

# Package: MrDAG (via r-universe)

September 22, 2024

**Type** Package

**Title** MrDAG: Mendelian randomization (MR) with Bayesian Directed Acyclic (DAG) Graphs exploration and causal effects estimation

**Version** 0.1.0

**Date** 2024-05-01

**Maintainer** Leonardo Bottolo <lb664@cam.ac.uk>

**Description** This package performs Mendelian randomization for multiple exposures and outcomes with Bayesian Directed Acyclic Graphs exploration and causal effects estimation.

**License** GPL-2 | file LICENSE

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.3.1

**RdMacros** Rdpack

**Imports** gRbase, mvtnorm, Rdpack, utils

**Depends** R (>= 2.10)

**LazyData** true

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

**RemoteUrl** <https://github.com/lb664/MrDAG>

**RemoteRef** HEAD

**RemoteSha** 06d8ba420315136f7cd8d8e12829e496efdc2ac2

## Contents

get_causaleffect . . . . .	2
get_causaleffects . . . . .	3
get_edgeprob . . . . .	5
LBT2MD_data . . . . .	7
MD2LBT_data . . . . .	7
MrDAG . . . . .	8

<b>Index</b>	<b>12</b>
--------------	-----------

---

get_causaleffect	<i>Estimation of the (average) causal effects under intervention on an exposure</i>
------------------	---

---

### Description

Estimation of the (average) causal effect under intervention on a trait (target) and measured on another one (response) based on the Directed Acyclic Graphs (DAGs) explored by MrDAG algorithm

### Usage

```
get_causaleffect(output, response, target, BMA = TRUE, CI = 0.95)
```

### Arguments

output	Output produced by MrDAG algorithm
response	Trait (response) where the effect of the intervention on another trait (target) is measured
target	Trait (target) under intervention
BMA	If TRUE, Bayesian Model Averaging of the estimated (average) causal effect across all explored DAGs in the visited Completed Partially DAG (CPDAG) (Chickering (2002)) or Essential Graph (EG) (Andersson et al. (1997)) is performed
CI	Level (0.95 default) of the credible interval (CI) of the (average) causal effect. It is calculated based on suitable quantiles of the estimated (average) causal effect across all explored DAGs

### Value

The value returned is a list object `list(causaleffect, causaleffect_LL, causaleffect_UL, group, BMA, CI)`

- `causaleffect` Estimate of the (average) causal effect under intervention on a trait (target) on another one (response) based on the DAGs explored by MrDAG algorithm. Feedback loop causal effect of the same trait is not allowed (NA)
- `causaleffect_LL` Lower limit of the CI of the (average) causal effect. It is calculated as the  $(1-CIboot)/2$  % quantile of the (average) causal effect across all explored DAGs
- `causaleffect_UL` Upper limit of the CI of the (average) causal effect. It is calculated as the  $1-[(1-CIboot)/2]$  % quantile of the (average) causal effect across all explored DAGs
- `group` If MrDAGcheck in MrDAG algorithm contains the indices of the traits in data that define the outcomes and the exposures, group coincides with this list
- BMA Logical option
- CI Level of the credible interval option

## References

Andersson SA, Madigan D, Perlman MD (1997). “A characterization of Markov equivalence classes for acyclic digraphs.” *Annals of Statistics*, **25**(2), 505–541. doi:10.1214/aos/1031833662.

Chickering DM (2002). “Learning equivalence classes of bayesian-network structures.” *Journal of Machine Learning Research*, **2**, 445–498. <http://www.ai.mit.edu/projects/jmlr/papers/volume2/chickering02a/chickering02a.pdf>.

## Examples

```
# Example: Estimation of the (average) causal effects under intervention on lifestyle and
# behavioural exposures, and measured on mental health phenotypes. 708 independent Instrumental
# Variables (IVs) were selected to be associated at genome-wide significance with six lifestyle
# and behavioural traits after pruning or clumping which are considered exposures of the risk
# of seven mental health outcome phenotypes

# After loading the data set

data(LBT2MD_data)

# the indices of the traits that define the outcomes and exposures are provided in a list object

MrDAGcheck <- NULL
MrDAGcheck$Y_idx <- 1 : 7 # Mental health phenotypes
MrDAGcheck$X_idx <- 8 : 13 # Lifestyle and behavioural traits

# MrDAG algorithm is run to generate 1,000 posterior samples of all unknowns

output <- MrDAG(data = LBT2MD_data,
                niter = 5000, burnin = 2500, thin = 5, tempmax = 20, w = 0.01,
                MrDAGcheck = MrDAGcheck, filename = NULL)

# Finally, the (average) causal effects and credible intervals (CI) of the intervention on ALC
# and measured on SCZ are estimated

causaleffects <- get_causaleffect(output, 7, 8) # Intervention on ALC and measured on SCZ
```

