

# ‘BWMM’ Package for causal inference based on summary statistics

Jia Zhao <sup>1,2</sup>, Jingsi Ming <sup>1</sup>, Xianghong Hu <sup>3,4</sup>,  
Gang Chen <sup>5</sup>, Jin Liu <sup>6</sup>, and Can Yang <sup>1</sup>

<sup>1</sup> Department of Mathematics, The Hong Kong University of Science and Technology,  
Hong Kong.

<sup>2</sup> School of Mathematical Sciences, Beijing Normal University, Beijing, China.

<sup>3</sup> Department of Mathematics, Hong Kong Baptist University, Hong Kong.

<sup>4</sup> Department of Mathematics, Southern University of Science and Technology,  
Shenzhen, China.

<sup>5</sup> The WeGene Company, Shenzhen, China.

<sup>6</sup> Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore.

September 10, 2024

## 1 Overview

This vignette provides an introduction to the ‘BWMM’ package. R package ‘BWMM’ implements BWMM (Bayesian Weighted Mendelian Randomization), an efficient statistical approach for causal inference based on summary statistics. It provides model parameter estimation as well as statistical inference.

The package can be loaded with the command:

```
> library("BWMM")
```

This vignette is organized as follows. Section 2.1 discusses the data processing procedure for the BWMM. Section 2.2 explains command lines for performing MR analysis and visualizing data and MR result using the ‘BWMM’ package.

Please feel free to contact Jia Zhao at [jiazh0103@163.com](mailto:jiazh0103@163.com) for any questions or suggestions regarding the ‘BWMM’ package.

## 2 Workflow

### 2.1 Data preprocessing for the BWMM

In this vignette, we use an `ExampleData` (LDL cholesterol - dyslipidemia) from our real data analysis in the package. Users can find the SNP-exposure effects in the `ExampleData$beta.exposure`, the SNP-outcome effects in the `ExampleData$beta.outcome`, standard errors of SNP-exposure effects and SNP-outcome effects in `ExampleData$se.exposure` and `ExampleData$se.outcome` respectively.

```
> data(ExampleData)
> head(ExampleData)
```

	SNP	beta.exposure	beta.outcome	se.exposure	se.outcome
1	rs10102164	0.03147495	0.03342949	0.004482193	0.007697326
2	rs10195252	0.02206123	0.01835167	0.003615076	0.006201982
3	rs10401969	0.10908016	0.04495502	0.006633253	0.011908614
4	rs10490626	0.04634563	0.04448907	0.006294977	0.011214789
5	rs10832962	0.03025899	0.01336614	0.003782374	0.006907564
6	rs10893499	0.04761681	0.03467897	0.004843937	0.009269974

When we analyze real data, we need to ensure that the summary statistics for ‘BWMM’ are obtained through the following data preprocessing process:

1. Select SNPs that are significant for the exposure as instrument variables (IVs) by a  $p$ -value threshold (e.g. GWAS significant threshold  $5 \times 10^{-8}$ ).
2. Perform LD clumping to ensure that the IVs are independent.
3. Extract the IVs from summary-level database for the outcomes of interest.
4. Perform harmonisation of effects to ensure that SNP-exposure effect and SNP-outcome effect correspond to the same allele.
5. Scale the summary-level data if necessary.

## 2.2 Fitting the BWMM

We are now ready to fit BWMM using the data described above:

```
> fit.BWMM <- BWMM(gammahat = ExampleData$beta.exposure,
+                 Gammahat = ExampleData$beta.outcome,
+                 sigmaX = ExampleData$se.exposure,
+                 sigmaY = ExampleData$se.outcome)
```

The MR results are as follows:

```
> cat("Effect of risk exposure on the outcome of interest: ", fit.BWMM$beta)
Effect of risk exposure on the outcome of interest: 0.6177372
> cat("Standard error of the effect of exposure on outcome: ", fit.BWMM$se_beta)
Standard error of the effect of exposure on outcome: 0.0256347
> cat("p-value of the effect of exposure on outcome: ", fit.BWMM$P_value)
p-value of the effect of exposure on outcome: 2.642837e-128
```

Besides, the ‘BWMM’ function also provides the visualization of data and MR result.

