MR_analysis

Namratha Shivani Chalasani

Installing and Loading packages

```
#install.packages("remotes")
#install.packages("forestplot")
#remotes::install_github("MRCIEU/TwoSampleMR")

library(TwoSampleMR)

## TwoSampleMR version 0.5.11
## [>] New: Option to use non-European LD reference panels for clumping etc
## [>] Some studies temporarily quarantined to verify effect allele
## [>] See news(package = 'TwoSampleMR') and https://gwas.mrcieu.ac.uk for
further details

library(MRPRESSO)
library(forestplot)

## Loading required package: grid

## Loading required package: checkmate

## Loading required package: abind

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)
```

Reading GWAS summary statistics

The data utilized for Mendelian Randomization (MR) analysis, comprising the exposure Body Mass Index (BMI) dataset and the outcome Type 2 Diabetes (T2D) dataset, is sourced from the IEU OpenGWAS project.

```
BMI <- extract_instruments(outcomes = 'ieu-a-2')
Diabetes <- extract_outcome_data(BMI$SNP, outcomes = 'ieu-a-976')

## Extracting data for 79 SNP(s) from 1 GWAS(s)
```

Harmonizing the Data

In Mendelian Randomization (MR), data harmonization involves ensuring consistency between genetic variants used as instruments for the exposure and outcome variables. This often includes flipping the reference (Ref) and alternate (Alt) alleles in the outcome dataset to match those in the exposure dataset. By aligning the alleles, the effect sizes (betas) estimated for the genetic variants remain consistent across both datasets. This harmonization process helps mitigate potential biases introduced by allele inconsistencies and ensures the validity of causal inference in MR.

```
data <- harmonise_data(exposure_dat = BMI,
                       outcome_dat = Diabetes)

## Harmonising Body mass index || id:ieu-a-2 (ieu-a-2) and Type 2 diabetes ||
id:ieu-a-976 (ieu-a-976)

## Removing the following SNPs for being palindromic with intermediate allele
frequencies:
## rs14810, rs17001654, rs9304665, rs9579083

head(data)

##          SNP effect_allele.exposure other_allele.exposure
effect_allele.outcome
## 1  rs1000940                      G                     A
G
## 2 rs10132280                      A                     C
A
## 3  rs1016287                      C                     T
C
## 4 rs10182181                      G                     A
G
## 5 rs10733682                      G                     A
G
## 6 rs10840100                      G                     A
G
##   other_allele.outcome beta.exposure beta.outcome eaf.exposure eaf.outcome
## 1                    A        0.0184   0.02316750       0.2250          NA
## 2                    C       -0.0221  -0.03208180       0.3333          NA
## 3                    T       -0.0228  -0.02884500       0.6750          NA
## 4                    A        0.0309  -0.02292890       0.5000          NA
## 5                    A       -0.0188   0.00131613       0.5750          NA
## 6                    A        0.0206  -0.00915982       0.7250          NA
##   remove palindromic ambiguous id.outcome chr       pos se.outcome
## 1  FALSE       FALSE     FALSE  ieu-a-976  17   5283252  0.0241727
```

```
## 2  FALSE       FALSE     FALSE  ieu-a-976  14  25928179  0.0246112
## 3  FALSE       FALSE     FALSE  ieu-a-976   2  59305625  0.0250353
## 4  FALSE       FALSE     FALSE  ieu-a-976   2  25150296  0.0226273
## 5  FALSE       FALSE     FALSE  ieu-a-976   9 129460914  0.0228711
## 6  FALSE       FALSE     FALSE  ieu-a-976  11   8669437  0.0234505
##   samplesize.outcome pval.outcome                        outcome
## 1              64171     0.337840 Type 2 diabetes || id:ieu-a-976
## 2              64171     0.192367 Type 2 diabetes || id:ieu-a-976
## 3              64171     0.249232 Type 2 diabetes || id:ieu-a-976
## 4              64171     0.310901 Type 2 diabetes || id:ieu-a-976
## 5              64171     0.954081 Type 2 diabetes || id:ieu-a-976
## 6              64171     0.696105 Type 2 diabetes || id:ieu-a-976
##   originalname.outcome      outcome.deprecated mr_keep.outcome
## 1      Type 2 diabetes Type 2 diabetes ||  ||             TRUE
## 2      Type 2 diabetes Type 2 diabetes ||  ||             TRUE
## 3      Type 2 diabetes Type 2 diabetes ||  ||             TRUE
## 4      Type 2 diabetes Type 2 diabetes ||  ||             TRUE
## 5      Type 2 diabetes Type 2 diabetes ||  ||             TRUE
## 6      Type 2 diabetes Type 2 diabetes ||  ||             TRUE
##   data_source.outcome pval.exposure samplesize.exposure chr.exposure
## 1                igd   1.81201e-08              338903           17
## 2                igd   1.40088e-11              338856           14
## 3                igd   4.35512e-12              339033            2
## 4                igd   8.07049e-26              338829            2
## 5                igd   2.45499e-10              336886            9
## 6                igd   6.66653e-12              339135           11
##   se.exposure pos.exposure id.exposure                     exposure
## 1      0.0033      5283252     ieu-a-2 Body mass index || id:ieu-a-2
## 2      0.0033     25928179     ieu-a-2 Body mass index || id:ieu-a-2
## 3      0.0033     59305625     ieu-a-2 Body mass index || id:ieu-a-2
## 4      0.0029     25150296     ieu-a-2 Body mass index || id:ieu-a-2
## 5      0.0030    129460914     ieu-a-2 Body mass index || id:ieu-a-2
## 6      0.0030      8669437     ieu-a-2 Body mass index || id:ieu-a-2
##   mr_keep.exposure pval_origin.exposure data_source.exposure action
SNP_index
## 1             TRUE             reported                  igd      2
1
## 2             TRUE             reported                  igd      2
1
## 3             TRUE             reported                  igd      2
1
## 4             TRUE             reported                  igd      2
1
## 5             TRUE             reported                  igd      2
1
## 6             TRUE             reported                  igd      2
1
##   mr_keep
## 1    TRUE
## 2    TRUE
```

```
## 3    TRUE
## 4    TRUE
## 5    TRUE
## 6    TRUE
```

