

# Package ‘multibias’

January 28, 2024

**Type** Package

**Title** Simultaneous Multi-Bias Adjustment

**Version** 1.4.0

**Maintainer** Paul Brendel <pcbrendel@gmail.com>

**Description** Quantify the causal effect of a binary exposure on a binary outcome with adjustment for multiple biases. The functions can simultaneously adjust for any combination of uncontrolled confounding, exposure/outcome misclassification, and selection bias. The underlying method generalizes the concept of combining inverse probability of selection weighting with predictive value weighting. Simultaneous multi-bias analysis can be used to enhance the validity and transparency of real-world evidence obtained from observational, longitudinal studies. Based on the work from Paul Brendel, Aracelis Torres, and Onyebuchi Arah (2023) <[doi:10.1093/ije/dyad001](https://doi.org/10.1093/ije/dyad001)>.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**Depends** R (>= 2.10)

**RoxygenNote** 7.2.3

**Imports** dplyr (>= 1.1.3), magrittr (>= 2.0.3), rlang (>= 1.1.1)

**Suggests** knitr, rmarkdown, testthat (>= 3.0.0)

**URL** <https://github.com/pcbrendel/multibias>

**BugReports** <https://github.com/pcbrendel/multibias/issues>

**Config/testthat/edition** 3

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Paul Brendel [aut, cre, cph]

**Repository** CRAN

**Date/Publication** 2024-01-27 23:40:02 UTC

**R topics documented:**

adjust_emc	3
adjust_emc_omc	4
adjust_emc_sel	5
adjust_multinom_emc_omc	7
adjust_multinom_uc_emc	8
adjust_multinom_uc_emc_sel	10
adjust_multinom_uc_omc	12
adjust_omc	14
adjust_omc_sel	15
adjust_sel	16
adjust_uc	18
adjust_uc_emc	19
adjust_uc_emc_sel	20
adjust_uc_omc	22
adjust_uc_sel	23
df_emc	25
df_emc_omc	25
df_emc_omc_source	26
df_emc_sel	27
df_emc_sel_source	27
df_emc_source	28
df_omc	29
df_omc_sel	29
df_omc_sel_source	30
df_omc_source	30
df_sel	31
df_sel_source	32
df_uc	32
df_uc_emc	33
df_uc_emc_sel	34
df_uc_emc_sel_source	34
df_uc_emc_source	35
df_uc_omc	36
df_uc_omc_source	36
df_uc_sel	37
df_uc_sel_source	37
df_uc_source	38
evans	39

---

adjust_emc	<i>Adjust for exposure misclassification.</i>
------------	---

---

### Description

adjust\_emc returns the exposure-outcome odds ratio and confidence interval, adjusted for exposure misclassification.

### Usage

```
adjust_emc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  x_model_coefs,
  level = 0.95
)
```

### Arguments

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
x_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(X = 1)) = \delta_0 + \delta_1 X^* + \delta_2 Y + \delta_{2+j} C_j$ , where X represents the binary true exposure, $X^*$ is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters is therefore $3 + j$ .
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

### Details

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

### Value

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

**Examples**

```
adjust_emc(
  evans,
  exposure = "SMK",
  outcome = "CHD",
  confounders = "HPT",
  x_model_coefs = c(qlogis(0.01), log(6), log(2), log(2))
)
```

---

adjust_emc_omc	<i>Adust for exposure misclassification and outcome misclassification.</i>
----------------	--

---

**Description**

adjust\_emc\_omc returns the exposure-outcome odds ratio and confidence interval, adjusted for exposure misclassification and outcome misclassification.

**Usage**

```
adjust_emc_omc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  x_model_coefs,
  y_model_coefs,
  level = 0.95
)
```

**Arguments**

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
x_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(X = 1)) = \delta_0 + \delta_1 X^* + \delta_2 Y^* + \delta_2 + jC_j$ , where $X$ represents the binary true exposure, $X^*$ is the binary misclassified exposure, $Y^*$ is the binary misclassified outcome, $C$ represents the vector of binary measured confounders (if any), and $j$ corresponds to the number of measured confounders. The number of parameters is therefore $3 + j$ .

y_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(Y = 1)) = \beta_0 + \beta_1 X + \beta_2 Y^* + \beta_3 + \sum_j C_j$ , where Y represents the binary true exposure, X is the binary exposure, Y is the binary misclassified outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters is therefore 3 + j.
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

### Details

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

### Value

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

### Examples

```
adjust_emc_omc(
  df_emc_omc,
  exposure = "Xstar",
  outcome = "Ystar",
  confounders = "C1",
  x_model_coefs = c(-2.15, 1.64, 0.35, 0.38),
  y_model_coefs = c(-3.10, 0.63, 1.60, 0.39)
)
```

---

adjust\_emc\_sel

*Adjust for exposure misclassification and selection bias.*

---

### Description

adjust\_emc\_sel returns the exposure-outcome odds ratio and confidence interval, adjusted for exposure misclassification and selection bias.

