

Online Supplementary Material: Bayesian estimation of IVW and MR-Egger models for two-sample Mendelian randomization studies

Okezie Uche-Ikonne*

Frank Dondelinger[†]

Tom Palmer[‡]

Online Supplementary Material

We introduce the methods implemented in our package and include some prior distributions. We also show the R syntax and results when the models are fitted using personalized informative prior distributions on example datasets. Further investigations were conducted into the joint prior distribution in relation to the InSIDE assumption and we introduce some strategies when choosing prior distributions.

Methods

- The instrument-phenotype associations are denoted by $\hat{\gamma}_j$.
- The instrument-outcome associations are denoted by $\hat{\Gamma}_j$.

In one-sample MR analyses, the Wald estimator for ratio estimates is

$$\beta_j = \frac{\hat{\Gamma}_j}{\hat{\gamma}_j}. \quad (1)$$

In two-sample MR analyses the numerator and denominator are obtained from different samples.¹

When there are multiple instruments, the IVW estimator is used to estimate the causal effect in a summary level dataset. Equation (2) shows the linear model, with no intercept, from which we derive the IVW estimator in equation (3).

$$\hat{\Gamma}_j = \beta \hat{\gamma}_j + \varepsilon_j; \quad \varepsilon_j \sim N(0, \sigma_{y_j}^2). \quad (2)$$

The IVW estimator is given by,

$$\beta_{\text{IVW}} = \frac{\sum_{j=1}^J w_j \hat{\beta}_j}{\sum_{j=1}^J w_j}, \quad (3)$$

The variable w_j in equation (3) represents the weights denoted as the inverse variance of the ratio estimate $\left(\frac{1}{\text{var}(\hat{\beta}_j)} = \frac{1}{\sigma_{y_j}^2} \right)$.

Equation (4) denotes the MR-Egger model for estimating causal inference. MR-Egger models assume instrument strength independent of direct effects (InSIDE), which means irrespective of the magnitude of

*Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

[†]Faculty of Health and Medicine, Lancaster University, Lancaster, UK

[‡]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

pleiotropic effects valid estimates can be obtained given they are independent of SNP-exposure associations $(\hat{\gamma}_j)^2$. For a summary level dataset, if the value of the mean pleiotropic effect ($\hat{\alpha}$) is further from zero, the larger the difference between the causal effect from the true effect

$$\hat{\Gamma}_j = \alpha + \beta \hat{\gamma}_j + \varepsilon_j, \quad \varepsilon_j \sim N(0, \sigma^2) \quad (4)$$

The radial MR-Egger model, as seen in equation (5). The variable w_j denotes the radial weights which differ from the weights in MR-Egger, assuming first order weights $\frac{\gamma_j^2}{\sigma_{y_j}^2}$ and substituting in the model we see the intercept is unweighted, the IVW estimator (equation (3)) is its submodel.

$$\hat{\beta}_j \sqrt{w_j} = \alpha + \beta \sqrt{w_j} + \varepsilon_j, \quad \varepsilon_j \sim N(0, \sigma^2), \quad (5)$$

Assuming known variance, the likelihood of the MR-Egger estimator follows a univariate Gaussian distribution in equation (6), again w denotes the weights.

$$P(\hat{\Gamma}|\alpha, \beta, \sigma, \hat{\gamma}_j) = \prod_{j=1}^J N(\alpha + \hat{\gamma}_j \beta, \sigma^2 w) \quad (6)$$

Equation (7) denotes the Bayesian posterior distribution;

$$P(\alpha, \beta, \sigma | \hat{\Gamma}_j, \hat{\gamma}_j) \propto P(\hat{\Gamma}_j, \hat{\gamma}_j | \alpha, \beta, \sigma) P(\alpha, \beta, \sigma) \quad (7)$$

Prior distributions

The choice of prior distributions is an important factor in Bayesian estimation. This section gives a brief description on the formulation of the different prior distributions included in this package and the rjags syntax used for implementing them.³

