE estimator performed better than the BMRE method. The InSIDE assumption may be violated in empirical data, for example when the pleiotropy effects of different variants affect the outcome via the same set of confounders. Pickrell *et al.*, however, present evidence that pleiotropic SNPs often work via independent pathways, suggesting the InSIDE assumption may hold more generally.[18]

The analyses presented here are naturally limited and the following deserves consideration. First, we chose to implement the BMRE using conjugate priors because these have closed form solutions which increase ease of use and provide exact solutions. In most empirical settings, conjugate priors seem sufficient and are a natural way to encode prior knowledge. Furthermore, the normal distribution is not sharply peaked at its mean value, allowing a reasonable range of values to be given high prior probability, while still discounting unreasonably large values. Second, whereas the IVW method is susceptible to directional pleiotropy, this estimator generally has more precision and power and is more robust to uncertainty in the SNP-exposure association.[19] As such, the IVW method should,

in our opinion, remain the starting point of any MR analysis, with other approaches including the WM, MRE and BMRE used as informative sensitivity analyses. Third, the BMRE methods were evaluated on frequentist concepts of power and type 1 error. Given that MR analyses are often used to test whether a biomarker has a causal effect on disease, we feel these metrics are relevant. Fourth, whereas the weakly informative hyperparameter of $\mu_\phi = \{0.00, 0.05\}$ and $\sigma^2_{\mu_\phi} \leq 10^{-2.4}$ had the desired property of increasing power while maintaining type 1 error rates close to nominal, this is specific to the scenarios considered. Indeed, as we show in our empirical example, these prior hyperparameters should be tailored to the data under consideration. We encourage empirical researchers to use our example as a blueprint to explore observed pleiotropy and to tailor the hyperparameters. In practice, analyses should be repeated under a range of variance hyperparameters, to gain a sense of how precise the prior beliefs must be to maintain significant evidence of causality. Additionally, and similar to designing a Bayesian randomized controlled trial, one may wish to repeat the simulations using scenarios based on the available empirical data and explore performance (see Appendix 2 for the simulation code which took 42 s to run 500 replications of scenario II).

The BMRE method can be used to explore the importance of the balanced pleiotropy assumption of the IVW estimator, and may be particularly useful for reconciling conflicting results from IVW and MRE methods, as we have shown in our example of urate and CHD. Applied researchers may wish to look to a recent framework[14] reviewing the underlying assumptions of the IVW and MRE methods, as well as describing a number of goodness-of-fit statistics and sensitivity analyses. By using a conjugate Bayesian prior, the same framework can readily be applied to the BMRE method presented here. Similarly, the SIMEX[19] adjustment for uncertainty in the SNP-exposure association can be readily applied to our BMRE method as well.

In addition to MRE and WM methods, several other approaches to deal with pleiotropy have recently been proposed, each with its own assumptions, including a weighted mode estimator[20] and a Bayesian model averaging[21] approach conceptually similar to ours. Furthermore, detection and removal of SNPs yielding outlier IV estimates is an important step that can be combined with the pleiotropy robust estimators.[14] A full comparison of methods under realistic settings is beyond the scope of this paper, but a sensible strategy in general is to perform a series of complementary sensitivity analyses in addition to the standard IVW analysis. In this regard, our BMRE method can increase the precision of the MRE estimator and provide insight into discrepancies between IVW and MRE analyses. Further, our BMRE method may be especially useful when candidate instruments show likely pleiotropic effects, but there are too few strong instruments to exclude these pleiotropic SNPs.

In conclusion, we introduce a Bayesian version of the MR-Egger method, which, by placing weakly informative priors on the intercept term increases power over MR-Egger while retaining acceptable type 1 error rates. Violations of the InSIDE assumption increase bias and type 1 error rates beyond those of the MR-Egger method. We suggest that Bayesian MR-Egger is a useful sensitivity analysis that can strengthen the evidence for causal effects in MR studies, particularly in the presence of observed pleiotropy.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

Acknowledgement

Author Contributions

A.F.S. and F.D. contributed to the idea and design of the study. A.F.S. performed the analyses. A.F.S. and F.D. drafted the manuscript. A.F.S. had full access to all of the data and takes responsibility for the integrity of the data presented.

Conflict of interest: Neither of the authors of this paper has a financial or personal relationship with other people or organizations that could inappropriately influence or bias the content of the paper.

References

1. Hingorani A, Humphries S. Nature's randomised trials. *Lancet* 2005;**366**:1906–08.
2. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32**:1–22.
3. Holmes MV, Asselbergs FW, Palmer TM *et al*. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* 2015;**36**:539–50.
4. White J, Swerdlow DI, Preiss D *et al*. Association of lipid fractions with risks for coronary artery disease and diabetes. *JAMA Cardiol* 2016;**366**:1108–18.

5. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;**44**:512–25.

6. Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Br Med J* 1997;**315**:629–34.

7. White J, Sofat R, Hemani G *et al*. Plasma urate concentration and risk of coronary heart disease: A Mendelian randomisation analysis. *Lancet Diabetes Endocrinol* 2016;**4**:327–36.

8. Tyrrell J, Jones SE, Beaumont R *et al*. Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank. *BMJ* 2016;**352**:i582.

9. Dekkers KF, Iterson M van, Slieker RC *et al*. Blood lipids influence DNA methylation in circulating cells. *Genome Biol* 2016;**17**:138.

10. Hampson LV, Whitehead J, Eleftheriou D, Brogan P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. *Stat Med* 2014;**33**:4186–201.

11. Hemani Gibran, Zheng Jie, Wade KH *et al*. MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. *bioRxiv* 2016, December 16. doi: https://doi.org/10.1101/078972.

12. Burgess S, Thompson SG. Avoiding bias from weak instruments in mendelian randomization studies. *Int J Epidemiol* 2011;**40**:755–64.

13. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods. *Stat Med* 2016;**35**:1880–906.

14. Bowden J, Greco M F Del, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* 2017;**36**:1783–802.

15. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: A comparison of methods. *Stat Med* 1999;**18**:2693–708.

16. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016;**40**:304–14.

17. R Core Development Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2017.

18. Pickrell JK, Berisa T, Liu JZ, Segurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* 2016;**48**:709–17.

19. Bowden J, Greco M F Del, Minelli C *et al*. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol* 2016;**45**:1961–74.

20. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017;**46**:1985–98.

21. Thompson JR, Minelli C, Bowden J *et al*. Mendelian randomization incorporating uncertainty about pleiotropy. *Stat Med* 2017, Aug 28. doi: 10.1002/sim.7442.