---

get_causaleffects	<i>Estimation of the (average) causal effects under intervention on the exposures</i>
-------------------	---

---

## Description

Estimation of the (average) causal effects under intervention on the exposures (targets) and measured on the outcomes (responses) based on the Directed Acyclic Graphs (DAGs) explored by MrDAG algorithm

**Usage**

```

get_causaleffects(
  output,
  ord = NULL,
  BMA = TRUE,
  CI = 0.95,
  progress = FALSE,
  plot = FALSE
)

```

**Arguments**

<code>output</code>	Output produced by MrDAG algorithm
<code>ord</code>	Indices of the traits in data that define the outcomes and the exposures. They specify the order of appearance of the traits when printing the (average) causal effect. If <code>ord = NULL</code> , the order is the same as <code>MrDAGcheck</code> in <code>MrDAG</code> output
<code>BMA</code>	If <code>TRUE</code> , Bayesian Model Averaging of the estimated (average) causal effect across all explored DAGs in the visited Completed Partially DAG (CPDAG) (Chickering (2002)) or Essential Graph (EG) (Andersson et al. (1997)) is performed
<code>CI</code>	Level (0.95 default) of the credible interval (CI) of the (average) causal effect. It is calculated based on suitable quantiles of the estimated (average) causal effect across all explored DAGs
<code>progress</code>	Logical value set as <code>FALSE</code> by default to print on the screen the progress of the causal effects estimation
<code>plot</code>	Logical value (default <code>FALSE</code> ). If <code>TRUE</code> , <code>get_causaleffects</code> is used to generate a plot

**Value**

The value returned is a list object `list(causaleffects, causaleffects_LL, causaleffects_UL, group, ord, BMA, CI)`

- `causaleffects` Estimate of the (average) causal effects under intervention on the exposures based on the DAGs explored by MrDAG algorithm. Feedback loop causal effect of the same outcome is not allowed (NA)
- `causaleffects_LL` Lower limit of the CI of the (average) causal effects. It is calculated as the  $(1-CI_{boot})/2$  % quantile of the (average) causal effect across all explored DAGs
- `causaleffects_UL` Upper limit of the CI of the (average) causal effects. It is calculated as the  $1-[(1-CI_{boot})/2]$  % quantile of the (average) causal effect across all explored DAGs
- `group` If `MrDAGcheck` in `MrDAG` algorithm contains the indices of the traits in data that define the outcomes and the exposures, `group` coincides with this list
- `ord` Indices of the traits in data that specify the outcomes and the exposures. It might differ from `group` if `ord` has been specified and it is different from `MrDAGcheck`
- `BMA` Logical option
- `CI` Level of the credible interval option

## References

Andersson SA, Madigan D, Perlman MD (1997). “A characterization of Markov equivalence classes for acyclic digraphs.” *Annals of Statistics*, **25**(2), 505–541. doi:10.1214/aos/1031833662.

Chickering DM (2002). “Learning equivalence classes of bayesian-network structures.” *Journal of Machine Learning Research*, **2**, 445–498. <http://www.ai.mit.edu/projects/jmlr/papers/volume2/chickering02a/chickering02a.pdf>.

## Examples

```
# Example: Estimation of the (average) causal effects under intervention on lifestyle and
# behavioural exposures, and measured on mental health phenotypes. 708 independent Instrumental
# Variables (IVs) were selected to be associated at genome-wide significance with six lifestyle
# and behavioural traits after pruning or clumping which are considered exposures of the risk
# of seven mental health outcome phenotypes

# After loading the data set

data(LBT2MD_data)

# the indices of the traits that define the outcomes and exposures are provided in a list object

MrDAGcheck <- NULL
MrDAGcheck$Y_idx <- 1 : 7    # Mental health phenotypes
MrDAGcheck$X_idx <- 8 : 13  # Lifestyle and behavioural traits

# MrDAG algorithm is run to generate 1,000 posterior samples of all unknowns

output <- MrDAG(data = LBT2MD_data,
                niter = 5000, burnin = 2500, thin = 5, tempmax = 20, w = 0.01,
                MrDAGcheck = MrDAGcheck, filename = NULL)

# Finally, the (average) causal effects and 90% credible intervals are estimated

causaleffects <- get_causaleffects(output, CI = 0.90, progress = TRUE)
```

