# Practical MR session using R

Daniele Bottigliengo

Institute for Biomedicine, Eurac Research, Bolzano, Italy

July 22th, 2021


# Outline

1. **Introduction**
    - *Two-sample MR framework*
    - *R packages for Two-sample MR*
2. **Data preparation**
    - *Summary GWAS data*
    - *Data harmonization*
3. **Performing MR analysis**
    - `MendelianRandomization` *and* `TwoSampleMR` *functions*
    - *Displaying results*
4. **Sensitivity analyses**
    - *Evaluating MR assumptions*
    - *Using MR robust methods*
5. **Case study**

# Introduction

- Two sample MR analysis
- Summary instrument-exposure and instrument-outcome association results
- Summary-level data from GWAS
- Non-overlapping sets of individuals

# R packages

## `TwoSampleMR`

- GWAS database
- Data harmonization
- Most commonly used MR methods
- Visualization and sensitivity analyses
- Installation from Github:

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

## `MendelianRandomization`

- GWAS database
- Several MR methods are implemented
- Visualization and sensitivity analyses
- Installation from CRAN:

```
install.packages("MendelianRandomization", dependencies
```

# Data sources

- Retrieving summary-level GWAS data:
  - Phenoscanner `→ MendelianRandomization`
  - IEU GWAS database `→ TwoSampleMR`

# IVs selection

- Several strategies for selecting genetic instruments

- Accounting for:
  - weak instrument bias
  - linkage disequilibrium (LD), i.e. correlations between instruments
  - horizontal pleiotropy

# IVs selection

1. Selection of genome-wide significant instruments from GWAS exposure data $(p - value < 5 \times 10^{-8})$
    - One can select functional genetic variants, i.e. genome-wide significant SNPs in the encoding gene
    - Check $F - statistic$ to evaluate the strength of the instrument
2. Perform LD clumping to select independent genetic instruments
    - Default at $r^2 < 0.001$
3. Proxies can be used to replace exposure variants that are missing in the outcome dataset
    - The proxy should be in high LD with the missing genetic instrument
4. Applying Steiger filter to remove IVs for which there might evidence of reverse causation

