

# Package: winnerscurse (via r-universe)

October 3, 2024

**Type** Package

**Title** Winner's Curse Adjustment Methods for Summary Statistics from Genome-Wide Association Studies

**Version** 0.1.1

**Description** Designed to provide users with easy access to published methods which aim to correct for Winner's Curse using only summary statistics from genome-wide association studies. With merely estimates of effect size and associated standard error for each genetic variant, users are able to implement these methods to obtain more accurate estimates of the true effect sizes. These methods can be applied to data from both quantitative and binary traits.

**License** MIT + file LICENSE

**URL** <https://amandaforde.github.io/winnerscurse/>

**Encoding** UTF-8

**LazyData** true

**Suggests** testthat, knitr (>= 1.34), markdown, rmarkdown (>= 2.11), RColorBrewer, ggplot2

**RoxygenNote** 7.2.3

**Imports** stats, dplyr, scam, mgcv

**VignetteBuilder** knitr

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

**RemoteUrl** <https://github.com/amandaforde/winnerscurse>

**RemoteRef** HEAD

**RemoteSha** 2ed00bb119a5445e6a2985c0b6bf37a3068b9560

## Contents

winnerscurse-package . . . . .	2
BR_ss . . . . .	2

cl_interval . . . . .	3
conditional_likelihood . . . . .	4
condlike_rep . . . . .	5
empirical_bayes . . . . .	7
FDR_IQT . . . . .	8
MSE_minimizer . . . . .	9
se_adjust . . . . .	10
sim_stats . . . . .	11
UMVCUE . . . . .	12

<b>Index</b>	<b>14</b>
--------------	-----------

---

winnercurse-package	<i>winnercurse: Winner's Curse Adjustment Methods for Summary Statistics from Genome-Wide Association Studies</i>
---------------------	---

---

## Description

A package designed to provide users with easy access to published methods which aim to correct for Winner's Curse using only summary statistics from genome-wide association studies. With merely estimates of effect size and associated standard error for each genetic variant, users are able to implement these methods to obtain more accurate estimates of the true effect sizes. These methods can be applied to data from both quantitative and binary traits.

## Details

**Full documentation available here:** <https://amandaforde.github.io/winnercurse/>

---

BR_ss	<i>Bootstrap method for use with discovery GWAS</i>
-------	---

---

## Description

BR\_ss is a function which aims to use summary statistics to alleviate Winner's Curse bias in SNP-trait association estimates, obtained from a discovery GWAS. The function implements a parametric bootstrap approach, proposed by [Forde \*et al.\* \(2023\)](#). This approach was inspired by the bootstrap resampling method detailed in [Faye \*et al.\* \(2011\)](#), which requires original individual-level data.

## Usage

```
BR_ss(summary_data, seed_opt = FALSE, seed = 1998)
```

**Arguments**

summary_data	A data frame containing summary statistics from the discovery GWAS. It must have three columns with column names rsid, beta and se, respectively, and columns beta and se must contain numerical values. Each row must correspond to a unique SNP, identified by rsid. The function requires that there must be at least 5 SNPs as any less will result in issues upon usage of the smoothing spline.
seed_opt	A logical value which allows the user to choose if they wish to set a seed, in order to ensure reproducibility of adjusted estimates. Small differences can occur between iterations of the function with the same data set due to the use of parametric bootstrapping. The default setting is seed_opt=FALSE.
seed	A numerical value which specifies the seed used if seed_opt=TRUE. The default setting is the arbitrary value of 1998.

**Value**

A data frame with the inputted summary data occupying the first three columns. The new adjusted association estimates for each SNP are returned in the fourth column, namely beta\_BR\_ss. The SNPs are contained in this data frame according to their significance, with the most significant SNP, i.e. the SNP with the largest absolute  $z$ -statistic, now located in the first row of the data frame.

**References**

Forde, A., Hemani, G., & Ferguson, J. (2023). Review and further developments in statistical corrections for Winner's Curse in genetic association studies. *PLoS Genetics*, **19(9)**, e1010546.

