

Package: simmrd (via r-universe)

October 14, 2026

Title Simulation of Mendelian Randomization data

Version 0.0.1

Description This package generates simulation data to use in the evaluation of univariable or multivariable Mendelian Randomization methods. MR scenarios can include uncorrelated horizontal pleiotropy, correlated horizontal pleiotropy, weak instruments, winner's curse, and correlated SNP instruments.

License MIT + file LICENSE

URL <https://github.com/noahlorinczcomi/simmrd>

BugReports <https://github.com/noahlorinczcomi/simmrd/issues>

Imports ggplot2, ggpubr, mvnfast

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

Config/pak/sysreqs cmake make libicu-dev

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

Date/Publication 2026-05-17 19:32:04 UTC

RemoteUrl <https://github.com/noahlorinczcomi/simmrd>

RemoteRef HEAD

RemoteSha 75a296296a91160d5990451d33c900114e5f0ea8

Contents

adj_overlap	2
generate_individual	3
generate_summary	4

list_presets	5
load_preset	6
plot_simdata	7
plot_simdata_lower	8
set_params	8

Index	12
--------------	-----------

adj_overlap	<i>Construct a GWAS overlap proportion matrix</i>
-------------	---

Description

Returns the (p+1) x (p+1) matrix of pairwise GWAS sample overlap proportions for use as the prop_gwas_overlap_Xs argument of [set_params](#) when exposure GWAS samples partially overlap each other and the outcome GWAS.

Usage

```
adj_overlap(  
  exposure_overlap_proportions,  
  prop_gwas_overlap_Xs_and_Y,  
  number_of_exposures  
)
```

Arguments

- exposure_overlap_proportions
Scalar or matrix of overlap proportions between exposure GWAS.
- prop_gwas_overlap_Xs_and_Y
Scalar or vector of overlap proportions between exposures and outcome GWAS.
- number_of_exposures
Number of exposures.

Value

A named (p+1) x (p+1) matrix where rows/columns are labelled "Outcome", "Exposure1", etc.

Examples

```
adj_overlap(  
  exposure_overlap_proportions = 0.2,  
  prop_gwas_overlap_Xs_and_Y = 0.1,  
  number_of_exposures = 3  
)
```

generate_individual	<i>Generate individual-level simulated GWAS data</i>
---------------------	--

Description

Generates simulated individual-level GWAS data for Mendelian Randomization evaluation given a parameter list produced by [set_params](#).

Usage

```
generate_individual(params, seed = 1)
```

Arguments

params	Named parameter list from set_params .
seed	Integer seed passed to <code>set.seed()</code> , or NULL to use the current RNG state. Defaults to 1.

Value

A named list with elements:

bx $m \times p$ matrix of IV-exposure associations

bxse Standard errors for bx

by $m \times 1$ vector of IV-outcome associations

byse Standard errors for by

RhoME $(p+1) \times (p+1)$ measurement-error correlation matrix

LDMatrix True LD correlation matrix among IVs

LDhatMatrix Estimated LD correlation matrix among IVs

theta True causal effects

IVtype Per-IV classification: "valid", "UHP", or "CHP"

bx_unstd Unstandardized version of bx

bxse_unstd Standard errors for bx_unstd

by_unstd Unstandardized version of by

byse_unstd Standard errors for by_unstd

See Also

[set_params](#), [generate_summary](#), [plot_simdata](#)

Examples

```
## Not run:
params <- set_params(
  type = "individual",
  number_of_exposures = 2,
  Y_variance_explained_by_Xs = c(0, 0.5),
  signs_of_causal_effects = c(1, 1),
  Xs_variance_explained_by_U = 0.12,
  Y_variance_explained_by_U = 0.10,
  simtype = "weak",
  fix_Fstatistic_at = 10
)
data <- generate_individual(params)

## End(Not run)
```

generate_summary

Generate summary-level simulated GWAS data

Description

Generates simulated GWAS summary statistics for Mendelian Randomization evaluation given a parameter list produced by [set_params](#).

Usage

```
generate_summary(params, seed = NULL)
```

Arguments

params	Named parameter list from set_params .
seed	Integer seed passed to <code>set.seed()</code> , or NULL to use the current RNG state. Defaults to NULL.