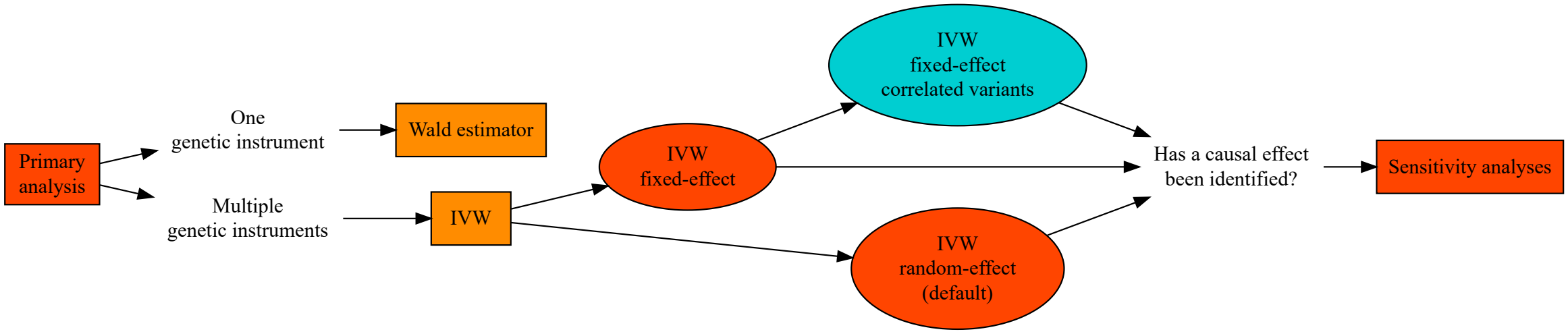
# Data harmonization

**Harmonise exposure and outcome effects**

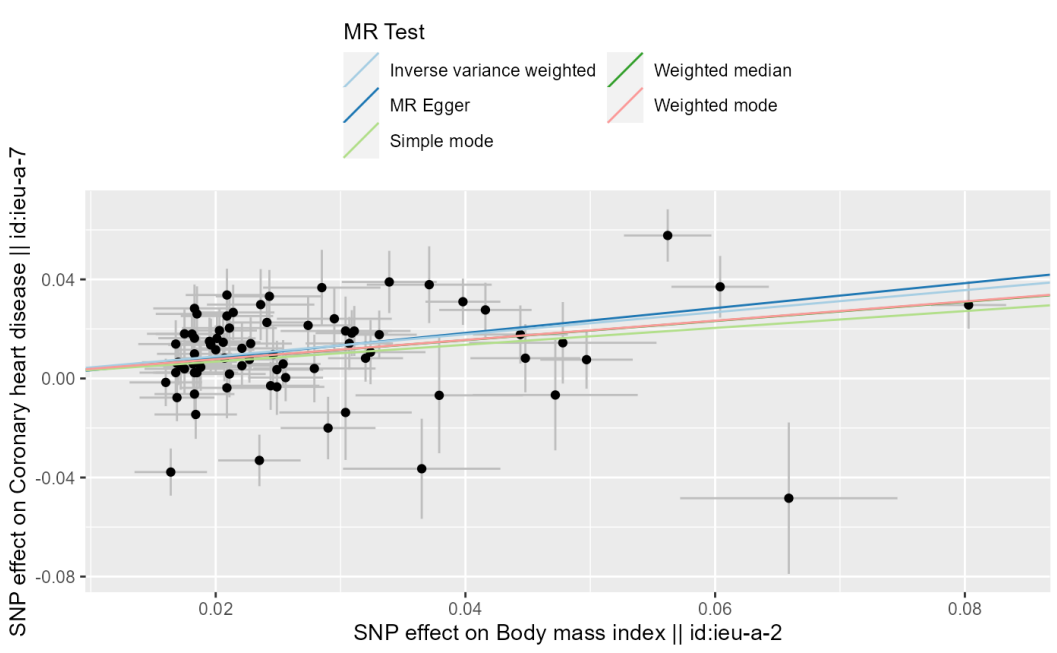
| | Exposure GWAS | | | | Outcome GWAS | | | |
|---|---|---|---|---|---|---|---|---|
| *SNP* | *Effect* | *Effect allele* | *Other allele* | *Effect allele frequency* | *Effect* | *Effect allele* | *Other allele* | *Effect allele frequency* |
| *rs12345* | 0.132 | A | G | 0.28 | 0.022 | A | G | 0.26 |
| *rs23456* | -0.485 | G | T | 0.41 | **0.056** | **T** | **G** | **0.61** |
| *rs34567* | 0.203 | G | C | 0.11 | **-0.046** | **G** | **C** | **0.88** |

| | Exposure GWAS | | | | Outcome GWAS | | | |
|---|---|---|---|---|---|---|---|---|
| *SNP* | *Effect* | *Effect allele* | *Other allele* | *Effect allele frequency* | *Effect* | *Effect allele* | *Other allele* | *Effect allele frequency* |
| *rs12345* | 0.132 | A | G | 0.28 | 0.022 | A | G | 0.26 |
| *rs23456* | -0.485 | G | T | 0.41 | **-0.056** | **G** | **T** | **0.39** |
| *rs34567* | 0.203 | G | C | 0.11 | **0.046** | **G** | **C** | **0.12** |