**See Also**

[https://amandaforde.github.io/winnerscurse/articles/winners\\_curse\\_methods.html](https://amandaforde.github.io/winnerscurse/articles/winners_curse_methods.html) for illustration of the use of BR\_ss with a toy data set and further information regarding the computation of the adjusted SNP-trait association estimates.

---

cl_interval	<i>Confidence interval associated with discovery GWAS conditional likelihood methods</i>
-------------	--

---

**Description**

cl\_interval is a function that allows the user to obtain a confidence interval for the adjusted association estimates of significant SNPs, which have been obtained through the implementation of [conditional\\_likelihoood](#). This function produces one confidence interval for each significant SNP, based on the approach suggested in [Ghosh et al. \(2008\)](#). Note that in order for an appropriate confidence interval to be outputted for each significant SNP, the absolute value of the largest  $z$ -statistic in the data set must be less than 150.

**Usage**

```
cl_interval(summary_data, alpha = 5e-08, conf_level = 0.95)
```

### Arguments

summary_data	A data frame containing summary statistics from the discovery GWAS. It must have three columns with column names rsid, beta and se, respectively, and columns beta and se must contain numerical values. Each row must correspond to a unique SNP, identified by rsid.
alpha	A numerical value which specifies the desired genome-wide significance threshold. The default is given as $5e-8$ .
conf_level	A numerical value between 0 and 1 which determines the confidence interval to be computed. The default setting is 0.95 which results in the calculation of a 95% confidence interval for the adjusted association estimate for each SNP.

### Value

A data frame which combines the output of conditional\_likelihood with two additional columns, namely lower and upper, containing the lower and upper bounds of the required confidence interval for each significant SNP, respectively. However, if no SNPs are detected as significant in the data set, cl\_interval returns a warning message: "WARNING: There are no significant SNPs at this threshold."

### References

Ghosh, A., Zou, F., & Wright, F. A. (2008). Estimating odds ratios in genome scans: an approximate conditional likelihood approach. *American journal of human genetics*, **82**(5), 1064–1074. doi:10.1016/j.ajhg.2008.03.002

### See Also

[conditional\\_likelihood](#) for details on operation of conditional likelihood methods with summary statistics from discovery GWAS.

[https://amandaforde.github.io/winnerscurse/articles/standard\\_errors\\_confidence\\_intervals.html](https://amandaforde.github.io/winnerscurse/articles/standard_errors_confidence_intervals.html) for illustration of the use of cl\_interval with a toy data set and further information regarding the manner in which the confidence interval is computed.

---

conditional\_likelihood

*Conditional likelihood methods for use with discovery GWAS*

---

### Description

conditional\_likelihood is a function which uses summary statistics to correct bias created by the Winner's Curse phenomenon in the SNP-trait association estimates, obtained from a discovery GWAS, of SNPs which are considered significant. The function implements the approximate conditional likelihood approach, discussed in Ghosh *et al.* (2008), which suggests three different forms of a less biased association estimate. Note that if the  $z$ -statistic of a particular SNP is greater than 100, then merely the original naive estimate will be outputted for the second form of the adjusted estimate, namely beta.c12, for that SNP.

## Usage

```
conditional_likelihood(summary_data, alpha = 5e-08)
```

## Arguments

- `summary_data` A data frame containing summary statistics from the discovery GWAS. It must have three columns with column names `rsid`, `beta` and `se`, respectively, and columns `beta` and `se` must contain numerical values. Each row must correspond to a unique SNP, identified by `rsid`.
- `alpha` A numerical value which specifies the desired genome-wide significance threshold. The default is given as  $5e-8$ .

## Value

A data frame with summary statistics and adjusted association estimates of only those SNPs which have been deemed significant according to the specified threshold, `alpha`, i.e. SNPs with  $p$ -values less than `alpha`. The inputted summary data occupies the first three columns. The new adjusted association estimates for each SNP, as defined in the aforementioned paper, are contained in the next three columns, namely `beta.c11`, `beta.c12` and `beta.c13`. The SNPs are contained in this data frame according to their significance, with the most significant SNP, i.e. the SNP with the largest absolute  $z$ -statistic, now located in the first row of the data frame. However, if no SNPs are detected as significant in the data set, `conditional_likelihood` returns a warning message: "WARNING: There are no significant SNPs at this threshold."