Non-informative prior distributions

We use non-informative prior distributions when we have no prior beliefs about the distribution of a parameter. This type of prior distribution is expected to produce estimates similar to frequentist estimates. There is no “better” choice of an uninformative prior but Equation (8) denotes some possible non-informative prior distributions, these have large variances for the average pleiotropic effect (α) and the causal effect (β). Although an improper prior density was set for the σ , given a large number of instruments ($J > 3$) the prior yields proper posterior densities. 4 Although in the presence of pleiotropic instruments the use of vague/uninformative prior distributions may lead to estimates with low precision.⁵

$$\alpha \sim N(0, 1000), \quad \beta \sim N(0, 1000), \quad \sigma \sim U(0.0001, 10) \quad (8)$$

Weakly informative prior distributions

The idea of a weakly informative prior, equation (9), is to provide partial information on the variables, which is ideal for regularization. Weakly informative priors could mitigate the effects of winners curse. The prior distributions are described in equation (9), where the variance is reduced for α and β compared to the non-informative prior distributions.

$$\alpha \sim N(0, 1), \quad \beta \sim N(0, 1), \quad \sigma \sim U(0.0001, 10) \quad (9)$$

Pseudo-Horseshoe prior distribution

We extend the MR-Egger estimator by placing a cauchy distribution prior on the causal effect $\beta \sim c(0, 1)$. The Cauchy distribution was chosen as the prior distribution due to some appealing properties, for example the divergence property of no mean and infinite variances, rather mode and median which are equal. An investigation into the direction of causality through Bayesian models showed that pleiotropic instruments can give the causal effect a multimodal distribution.⁶ Assuming there are more strong instruments, the divergence property of the Cauchy distribution weighs more on the effect of the strong instruments and reduces the effect of outlying instruments. The convergence towards the Gaussian distribution in the presence of a large number of instruments is another useful property of the Cauchy distribution as a shrinkage prior. This is also a good test for reliable strategies designed to function well under a wide variety of distributional assumptions. The default prior distributions for our `prior = "pseudo"` option, in the `mr_egger_rjags` and `mr_radialegger_rjags` functions, are as in equation (10).

$$\alpha \sim N(0, 1), \beta \sim C(0, 1), \sigma \sim IG(0.5, 0.5) \quad (10)$$

Joint prior distribution

A conjugate bivariate normal prior distribution on the slope and intercept in the MR-Egger model has been shown to have good properties.⁷ We assume α , β and σ^2 follow a bivariate prior distribution which have marginal normal distributions, the proposed priors are denoted in (11).

$$\begin{aligned} \alpha|\sigma^2 &\sim N(\mu_\alpha, \sigma^2\sigma_\alpha) \\ \beta|\sigma^2 &\sim N(\mu_\beta, \sigma^2\sigma_\beta) \\ \sigma^2 &\sim U(1, 10) \\ Cov(\alpha, \beta|\sigma^2) &= \sigma^2\rho_{\alpha\beta} \end{aligned} \quad (11)$$

Under its accompanying InSIDE assumption, the correlation coefficient can be described as the degree of InSIDE violation when $\sigma_\alpha\sigma_\beta \geq 0$ within the MR-Egger model. We investigate the InSIDE assumption using assumed external information on the values for the hyperparameters in equation (11) denoted below;

$$\begin{aligned} \mu_\alpha, \mu_\beta &= 0 \\ \sigma_\alpha, \sigma_\beta &= 10 \end{aligned} \quad (12)$$

We are interested in whether the magnitude of the correlation coefficient between the intercept and slope of the joint prior distribution influences the estimates, which can help us determine the ideal value of ρ while conducting an MR analysis. We fitted the MR-Egger and radial MR-Egger models for values of the correlation coefficient (ρ) between -0.99 upto 0.99 under the null and alternative hypothesis. The simulated datasets consist of two-sample study design showing directional pleiotropy when the InSIDE assumption is violated ($\beta = 0.5$).