### Usage

```
adjust_emc_sel(
  data,
  exposure,
  outcome,
```

```

    confounders = NULL,
    x_model_coefs,
    s_model_coefs,
    level = 0.95
  )

```

### Arguments

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
x_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(X = 1)) = \delta_0 + \delta_1 X^* + \delta_2 Y + \delta_2 + jC_j$ , where X represents the binary true exposure, X* is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters is therefore 3 + j.
s_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(S = 1)) = \beta_0 + \beta_1 X^* + \beta_2 Y + \beta_2 + jC_j$ , where S represents binary selection, X* is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters is therefore 3 + j.
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

### Details

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

### Value

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

### Examples

```

adjust_emc_sel(
  df_emc_sel,
  exposure = "Xstar",
  outcome = "Y",
  confounders = "C1",
  x_model_coefs = c(-2.78, 1.62, 0.58, 0.34),

```

```
s_model_coefs = c(0.04, 0.18, 0.92, 0.05)
)
```

---

```
adjust_multinom_emc_omc
```

*Adjust for exposure misclassification and outcome misclassification*

---

## Description

adjust\_multinom\_emc\_omc returns the exposure-outcome odds ratio and confidence interval, adjusted for exposure misclassification and outcome misclassification.

## Usage

```
adjust_multinom_emc_omc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  x1y0_model_coefs,
  x0y1_model_coefs,
  x1y1_model_coefs,
  level = 0.95
)
```

## Arguments

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
x1y0_model_coefs	The regression coefficients corresponding to the model: $\log(P(X = 1, Y = 0)/P(X = 0, Y = 0)) = \gamma_{1,0} + \gamma_{1,1}X^* + \gamma_{1,2}Y^* + \gamma_{1,2+j}C_j$ , where X is the binary true exposure, Y is the binary true outcome, X* is the binary misclassified exposure, Y* is the binary misclassified outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.
x0y1_model_coefs	The regression coefficients corresponding to the model: $\log(P(X = 0, U = 1)/P(X = 0, U = 0)) = \gamma_{2,0} + \gamma_{2,1}X^* + \gamma_{2,2}Y^* + \gamma_{2,2+j}C_j$ , where X is the binary true exposure, Y is the binary true outcome, X* is the binary misclassified exposure, Y* is the binary misclassified outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.

x1y1\_model\_coefs

The regression coefficients corresponding to the model:  $\log(P(X = 1, Y = 1)/P(X = 0, Y = 0)) = \gamma_{3,0} + \gamma_{3,1}X^* + \gamma_{3,2}Y^* + \gamma_{3,2+j}C_j$ , where X is the binary true exposure, Y is the binary true outcome, X\* is the binary misclassified exposure, Y\* is the binary misclassified outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.

level

Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

### Details

This function uses one bias model, a multinomial logistic regression model, to predict the exposure (X) and outcome (Y). If separate bias models for X and Y are desired, use `adjust_emc_omc`.

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

### Value

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

### Examples

```
adjust_multinom_emc_omc(
  df_emc_omc,
  exposure = "Xstar",
  outcome = "Ystar",
  confounders = c("C1", "C2", "C3"),
  x1y0_model_coefs = c(-2.86, 1.63, 0.23, 0.37, -0.22, 0.87),
  x0y1_model_coefs = c(-3.26, 0.22, 1.60, 0.41, -0.93, 0.28),
  x1y1_model_coefs = c(-5.62, 1.83, 1.83, 0.74, -1.15, 1.19)
)
```

---

adjust\_multinom\_uc\_emc

*Adjust for uncontrolled confounding and exposure misclassification.*

---

### Description

`adjust_multinom_uc_emc` returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding and exposure misclassification.



**Usage**

```
adjust_multinom_uc_emc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  x1u0_model_coefs,
  x0u1_model_coefs,
  x1u1_model_coefs,
  level = 0.95
)
```

**Arguments**

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
x1u0_model_coefs	The regression coefficients corresponding to the model: $\log(P(X = 1, U = 0)/P(X = 0, U = 0)) = \gamma_{1,0} + \gamma_{1,1}X^* + \gamma_{1,2}Y + \gamma_{1,2+j}C_j$ , where X is the binary true exposure, U is the binary unmeasured confounder, X* is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.
x0u1_model_coefs	The regression coefficients corresponding to the model: $\log(P(X = 0, U = 1)/P(X = 0, U = 0)) = \gamma_{2,0} + \gamma_{2,1}X^* + \gamma_{2,2}Y + \gamma_{2,2+j}C_j$ , where X is the binary true exposure, U is the binary unmeasured confounder, X* is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.
x1u1_model_coefs	The regression coefficients corresponding to the model: $\log(P(X = 1, U = 1)/P(X = 0, U = 0)) = \gamma_{3,0} + \gamma_{3,1}X^* + \gamma_{3,2}Y + \gamma_{3,2+j}C_j$ , where X is the binary true exposure, U is the binary unmeasured confounder, X* is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

**Details**

This function uses one bias model, a multinomial logistic regression model, to predict the uncontrolled confounder (U) and exposure (X). If separate bias models for X and U are desired, use adjust\_uc\_emc.

