

Package: MRAPSS (via r-universe)

October 8, 2026

Title Mendelian Randomization Accounting for Pleiotropy and Sample Structure

Version 0.0.0.9000

Description Implements the MR-APSS approach to test for the causal effects between an exposure and an outcome disease. MR-APSS is a unified approach to Mendelian Randomization accounting for polygenicity, pleiotropy and sample structure using genome-wide summary statistics. Specifically, MR-APSS uses a background-foreground model to characterize both SNP-exposure effects and SNP-outcome effects estimates, where the background model accounts for confounding from genetic correlation and sample structure and the foreground model captures the valid signal for causal inference.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

LazyDataCompression xz

Imports ggplot2, dplyr, magrittr, ieugwasr (>= 0.1.4), readr, expm, data.table, mvtnorm, stats, utils

Remotes mrcieu/ieugwasr

Depends R (>= 2.10)

Suggests knitr, rmarkdown

Config/roxygen2/version 8.1.0

Config/pak/sysreqs libssl-dev libx11-dev

Repository <https://mrcieu.r-universe.dev>

Date/Publication 2026-08-11 13:12:52 UTC

RemoteUrl <https://github.com/remlapmot/MR-APSS>

RemoteRef suggestions

RemoteSha ac1e457511b91c8a0b555a048dfd963633ac89a9

Contents

C	2
clump	3
est_paras	4
format_data	5
MHC.SNPs	6
MR_EM_fix_rf_func	7
MR_fix_rf	8
MRAPSS	10
MRdat	11
MRplot	12
Omega	13
sensitivity	13
w_hm3.snplist	14
Index	15

C	<i>Estimated C matrix for the BMI–T2D example</i>
---	---

Description

The 2x2 matrix capturing the effects of sample structure for the example analysis of BMI (UK Biobank) on type 2 diabetes, as returned by [est_paras()].

Usage

C

Format

A 2 by 2 numeric matrix. Entry '[1, 1]' corresponds to the exposure, entry '[2, 2]' to the outcome and the off-diagonal entries to the correlation induced by sample overlap.

See Also

[Omega], [MRdat]

clump	<i>Perform LD clumping</i>
-------	----------------------------

Description

Perform LD clumping, to prune SNPs in LD within a window. Keep the most significant ones.

Usage

```
clump(
  dat,
  IV.Threshold = 5e-05,
  SNP_col = "SNP",
  pval_col = "pval.exp",
  clump_kb = 1000,
  clump_r2 = 0.001,
  clump_p = 0.999,
  pop = "EUR",
  bfile = NULL,
  plink_bin = NULL
)
```

Arguments

dat	a data frame must have columns with information about SNPs and p values
IV.Threshold	p-value threshold for IV selection. SNPs with p value larger than 'IV.Threshold' are discarded before clumping. Default = 5e-05.
SNP_col	column with SNP rsid. The default is "SNP"
pval_col	column with p value. The default is "pval"
clump_kb	clumping window in kb. Default is 1000.
clump_r2	clumping r2 threshold. Default is 0.001.
clump_p	clumping significance level for index variants. Default = 5e-05
pop	super-population to use as the LD reference panel when clumping with the remote API. One of "EUR", "SAS", "EAS", "AFR" or "AMR". Default = "EUR".
bfile	bfile as LD reference panel. If this is provided, then will use local PLINK. Default = NULL.
plink_bin	path to local plink binary. Default = NULL.

Value

data frame of clumped SNPs

est_paras	<i>A function harmonizing datasets and estimate background parameters by LD score regression.</i>
-----------	---

Description

A function harmonizing datasets and estimate background parameters by LD score regression.

Usage

```
est_paras(
  dat1,
  dat2,
  trait1.name = "exposure",
  trait2.name = "outcome",
  LDSC = TRUE,
  h2.fix.intercept = FALSE,
  ldscore.dir = NULL,
  ld = NULL,
  M = NULL
)
```

Arguments

dat1	formatted GWAS summary-level data for trait 1.
dat2	formatted GWAS summary-level data for trait 2.
trait1.name	specify the name of trait 1, default 'exposure'.
trait2.name	specify the name of trait 2, default 'outcome'.
LDSC	logical, whether to run LD score regression, default 'TRUE'. If 'FALSE', the function will not give the parameter estimates but will do harmonizing.
h2.fix.intercept	logical, whether to fix LD score regression intercept to 1, default 'FALSE'.
ldscore.dir	specify the path to the LD score files.
ld	specify the data.frame for the ld scores of SNPs if available and setting ld-score.dir=NULL.
M	specify the number of SNPs used for calculating LD score.