MR analysis

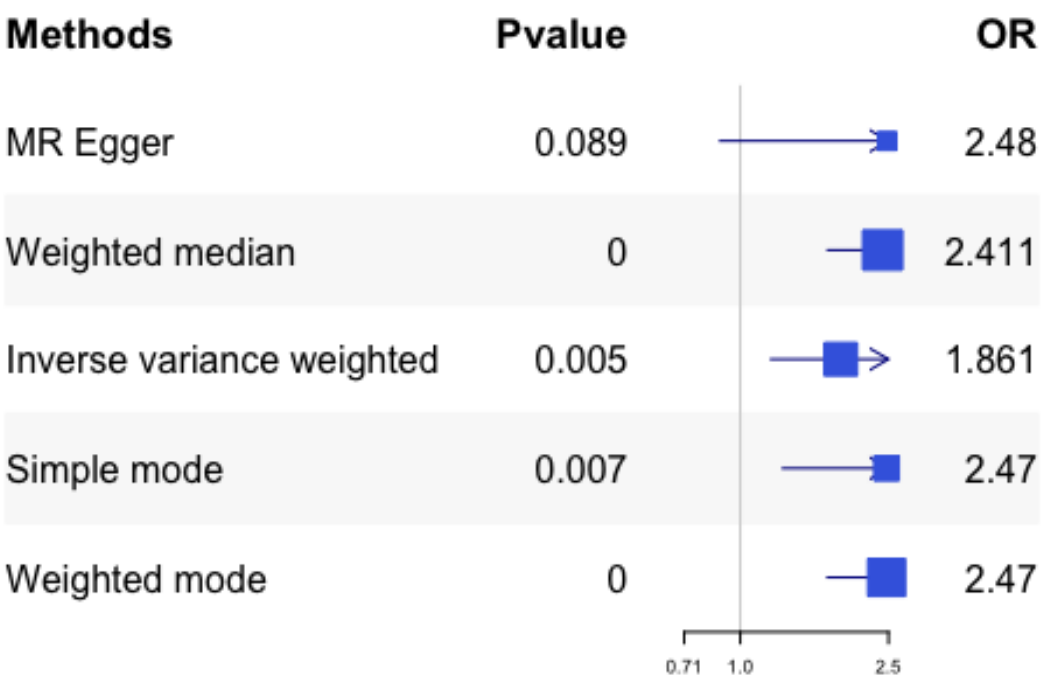
Conducting MR analysis and generating odds ratios using TwoSample MR using MR-Egger, IVW, Weighted-mode, Simple-mode methods.

```
mr_result <- generate_odds_ratios(mr(data))

## Analysing 'ieu-a-2' on 'ieu-a-976'

mr_result

##   id.exposure id.outcome                          outcome
## 1     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
## 2     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
## 3     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
## 4     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
## 5     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
##                         exposure                    method nsnp         b
## 1 Body mass index || id:ieu-a-2                  MR Egger   75 0.9083682
## 2 Body mass index || id:ieu-a-2           Weighted median   75 0.8800713
## 3 Body mass index || id:ieu-a-2 Inverse variance weighted   75 0.6209116
## 4 Body mass index || id:ieu-a-2               Simple mode   75 0.9042006
## 5 Body mass index || id:ieu-a-2             Weighted mode   75 0.9042006
##          se          pval      lo_ci    up_ci       or  or_lci95 or_uci95
## 1 0.5276226 8.937277e-02 -0.1257721 1.942508 2.480272 0.8818158 6.976228
## 2 0.1733182 3.818724e-07  0.5403677 1.219775 2.411072 1.7166380 3.386425
## 3 0.2200615 4.779424e-03  0.1895910 1.052232 1.860623 1.2087551 2.864037
## 4 0.3275069 7.264578e-03  0.2622870 1.546114 2.469957 1.2998996 4.693198
## 5 0.1863153 6.573874e-06  0.5390226 1.269379 2.469957 1.7143305 3.558641
```

