Bayesian estimation of IVW and MR-Egger models for two-sample Mendelian randomization studies

Okezie Uche-Ikonne[*]

Frank Dondelinger[†]

Tom Palmer[‡]

Abstract

We present our package, `mrbayes`, for the open source software environment R. The package implements Bayesian estimation of IVW and MR-Egger models, including the radial MR-Egger model, for summary-level data Mendelian randomization analyses. We have implemented a choice of prior distributions for the model parameters, namely; weakly informative, non-informative, a joint prior for the MR-Egger model slope and intercept, and a pseudo-horseshoe prior, or the user can specify their own prior. We show how to use the package through an applied example investigating the causal effect of BMI on insulin resistance. In future work, we plan to provide functions for alternative MCMC estimation software such as Stan and OpenBugs.

Availability

The package is freely available, under the MIT license, on GitHub here https://github.com/okezie94/mrbayes.

It can be installed in R using the following commands.

```
# install.packages("remotes") # uncomment if remotes not installed
remotes::install_github("okezie94/mrbayes")
```

There is a website of the package helpfiles at https://okezie94.github.io/mrbayes/.

Introduction

Observational epidemiology is limited by possible bias due to unmeasured confounding, reverse causation, and other issues. Mendelian randomization (MR) is a method of testing and estimating causal effects for the

[*]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK
[†]Faculty of Health and Medicine, Lancaster University, Lancaster, UK
[‡]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

aetiology of diseases.[1] MR uses genetic variants as instrumental variables related to a modifiable phenotype to estimate a causal effect of the phenotype on a disease outcome.

By including multiple instruments, we can increase power for hypothesis testing. Genome wide association studies (GWAS) provide many potential instruments, and we can obtain summary-level datasets for MR analyses. For two-sample MR, instrument-phenotype associations and instrument-outcome associations are obtained from different samples. The trade-off with this approach is the risk of violating the second and third instrumental variable assumptions due to pleiotropy.

The inverse variance weighted (IVW) model estimates the causal effect for multiple independent instruments in summary data. However, in the presence of pleiotropy its estimates are biased. Methods have been derived which estimate causal effects that are robust to pleiotropy; such as the MR-Egger model.[2] The MR-Egger model relies on its InSIDE (Instrument strength is independent of direct effect) assumption. The MR-Egger model has recently been adapted with its radial formulation which has the advantage that IVW is its sub-model.[3]

This paper introduces our `mrbayes` R package that implements Bayesian estimation of the IVW, MR-Egger, and Radial MR-Egger models. The Bayesian analysis is performed via the `rjags` package which provides an interface to the JAGS software, which performs Markov chain Monte Carlo estimation (MCMC), through R.[4]. Our package includes some specified prior distributions; non-informative, weakly informative, a shrinkage prior on the causal effect estimate (Pseudo-Horseshoe prior), and a joint prior on the intercept and causal effect estimate in the MR-Egger and radial MR-Egger models. The package also allows users to specify their own prior distributions.

The next section introduces the features of our `mrbayes` package, we then show how to use the package through an applied example. The supplementary material includes the methodology and additional examples.

Implementation

Our mrbayes package provides the following functions:

- `mr_format`, a function for setting up the summary-level dataset for analysis;
- `mr_ivw_rjags`, a function for estimating causal effects using the Bayesian IVW model, with a choice of prior distributions;
- `mr_egger_rjags`, a function for estimating causal effects through the Bayesian MR-Egger mode, with a choice of prior distributions;
- `mr_radialegger_rjags`, a function for performing Bayesian analysis under the radial formulation of MR-Egger.

The package allows users:

- to specify custom prior distributions for the estimate of the causal effect (`betaprior`) and optionally for the residual standard error (`sigmaprior`) for the MR-Egger models (original and radial). The prior distributions are written in the JAGS syntax. For more information on how to specify prior distributions see page 34 of JAGS manual;[5]
- to choose a random seed for reproducible results and to choose the number of chains for MCMC, each chain should have a different seed;
- to set parameter `rho`, the correlation coefficient between the average pleiotropic effect and causal estimate. This option is only relevant when using the *joint* prior method;
- to plot the posterior density and investigate the MCMC diagnostics.

The package also includes two summary-level datasets containing:

- 185 SNPs with multiple instrument-phenotype associations for low-density lipoprotein cholesterol (LDL-c), high-density lipoprotein (HDL-c) and triglycerides (trig) while the instrument-outcome associations for coronary heart disease (CHD);[6]

- 14 SNPs with instrument-phenotype associations of body mass index (BMI) and instrument-outcome associations of insulin resistance .[7]

The next section shows an applied example with step-by-step instructions on how to use the package.

Example: Investigating the effect of BMI on insulin resistance

We demonstrate the package using an investigation of the role of increased adiposity on insulin resistance in an MR study.[7] We compare frequentist estimates with Bayesian estimates using weakly informative and pseudo-horseshoe prior distributions.