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

### Value

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

### Examples

```
adjust_multinom_uc_emc(
  df_uc_emc,
  exposure = "Xstar",
  outcome = "Y",
  confounders = "C1",
  x1u0_model_coefs = c(-2.82, 1.62, 0.68, -0.06),
  x0u1_model_coefs = c(-0.20, 0.00, 0.68, -0.05),
  x1u1_model_coefs = c(-2.36, 1.62, 1.29, 0.27)
)
```

---

adjust\_multinom\_uc\_emc\_sel

*Adjust for uncontrolled confounding, exposure misclassification, and selection bias.*

---

### Description

`adjust_multinom_uc_emc_sel` returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding, exposure misclassification, and selection bias.

### Usage

```
adjust_multinom_uc_emc_sel(
  data,
  exposure,
  outcome,
  confounders = NULL,
  x1u0_model_coefs,
  x0u1_model_coefs,
  x1u1_model_coefs,
  s_model_coefs,
  level = 0.95
)
```

**Arguments**

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
x1u0_model_coefs	The regression coefficients corresponding to the model: $\log(P(X = 1, U = 0)/P(X = 0, U = 0)) = \gamma_{1,0} + \gamma_{1,1}X^* + \gamma_{1,2}Y + \gamma_{1,2+j}C_j$ , where X is the binary true exposure, U is the binary unmeasured confounder, X* is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.
x0u1_model_coefs	The regression coefficients corresponding to the model: $\log(P(X = 0, U = 1)/P(X = 0, U = 0)) = \gamma_{2,0} + \gamma_{2,1}X^* + \gamma_{2,2}Y + \gamma_{2,2+j}C_j$ , where X is the binary true exposure, U is the binary unmeasured confounder, X* is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.
x1u1_model_coefs	The regression coefficients corresponding to the model: $\log(P(X = 1, U = 1)/P(X = 0, U = 0)) = \gamma_{3,0} + \gamma_{3,1}X^* + \gamma_{3,2}Y + \gamma_{3,2+j}C_j$ , where X is the binary true exposure, U is the binary unmeasured confounder, X* is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.
s_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(S = 1)) = \beta_0 + \beta_1X^* + \beta_2Y + \beta_{2+j}C_j$ , where S represents binary selection, X* is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

**Details**

This function uses one bias model, a multinomial logistic regression model, to predict the uncontrolled confounder (U) and exposure (X). If separate bias models for X and U are desired, use `adjust_uc_emc_sel`.

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

**Value**

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

**Examples**

```
adjust_multinom_uc_omc_sel(
  df_uc_omc_sel,
  exposure = "Xstar",
  outcome = "Y",
  confounders = c("C1", "C2", "C3"),
  x1u0_model_coefs = c(-2.78, 1.62, 0.61, 0.36, -0.27, 0.88),
  x0u1_model_coefs = c(-0.17, -0.01, 0.71, -0.08, 0.07, -0.15),
  x1u1_model_coefs = c(-2.36, 1.62, 1.29, 0.25, -0.06, 0.74),
  s_model_coefs = c(0.00, 0.26, 0.78, 0.03, -0.02, 0.10)
)
```

---

adjust\_multinom\_uc\_omc

*Adjust for uncontrolled confounding and outcome misclassification.*

---

**Description**

adjust\_multinom\_uc\_omc returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding and outcome misclassification.

**Usage**

```
adjust_multinom_uc_omc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  u1y0_model_coefs,
  u0y1_model_coefs,
  u1y1_model_coefs,
  level = 0.95
)
```

**Arguments**

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.

**u1y0\_model\_coefs**

The regression coefficients corresponding to the model:  $\log(P(U = 1, Y = 0)/P(U = 0, Y = 0)) = \gamma_{1,0} + \gamma_{1,1}X + \gamma_{1,2}Y^* + \gamma_{1,2+j}C_j$ , where U is the binary unmeasured confounder, Y is the binary true outcome, X is the binary exposure, Y\* is the binary misclassified outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.

**u0y1\_model\_coefs**

The regression coefficients corresponding to the model:  $\log(P(U = 0, Y = 1)/P(U = 0, Y = 0)) = \gamma_{2,0} + \gamma_{2,1}X + \gamma_{2,2}Y^* + \gamma_{2,2+j}C_j$ , where U is the binary unmeasured confounder, Y is the binary true outcome, X is the binary exposure, Y\* is the binary misclassified outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.