---

get\_edgeprob

*Estimation of the Posterior Probability of Edge Inclusion*

---

## Description

Estimation of the Posterior Probability of Edge Inclusion (PPEI) based on the Directed Acyclic Graphs (DAGs) explored by MrDAG algorithm

## Usage

```
get_edgeprob(output, ord = NULL)
```

**Arguments**

output	Output produced by MrDAG algorithm
ord	Indices of the traits in data that define the outcomes and the exposures. They specify the order of appearance of the traits when printing the PPEI. If ord = NULL, the order is the same as MrDAGcheck in MrDAG output

**Value**

The value returned is a list object `list(edgeprob, group, ord)`

- effect Estimate of the PPEI between each trait. Feedback loop of the same trait is not allowed (NA). If a partition of the traits between outcomes and exposures is specified, the PPEIs between outcomes and exposures are not considered (NA)
- group If MrDAGcheck in MrDAG algorithm contains the indices of the traits in data that define the outcomes and the exposures, group coincides with this list
- ord Indices of the traits in data that specify the outcomes and the exposures. It might differ from group if ord has been specified and it is different from MrDAGcheck

**Examples**

```
# Example: Estimation of the Posterior Probabilities of Edge Inclusion (PPEIs) of lifestyle and
# behavioural traits, and mental health phenotypes. 708 independent Instrumental Variables
# (IVs) were selected to be associated at genome-wide significance with six lifestyle and
# behavioural traits after pruning or clumping which are considered exposures of the risk of
# seven mental health outcome phenotypes

# After loading the data set

data(LBT2MD_data)

# the indices of the traits that define the outcomes and exposures are provided in a list object

MrDAGcheck <- NULL
MrDAGcheck$Y_idx <- 1 : 7   # Mental health phenotypes
MrDAGcheck$X_idx <- 8 : 13 # Lifestyle and behavioural traits

# MrDAG algorithm is run to generate 1,000 posterior samples of all unknowns

output <- MrDAG(data = LBT2MD_data,
                niter = 5000, burnin = 2500, thin = 5, tempmax = 20, w = 0.01,
                MrDAGcheck = MrDAGcheck, filename = NULL)

# Finally, PPEIs are calculated and presented with lifestyle and behavioural traits first,
# followed by mental health phenotypes

ord <- c(8 : 13, 1 : 7)
PPEI <- get_edgeprob(output, ord = ord)
```

---

LBT2MD_data	<i>MrDAG data set: Lifestyle and behavioural exposures that might impact mental health phenotypes</i>
-------------	---

---

## Description

The data set contains lifestyle and behavioural traits that are considered exposures of the risk of mental health phenotypes. As outcomes, seven mental health phenotypes are considered, including (in alphabetic order) attention deficit hyperactivity disorder (ADHD), anorexia nervosa (AN), autism spectrum disorder (ASD), bipolar disorder (BD), cognition (COG), major depressive disorder (MDD) and schizophrenia (SCZ). As exposures, six lifestyle and behavioural traits that have previously been investigated for their protective/risk effects on mental health are considered, including (in alphabetic order) alcohol consumption (ALC), education (in years) (EDU), leisure screen time (LST), physical activity (PA), lifetime smoking index (SM) and sleep duration (SP)

## Usage

LBT2MD\_data

## Format

A data frame consisting of 708 independent Instrumental Variables (IVs) selected to be associated at genome-wide significance with the exposures after pruning or clumping. For details, see Zuber et al. (2024)

## References

Zuber V, Cronjé T, Cai N, Gill D, Bottolo L (2024). “Mendelian randomization for multiple exposures and outcomes with Bayesian Directed Acyclic Graphs exploration and causal effects estimation.” *Submitted*.