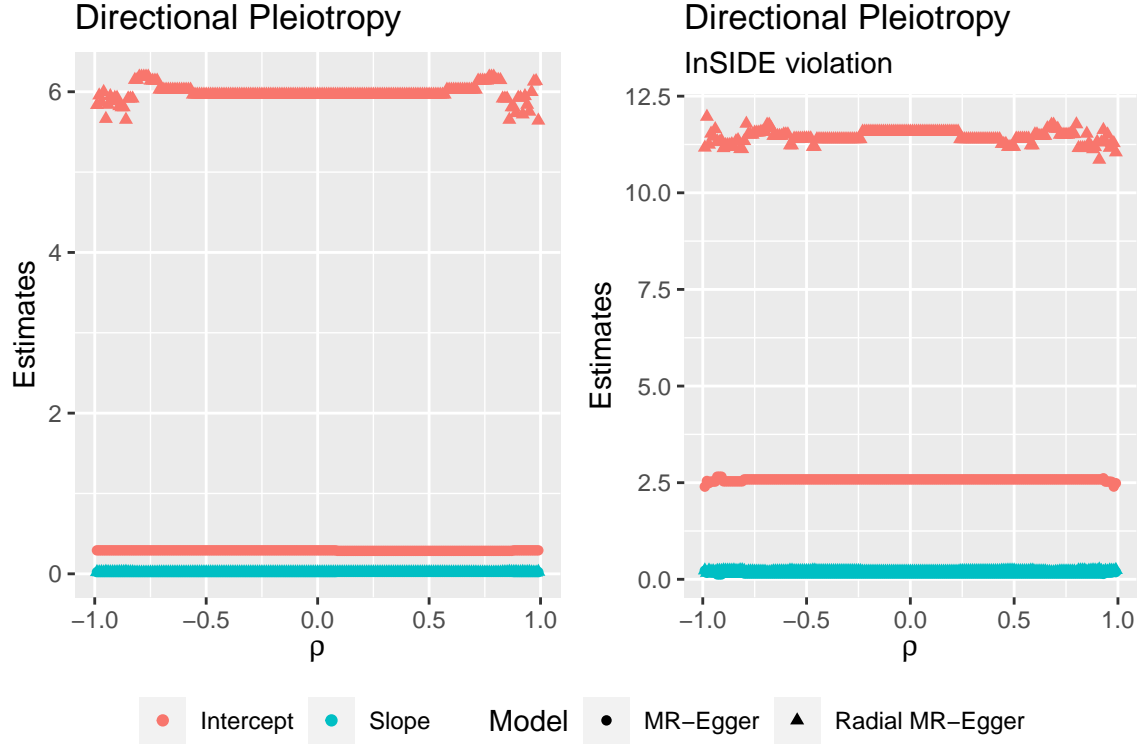


Figure S1: Estimates of the causal effect and average pleiotropic effect for different values of ρ in the joint prior distribution.

Figure S1 shows the results generated for α and β from the different values of ρ , the values show a similar pattern when the InSIDE assumption is valid or violated. The values of the parameters within the MR-Egger model show no difference when the correlation coefficient changes. The Radial formulation shows a pattern in the intercept parameter where we notice a little change when ρ gets closer to ± 1 .

Strategies for choosing priors

Informative prior distributions can help to account for pleiotropy in Bayesian MR analyses⁵, an approach for informative prior distribution is to use the result from a previous study. Alternatively, for cases where prior estimates cannot be obtained, we can use regularized priors similar to weakly informative prior distributions. We give some strategies when considering informative prior distribution. The emphasis would be on the slope parameter (which is the causal effect estimate).

The choice of a prior distribution with small standard deviation (e.g. $\beta \sim N(0, 1)$) can be regarded as an ideal option when $\hat{\Gamma}_j$ and $\hat{\gamma}_j$ are standardized which is comparable to the IVW and the original formulation of MR-Egger models. We can set prior distributions for the slope and its standard deviation independently an example is the normal gamma distribution $\beta \sim N(0, 1/\sigma_\beta)$; $\sigma_\beta \sim G(a, b)$. However, this prior distribution applies to non-standardized error terms similar to the radial MR-Egger. The selection of hyperparameters (a, b) can make the normal-gamma distribution have a similar shape as the laplace distribution which has stronger regularization. This prior distribution can also be considered a Bayesian version of frequentist LASSO regression⁸.

Example: Estimates from some informative prior distributions

Data description

The data is an excerpt from the `mr.raps` package, details of the instrument-exposure and instrument-outcome associations are within the package documentation. The outcome for this dataset is acute ischemic stroke and the exposure for this analysis is body mass index (BMI). This dataset is created from three genome-wide association studies (GWAS). GWAS on BMI was used for SNP selection by⁹, The UK BioBank GWAS of BMI was applied to estimate the SNPs' effect on BMI. The third GWAS study estimates the SNPs' effect on AIS¹⁰.