**u1y1\_model\_coefs**

The regression coefficients corresponding to the model:  $\log(P(U = 1, Y = 1)/P(U = 0, Y = 0)) = \gamma_{3,0} + \gamma_{3,1}X + \gamma_{3,2}Y^* + \gamma_{3,2+j}C_j$ , where U is the binary unmeasured confounder, Y is the binary true outcome, X is the binary exposure, Y\* is the binary misclassified outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.

**level**

Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

**Details**

This function uses one bias model, a multinomial logistic regression model, to predict the uncontrolled confounder (U) and outcome (Y). If separate bias models for X and U are desired, use `adjust_uc_omc`.

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

**Value**

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

**Examples**

```
adjust_multinom_uc_omc(
  df_uc_omc,
  "X",
  "Ystar",
  "C1",
  u1y0_model_coefs = c(-0.19, 0.61, 0.00, -0.07),
```

```

u0y1_model_coefs = c(-3.21, 0.60, 1.60, 0.36),
u1y1_model_coefs = c(-2.72, 1.24, 1.59, 0.34)
)

```

---

adjust\_omc

*Adjust for outcome misclassification.*


---

### Description

adjust\_omc returns the exposure-outcome odds ratio and confidence interval, adjusted for outcome misclassification.

### Usage

```

adjust_omc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  y_model_coefs,
  level = 0.95
)

```

### Arguments

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
y_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(Y = 1)) = \_delta_0 + \_delta_1 X + \_delta_2 Y^* + \_delta_{2+j} C_j$ , where Y represents the binary true outcome, X is the binary exposure, Y* is the binary misclassified outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters is therefore 3 + j.
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

### Details

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

**Value**

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

**Examples**

```
adjust_omc(
  evans,
  exposure = "SMK",
  outcome = "CHD",
  confounders = "HPT",
  y_model_coefs = c(qlogis(0.01), log(1.5), log(5), log(1.5))
)
```

---

adjust_omc_sel	<i>Adjust for outcome misclassification and selection bias.</i>
----------------	---

---

**Description**

adjust\_omc\_sel returns the exposure-outcome odds ratio and confidence interval, adjusted for outcome misclassification and selection bias.

**Usage**

```
adjust_omc_sel(
  data,
  exposure,
  outcome,
  confounders = NULL,
  y_model_coefs,
  s_model_coefs,
  level = 0.95
)
```

**Arguments**

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
y_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(Y = 1)) = \delta_0 + \delta_1 X + \delta_2 Y^* + \delta_{2+j} C_j$ , where Y represents the binary true outcome, X is the binary exposure, Y* is the binary misclassified outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters is therefore 3 + j.

s_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(S = 1)) = \beta_0 + \beta_1 X + \beta_2 Y^* + \beta_{2+j} C_j$ , where S represents binary selection, X is the binary exposure, Y* is the binary misclassified outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters is therefore 3 + j.
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

### Details

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

### Value

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

### Examples

```
adjust_omc_sel(
  df_omc_sel,
  exposure = "X",
  outcome = "Ystar",
  confounders = "C1",
  y_model_coefs = c(-3.24, 0.58, 1.59, 0.45),
  s_model_coefs = c(0.03, 0.92, 0.12, 0.05)
)
```

---

adjust_sel	<i>Adust for selection bias.</i>
------------	----------------------------------

---

### Description

adjust\_sel returns the exposure-outcome odds ratio and confidence interval, adjusted for selection bias.

### Usage

```
adjust_sel(
  data,
  exposure,
  outcome,
```



```

    confounders = NULL,
    s_model_coefs,
    level = 0.95
  )

```

### Arguments

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
s_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(S = 1)) = \beta_0 + \beta_1 X + \beta_2 Y$ , where S represents binary selection, X is the binary exposure, Y is the binary outcome. The number of parameters is therefore 3.
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

### Details

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

### Value

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

### Examples

```

adjust_sel(
  evans,
  exposure = "SMK",
  outcome = "CHD",
  confounders = "HPT",
  s_model_coefs = c(qlogis(0.25), log(0.75), log(0.75))
)

```

---

adjust\_uc                      *Adjust for uncontrolled confounding.*

---

### Description

adjust\_uc returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding.

### Usage

```
adjust_uc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  u_model_coefs,
  level = 0.95
)
```

### Arguments

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
u_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(U = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y + \alpha_{2+j} C_j$ , where U is the (binary) unmeasured confounder, X is the (binary) exposure, Y is the (binary) outcome, C represents the vector of (binary) measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters therefore equals $3 + j$ .
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

### Details

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

### Value

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

**Examples**

```
adjust_uc(
  evans,
  exposure = "SMK",
  outcome = "CHD",
  confounders = "HPT",
  u_model_coefs = c(qlogis(0.25), log(0.5), log(2.5), log(2)),
)
```

adjust\_uc\_emc

*Adjust for uncontrolled confounding and exposure misclassification.***Description**

adjust\_uc\_emc returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding and exposure misclassification.