## References

Ghosh, A., Zou, F., & Wright, F. A. (2008). Estimating odds ratios in genome scans: an approximate conditional likelihood approach. *American journal of human genetics*, **82**(5), 1064–1074. doi:10.1016/j.ajhg.2008.03.002

## See Also

[https://amandaforde.github.io/winnerscurse/articles/winners\\_curse\\_methods.html](https://amandaforde.github.io/winnerscurse/articles/winners_curse_methods.html) for illustration of the use of `conditional_likelihood` with a toy data set and further information regarding the computation of the adjusted SNP-trait association estimates for significant SNPs.

---

condlike\_rep

*Conditional likelihood method for use with discovery and replication GWASs*

---

## Description

`condlike_rep` is a function which attempts to produce less biased SNP-trait association estimates for SNPs deemed significant in the discovery GWAS, using summary statistics from both discovery and replication GWASs. The function computes three new association estimates for each SNP in a manner based closely on the method described in [Zhong and Prentice \(2008\)](#). It also returns confidence intervals for each new association estimate, if desired by the user.

**Usage**

```
condlike_rep(
  summary_disc,
  summary_rep,
  alpha = 5e-08,
  conf_interval = FALSE,
  conf_level = 0.95
)
```

**Arguments**

- |               |  |
|---------------|--|
| summary_disc  | A data frame containing summary statistics from the <i>discovery</i> GWAS. It must have three columns with column names <code>rsid</code> , <code>beta</code> and <code>se</code> , respectively, and columns <code>beta</code> and <code>se</code> must contain numerical values. Each row must correspond to a unique SNP, identified by <code>rsid</code> .   |
| summary_rep   | A data frame containing summary statistics from the <i>replication</i> GWAS. It must have three columns with column names <code>rsid</code> , <code>beta</code> and <code>se</code> , respectively, and all columns must contain numerical values. Each row must correspond to a unique SNP, identified by the numerical value <code>rsid</code> . SNPs must be ordered in the exact same manner as those in <code>summary_disc</code> , i.e. <code>summary_rep\$rsid</code> must be equivalent to <code>summary_disc\$rsid</code> . |
| alpha         | A numerical value which specifies the desired genome-wide significance threshold for the discovery GWAS. The default is given as $5e-8$ .  |
| conf_interval | A logical value which determines whether or not confidence intervals for each form of adjusted association estimate is also to be computed and outputted. The default is <code>conf_interval=FALSE</code> .  |
| conf_level    | A numerical value between 0 and 1 which specifies the confidence interval to be computed. The default setting is $0.95$ which results in the calculation of a 95% confidence interval for the adjusted association estimate for each SNP.  |

**Value**

A data frame with summary statistics and adjusted association estimates of only those SNPs which have been deemed significant in the discovery GWAS according to the specified threshold, `alpha`, i.e. SNPs with  $p$ -values less than `alpha`. The inputted summary data occupies the first five columns, in which the columns `beta_disc` and `se_disc` contain the statistics from the discovery GWAS and columns `beta_rep` and `se_rep` hold the replication GWAS statistics. For the default setting of `conf_interval=FALSE`, the new adjusted association estimates for each SNP, as defined in the aforementioned paper, are contained in the next three columns, namely `beta_com`, `beta_MLE` and `beta_MSE`. For the case when `conf_interval=TRUE`, the lower and upper boundaries of each confidence interval for each form of adjusted estimate are included in the data frame as well as the adjusted estimates for each SNP. The SNPs are contained in this data frame according to their significance, with the most significant SNP, i.e. the SNP with the largest absolute  $z$ -statistic, now located in the first row of the data frame. If no SNPs are detected as significant in the discovery GWAS, `condlike_rep` merely returns a data frame which combines the two inputted data sets.