To obtain this dataset, the Akiyama study is used for SNP selection (column `pval.selection`). The UK BioBank dataset estimates and the Malik dataset estimates the SNPs' effect on AIS.

Application of some informative and Hierarchical priors

From the example dataset, we applied some informative prior distributions discussed above on the parameter for the causal effect estimate (β), the R-syntax is given below and table S1 results of the estimates for illustration.

```
# restrictive prior distribution
user_egger <-
  mr_egger_rjags(
    dat,
    betaprior = "dnorm(0,1)",
    seed = c(123456, 456789, 342564),
    n.chains = 3
  )
```

```
# normal gamma prior distribution
user_radialegger <-
  mr_radialegger_rjags(
    dat,
    betaprior = "dnorm(0,1)",
    seed = c(123456, 456789, 342564),
    n.chains = 3
  )
```

```
user_trialegger <-
  mr_egger_rjags(
    dat,
    betaprior = "dnorm(0,phi) phi ~ dgamma(0.5,0.5)",
    sigmaprior = "dgamma(1E-4,1E-4)",
    seed = c(123456, 456789, 342564),
    n.chains = 3
  )

user_trialradialegger <-
  mr_radialegger_rjags(
    dat,
    betaprior = "dnorm(0,tau/lambda) lambda ~ dgamma(0.5,0.5) tau = 0.025",
    sigmaprior = "dgamma(1E-4,1E-4)",
    seed = c(123456, 456789, 342564),
```

Table S1: Estimates from informative prior distributions

| Model | Coefficient | Estimate | CrI |
|--------------------------|-------------|----------|-----------------|
| Bayesian MR-Egger | Intercept | 0.0046 | -0.0066, 0.0162 |
| Bayesian MR-Egger | Slope | 0.3225 | 0.0052, 0.6406 |
| Bayesian MR-Egger Radial | Intercept | 0.3847 | -0.4761, 1.2558 |
| Bayesian MR-Egger Radial | Slope | 0.3083 | -0.0044, 0.6143 |

```
n.chains = 3
)
```

Estimates in table S1 show shrinkage towards the null and the credible interval cuts across zero for the slope parameter. The estimates show the effects of different priors when using Bayesian models in MR for summary-level dataset.

In summary, it is helpful to compare estimates from models fitted with both uninformative and partially informative prior distributions.

References

1. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data mendelian randomization. *Statistics in Medicine*. Wiley Online Library; 2017;**36**(11):1783–1802.
2. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* [Internet]. 2015 Jun;**44**(2):512–525. Available from: <https://dx.doi.org/10.1093/ije/dyv080>
3. Plummer M. Rjags: Bayesian graphical models using mcmc [Internet]. 2018. Available from: <https://CRAN.R-project.org/package=rjags>
4. Gelman A, others. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*. International Society for Bayesian Analysis; 2006;**1**(3):515–534.
5. Jones E, Thompson J, Didelez V, Sheehan N. On the choice of parameterisation and priors for the Bayesian analyses of Mendelian randomisation studies. *Statistics in Medicine*. Wiley Online Library; 2012;**31**(14):1483–1501.
6. Bucur IG, Claassen T, Heskes T. Inferring the direction of a causal link and estimating its effect via a bayesian mendelian randomization approach. *Statistical Methods in Medical Research*. SAGE Publications Sage UK: London, England; 2019;0962280219851817.
7. Schmidt A, Dudbridge F. Mendelian randomization with Egger pleiotropy correction and weakly informative Bayesian priors. *International Journal of Epidemiology*. Oxford University Press; 2017;**47**(4):1217–1228.
8. Griffin JE, Brown PJ, others. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*. International Society for Bayesian Analysis; 2010;**5**(1):171–188.
9. Akiyama M, Okada Y, Kanai M, et al. Genome-wide association study identifies 112 new loci for body mass index in the japanese population. *Nature genetics*. Nature Publishing Group; 2017;**49**(10):1458.
10. Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature genetics*. Nature Publishing Group; 2018;**50**(4):524–537.