plotiing Forest plots based on the odds ratios generated

```
forest_plot_data <- tibble::tibble(mean = round(mr_result$or, digits = 3),
                                   methods = c(mr_result$method),
                                   lower = round(mr_result$or_lci95, digits =
3),
                                   upper = round(mr_result$or_uci95, digits =
3),
                                   pvalue = round(mr_result$pval, digits =3
),
                                   Beta = mr_result$b,
                                   OR = round(mr_result$or, digits = 3)
                                   )

forest_plot_data |>
```

```
forestplot(labeltext = c(methods, pvalue, OR),
           clip = c(0.1, 2.5),
           xlog = TRUE) |>
fp_set_style(box = "royalblue",
             line = "darkblue") |>
fp_add_header(methods = c("", "Methods"),
              pvalue = c("", "Pvalue"),
              OR = c("", "OR")) |>
fp_decorate_graph(graph.pos = 3) |>
fp_set_zebra_style("#f9f9f9")
```



Interpretation : Collectively, the findings strongly support the conclusion that alterations in body mass index may causally impact the risk of developing Type 2 diabetes.

Calculating the Heterogeneity and pleiotropy

Heterogeneity in Mendelian Randomization (MR) refers to the variability in causal estimates across different genetic variants used as instrumental variables or across different studies.

Addressing heterogeneity in MR involves careful consideration of potential sources of variability and the implementation of appropriate statistical methods to assess and account for heterogeneity. Sensitivity analyses, meta-analyses, and robust statistical modeling techniques are commonly employed to explore and mitigate heterogeneity in MR studies, enhancing the reliability and generalizability of causal estimates.

```
mr_heterogeneity(data)
```

```
##   id.exposure id.outcome                          outcome
## 1     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
## 2     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
##                         exposure                    method        Q Q_df
## 1 Body mass index || id:ieu-a-2                  MR Egger 309.6789   73
## 2 Body mass index || id:ieu-a-2 Inverse variance weighted 311.2061   74
##         Q_pval
## 1 6.546912e-31
## 2 7.547105e-31
```

```
mr_pleiotropy_test(data)
```

```
##   id.exposure id.outcome                          outcome
## 1     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
##                         exposure egger_intercept         se      pval
## 1 Body mass index || id:ieu-a-2    -0.008761857 0.01460324 0.5503683
```

Heterogeneity Result Interpretation : both MR Egger and Inverse variance weighted methods detected significant heterogeneity ($p < 0.05$) between the exposure (Body mass index) and the outcome (Type 2 diabetes), as evidenced by the low p-values and elevated test statistics.

Pleiotropy result interpretation : The Egger intercept is approximately -0.0088 with a standard error of 0.0146 and a p-value of 0.5504. Since the p-value is greater than the typical significance threshold of 0.05, there is no strong evidence to reject the null hypothesis of no directional pleiotropy. Therefore, based on this test, there is no significant evidence of pleiotropy between Body mass index and Type 2 diabetes in the analyzed data.