**Usage**

```
adjust_uc_emc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  u_model_coefs,
  x_model_coefs,
  level = 0.95
)
```

**Arguments**

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
u_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(U = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y$ , where U is the binary unmeasured confounder, X is the binary true exposure, Y is the binary outcome. The number of parameters therefore equals 3.
x_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(X = 1)) = \delta_0 + \delta_1 X^* + \delta_2 Y + \delta_{2+j} C_j$ , where X represents the binary true exposure, X* is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters therefore equals 3 + j.
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

**Details**

This function uses two separate logistic regression models to predict the uncontrolled confounder (U) and exposure (X). If a single bias model for jointly modeling X and U is desired use `adjust_multinom_uc_emc`.

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

**Value**

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

**Examples**

```
adjust_uc_emc(
  df_uc_emc,
  exposure = "Xstar",
  outcome = "Y",
  confounders = "C1",
  u_model_coefs = c(-0.23, 0.63, 0.66),
  x_model_coefs = c(-2.47, 1.62, 0.73, 0.32)
)
```

---

<code>adjust_uc_emc_sel</code>	<i>Adjust for uncontrolled confounding, exposure misclassification, and selection bias.</i>
--------------------------------	---

---

**Description**

`adjust_uc_emc_sel` returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding, exposure misclassification, and selection bias.

**Usage**

```
adjust_uc_emc_sel(
  data,
  exposure,
  outcome,
  confounders = NULL,
  u_model_coefs,
  x_model_coefs,
  s_model_coefs,
  level = 0.95
)
```

**Arguments**

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
u_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(U = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y$ , where U is the binary unmeasured confounder, X is the binary true exposure, and Y is the binary outcome. The number of parameters therefore equals 3.
x_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(X = 1)) = \delta_0 + \delta_1 X^* + \delta_2 Y + \delta_{2+j} C_j$ , where X represents binary true exposure, X* is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters therefore equals 3 + j.
s_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(S = 1)) = \beta_0 + \beta_1 X^* + \beta_2 Y + \beta_{2+j} C_j$ , where S represents binary selection, X* is the binary misclassified exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters therefore equals 3 + j.
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

**Details**

This function uses two separate logistic regression models to predict the uncontrolled confounder (U) and exposure (X). If a single bias model for jointly modeling X and U is desired use `adjust_multinom_uc_emc_sel`.

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

**Value**

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

**Examples**

```
adjust_uc_emc_sel(
  df_uc_emc_sel,
  exposure = "Xstar",
  outcome = "Y",
  confounders = c("C1", "C2", "C3"),
  u_model_coefs = c(-0.32, 0.59, 0.69),
```

```

x_model_coefs = c(-2.44, 1.62, 0.72, 0.32, -0.15, 0.85),
s_model_coefs = c(0.00, 0.26, 0.78, 0.03, -0.02, 0.10)
)

```

---

adjust\_uc\_omc

*Adjust for uncontrolled confounding and outcome misclassification.*


---

## Description

adjust\_uc\_omc returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding and outcome misclassification.

## Usage

```

adjust_uc_omc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  u_model_coefs,
  y_model_coefs,
  level = 0.95
)

```

## Arguments

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
u_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(U = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y$ , where U is the binary unmeasured confounder, X is the binary true exposure, Y is the binary true outcome. The number of parameters therefore equals 3.
y_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(Y = 1)) = \delta_0 + \delta_1 X + \delta_2 Y^* + \delta_{2+j} C_j$ , where Y represents binary true outcome, X is the binary exposure, Y* is the binary misclassified outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters therefore equals 3 + j.
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

**Details**

This function uses two separate logistic regression models to predict the uncontrolled confounder (U) and outcome (Y). If a single bias model for jointly modeling Y and U is desired use `adjust_multinom_uc_omc`.

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

**Value**

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

**Examples**

```
adjust_uc_omc(
  df_uc_omc,
  "X",
  "Ystar",
  "C1",
  u_model_coefs = c(-0.22, 0.61, 0.70),
  y_model_coefs = c(-2.85, 0.73, 1.60, 0.38)
)
```

---

 adjust\_uc\_sel

*Adust for uncontrolled confounding and selection bias.*


---

**Description**

`adjust_uc_sel` returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding and exposure misclassification.

