# Package: mrSampleOverlap (via r-universe)

March 2, 2025

**Title** Estimate Bias Due To Sample Overlap In Mendelian Randomization Studies

**Version** 0.1.1

**Description** A function to estimate bias due to sample overlap in Mendelian Randomization studies.

**License** MIT + file LICENSE

**URL** https://github.com/mglev1n/mrSampleOverlap, https://mglev1n.github.io/mrSampleOverlap/

**BugReports** https://github.com/mglev1n/mrSampleOverlap/issues

**Imports** dplyr, magrittr, stats, tibble

**Suggests** knitr, rmarkdown, testthat

**VignetteBuilder** knitr

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.2.1

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

**RemoteUrl** https://github.com/mglev1n/mrSampleOverlap

**RemoteRef** HEAD

**RemoteSha** e59d2794c7c0097081bcd4a0d960aded4ef4a9ca

## Contents

## estimate_f *Estimate F-statistic*

### Description

Estimate F-statistic

### Usage

```
estimate_f(samplesize_exposure, n_variants, rsq_exposure, lci_95 = FALSE)
```

### Arguments

| | |
|---|---|
| samplesize_exposure | (numeric) Sample size of population used to define genetic instrument for the exposure of interest |
| n_variants | (numeric) Number of genetic variants included in genetic instrument for the exposure of interest |
| rsq_exposure | (numeric) $R^2$ value (coefficient of determination) of genetic instrument for the exposure of interest; used to estimate |
| lci_95 | (logical; default = FALSE) If TRUE, the function will return the lower limit of the one-sided 95% confidence interval of the F-statistic, which may represent a more conservative/less optimistic estimate |

### Value

Numeric vector containing the estimated F-statistic

### Examples

```
estimate_f(samplesize_exposure = 361194, n_variants = 196, rsq_exposure = 0.068, lci_95 = FALSE)

# return lower bound of one-sided 95% confidence interval of F-statistic
estimate_f(samplesize_exposure = 361194, n_variants = 196, rsq_exposure = 0.068, lci_95 = TRUE)
```

## estimate_overlap_bias *Estimate bias due to sample overlap*

### Description

Estimate bias due to sample overlap

## Usage

```
estimate_overlap_bias(
  samplesize_exposure,
  samplesize_outcome,
  n_variants,
  rsq_exposure,
  exp_f = NULL,
  lci_95 = FALSE,
  case_prop = 0,
  ols_bias,
  overlap_prop,
  var_x = 1,
  var_y = 1
)
```

## Arguments

| | |
|---|---|
| samplesize_exposure | (numeric) Sample size of population used to define genetic instrument for the exposure of interest |
| samplesize_outcome | (numeric) Sample size of population used for the outcome of interest |
| n_variants | (numeric) Number of genetic variants included in genetic instrument for the exposure of interest |
| rsq_exposure | (numeric) $R^2$ value (coefficient of determination) of genetic instrument for the exposure of interest; used to estimate F-statistic |
| exp_f | (numeric; optional) F-statistic for the genetic instrument (if provided, this value will be used, rather than an estimate based on the $R^2$) |
| lci_95 | (logical; default = FALSE) If TRUE, the function will return estimates of bias and type 1 error based on the lower limit of the one-sided 95% confidence interval of the F-statistic, which may represent a more conservative/less optimistic estimate of bias |
| case_prop | (numeric; optional) Proportion of cases (eg. cases/total samplesize) if outcome is binary |
| ols_bias | (numeric) Observational (biased) effect estimate (if known); otherwise, provide a hypothetical value |
| overlap_prop | (numeric; range = 0 to 1) Proportion of overlapping samples between exposure and outcome studies (if known); otherwise, provide a hypothetical value |
| var_x | (numeric) Variance in the exposure; default is 1 when the exposure is reported in standard deviation units |
| var_y | (numeric) Variance in the exposure; default is 1 when the exposure is reported in standard deviation units |

## Value

A tibble containing columns for the bias and type1_error

## Examples

```
# Binary outcome
estimate_overlap_bias(
  samplesize_exposure = 361194,
  samplesize_outcome = 1125328,
  case_prop = 0.035,
  rsq_exposure = 0.068,
  n_variants = 196,
  ols_bias = 0.2,
  overlap_prop = 0.3
)

# Continuous outcome
estimate_overlap_bias(
  samplesize_exposure = 361194,
  samplesize_outcome = 1125328,
  rsq_exposure = 0.068,
  n_variants = 196,
  ols_bias = 0.2,
  overlap_prop = 0.3
)
```

# Index