## References

Zhong, H., & Prentice, R. L. (2008). Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics (Oxford, England)*, **9**(4), 621–634. doi:10.1093/biostatistics/kxn001

## See Also

[https://amandaforde.github.io/winnerscurse/articles/discovery\\_replication.html](https://amandaforde.github.io/winnerscurse/articles/discovery_replication.html) for illustration of the use of `condlike_rep` with toy data sets and further information regarding computation of the adjusted SNP-trait association estimates and their corresponding confidence intervals for significant SNPs.

---

 empirical\_bayes

*Empirical Bayes method for use with discovery GWAS*


---

## Description

`empirical_bayes` is a function which uses summary statistics to correct for bias induced by Winner's Curse in SNP-trait association estimates, obtained from a discovery GWAS. The function is strongly based on the method originally detailed in [Ferguson \*et al.\* \(2013\)](#). However, the function also includes all potential adaptations to the empirical Bayes method discussed in [Forde \*et al.\* \(2023\)](#).

## Usage

```
empirical_bayes(summary_data, method = "AIC")
```

## Arguments

<code>summary_data</code>	A data frame containing summary statistics from the discovery GWAS. It must have three columns with column names <code>rsid</code> , <code>beta</code> and <code>se</code> , respectively, and columns <code>beta</code> and <code>se</code> must contain numerical values. Each row must correspond to a unique SNP, identified by <code>rsid</code> . The function requires that there must be at least 50 SNPs as any less will result in very poor performance of the method.
<code>method</code>	A string which allows the user to choose what modelling approach to take for the purpose of estimating the log density function. The default setting is <code>method="AIC"</code> , which is the current published method. Other options include <code>method="fix_df"</code> , <code>method="scam"</code> , <code>method="gam_nb"</code> and <code>method="gam_po"</code> . If <code>method="fix_df"</code> , the degrees of freedom is set to 7. The other three options all enforce additional constraints on the shape of the estimated log density function.

## Value

A data frame with the inputted summary data occupying the first three columns. The new adjusted association estimates for each SNP are returned in the fourth column, namely `beta_EB`. The SNPs are contained in this data frame according to their significance, with the most significant SNP, i.e. the SNP with the largest absolute  $z$ -statistic, now located in the first row of the data frame.

## References

- Ferguson, J. P., Cho, J. H., Yang, C., & Zhao, H. (2013). Empirical Bayes correction for the Winner's Curse in genetic association studies. *Genetic epidemiology*, **37**(1), 60–68.
- Forde, A., Hemani, G., & Ferguson, J. (2023). Review and further developments in statistical corrections for Winner's Curse in genetic association studies. *PLoS Genetics*, **19**(9), e1010546.

## See Also

[https://amandaforde.github.io/winnerscurse/articles/winners\\_curse\\_methods.html](https://amandaforde.github.io/winnerscurse/articles/winners_curse_methods.html) for illustration of the use of `empirical_bayes` with a toy data set and further information regarding the computation of the adjusted SNP-trait association estimates.

---

FDR\_IQT

*FDR IQT method for use with discovery GWAS*

---

## Description

FDR\_IQT is a function which uses summary statistics to reduce Winner's Curse bias in SNP-trait association estimates, obtained from a discovery GWAS. The function implements the FDR Inverse Quantile Transformation method described in [Bigdeli et al. \(2016\)](#), which was established for this purpose.

## Usage

```
FDR_IQT(summary_data, min_pval = 1e-300)
```

## Arguments

- |                           |   |
|---------------------------|---|
| <code>summary_data</code> | A data frame containing summary statistics from the discovery GWAS. It must have three columns with column names <code>rsid</code> , <code>beta</code> and <code>se</code> , respectively, and columns <code>beta</code> and <code>se</code> must contain numerical values. Each row must correspond to a unique SNP, identified by the numerical value <code>rsid</code> .   |
| <code>min_pval</code>     | A numerical value whose purpose is to avoid zero $p$ -values as this introduces issues when <code>qnorm()</code> is applied. Any SNP for which its computed $p$ -value is found to be less than <code>min_pval</code> is merely re-assigned <code>min_pval</code> as its $p$ -value and the method proceeds. By definition, the method makes no adjustment to the association estimate of a SNP for which this has occurred with the presumption that in general, estimates of SNPs with $z > 37$ are not biased. The default value is <code>min_pval = 1e-300</code> . |