## Examples

```
# Example:  
  
data(LBT2MD_data)  
head(LBT2MD_data)
```

---

MD2LBT_data	<i>MrDAG data set: Mental health phenotypes that might impact lifestyle and behavioural traits</i>
-------------	--

---

## Description

The data set contains mental health phenotypes that are considered exposures of the risk of lifestyle and behavioural traits. As outcomes, six lifestyle and behavioural traits are considered, including (in alphabetic order) alcohol consumption (ALC), education (in years) (EDU), leisure screen time (LST), physical activity (PA), lifetime smoking index (SM) and sleep duration (SP). As exposures, seven mental health phenotypes are considered, including (in alphabetic order) attention deficit hyperactivity disorder (ADHD), anorexia nervosa (AN), autism spectrum disorder (ASD), bipolar disorder (BD), cognition (COG), major depressive disorder (MDD) and schizophrenia (SCZ)

## Usage

MD2LBT\_data

## Format

A data frame consisting of 470 independent Instrumental Variables (IVs) selected to be associated at genome-wide significance with the exposures after pruning or clumping. For details, see Zuber et al. (2024)

## References

Zuber V, Cronjé T, Cai N, Gill D, Bottolo L (2024). “Mendelian randomization for multiple exposures and outcomes with Bayesian Directed Acyclic Graphs exploration and causal effects estimation.” *Submitted*.

## Examples

```
# Example:  
  
data(MD2LBT_data)  
head(MD2LBT_data)
```

---

MrDAG

*MrDAG: Mendelian randomization for multiple outcomes and exposures with Bayesian Directed Acyclic Graphs (DAGs) exploration and causal effects estimation*

---

## Description

Markov chain Monte Carlo (MCMC) implementation of Bayesian multivariable, multi-response Mendelian randomization (MR) model for summary-level data with Directed Acyclic Graphs (DAGs) exploration and causal effects estimation



**Usage**

```
MrDAG(
  data,
  niter,
  burnin,
  thin,
  w = 0.05,
  a = NULL,
  U = NULL,
  tempmax = 10,
  MrDAGcheck = NULL,
  filename = "MrDAG_object",
  filepath = NULL,
  fastMCMC = TRUE,
  savememory = FALSE,
  seed = 31122021
)
```

**Arguments**

data	Number of observations (IVs in a summary-level MR design) times the number of traits (both outcomes and exposures). There is a restriction in the order of appearance of the traits in the data matrix: The group of outcomes ought to appear first, followed by the group of exposures. See also MrDAGcheck argument
niter	Number of MCMC iterations (excluding burn-in)
burnin	Number of MCMC iterations to be performed during burn-in
thin	Parameter that defines how often the MCMC output should be stored, i.e., at every thin-th iteration
w	Prior probability of edge inclusion (0.05 default) between each pair of nodes (vertices) in the graph
a	Degrees of freedom of the Wishart prior distribution on the precision matrix, i.e., the inverse of the covariance matrix between the traits (m default, where m is the total number of outcomes and exposures)
U	(Proportional to the) expected value of the Wishart prior distribution on the precision matrix, i.e., the inverse of the covariance matrix between the traits. Specifically, proportional to an m-dimensional diagonal matrix as default, where m is the total number of outcomes and exposures
tempmax	Annealing parameter T used to facilitate the convergence of the MCMC algorithm to the target distribution (10 default). Temperature 1/T increases linearly during the burn-in until T=1 at the end of the burn-in
MrDAGcheck	List object that contains the indices of the traits in data that are defined as outcomes and exposures. If NULL (default), no partition of the traits between outcomes and exposures is specified and MrDAG algorithm performs structure learning between all traits without constraints. Note that if so, this procedure does not correspond to an MR analysis

filename	Name of the file for MrDAG output object ("MrDAG_object" default). If NULL, the output is not saved
filepath	Directory where MrDAG object is saved. If the directory does not exist, MrDAG algorithm will create it by using the current working directory as the root directory. If the directory is not specified (NULL default), MrDAG object is saved in the current working directory
fastMCMC	If logical TRUE (default), the first valid DAG proposed by randomly selecting either to add, delete or swap a directed edge from the current DAG is used in the proposal distribution. Otherwise, all possible DAGs which differ from the current one by adding, deleting or swapping a directed edge are generated and, then, a DAG is sampled at random for the proposal distribution
savememory	If logical TRUE (default), the visited graph and the posterior draws from the modified Cholesky decomposition (L,D) (Zuber et al. (2024) and Castelletti and Consonni (2021)) are stored as a list object, otherwise they are stored as an array
seed	Seed to be used in the initialisation of the MCMC algorithm (3112021 default)