Value

A named list with elements:

bx $m \times p$ matrix of IV-exposure associations

bxse Standard errors for bx

by $m \times 1$ vector of IV-outcome associations

byse Standard errors for by

RhoME $(p+1) \times (p+1)$ measurement-error correlation matrix

LDMatrix True LD correlation matrix among IVs

LDhatMatrix Estimated LD correlation matrix among IVs

theta True causal effects

IVtype Per-IV classification: "valid", "UHP", or "CHP"

bx_unstd Unstandardized version of bx

bxse_unstd Standard errors for bx_unstd

by_unstd Unstandardized version of by

byse_unstd Standard errors for by_unstd

beta_true True SNP-exposure effect sizes (all SNPs, before IV selection)

alpha_true True SNP-outcome associations (all SNPs, before IV selection)

u Exposure GWAS estimation errors (bx_unstd - beta_true, all SNPs)

v Outcome GWAS estimation errors (by_unstd - alpha_true, all SNPs)

iv_index Integer indices of the selected IVs within the full SNP set

See Also

[set_params](#), [generate_individual](#), [plot_simdata](#)

Examples

```
## Not run:
# Two exposures with CHP, no GWAS overlap
params <- set_params(
  number_of_exposures      = 2,
  true_causal_effects      = c(0.3, 0.1),
  prop_gwas_overlap_Xs_and_Y = 0,
  number_of_CHP_causal_SNPs = 20,
  ratio_of_CHP_variance    = 0.25,
  CHP_correlation          = -0.5
)
data <- generate_summary(params)

## End(Not run)
```

list_presets	<i>List available simulation presets</i>
--------------	--

Description

Prints the valid values for each argument of [load_preset](#).

Usage

```
list_presets()
```

Value

Invisibly returns a named list of valid values.

Examples

```
list_presets()
```

load_preset	<i>Load a named simulation preset</i>
-------------	---------------------------------------

Description

Returns a ready-to-use parameter list corresponding to one of the built-in simulation scenarios. The list is identical to what `set_params()` produces, so every element can be overridden afterwards.

Usage

```
load_preset(  
  bias = "none",  
  n = 1e+05,  
  snps = 100,  
  exposures = 1,  
  overlap = "full"  
)
```

Arguments

bias	Bias scenario. One of "none", "UHP", "CHP", "UHP_CHP", "UHP_CHP_WEAK", or "WEAK".
n	GWAS sample size for both exposures and outcome. 3e4 or 1e5.
snps	Number of causal SNPs per exposure. 100 or 500.
exposures	Number of exposures. 1 or 3.
overlap	"full" (complete exposure–outcome overlap) or "none" (no exposure–outcome overlap).

Value

A named parameter list, identical in structure to `set_params()` output.

See Also

[set_params](#), [list_presets](#)

Examples

```
# CHP scenario, small GWAS, no overlap
params <- load_preset("CHP", n = 3e4, snps = 100, exposures = 1, overlap = "none")
data <- generate_summary(params)

# Start from a preset, then tweak one thing
params <- load_preset("UHP_CHP", n = 1e5, snps = 500, exposures = 3, overlap = "full")
params$true_causal_effects <- c(0.1, 0.2, 0.3)
data <- generate_summary(params)
```

plot_simdata	<i>Plot simulated data</i>
--------------	----------------------------

Description

Plot simulated data.

Usage

```
plot_simdata(
  data,
  params = params,
  exposure_specific_plot = "total",
  verbose = TRUE
)
```

Arguments

data	direct output from generate()
params	Named list of parameters
exposure_specific_plot	One of 'total', 'joint', 'conditional'
verbose	Logical, default TRUE

Examples

```
## Not run:
# If you used generate_summary(), execute the following
plot_simdata(gwas_data, summary_params)

# If you used generate_individual(), execute the following
plot_simdata(gwas_data, individual_params)

## End(Not run)
```

plot_simdata_lower	<i>Helper function</i>
--------------------	------------------------

Description

Helper function

Usage

```
plot_simdata_lower(data, params = params, showFstat = TRUE)
```

Arguments

data	direct output from generate()
params	Named list of parameters
showFstat	Logical, default TRUE

Examples

```
## Not run:
plot_simdata_lower()

## End(Not run)
```

set_params	<i>Set simulation parameters</i>
------------	----------------------------------

Description

Constructs a named parameter list for use with generate_summary() or generate_individual(). Every argument has a sensible default so you only need to specify the values you want to change.

Usage

```
set_params(
  type = "summary",
  sample_size_Xs = 1e+05,
  sample_size_Y = 1e+05,
  number_of_exposures = 1,
  number_of_causal_SNPs = 100,
  prop_gwas_overlap_Xs_and_Y = 1,
  prop_gwas_overlap_Xs = 1,
  number_of_UHP_causal_SNPs = 0,
  number_of_CHP_causal_SNPs = 0,
  ratio_of_UHP_variance = 0,
```

```

ratio_of_CHP_variance = 0,
CHP_correlation = 0,
Y_variance_explained_by_UHP = 0,
U_variance_explained_by_CHP = 0,
true_causal_effects = 0.3,
Y_variance_explained_by_Xs = 0.3,
signs_of_causal_effects = 1,
phenotypic_correlation_Xs = 0.3,
genetic_correlation_Xs = 0.15,
phenotypic_correlations_Xs_and_Y = 0.3,
Xs_variance_explained_by_g = 0.1,
LD_causal_SNPs = "I",
number_of_LD_blocks = 1,
Xs_variance_explained_by_U = 0,
Y_variance_explained_by_U = 0,
simtype = "winners",
IV_Pvalue_threshold = 5e-08,
fix_Fstatistic_at = 10,
MVMR_IV_selection_type = "union",
LD_pruning_r2 = 1,
MR_standardization = "none",
N_of_LD_ref = Inf
)