## Value

A data frame with the inputted summary data occupying the first three columns. The new adjusted association estimates for each SNP are returned in the fourth column, namely `beta_FIQT`. The SNPs are contained in this data frame according to their significance, with the most significant SNP, i.e. the SNP with the largest absolute  $z$ -statistic, now located in the first row of the data frame.



## References

Bigdeli, T. B., Lee, D., Webb, B. T., Riley, B. P., Vladimirov, V. I., Fanous, A. H., Kendler, K. S., & Bacanu, S. A. (2016). A simple yet accurate correction for winner's curse can predict signals discovered in much larger genome scans. *Bioinformatics (Oxford, England)*, **32(17)**, 2598–2603. doi:10.1093/bioinformatics/btw303

## See Also

[https://amandaforde.github.io/winnerscurse/articles/winners\\_curse\\_methods.html](https://amandaforde.github.io/winnerscurse/articles/winners_curse_methods.html) for illustration of the use of FDR\_IQT with a toy data set and further information regarding the computation of the adjusted SNP-trait association estimates.

---

MSE_minimizer	<i>MSE minimization method for use with discovery and replication GWASs</i>
---------------	---

---

## Description

MSE\_minimizer is a function which implements an approach that combines the association estimates obtained from discovery and replication GWASs to form a new combined estimate for each SNP. The method used by this function is inspired by that detailed in [Ferguson \*et al.\* \(2017\)](#).

## Usage

```
MSE_minimizer(summary_disc, summary_rep, alpha = 5e-08, spline = TRUE)
```

## Arguments

summary_disc	A data frame containing summary statistics from the <i>discovery</i> GWAS. It must have three columns with column names rsid, beta and se, respectively, and columns beta and se must contain numerical values. Each row must correspond to a unique SNP, identified by rsid. The function requires that there must be at least 5 SNPs as any less will result in issues upon usage of the smoothing spline.
summary_rep	A data frame containing summary statistics from the <i>replication</i> GWAS. It must have three columns with column names rsid, beta and se, respectively, and all columns must contain numerical values. Each row must correspond to a unique SNP, identified by the numerical value rsid. SNPs must be ordered in the exact same manner as those in summary_disc, i.e. summary_rep\$rsid must be equivalent to summary_disc\$rsid.
alpha	A numerical value which specifies the desired genome-wide significance threshold for the discovery GWAS. The default is given as 5e-8.
spline	A logical value which determines whether or not a cubic smoothing spline is to be used. When spline=FALSE, the value for $B$ in the formula detailed in the aforementioned paper is merely calculated as $B = \text{summary\_disc}\$beta - \text{summary\_rep}\$beta$ for each SNP. Alternatively, spline=TRUE applies a cubic smoothing spline to predict values for $B$ when $B = \text{summary\_disc}\$beta - \text{summary\_rep}\$beta$ is regressed on $z = \text{summary\_disc}\$beta / \text{summary\_disc}\$se$ , and it is these predicted values that are then used for $B$ .

**Value**

A data frame with summary statistics and adjusted association estimate of only those SNPs which have been deemed significant in the discovery GWAS according to the specified threshold,  $\alpha$ , i.e. SNPs with  $p$ -values less than  $\alpha$ . The inputted summary data occupies the first five columns, in which the columns `beta_disc` and `se_disc` contain the statistics from the discovery GWAS and columns `beta_rep` and `se_rep` hold the replication GWAS statistics. The new combination estimate for each SNP is contained in the final column, namely `beta_joint`. The SNPs are contained in this data frame according to their significance, with the most significant SNP, i.e. the SNP with the largest absolute  $z$ -statistic, now located in the first row of the data frame. If no SNPs are detected as significant in the discovery GWAS, `MSE_minimizer` merely returns a data frame which combines the two inputted data sets.