Value

List with the following elements:

Mdat Homonised data set

C the estimated C matrix capturing the effects of sample structure

Omega the estimated variance-covariance matrix for polygenic effects

format_data	<i>Format GWAS summary data for MR-APSS.</i>
-------------	--

Description

Format GWAS summary data for MR-APSS. This function is adapted from the `format_data()` function in `MRCIEU/TwoSampleMR`.

Usage

```
format_data(
  dat,
  snps.merge = w_hm3.snplist,
  snps.remove = MHC.SNPs,
  snp_col = "SNP",
  b_col = "b",
  or_col = "or",
  se_col = "se",
  freq_col = "freq",
  A1_col = "A1",
  A2_col = "A2",
  p_col = "p",
  ncase_col = "ncase",
  ncontrol_col = "ncontrol",
  n_col = "n",
  n = NULL,
  z_col = "z",
  info_col = "INFO",
  log_pval = FALSE,
  chi2_max = NULL,
  min_freq = 0.05
)
```

Arguments

<code>dat</code>	data.frame with at least SNP A1 A2 signed statistics for calculating z scores and sample size.
<code>snps.merge</code>	data.frame with SNPs to extract. The data frame must have headers: SNP A1 and A2. For example, the hapmap3 SNP list.
<code>snps.remove</code>	a set of SNPs that needed to be removed. For example, SNPs in MHC region.
<code>snp_col</code>	name of a column with rs numbers. The default is SNP.
<code>b_col</code>	name of a column with effect sizes. The default is b.
<code>or_col</code>	name of a column with odds ratios. The default is or.
<code>se_col</code>	name of a column with standard errors. The default is se.
<code>freq_col</code>	name of a column with effect allele frequencies. The default is freq.

A1_col	name of a column with effect alleles. The default is A1.
A2_col	name of a column with the other alleles. The default is A2.
p_col	name of a column with p-values. The default is p.
ncase_col	name of a column with the number of cases. The default is ncase.
ncontrol_col	name of a column with the number of controls. The default is ncontrol.
n_col	name of a column with sample sizes. The default is n.
n	sample size. If the column for sample size is not available, users can use argument "n" to specify the total sample size for each SNP.
z_col	name of a column with z scores. The default is z.
info_col	name of a column with imputation INFO. The default is info.
log_pval	logical, whether the p-value is in -log10 scale. The default is FALSE.
chi2_max	SNPs with tested chi^2 statistics large than chi2_max will be removed. The default is 80
min_freq	SNPs with allele frequency less than min_freq will be removed. The default is 0.05.

Value

a data frame with headers: SNP: rs number; A1: effect allele; A2: the other allele; Z: Z score; N: sample size; chi2: chi-square statistics; P: p-value.

MHC.SNPs	<i>SNPs in the MHC region</i>
----------	-------------------------------

Description

rs numbers of SNPs lying in the major histocompatibility complex (MHC) region of chromosome 6. The extended LD structure of this region makes it unsuitable for LD score regression, so these SNPs are removed by [format_data()] by default.

Usage

MHC.SNPs

Format

A character vector of 109503 rs numbers.

See Also

[format_data()]

MR_EM_fix_rf_func	<i>Variational EM algorithm for MR-APSS with a correlated foreground model.</i>
-------------------	---

Description

Fits the MR-APSS background-foreground model in which the foreground effects on the exposure and on the outcome are allowed to be correlated. This is the workhorse called by [MR_fix_rf()] and [sensitivity()].

Usage

```
MR_EM_fix_rf_func(
  data = NULL,
  beta = NULL,
  Sigma = NULL,
  pi0 = NULL,
  fix.beta = TRUE,
  fix.r = TRUE,
  r = 0,
  C = diag(2),
  Omega = matrix(0, 2, 2),
  tol = 1e-06,
  Threshold = 1,
  ELBO = FALSE
)
```

Arguments

data	data.frame that at least contains the following columns: b.exp b.out se.exp se.out L2. L2: LD score.
beta	initial value for the causal effect, default 'NULL' uses 0.
Sigma	initial value for Sigma, the variance-covariance matrix of the foreground effects. Default 'NULL' uses the default initialization procedure.
pi0	initial value for pi0, default 'NULL' uses 0.5.
fix.beta	logical, whether to hold 'beta' fixed at its initial value. Used to fit the null model in the likelihood ratio test. Default 'TRUE'.
fix.r	logical, whether to hold the foreground correlation 'r' fixed. Default 'TRUE'.
r	the correlation between IV strength and direct effect in the foreground model. Default '0'.
C	the estimated C matrix capturing the effects of sample structure. Default 'diag(2)'.
Omega	the estimated variance-covariance matrix of polygenic effects. Default 'matrix(0, 2, 2)'.
tol	tolerance for the relative change in the likelihood. Default '1e-06'.