Plot of data with standard error bar:

Plot of evidence lower bound (elbo):

Posterior mean of weight of each observation:

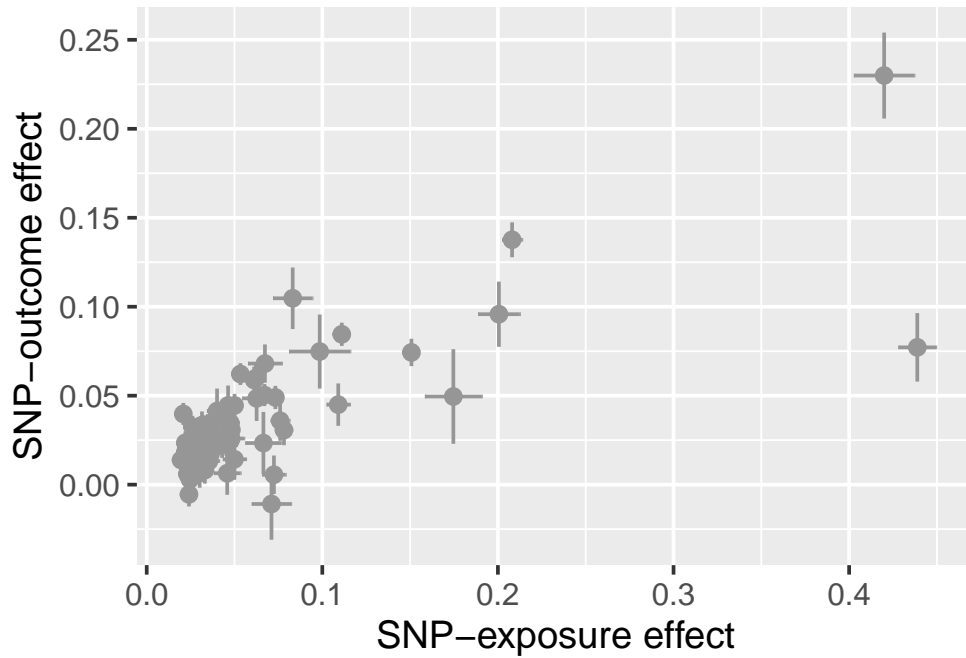
Plot of weighted data and its regression result:

If we are only interested in the MR results, we can run the ‘BWMM\_quick’ function without visualization, which is more time-efficient.

```
> fit.BWMM_quick <- BWMM_quick(gammahat = ExampleData$beta.exposure,
+                             Gammahat = ExampleData$beta.outcome,
+                             sigmaX = ExampleData$se.exposure,
+                             sigmaY = ExampleData$se.outcome)
```