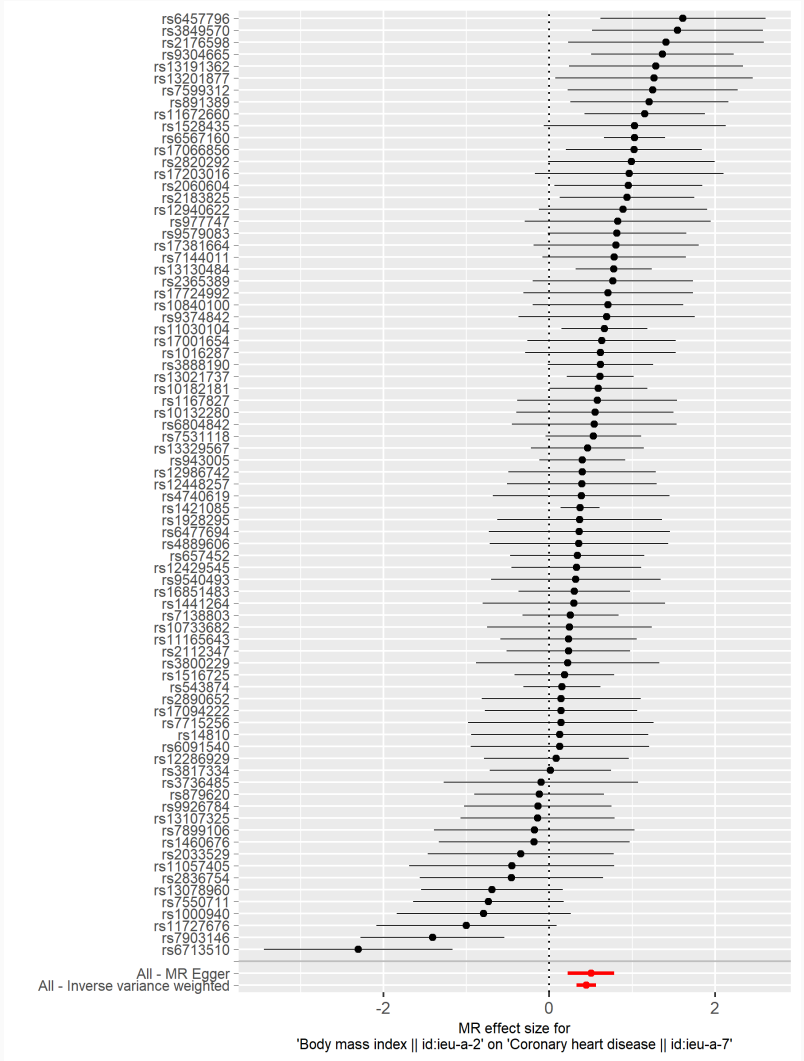
https://elifesciences.org/articles/34408


Primary analysis
One genetic instrument
Wald estimator
Multiple genetic instruments
IVW
IVW fixed-effect
IVW fixed-effect correlated variants
IVW random-effect (default)
Has a causal effect been identified?
Sensitivity analyses

# MR analysis




https://mrcieu.github.io/TwoSampleMR/articles/perform_mr.html



# Sensitivity analyses

- Evaluation of MR assumptions
- Check for the presence of pleiotropic effects
  - Heterogeneity of Wald estimates using Q-statistic or $I^2$
  - Intercept of MR-Egger
- Leave-one-out analysis to assess the reliance of MR analysis on a particular variant

# Robust MR methods

- Valid causal inferences under weaker assumptions than the standard IVW methods
- Using multiple methods that make different assumptions about the nature of the underlying pleiotropy
- Most commonly used robust methods:
  - *MR-Egger* $\longrightarrow$ `TwoSampleMR` and `MendelianRandomization`
  - *Median based method* $\longrightarrow$ `TwoSampleMR` and `MendelianRandomization`
  - *Mode based method* $\longrightarrow$ `TwoSampleMR` and `MendelianRandomization`
  - *MR-PRESSO* $\longrightarrow$ `MR-PRESSO` R package
- A simulation study[1] has shown that two recent methods have good statistical properties in a wide range of scenarios:
  - *Contamination mixture* $\longrightarrow$ implemented in `MendelianRandomization` with `mr_conmix` function
  - *MR-Mix* $\longrightarrow$ `MR-Mix` R package

[1] Slob EAW, Burgess S. **A comparison of robust Mendelian randomization methods using summary data**. *Genetic Epidemiology*. 2020;44:313–29.

# Other approaches to sensitivity analysis

- Using different datasets
- Positive and negative control outcomes
- Subgroup analyses
- Evaluate association with potentially pleiotropic variables

# Case study

- **Aim**: to evaluate the causal relationship between kidney function and blood pressure (BP)
- **Exposure**: estimated glomerular filtration rate from serum creatinine (eGFRcr)
- **Outcome**: systolic blood pressure (SBP)

# Case study

- Two sets of IVs:

- 256 SNPs associated with eGFRcr at genome-wide threshold $\left(p - value < 5 \times 10^{-8}\right)$

- 40 SNPs associated with both eGFRcr and blood urea nitrogen (BUN) at genome-wide threshold $\left(p - value < 5 \times 10^{-8}\right)$ $\rightarrow$ prioritizing IVs most likely relevant for kidney function