Threshold	the modified IV selection threshold used to correct for selection bias. 'Threshold = 1' disables the correction. Default '1'.
ELBO	logical, whether to monitor the evidence lower bound instead of the maximum likelihood. Default 'FALSE'.

Value

a list with the following elements:

beta: causal effect estimate

Sigma: estimated variance-covariance matrix of the foreground effects

pi0: the probability of an SNP with foreground signals after selection

likeli: the maximised log likelihood

Threshold: the IV selection threshold used

MR_fix_rf	<i>A function for causal effect estimation and inference in sensitivity analysis.</i>
-----------	---

Description

A function for causal effect estimation and inference in sensitivity analysis.

Usage

```
MR_fix_rf(
  MRdat = NULL,
  exposure = "exposure",
  outcome = "outcome",
  pi0 = NULL,
  Sigma = NULL,
  r = 0,
  C = matrix(c(1, 0, 0, 1), 2, 2),
  Omega = matrix(0, 2, 2),
  Cor.SelectionBias = TRUE,
  tol = 1e-08
)
```

Arguments

MRdat	data.frame at least contains the following columns: b.exp b.out se.exp se.out L2 Threshold. L2:LD score, Threshold: modified IV selection threshold for correction of selection bias
exposure	exposure name
outcome	outcome name

pi0	initial value for pi0, default 'NULL' will use the default initialize procedure.
Sigma	initial value for Sigma (the variance-covariance matrix for foreground effects), default 'NULL' will use the default initialize procedure.
r	The value of r (correlation between IV strength and direct effect in the foreground model), default '0', r is fixed.
C	the estimated C matrix capturing the effects of sample structure. default 'diag(2)'.
Omega	the estimated variance-covariance matrix of polygenic effects. default 'matrix(0,2,2)'.
Cor.SelectionBias	logical, whether to correct selection bias or not. If FALSE, the model won't correct for selection bias.
tol	tolerance, default '1e-12'.

Value

a list with the following elements:

exposure: exposure of interest

outcome: outcome of interest

beta: causal effect estimate

beta.se: standard error

pval: p-value

Examples

```
library(MRAPSS)
exposure = "BMI"
outcome = "T2D"
Threshold = 5e-05 # IV selection Threshold
data(C)
data(Omega)
data(MRdat)
MRres = MR_fix_rf(MRdat,
  exposure = "BMI",
  outcome = "T2D",
  r=0,
  C = C,
  Omega = Omega ,
  Cor.SelectionBias = TRUE)
```

MRAPSS

*A function for implementing MR-APSS.***Description**

MR-APSS: a unified approach to Mendelian Randomization accounting for Pleiotropy and Sample Structure using genome-wide summary statistics. MR-APSS uses a variational EM algorithm for the estimation of parameters. MR-APSS uses likelihood ratio test for inference.

Usage

```
MRAPSS(
  MRdat = NULL,
  exposure = "exposure",
  outcome = "outcome",
  pi0 = NULL,
  sigma.sq = NULL,
  tau.sq = NULL,
  C = matrix(c(1, 0, 0, 1), 2, 2),
  Omega = matrix(0, 2, 2),
  Cor.SelectionBias = TRUE,
  tol = 1e-08,
  ELBO = FALSE
)
```

Arguments

MRdat	data.frame at least contains the following columns: b.exp b.out se.exp se.out L2 Threshold. L2:LD score, Threshold: modified IV selection threshold for correction of selection bias
exposure	exposure name
outcome	outcome name
pi0	initial value for pi0, default 'NULL' will use the default initialize procedure.
sigma.sq	initial value for sigma.sq , default 'NULL' will use the default initialize procedure.
tau.sq	initial value for tau.sq , default 'NULL' will use the default initialize procedure.
C	the estimated C matrix capturing the effects of sample structure. default 'diag(2)'.
Omega	the estimated variance-covariance matrix of polygenic effects. default 'matrix(0,2,2)'.
Cor.SelectionBias	logical, whether to correct selection bias or not. If FALSE, the model won't correct for selection bias.
tol	tolerance, default '1e-08'.
ELBO	logical, whether to the evidence lower bound or not. If 'FALSE', check the maximum likelihood instead. default 'FALSE'.