Firstly, we load the package into our R session (See the Availability section for installation instructions).

```
library(mrbayes)
```

The next stage involves the setting up the dataset by using the `mr_format()` function;

```
dat <-
    mr_format(
    rsid = bmi_insulin$rsid,
    xbeta = bmi_insulin$beta.exposure,
    ybeta = bmi_insulin$beta.outcome,
```

```
    xse = bmi_insulin$se.exposure,
    yse = bmi_insulin$se.outcome
  )
```

The R syntax to use the weakly informative prior distributions is shown below.

```
vague_ivw <-
  mr_ivw_rjags(
    dat,
    prior = "weak",
    seed = c(123456, 456789, 342564),
    n.chains = 3)

vague_mregger <-
  mr_egger_rjags(
    dat,
    prior = "weak",
    seed = c(123456, 456789, 342564),
    n.chains = 3)

vague_radmregger <-
  mr_radialegger_rjags(
    dat,
    prior = "weak",
    seed = c(123456, 456789, 342564),
    n.chains = 3)
```

Table 1 shows that the frequentist and these Bayesian estimates are similar. The intercept estimate for the frequentist and Bayesian MR-Egger models is indicative of negative horizontal pleiotropy. Whereas, the radial MR-Egger model confidence and credible intervals provide stronger evidence of horizontal pleiotropy.

Table 1: Comparing estimates of frequentist method and weakly informative prior distribution.

Model	Coefficient	Estimate	95% Confidence/Credible Interval
Frequentist Model			
IVW	Slope	0.58	0.48, 0.68
MR-Egger	Intercept	-0.05	-1.49, 0.16
MR-Egger	Slope	3.76	-0.48, 8
MR-Egger Radial	Intercept	-32.09	-60.36, -3.83
MR-Egger Radial	Slope	6.66	1.25, 12.06
Weakly Informative Prior			
Bayesian IVW	Slope	0.58	0.48, 0.68
Bayesian MR-Egger	Intercept	-0.05	-0.12, 0.02
Bayesian MR-Egger	Slope	3.75	-0.94, 7.65
Bayesian MR-Egger Radial	Intercept	-33.77	-57.88, -7.74
Bayesian MR-Egger Radial	Slope	6.98	1.97, 11.56

We compare the results with estimates derived using shrinkage prior distributions. The syntax for the pseudo horseshoe prior is given below.

```
pseudo_ivw <-
  mr_ivw_rjags(dat,
               prior = "pseudo",
               seed = c(123456, 456789, 342564),
               n.chains = 3)
pseudo_mregger <-
  mr_egger_rjags(dat,
                 prior = "pseudo",
                 seed = c(123456, 456789, 342564),
                 n.chains = 3)
pseudo_radmregger <-
  mr_radialegger_rjags(dat,
                       prior = "pseudo",
                       seed = c(123456, 456789, 342564),
                       n.chains = 3)
```

The estimates from Table 2, compared to the estimates in Table 1, show the shrinkage of the intercept and slope for both the Bayesian MR-Egger and Radial MR-Egger models. The shrinkage prior has minimal effect

on Bayesian IVW mean estimate.

Table 2: Estimates from MR models using pseudo-horseshoe prior distributions.

Model	Coefficient	Estimate	95% Credible Interval
Bayesian IVW	Slope	0.58	0.48, 0.68
Bayesian MR-Egger	Intercept	-0.01	-0.08, 0.03
Bayesian MR-Egger	Slope	1.34	-1.14, 5.24
Bayesian MR-Egger Radial	Intercept	-11.61	-39.11, 8.53
Bayesian MR-Egger Radial	Slope	2.72	-1.01, 7.97

```
joint_mregger <-
  mr_egger_rjags(
    dat,
    prior = "joint",
    rho = 0.75,
    seed = c(123456, 456789, 342564),
    n.chains = 3
  )

joint_mrradegger <-
  mr_radialegger_rjags(
    dat,
    prior = "joint",
    rho = 0.75,
    seed = c(123456, 456789, 342564),
    n.chains = 3
  )
```

Figure 1 shows density plots for model estimates using three sets of priors; weakly informative, joint, and pseudo-horseshoe. The joint prior distributions are described in the Supplmentary material. The estimates using the joint prior distribution are similar to those from the weakly informative prior distribution. The estimates using the pseudo-horseshoe prior distribution are shrunk towards the null, which is arguably desirable in this example where we have evidence of pleiotropy.