**Usage**

```
adjust_uc_sel(
  data,
  exposure,
  outcome,
  confounders = NULL,
  u_model_coefs,
  s_model_coefs,
  level = 0.95
)
```

**Arguments**

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders are allowed.
u_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(U = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y + \alpha_{2+j} C_j$ , where U is the binary unmeasured confounder, X is the binary exposure, Y is the binary outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders. The number of parameters therefore equals 3 + j.
s_model_coefs	The regression coefficients corresponding to the model: $\text{logit}(P(S = 1)) = \beta_0 + \beta_1 X + \beta_2 Y$ , where S represents binary selection, X is the binary exposure, and Y is the binary outcome. The number of parameters therefore equals 3.
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

**Details**

Values for the regression coefficients can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

**Value**

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

**Examples**

```
adjust_uc_sel(
  df_uc_sel,
  exposure = "X",
  outcome = "Y",
  confounders = "C1",
  u_model_coefs = c(-0.36, 0.96, 1.11, -0.12),
  s_model_coefs = c(0.02, 0.89, 0.90)
)
```



df\_emc

*Simulated data with exposure misclassification***Description**

Data containing one source of bias, three known confounders, and 100,000 observations. This data is obtained from df\_emc\_source by removing the column X. The resulting data corresponds to what a researcher would see in the real-world: a misclassified exposure, Xstar, and no data on the true exposure. As seen in df\_emc\_source, the true, unbiased exposure-outcome odds ratio = 2.

**Usage**

df\_emc

**Format**

A dataframe with 100,000 rows and 5 columns:

**Xstar** misclassified exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

df\_emc\_omc

*Simulated data with exposure misclassification and outcome misclassification***Description**

Data containing two sources of bias, three known confounders, and 100,000 observations. This data is obtained from df\_emc\_omc\_source by removing the columns X and Y. The resulting data corresponds to what a researcher would see in the real-world: a misclassified exposure, Xstar, and a misclassified outcome, Ystar. As seen in df\_emc\_omc\_source, the true, unbiased exposure-outcome odds ratio = 2.

**Usage**

df\_emc\_omc

**Format**

A dataframe with 100,000 rows and 5 columns:

**Xstar** misclassified exposure, 1 = present and 0 = absent

**Ystar** misclassified outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

---

df_emc_omc_source	<i>Data source for df_emc_omc</i>
-------------------	-----------------------------------

---

**Description**

Data with complete information on the two sources of bias, three known confounders, and 100,000 observations. This data is used to derive df\_emc\_omc and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df\_emc\_omc. With this source data, the fitted regression  $\text{logit}(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3$  shows that the true, unbiased exposure-outcome odds ratio = 2.

**Usage**

df\_emc\_omc\_source

**Format**

A dataframe with 100,000 rows and 7 columns:

**X** true exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

**Xstar** misclassified exposure, 1 = present and 0 = absent

**Ystar** misclassified outcome, 1 = present and 0 = absent

---

df_emc_sel	<i>Simulated data with exposure misclassification and selection bias</i>
------------	--

---

### Description

Data containing two sources of bias, three known confounders, and 100,000 observations. This data is obtained by sampling with replacement with probability = S from df\_emc\_sel\_source then removing the columns X and S. The resulting data corresponds to what a researcher would see in the real-world: a misclassified exposure, Xstar, and missing data for those not selected into the study (S=0). As seen in df\_emc\_sel\_source, the true, unbiased exposure-outcome odds ratio = 2.

### Usage

```
df_emc_sel
```

### Format

A dataframe with 100,000 rows and 5 columns:

**Xstar** misclassified exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

---

df_emc_sel_source	<i>Data source for df_emc_sel</i>
-------------------	-----------------------------------

---

### Description

Data with complete information on the two sources of bias, three known confounders, and 100,000 observations. This data is used to derive df\_emc\_sel and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df\_emc\_sel. With this source data, the fitted regression  $\text{logit}(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3$  shows that the true, unbiased exposure-outcome odds ratio = 2.

### Usage

```
df_emc_sel_source
```

**Format**

A dataframe with 100,000 rows and 7 columns:

**X** true exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

**Xstar** misclassified exposure, 1 = present and 0 = absent

**S** selection, 1 = selected into the study and 0 = not selected into the study

---

df_emc_source	<i>Data source for df_emc</i>
---------------	-------------------------------

---

**Description**

Data with complete information on one sources of bias, three known confounders, and 100,000 observations. This data is used to derive df\_emc and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df\_emc. With this source data, the fitted regression  $\text{logit}(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3$  shows that the true, unbiased exposure-outcome odds ratio = 2.

**Usage**

df\_emc\_source

**Format**

A dataframe with 100,000 rows and 6 columns:

**X** exposure, 1 = present and 0 = absent

**Y** true outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

**Xstar** misclassified exposure, 1 = present and 0 = absent

---

df_omc	<i>Simulated data with outcome misclassification</i>
--------	--

---

**Description**

Data containing one source of bias, three known confounders, and 100,000 observations. This data is obtained from `df_omc_source` by removing the column `Y`. The resulting data corresponds to what a researcher would see in the real-world: a misclassified outcome, `Ystar`, and no data on the true outcome. As seen in `df_omc_source`, the true, unbiased exposure-outcome odds ratio = 2.