**References**

Ferguson, J., Alvarez-Iglesias, A., Newell, J., Hinde, J., & O'Donnell, M. (2017). Joint incorporation of randomised and observational evidence in estimating treatment effects. *Statistical Methods in Medical Research*, **28**(1), 235–247. doi:10.1177/0962280217720854

**See Also**

[https://amandaforde.github.io/winnerscurse/articles/discovery\\_replication.html](https://amandaforde.github.io/winnerscurse/articles/discovery_replication.html) for illustration of the use of `MSE_minimizer` with toy data sets and further information regarding computation of the combined SNP-trait association estimates for significant SNPs.

---

se\_adjust

*Standard errors of adjusted discovery GWAS estimates via parametric bootstrap*

---

**Description**

`se_adjust` is a function which allows the user to obtain approximate standard errors of adjusted association estimates, by means of parametric bootstrapping. Standard errors can be evaluated for estimates which have been corrected with the Empirical Bayes method, FDR Inverse Quantile Transformation method or the bootstrap method. Note that in comparison to the other functions in this package, this function can be computationally intensive and take a several minutes to run, depending on the size of the data set, the method and the number of bootstraps chosen.

**Usage**

```
se_adjust(summary_data, method, n_boot = 100)
```

**Arguments**

`summary_data` A data frame containing summary statistics from the discovery GWAS. It must have three columns with column names `rsid`, `beta` and `se`, respectively, and columns `beta` and `se` must contain numerical values. Each row must correspond to a unique SNP, identified by `rsid`.

method	A string specifying the function to be implemented on each of the bootstrap samples. It should take the form "BR_ss", "empirical_bayes" or "FDR_IQT".
n_boot	A numerical value which determines the number of bootstrap repetitions to be used. it must be greater than 1. The default value is 100.

**Value**

A data frame which combines the output of the chosen method with an additional column, namely `adj_se`. This column provides the standard errors of the adjusted association estimates for each SNP.

**See Also**

[empirical\\_bayes](#), [BR\\_ss](#) and [FDR\\_IQT](#) for details on operation of these methods with summary statistics from discovery GWAS.

[https://amandaforde.github.io/winnerscurse/articles/standard\\_errors\\_confidence\\_intervals.html](https://amandaforde.github.io/winnerscurse/articles/standard_errors_confidence_intervals.html) for illustration of the use of `se_adjust` with a toy data set and further information regarding the manner in which the standard errors are computed.

---

sim_stats	<i>Simulating GWAS summary statistics</i>
-----------	---

---

**Description**

`sim_stats` is a function which can be used to simulate summary statistics for a set of independent SNPs for both discovery and replication GWASs. This function allows the user to create toy datasets with which they can explore the implementation of the Winner's Curse correction methods.

**Usage**

```
sim_stats(
  nsnp = 10^6,
  h2 = 0.4,
  prop_effect = 0.01,
  nid = 50000,
  rep = FALSE,
  rep_nid = 50000
)
```

**Arguments**

nsnp	A numerical value which specifies the total number of SNPs that the user wishes to simulate summary statistics for. The default is 1,000,000 SNPs, i.e. $nsnp=10^6$ .
h2	A numerical value between 0 and 1 which represents the desired heritability of the trait of interest, or in other words, the total variance explained in the trait by all SNPs. The default is a moderate heritability value of 0.4, $h2=0.4$ .

prop_effect	A numerical value between 0 and 1 which determines the trait's polygenicity, the fraction of the total number of SNPs which are truly associated with the trait. The default setting is prop_effect = 0.01.
nid	A numerical value which specifies the number of individuals that the discovery GWAS has been performed with. This value defaults to 50,000 individuals, nid=50000.
rep	A logical value which allows the user to state whether they would also like to simulate summary statistics for a replication GWAS based on the same parameters and true effect sizes. The default setting is rep=FALSE.
rep_nid	A numerical value which specifies the number of individuals that the replication GWAS has been performed with. Similar to nid, this value defaults to 50,000 individuals, nid=50000.