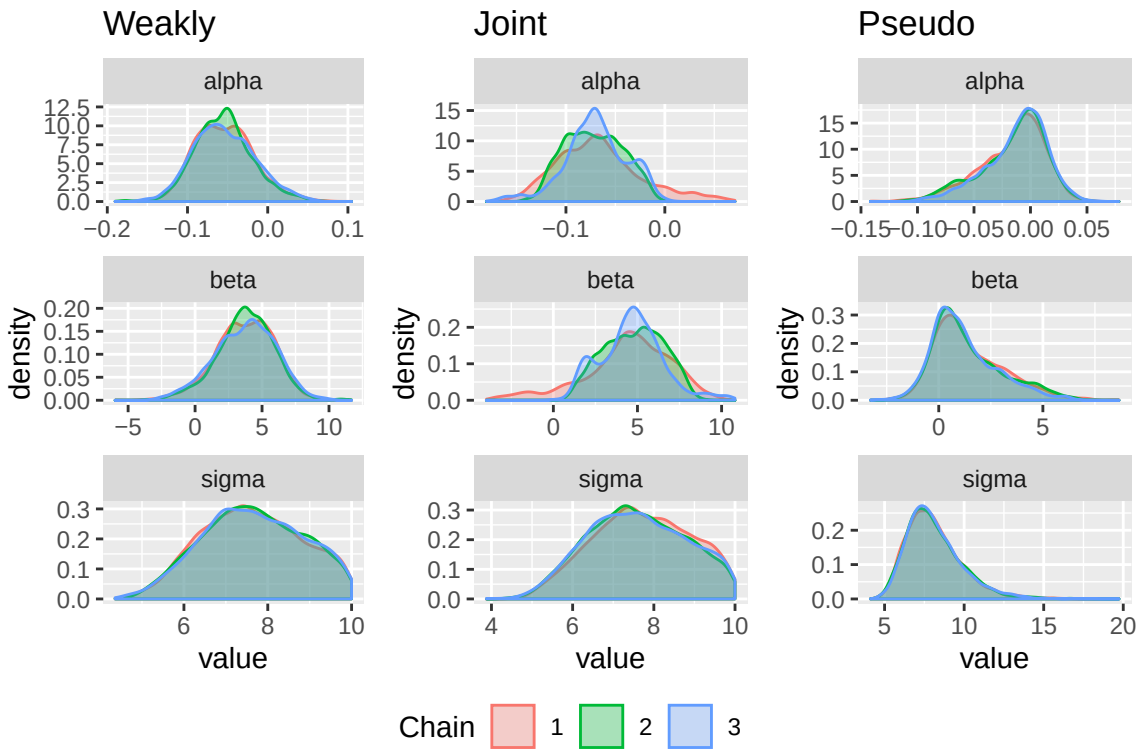

Figure 1: Density plots for MR-Egger model estimates using weakly informative, joint, and pseudo-horseshoe prior distributions.

Conclusion

We present an R package, mrbayes, to perform Bayesian estimation of the IVW and MR-Egger estimators implemented through the JAGS software. In our example we demonstrated the use of several different prior distributions to estimate the causal and average pleiotropic effects from these models.

Extending MR analysis with Bayesian estimation allows for the introduction of prior distributions on the model parameters. Several authors have considered Bayesian estimation of MR models including assessing different model parameterisations and Bayesian model averaging.[8,9] And the use of weakly informative prior distributions in the MR-Egger model has been shown to have good properties.[10]

There are several R packages providing functions for MR analyses. The `MendelianRandomization` and `TwoSampleMR` packages implement various two-sample MR methods.[11,12] The `RadialMR` R package implements the radial MR models and visualization of instruments through radial plots.[3,13] Our package complements these packages by offering the choice of four prior distributions for the causal effect; non-informative, weakly informative, pseudo-horseshoe, and a joint prior distribution for the MR-Egger model's intercept and slope.

In future work, we plan to provide functions for alternative MCMC estimation software such as Stan and OpenBugs.[14,15]

References

1. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology.* Oxford University Press; 2003;**32**(1):1–22.

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. Bowden J, Spiller W, Del Greco M F, et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *International Journal of Epidemiology* [Internet]. 2018 Jun;**47**(4):1264–1278. Available from: https://doi.org/10.1093/ije/dyy101

4. Plummer M. Rjags: Bayesian graphical models using mcmc [Internet]. 2018. Available from: https://CRAN.R-project.org/package=rjags

5. Plummer M. JAGS Version 3.3.0 user manual. Lyon, France: International Agency for Research on Cancer; 2012.

6. Do R, Willer CJ, Schmidt EM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nature Genetics.* Nature Publishing Group; 2013;**45**(11):1345.

7. Richmond R, Wade K, Corbin L, et al. Investigating the role of insulin in increased adiposity: Bi-directional Mendelian randomization study. *bioRxiv* [Internet]. Cold Spring Harbor Laboratory; 2017;155739. Available from: https://www.biorxiv.org/content/early/2017/06/28/155739

8. 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.

9. Burgess S, Thompson SG. Improving bias and coverage in instrumental variable analysis with weak instruments for continuous and binary outcomes. *Statistics in Medicine.* Wiley Online Library; 2012;**31**(15):1582–1600.

10. 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.

11. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *International Journal of Epidemiology*. Oxford University Press; 2017;**46**(6):1734–1739.

12. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. eLife Sciences Publications Limited; 2018;**7**:e34408.

13. Spiller W, Bowden J. RadialMR: A package for implementing radial inverse variance weighted and MR-Egger methods. [Internet]. 2019. Available from: https://github.com/WSpiller/RadialMR

14. Stan Development Team. RStan: the R interface to Stan [Internet]. 2018. Available from: http://mc-stan.org/

15. Thomas A, O'Hara B, Ligges U, Sturtz S. Making bugs open. *R News*. 2006;**6**(1):12–17.