**Usage**

`df_omc`

**Format**

A dataframe with 100,000 rows and 5 columns:

**X** exposure, 1 = present and 0 = absent

**Ystar** misclassified outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

---

df_omc_sel	<i>Simulated data with outcome misclassification and selection bias</i>
------------	---

---

**Description**

Data containing two sources of bias, a known confounder, and 100,000 observations. This data is obtained by sampling with replacement with probability = `S` from `df_omc_sel_source` then removing the columns `Y` and `S`. The resulting data corresponds to what a researcher would see in the real-world: a misclassified outcome, `Ystar`, and missing data for those not selected into the study (`S=0`). As seen in `df_omc_sel_source`, the true, unbiased exposure-outcome odds ratio = 2.

**Usage**

`df_omc_sel`

**Format**

A dataframe with 100,000 rows and 3 columns:

**X** exposure, 1 = present and 0 = absent

**Ystar** misclassified outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

---

df_omc_sel_source	<i>Data source for df_omc_sel</i>
-------------------	-----------------------------------

---

### Description

Data with complete information on the two sources of bias, a known confounder, and 100,000 observations. This data is used to derive df\_omc\_sel and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df\_omc\_sel. With this source data, the fitted regression  $\text{logit}(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1$  shows that the true, unbiased exposure-outcome odds ratio = 2.

### Usage

df\_omc\_sel\_source

### Format

A dataframe with 100,000 rows and 5 columns:

**X** exposure, 1 = present and 0 = absent

**Y** true outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**Ystar** misclassified outcome, 1 = present and 0 = absent

**S** selection, 1 = selected into the study and 0 = not selected into the study

---

df_omc_source	<i>Data source for df_omc</i>
---------------	-------------------------------

---

### Description

Data with complete information on one sources of bias, three known confounders, and 100,000 observations. This data is used to derive df\_omc and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df\_omc. With this source data, the fitted regression  $\text{logit}(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3$  shows that the true, unbiased exposure-outcome odds ratio = 2.

### Usage

df\_omc\_source

**Format**

A dataframe with 100,000 rows and 6 columns:

**X** exposure, 1 = present and 0 = absent

**Y** true outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

**Ystar** misclassified outcome, 1 = present and 0 = absent

---

df\_sel

*Simulated data with selection bias*

---

**Description**

Data containing one source of bias, three known confounders, and 100,000 observations. This data is obtained by sampling with replacement with probability = S from df\_sel\_source then removing the S column. The resulting data corresponds to what a researcher would see in the real-world: missing data for those not selected into the study (S=0). As seen in df\_sel\_source, the true, unbiased exposure-outcome odds ratio = 2.

**Usage**

df\_sel

**Format**

A dataframe with 100,000 rows and 5 columns:

**X** exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

---

df_sel_source	<i>Data source for df_sel</i>
---------------	-------------------------------

---

### Description

Data with complete information on study selection, three known confounders, and 100,000 observations. This data is used to derive df\_sel and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df\_sel. With this source data, the fitted regression  $\text{logit}(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3$  shows that the true, unbiased exposure-outcome odds ratio = 2.

### Usage

df\_sel\_source

### Format

A dataframe with 100,000 rows and 6 columns:

**X** true exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

**S** selection, 1 = selected into the study and 0 = not selected into the study

---

df_uc	<i>Simulated data with uncontrolled confounding</i>
-------	---

---

### Description

Data containing one source of bias, three known confounders, and 100,000 observations. This data is obtained from df\_uc\_source by removing the column U. The resulting data corresponds to what a researcher would see in the real-world: information on known confounders (C1, C2, and C3), but nothing for confounder U. As seen in df\_uc\_source, the true, unbiased exposure-outcome odds ratio = 2.

### Usage

df\_uc



**Format**

A dataframe with 100,000 rows and 5 columns:

**X** exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

---

df\_uc\_emc

*Simulated data with uncontrolled confounding and exposure misclassification*

---

**Description**

Data containing two sources of bias, a known confounder, and 100,000 observations. This data is obtained from df\_uc\_emc\_source by removing the columns X and U. The resulting data corresponds to what a researcher would see in the real-world: a misclassified exposure, Xstar, and missing data on a confounder U. As seen in df\_uc\_emc\_source, the true, unbiased exposure-outcome odds ratio = 2.

**Usage**

df\_uc\_emc

**Format**

A dataframe with 100,000 rows and 3 columns:

**Xstar** misclassified exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** confounder, 1 = present and 0 = absent

---

df_uc_emc_sel	<i>Simulated data with uncontrolled confounding, exposure misclassification, and selection bias</i>
---------------	---

---

### Description

Data containing three sources of bias, three known confounders, and 100,000 observations. This data is obtained by sampling with replacement with probability = S from df\_uc\_emc\_sel\_source then removing the columns X, U, and S. The resulting data corresponds to what a researcher would see in the real-world: a misclassified exposure, Xstar; missing data on a confounder U; and missing data for those not selected into the study (S=0). As seen in df\_uc\_emc\_sel\_source, the true, unbiased exposure-outcome odds ratio = 2.