Checking the Directionality

Directionality refers to the causal relationship between the exposure and the outcome, indicating whether the exposure variable influences the outcome or vice versa.

checking directionality in MR analysis helps ensure the validity and reliability of causal inference and enhances the interpretation of study findings.

```
directionality_test(data)
```

```
## r.exposure and/or r.outcome not present.
```

```
## Calculating approximate SNP-exposure and/or SNP-outcome correlations,
assuming all are quantitative traits. Please pre-calculate r.exposure and/or
r.outcome using get_r_from_lor() for any binary traits

##   id.exposure id.outcome                       exposure
## 1     ieu-a-2  ieu-a-976 Body mass index || id:ieu-a-2
##                          outcome snp_r2.exposure snp_r2.outcome
## 1 Type 2 diabetes || id:ieu-a-976      0.01580819    0.005398399
##   correct_causal_direction steiger_pval
## 1                     TRUE 2.403171e-34
```

Directionality Test interpretation : The p-value (2.403171e-34) associated with the Steiger test indicates a significant difference between the SNP-exposure and SNP-outcome correlations, supporting the correct specification of the causal direction between body mass index (exposure) and Type 2 diabetes (outcome).

Sensitivity Analysis

Sensitivity analysis involves testing the robustness of the MR results by assessing the impact of potential outliers or influential data points on the causal estimates.

Here MRPRESSO outlier test is used as it specifically identifies and addresses outliers in the genetic instrumental variable (IV) data, which may affect the validity of the MR analysis.

```
mr_presso_res <- mr_presso(BetaExposure = 'beta.exposure', BetaOutcome =
'beta.outcome', SdExposure = 'se.exposure', SdOutcome = 'se.outcome',
  data = data, OUTLIERtest = TRUE, DISTORTIONtest = TRUE, NbDistribution =
10000)

mr_presso_res

## $`Main MR results`
##        Exposure       MR Analysis Causal Estimate        Sd   T-stat
## 1 beta.exposure               Raw       0.6183634 0.2106576 2.935395
## 2 beta.exposure Outlier-corrected       0.7704031 0.1244539 6.190268
##        P-value
## 1 4.374980e-03
## 2 2.708571e-08
##
## $`MR-PRESSO results`
## $`MR-PRESSO results`$`Global Test`
## $`MR-PRESSO results`$`Global Test`$RSSobs
## [1] 319.0294
##
## $`MR-PRESSO results`$`Global Test`$Pvalue
## [1] "<1e-04"
##
##
## $`MR-PRESSO results`$`Outlier Test`
```

```
##          RSSobs  Pvalue
## 1  1.407968e-04       1
## 2  3.452897e-04       1
## 3  2.215059e-04       1
## 4  1.842399e-03       1
## 5  1.700148e-04       1
## 6  4.878327e-04       1
## 7  1.647816e-04       1
## 8  7.647421e-03       1
## 9  1.108316e-03       1
## 10 2.456651e-03       1
## 11 1.596339e-03       1
## 12 3.852070e-02  0.1343
## 13 9.542468e-05       1
## 14 6.717585e-03       1
## 15 2.512366e-04       1
## 16 1.073830e-03       1
## 17 5.475276e-04       1
## 18 5.518612e-06       1
## 19 1.831382e-04       1
## 20 4.354957e-04       1
## 21 8.378820e-05       1
## 22 1.218053e-05       1
## 23 1.158328e-03       1
## 24 2.926518e-04       1
## 25 1.184253e-03       1
## 26 2.388184e-03       1
## 27 2.571719e-05       1
## 28 1.171223e-04       1
## 29 5.774064e-03       1
## 30 5.890077e-06       1
## 31 2.237732e-03       1
## 32 3.950530e-05       1
## 33 6.093417e-04       1
## 34 5.108176e-05       1
## 35 1.322180e-04       1
## 36 5.763996e-04       1
## 37 8.650461e-04       1
## 38 8.476025e-04       1
## 39 1.097151e-03       1
## 40 1.492613e-04       1
## 41 3.699447e-04       1
## 42 3.226501e-03       1
## 43 3.786581e-04       1
## 44 2.268574e-04       1
## 45 2.708093e-03       1
## 46 1.247610e-03       1
## 47 6.604723e-04       1
## 48 6.814920e-06       1
## 49 1.419302e-04       1
```

```
## 50 3.841437e-05       1
## 51 5.756781e-07       1
## 52 1.961281e-05       1
## 53 2.143705e-04       1
## 54 1.290307e-03       1
## 55 1.951408e-04       1
## 56 3.487441e-07       1
## 57 1.669679e-04       1
## 58 2.288047e-03       1
## 59 4.303371e-05       1
## 60 1.903820e-03       1
## 61 1.648001e-03       1
## 62 9.279227e-05       1
## 63 3.147217e-04       1
## 64 1.360698e-04       1
## 65 7.705805e-05       1
## 66 1.419207e-02       1
## 67 2.381225e-05       1
## 68 1.360127e-03       1
## 69 5.096330e-03       1
## 70 1.268004e-01 <0.0079
## 71 6.226571e-04       1
## 72 9.262736e-05       1
## 73 4.607899e-04       1
## 74 5.927021e-03  0.1975
## 75 1.017119e-06       1
## 76 7.592083e-04       1
## 77 2.765990e-04       1
## 78 7.097085e-05       1
## 79 7.105848e-04       1
##
## $`MR-PRESSO results`$`Distortion Test`
## $`MR-PRESSO results`$`Distortion Test`$`Outliers Indices`
## [1] 70
##
## $`MR-PRESSO results`$`Distortion Test`$`Distortion Coefficient`
## beta.exposure
##     -19.73508
##
## $`MR-PRESSO results`$`Distortion Test`$Pvalue
## [1] 0.2752
```

