Package 'MRlap'

October 5, 2021

Title MRlap is an R-package to perform two-sample Mendelian Randomisation (MR) analyses using (potentially) overlapping samples

Version 0.0.1.0000

Description MR estimates can be subject to different types of biases due to the overlap between the exposure and outcome samples, the use of weak instruments and Winner's curse. Our approach simultaneously accounts and corrects for all these biases, using cross-trait LD-score regression (LDSC) to approximate the overlap. It requires only GWAS summary statistics. Estimating the corrected effect using our approach can be performed as a sensitivity analysis: if the corrected effect do not significantly differ from the observed effect, then IVW-MR estimate can be safely used. However, when there is a significant difference, corrected effects should be preferred as they should be less biased, independently of the sample overlap.

License GPL-2 | file LICENSE

Encoding UTF-8

LazyData true

Imports magrittr,
tibble,
dplyr,
data.table,
rlang,
stringr,
TwoSampleMR,
GenomicSEM

Remotes MRCIEU/TwoSampleMR,
GenomicSEM/GenomicSEM

RoxygenNote 6.1.1

R topics documented:

MRlap *MRlap - main function*

Description

Performs cross-trait LD score regression, IVW-MR analysis and provide a correction that simultaneously accounts for biases due to the overlap between the exposure and outcome samples, the use of weak instruments and Winner's curse.

Usage

```
MRlap(exposure, exposure_name = NULL, K_exposure = NA,
  P_exposure = NA, outcome, outcome_name = NULL, K_outcome = NA,
  P_outcome = NA, ld, hm3, MR_threshold = 5e-08,
  MR_pruning_dist = 500, MR_pruning_LD = 0, MR_reverse = 0.001,
  save_logfiles = FALSE, verbose = TRUE)
```

Arguments

exposure	The path to the file containing the GWAS summary statistics for the exposure, or a data.frame (character, or data.frame)
exposure_name	The name of the exposure trait, default="exposure" (character)
K_exposure	If case-control exposure, prevalence in the population, default=NA for continuous traits (numeric)
P_exposure	If case-control exposure, prevalence in the sample, default=NA for continuous traits (numeric)
outcome	The path to the file containing the GWAS summary statistics for the exposure, or a data.frame (character, or data.frame)
outcome_name	The name of the outcome trait, default="outcome" (character)
K_outcome	If case-control outcome, prevalence in the population, default=NA for continuous traits (numeric)
P_outcome	If case-control outcome, prevalence in the sample, default=NA for continuous traits (numeric)
ld	The path to the folder in which the LD scores used in the analysis are located. Expects LD scores formated as required by the original LD score regression software. (character)
hm3	The path to a file of SNPs with alt, ref alleles and rsid used to allign alleles across traits (character)
MR_threshold	The threshold used to select strong instruments for MR, should be lower than 1e-5, default=5e-8 (numeric)
MR_pruning_dist	The distance used for pruning MR instruments (in Kb), should be between 10 and 50000, default=500 (numeric)
MR_pruning_LD	The LD threshold (r2) used for pruning MR instruments, should be between 0 and 1 (if 0, distance-based pruning is used), default=0 (numeric)
MR_reverse	The p-value used to exclude MR instruments that are more strongly associated with the outcome than with the exposure,default=1e-3 (numeric)
save_logfiles	A logical indicating if log files from LDSC should be saved, default=FALSE
verbose	A logical indicating if information on progress should be reported, default=TRUE

Details

exposure and outcome are required arguments. The input file / data.frame should contain the following columns (lower or upper case) :
SNPID (rs numbers) should be : rs, rsid, snp, snpid, rnpid
CHR (chromosome) should be : chr
POS (position) should be : pos
ALT (effect allele) should be : a1, alt, alts
REF (reference allele) should be : a2, a0, ref
Z (z-score) should be : Z, zscore
N (sample size) should be : N
If Z is not present, it can be calculated from BETA and SE.
BETA should be : b, beta, beta1, or
SE should be : se, std
If (at least) one of the datasets is coming from a case-control GWAS:* The Sample size column should correspond to the effective sample size (not the total sample size). The number of cases (NCASES) and the number of controls (NCONTROLS) can also be provided (instead or in addition to the effective sample size). NCASES should be : n_cases, ncases, n_case, ncase
NCONTROLS should be : n_controls, ncontrols, n_control, ncontrol

---

SmallExposure_Data *Exposure*

Description

Subset of the original dataset containing the estimated effect of SNPs on the exposure

Usage

```
SmallExposure_Data
```

Format

A data frame with 750,000 rows and 13 variables:

chr chromosome

rsid rsid of the SNP

pos position

ref reference allele for the SNP

alt effect allele for the SNP

af allele frequency

info imputation quality

beta estimated effect size for the SNP

se standard error of the estimated effect size for the SNP

z z-score for the SNP

minuslog10p -log10(p) for the SNP

p p-value for the SNP

N sample size

`SmallOutcome_Data` *Outcome*

Description

Subset of the original dataset containing the estimated effect of SNPs on the outcome

Usage

```
SmallOutcome_Data
```

Format

A data frame with 750,000 rows and 13 variables:

chr chromosome

rsid rsid of the SNP

pos position

ref reference allele for the SNP

alt effect allele for the SNP

af allele frequency

info imputation quality

beta estimated effect size for the SNP

se standard error of the estimated effect size for the SNP

z z-score for the SNP

minuslog10p -log10(p) for the SNP

p p-value for the SNP

N sample size

Index