### Details

For details regarding the model and the algorithm, see Zuber et al. (2024)

### Value

The value returned is a list object `list(graphs, L,D, logmarglik, validpropMrDAG, acceptpropDAG, timeMrDAG, hyperpar, samplerpar, opt)`

- `graphs` Explored DAGs that belong to learned Completed Partially DAGs (CPDAGs) (Chickering (2002)) or Essential Graphs (EGs) (Andersson et al. (1997)). The class of graphs depends on the `savememory` option. The number of explored DAGs corresponds to the number of thinned (`thin`) MCMC iterations (excluding burn-in)
- `L` Posterior samples of the lower triangular matrix of the modified Cholesky decomposition (L,D) (Zuber et al. (2024) and Castelletti and Consonni (2021))
- `D` Posterior samples of the diagonal matrix of the modified Cholesky decomposition (L,D) (Zuber et al. (2024) and Castelletti and Consonni (2021))
- `logmarglik` Log-marginal likelihood of explored DAGs which belong to the Markov Equivalent Classes whose unique representative chain graphs are the EGs learned during MCMC iterations (including burn-in) without thinning
- `validpropMrDAG` If `MrDAGcheck` is different from NULL, the proportion of proposed DAGs that comply with the partial ordering (Perković et al. (2017)) implied by the partition of the traits between exposures and outcomes with directed causal effects only from the former to the latter
- `acceptpropDAG` Proportion of proposed DAGs that are accepted by the Metropolis-Hastings ratio after burn-in
- `timeMrDAG` Time in minutes employed by MrDAG algorithm to analyse the data. Considerable gains are obtained by specifying `fastMCMC = TRUE` option
- `hyperpar` List of the hyper-parameters `list(w, a, U, tempmax)` and, if specified, `MrDAGcheck` list

- samplerpar List of parameters used in the MCMC algorithm `list(niter, burnin, thin)`
- opt List of options used in MrDAG algorithm `list(fastMCMC, savememory, seed)`

## References

Andersson SA, Madigan D, Perlman MD (1997). “A characterization of Markov equivalence classes for acyclic digraphs.” *Annals of Statistics*, **25**(2), 505–541. doi:10.1214/aos/1031833662.

Castelletti F, Consonni G (2021). “Bayesian inference of causal effects from observational data in Gaussian graphical models.” *Biometrics*, **77**, 136–149. doi:10.1111/biom.13281.

Chickering DM (2002). “Learning equivalence classes of bayesian-network structures.” *Journal of Machine Learning Research*, **2**, 445–498. <http://www.ai.mit.edu/projects/jmlr/papers/volume2/chickering02a/chickering02a.pdf>.

Perković E, Kalisch M, Maathuis MH (2017). “Interpreting and using CPDAGs with background knowledge.” In *Proceedings UAI*. <http://auai.org/uai2017/proceedings/papers/120.pdf>.

Zuber V, Cronjé T, Cai N, Gill D, Bottolo L (2024). “Mendelian randomization for multiple exposures and outcomes with Bayesian Directed Acyclic Graphs exploration and causal effects estimation.” *Submitted*.

## Examples

```
# Example: Analysis of lifestyle and behavioural exposures that might impact mental health
# phenotypes. 708 independent Instrumental Variables (IVs) were selected to be associated at
# genome-wide significance with six lifestyle and behavioural traits after pruning or clumping,
# which are considered exposures of the risk of seven mental health outcome phenotypes

# After loading the data set

data(LBT2MD_data)

# the indices of the traits that define the outcomes and exposures are provided in a list object

MrDAGcheck <- NULL
MrDAGcheck$Y_idx <- 1 : 7   # Mental health phenotypes
MrDAGcheck$X_idx <- 8 : 13 # Lifestyle and behavioural traits

# MrDAG algorithm is run to generate 1,000 posterior samples of all unknowns

output <- MrDAG(data = LBT2MD_data,
                niter = 5000, burnin = 2500, thin = 5, tempmax = 20, w = 0.01,
                MrDAGcheck = MrDAGcheck, filename = NULL)
```

# Index

## \* datasets

LBT2MD\_data, [7](#)

MD2LBT\_data, [7](#)

get\_causaleffect, [2](#)

get\_causaleffects, [3](#)

get\_edgeprob, [5](#)

LBT2MD\_data, [7](#)

MD2LBT\_data, [7](#)

MrDAG, [8](#)