### Value

A list containing three different components in the form of data frames, true, disc and rep. The first element, true has two columns, rsid which contains identification numbers for each SNP and true\_beta which is each SNP's simulated true association value. disc has three columns representing the summary statistics one would obtain in a discovery GWAS. For each SNP, this data frame contains its ID number, its estimated effect size, beta, and associated standard error, se. Similarly, if the rep argument in the function has been set to TRUE, then the data frame, rep has three columns representing the summary statistics one would obtain in a replication GWAS. In this data frame, just as with disc, the values for beta have been simulated using the true association values, true\_beta, and the standard errors are reflective of the chosen sample size. If rep=FALSE, NULL is merely returned for this third element.

---

UMVCUE

*UMVCUE method for use with discovery and replication GWASs*

---

### Description

UMVCUE is a function which aims to produce less biased SNP-trait association estimates for SNPs deemed significant in the discovery GWAS, using summary statistics from both discovery and replication GWASs. The function implements the method described in [Bowden and Dudbridge \(2009\)](#), which was established for this purpose.

### Usage

```
UMVCUE(summary_disc, summary_rep, alpha = 5e-08)
```

### Arguments

summary_disc	A data frame containing summary statistics from the <i>discovery</i> GWAS. It must have three columns with column names rsid, beta and se, respectively, and columns beta and se must contain numerical values. Each row must correspond to a unique SNP, identified by rsid.
--------------	---

summary_rep	A data frame containing summary statistics from the <i>replication</i> GWAS. It must have three columns with column names <i>rsid</i> , <i>beta</i> and <i>se</i> , respectively, and all columns must contain numerical values. Each row must correspond to a unique SNP, identified by the numerical value <i>rsid</i> . SNPs must be ordered in the exact same manner as those in <i>summary_disc</i> , i.e. <code>summary_rep\$rsid</code> must be equivalent to <code>summary_disc\$rsid</code> .
alpha	A numerical value which specifies the desired genome-wide significance threshold for the discovery GWAS. The default is given as $5e-8$ .

### Value

A data frame with summary statistics and adjusted association estimate of only those SNPs which have been deemed significant in the discovery GWAS according to the specified threshold, *alpha*, i.e. SNPs with *p*-values less than *alpha*. The inputted summary data occupies the first five columns, in which the columns *beta\_disc* and *se\_disc* contain the statistics from the discovery GWAS and columns *beta\_rep* and *se\_rep* hold the replication GWAS statistics. The new adjusted association estimate for each SNP, as defined in the aforementioned paper, is contained in the final column, namely *beta\_UMVCUE*. The SNPs are contained in this data frame according to their significance, with the most significant SNP, i.e. the SNP with the largest absolute *z*-statistic, now located in the first row of the data frame. If no SNPs are detected as significant in the discovery GWAS, *UMVCUE* merely returns a data frame which combines the two inputted data sets.

### References

Bowden, J., & Dudbridge, F. (2009). Unbiased estimation of odds ratios: combining genomewide association scans with replication studies. *Genetic epidemiology*, **33**(5), 406–418. doi:10.1002/gepi.20394

### See Also

[https://amandaforde.github.io/winnerscurse/articles/discovery\\_replication.html](https://amandaforde.github.io/winnerscurse/articles/discovery_replication.html) for illustration of the use of *UMVCUE* with toy data sets and further information regarding computation of the adjusted SNP-trait association estimates for significant SNPs.

# Index

BR\_ss, [2](#), [11](#)

cl\_interval, [3](#)

conditional\_likelihood, [3](#), [4](#), [4](#)

condlike\_rep, [5](#)

empirical\_bayes, [7](#), [11](#)

FDR\_IQT, [8](#), [11](#)

MSE\_minimizer, [9](#)

se\_adjust, [10](#)

sim\_stats, [11](#)

UMVCUE, [12](#)

winnercurse (winnercurse-package), [2](#)

winnercurse-package, [2](#)