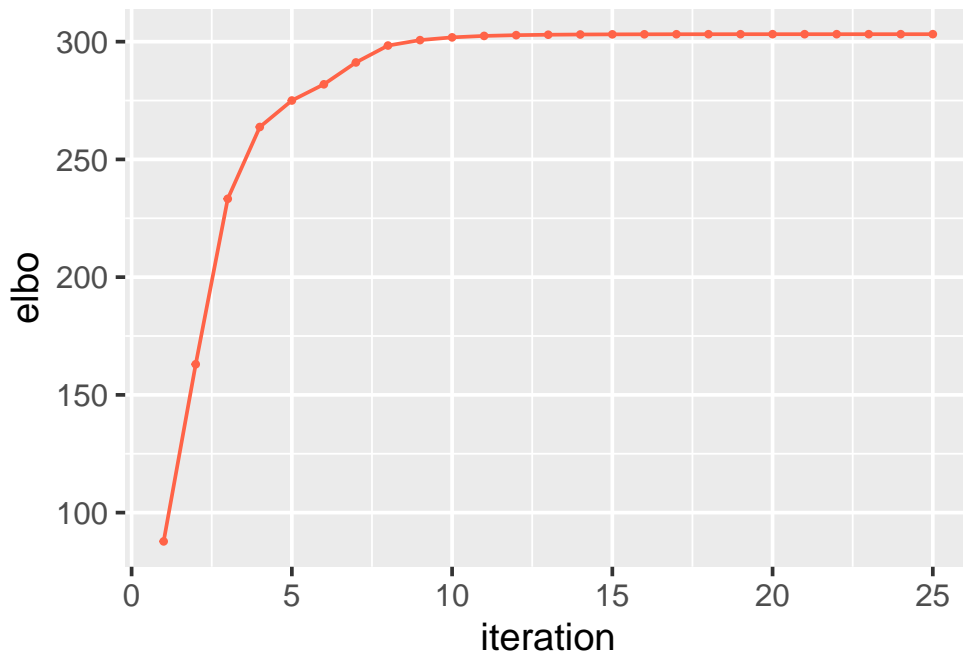
```
> fit.BWMR$plot1
```

Plot1: Plot of data with standard error bar

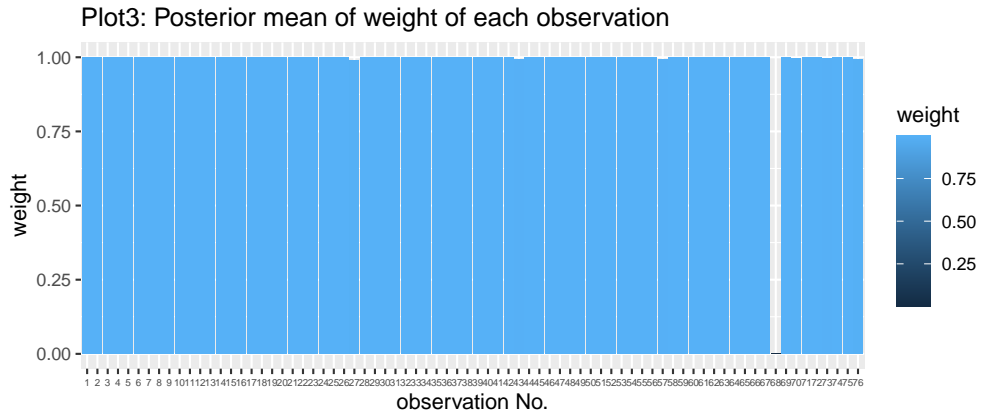


```
> fit.BWMR$plot2
```

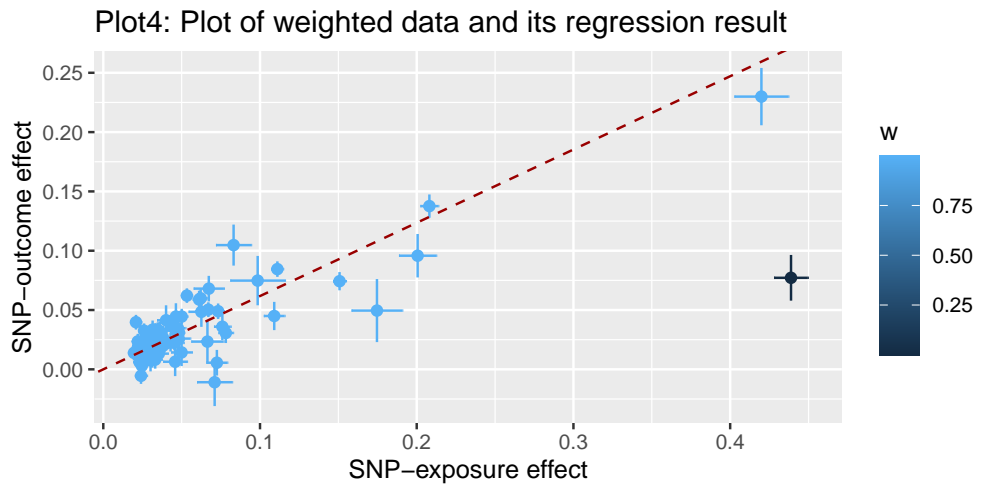
Plot2: Plot of evidence lower bound (elbo)



```
> fit.BWMR$plot3
```



```
> fit.BWMR$plot4
```



```
> fit.BWMR_quick$beta
[1] 0.6177372
> fit.BWMR_quick$se_beta
[1] 0.0256347
> fit.BWMR_quick$P_value
[1] 2.642837e-128
```

## References

Jia Zhao, Jingsi Ming, Xianghong Hu, Gang Chen, Jin Liu, Can Yang; Bayesian weighted Mendelian randomization for causal inference based on summary statistics.