Value

a list with the following elements:

MRdat: Input data frame

exposure: exposure of interest

outcome: outcome of interest

beta: causal effect estimate

beta.se: standard error

pval: p-value

sigma.sq: variance of foreground exposure effect

tau.sq: variance of foreground outcome effect

pi0: The probability of an SNP with foreground signals after selection

post: Posterior estimates of latent variables

method: "MR-APSS"

Examples

```
library(MRAPSS)
exposure = "BMI"
outcome = "T2D"
Threshold = 5e-05 # IV selection Threshold
data(C)
data(Omega)
data(MRdat)
MRres = MRAPSS(MRdat,
               exposure = "BMI",
               outcome = "T2D",
               C = C,
               Omega = Omega,
               Cor.SelectionBias = TRUE)
MRplot(MRres, exposure = "BMI", outcome = "T2D")
```

MRdat

Harmonized instrument data for the BMI–T2D example

Description

Clumped and harmonized GWAS summary statistics for BMI (UK Biobank) and type 2 diabetes, ready to be passed to [MRAPSS()].

Usage

```
MRdat
```

Format

A data frame with 1227 rows and 11 columns:

SNP rs number

A1 effect allele

A2 the other allele

b.exp SNP effect on the exposure

b.out SNP effect on the outcome

se.exp standard error of 'b.exp'

se.out standard error of 'b.out'

pval.exp p-value for the SNP-exposure association

pval.out p-value for the SNP-outcome association

L2 LD score

Threshold modified IV selection threshold used to correct for selection bias

See Also

[C], [Omega]

MRplot

Visualize the MR-APSS analysis results

Description

Visualize the MR-APSS analysis results

Usage

```
MRplot(MRres, exposure = "trait 1", outcome = "trait 2")
```

Arguments

MRres	MR-APSS fit results
exposure	exposure name
outcome	outcome name

Value

a scatter plot of SNP-exposure effect and SNP-outcome effect with the causal effect and 95% confidence interval.

Omega	<i>Estimated Omega matrix for the BMI–T2D example</i>
-------	---

Description

The 2x2 variance-covariance matrix of the polygenic (background) effects for the example analysis of BMI (UK Biobank) on type 2 diabetes, as returned by [est_paras()].

Usage

Omega

Format

A 2 by 2 numeric matrix. The diagonal entries are the per-SNP heritabilities of the exposure and the outcome and the off-diagonal entries are the per-SNP genetic covariance.

See Also

[C], [MRdat]

sensitivity	<i>Sensitivity analysis for the foreground correlation parameter r.</i>
-------------	---

Description

Refits MR-APSS over a grid of values for ‘r’, the correlation between IV strength and direct effect in the foreground model, and summarises how the causal effect estimate varies with ‘r’. The grid runs from 0 to the value of ‘r’ implied by an unconstrained fit of the foreground variance-covariance matrix.

Usage

```
sensitivity(  
  MRdat = NULL,  
  Omega = NULL,  
  C = NULL,  
  exposure = "exposure",  
  outcome = "outcome"  
)
```

Arguments

MRdat	data.frame that at least contains the following columns: b.exp b.out se.exp se.out L2 Threshold. L2: LD score, Threshold: modified IV selection threshold for correction of selection bias.
Omega	the estimated variance-covariance matrix of polygenic effects.
C	the estimated C matrix capturing the effects of sample structure.
exposure	exposure name.
outcome	outcome name.

Value

- a list with the following elements:
 - estimates:** data frame with columns r, beta, se and pval, one row per value of r
 - plot:** a ggplot object showing the causal effect estimate against r

w_hm3.snplist	<i>HapMap 3 SNP list</i>
---------------	--------------------------

Description

The HapMap 3 SNP list distributed with LDSC, used by [format_data()] as the default set of SNPs to retain.

Usage

w_hm3.snplist

Format

- A data frame with 1217311 rows and 3 columns:
 - SNP** rs number
 - A1** effect allele
 - A2** the other allele

See Also

[format_data()]

Index

* datasets

- C, [2](#)
- MHC.SNPs, [6](#)
- MRdat, [11](#)
- Omega, [13](#)
- w_hm3.snplist, [14](#)

- C, [2](#)
- clump, [3](#)

- est_paras, [4](#)

- format_data, [5](#)

- MHC.SNPs, [6](#)
- MR_EM_fix_rf_func, [7](#)
- MR_fix_rf, [8](#)
- MRAPSS, [10](#)
- MRdat, [11](#)
- MRplot, [12](#)

- Omega, [13](#)

- sensitivity, [13](#)

- w_hm3.snplist, [14](#)