MRPRESSO has identified one outlier SNP (rs7903146).

Removing the outlier and checking if the heterogeneity present in the analysis is reducing or not

Reanalysis for Heterogeneity

```
mr_heterogeneity(data[-mr_presso_res$`MR-PRESSO results`$`Distortion
Test`$`Outliers Indices`,])
```

```
##   id.exposure id.outcome                         outcome
## 1     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
## 2     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
##                        exposure                    method        Q Q_df
## 1 Body mass index || id:ieu-a-2                  MR Egger 105.1144   72
## 2 Body mass index || id:ieu-a-2 Inverse variance weighted 105.1971   73
##        Q_pval
## 1 0.006630215
## 2 0.008111148
```

The test statistic has indeed decreased significantly from around 309 to approximately 105 after removing the outlier detected by MR PRESSO. This reduction in the test statistic indicates a decrease in the overall heterogeneity in the Mendelian Randomization (MR) analysis.

However, despite the decrease, both MR Egger and Inverse Variance Weighted methods still show evidence of significant heterogeneity, suggesting that some level of variability remains in the causal estimates obtained from different genetic variants.

This suggests that there may be underlying differences in the causal estimates obtained from different genetic variants, indicating potential complexities or biases in the MR analysis.

checking if the removal of the outlier changes the MR result significantly

```
mr_result_out <- generate_odds_ratios(mr(data[-mr_presso_res$`MR-PRESSO
results`$`Distortion Test`$`Outliers Indices`,]))
```

```
## Analysing 'ieu-a-2' on 'ieu-a-976'
```

```
mr_result_out
```

```
##   id.exposure id.outcome                         outcome
## 1     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
## 2     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
## 3     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
## 4     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
## 5     ieu-a-2  ieu-a-976 Type 2 diabetes || id:ieu-a-976
##                        exposure                    method nsnp         b
## 1 Body mass index || id:ieu-a-2                  MR Egger   74 0.8458729
## 2 Body mass index || id:ieu-a-2           Weighted median   74 0.8805478
## 3 Body mass index || id:ieu-a-2 Inverse variance weighted   74 0.7790260
## 4 Body mass index || id:ieu-a-2               Simple mode   74 0.9271389
## 5 Body mass index || id:ieu-a-2             Weighted mode   74 0.8805538
##          se         pval     lo_ci    up_ci       or or_lci95 or_uci95
## 1 0.3095687 7.903902e-03 0.2391183 1.452627 2.330011 1.270129 4.274330
## 2 0.1687213 1.799406e-07 0.5498541 1.211241 2.412221 1.733000 3.357650
```

```
## 3 0.1294950 1.789150e-09 0.5252158 1.032836 2.179349 1.690824 2.809021
## 4 0.3102816 3.822373e-03 0.3189869 1.535291 2.527268 1.375733 4.642676
## 5 0.1857524 1.026963e-05 0.5164792 1.244628 2.412235 1.676116 3.471644
```

Interpretation : There is no significant changes observed in the results before and after the removal of outliers. The results remain constant, there is significant causal relation between BMI and Type 2 Diabetes.