```

Arguments

type	"summary" (default) or "individual". — Study design —
sample_size_Xs	Exposure GWAS sample size(s). Scalar or vector with one value per exposure.
sample_size_Y	Outcome GWAS sample size.
number_of_exposures	Number of exposures.
number_of_causal_SNPs	Number of SNPs with a direct effect on each exposure. — GWAS overlap —
prop_gwas_overlap_Xs_and_Y	Proportion of overlap between exposure and outcome GWAS. Scalar or vector.
prop_gwas_overlap_Xs	Proportion of overlap among the exposure GWAS (summary only). Scalar or numeric matrix. — Pleiotropy (summary data) —
number_of_UHP_causal_SNPs	Number of uncorrelated horizontal pleiotropy (UHP) SNPs.
number_of_CHP_causal_SNPs	Number of correlated horizontal pleiotropy (CHP) SNPs.
ratio_of_UHP_variance	Ratio of UHP variance to valid-IV variance.

ratio_of_CHP_variance
 Ratio of CHP variance to valid-IV variance.

CHP_correlation
 Correlation between CHP and valid-IV effect sizes (magnitude of CHP).
 — Pleiotropy (individual data) —

Y_variance_explained_by_UHP
 Outcome variance explained by UHP SNPs.

U_variance_explained_by_CHP
 Confounder variance explained by CHP SNPs.
 — Causal effects —

true_causal_effects
 True causal effect size(s). Scalar or vector (summary only).

Y_variance_explained_by_Xs
 Outcome variance explained by each exposure. Scalar or vector (individual only).

signs_of_causal_effects
 Signs of causal effects. Scalar or vector (individual only).
 — Correlations —

phenotypic_correlation_Xs
 Phenotypic correlations among exposures. Scalar, string ('ar1(0.5)', 'cs(0.3)', 'toeplitz'), or matrix.

genetic_correlation_Xs
 Genetic correlations among exposures. Same formats as above.

phenotypic_correlations_Xs_and_Y
 Phenotypic correlations between each exposure and the outcome. Scalar or vector (summary only).
 — Genetic architecture —

Xs_variance_explained_by_g
 Heritability of each exposure (variance explained by all causal SNPs). Scalar or vector.

LD_causal_SNPs
 LD structure among causal SNPs. Scalar, string ('I', 'toeplitz', 'ar1(0.5)'), or numeric matrix.

number_of_LD_blocks
 Number of independent LD blocks.
 — Confounding (individual only) —

Xs_variance_explained_by_U
 Exposure variance explained by the latent confounder.

Y_variance_explained_by_U
 Outcome variance explained by the latent confounder.
 — IV selection —

simtype
 "winners" (P-value-based selection) or "weak" (fix F-statistic).

IV_Pvalue_threshold
 P-value threshold for IV selection (used when simtype = "winners").

fix_Fstatistic_at
 Target mean F-statistic (used when simtype = "weak").

MVMR_IV_selection_type
 "union" or "joint" (multivariable MR only).

LD_pruning_r2 Upper r^2 threshold for LD pruning of IVs.
 — Output —

MR_standardization
 Standardization applied to GWAS summary statistics. "none", "Z", or "QC".

N_of_LD_ref Sample size of the LD reference panel (Inf = use true LD).

Value

A named list of parameters ready to pass to generate_summary() or generate_individual().

Examples

```
# Minimal: one exposure, default settings
params <- set_params()
data <- generate_summary(params)

# Two exposures with CHP, no GWAS overlap
params <- set_params(
  number_of_exposures      = 2,
  true_causal_effects      = c(0.3, 0.1),
  prop_gwas_overlap_Xs_and_Y = 0,
  number_of_CHP_causal_SNPs = 20,
  ratio_of_CHP_variance    = 0.25,
  CHP_correlation          = -0.5
)
data <- generate_summary(params)
```

Index

`adj_overlap`, [2](#)

`generate_individual`, [3](#), [5](#)

`generate_summary`, [3](#), [4](#)

`list_presets`, [5](#), [6](#)

`load_preset`, [5](#), [6](#)

`plot_simdata`, [3](#), [5](#), [7](#)

`plot_simdata_lower`, [8](#)

`set_params`, [2–6](#), [8](#)