### Usage

```
df_uc_emc_sel
```

### Format

A dataframe with 100,000 rows and 5 columns:

**Xstar** misclassified exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

---

df_uc_emc_sel_source	<i>Data source for df_uc_emc_sel</i>
----------------------	--------------------------------------

---

### Description

Data with complete information on the three sources of bias, three known confounders, and 100,000 observations. This data is used to derive df\_uc\_emc\_sel and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df\_uc\_emc\_sel. With this source data, the fitted regression  $\text{logit}(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3 + \alpha_5 U$  shows that the true, unbiased exposure-outcome odds ratio = 2.

### Usage

```
df_uc_emc_sel_source
```

**Format**

A dataframe with 100,000 rows and 8 columns:

**X** true exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

**U** unmeasured confounder, 1 = present and 0 = absent

**Xstar** misclassified exposure, 1 = present and 0 = absent

**S** selection, 1 = selected into the study and 0 = not selected into the study

---

df_uc_emc_source	<i>Data source for df_uc_emc</i>
------------------	----------------------------------

---

**Description**

Data with complete information on the two sources of bias, a known confounder, and 100,000 observations. This data is used to derive df\_uc\_emc and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df\_uc\_emc. With this source data, the fitted regression  $\text{logit}(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 U$  shows that the true, unbiased exposure-outcome odds ratio = 2.

**Usage**

df\_uc\_emc\_source

**Format**

A dataframe with 100,000 rows and 5 columns:

**X** true exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**U** unmeasured confounder, 1 = present and 0 = absent

**Xstar** misclassified exposure, 1 = present and 0 = absent

---

df_uc_omc	<i>Simulated data with uncontrolled confounding and outcome misclassification</i>
-----------	---

---

### Description

Data containing two sources of bias, a known confounder, and 100,000 observations. This data is obtained from df\_uc\_omc\_source by removing the columns Y and U. The resulting data corresponds to what a researcher would see in the real-world: a misclassified outcome, Ystar, and missing data on a confounder U. As seen in df\_uc\_omc\_source, the true, unbiased exposure-outcome odds ratio = 2.

### Usage

df\_uc\_omc

### Format

A dataframe with 100,000 rows and 3 columns:

**X** exposure, 1 = present and 0 = absent

**Ystar** misclassified outcome, 1 = present and 0 = absent

**C1** confounder, 1 = present and 0 = absent

---

df_uc_omc_source	<i>Data source for df_uc_omc</i>
------------------	----------------------------------

---

### Description

Data with complete information on the two sources of bias, a known confounder, and 100,000 observations. This data is used to derive df\_uc\_omc and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df\_uc\_omc. With this source data, the fitted regression  $\text{logit}(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 U$  shows that the true, unbiased exposure-outcome odds ratio = 2.

### Usage

df\_uc\_omc\_source

**Format**

A dataframe with 100,000 rows and 5 columns:

**X** true exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**U** unmeasured confounder, 1 = present and 0 = absent

**Ystar** misclassified outcome, 1 = present and 0 = absent

---

df\_uc\_sel

*Simulated data with uncontrolled confounding and selection bias*


---

**Description**

Data containing two sources of bias, a known confounder, and 100,000 observations. This data is obtained by sampling with replacement with probability = S from df\_uc\_sel\_source then removing the columns U and S. The resulting data corresponds to what a researcher would see in the real-world: missing data on confounder U; and missing data for those not selected into the study (S=0). As seen in df\_uc\_sel\_source, the true, unbiased exposure-outcome odds ratio = 2.

**Usage**

```
df_uc_sel
```

**Format**

A dataframe with 100,000 rows and 3 columns:

**X** exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

---

df\_uc\_sel\_source

*Data source for df\_uc\_sel*


---

**Description**

Data with complete information on the two sources of bias, a known confounder, and 100,000 observations. This data is used to derive df\_uc\_sel and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df\_uc\_sel. With this source data, the fitted regression  $\text{logit}(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 U$  shows that the true, unbiased exposure-outcome odds ratio = 2.

**Usage**

df\_uc\_sel\_source

**Format**

A dataframe with 100,000 rows and 5 columns:

**X** true exposure, 1 = present and 0 = absent

**Y** outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**U** unmeasured confounder, 1 = present and 0 = absent

**S** selection, 1 = selected into the study and 0 = not selected into the study

---

df_uc_source	<i>Data source for df_uc</i>
--------------	------------------------------

---

**Description**

Data with complete information on one sources of bias, three known confounders, and 100,000 observations. This data is used to derive df\_uc and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df\_uc. With this source data, the fitted regression  $\text{logit}(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3 + \alpha_5 U$  shows that the true, unbiased exposure-outcome odds ratio = 2.

**Usage**

df\_uc\_source

**Format**

A dataframe with 100,000 rows and 6 columns:

**X** exposure, 1 = present and 0 = absent

**Y** true outcome, 1 = present and 0 = absent

**C1** 1st confounder, 1 = present and 0 = absent

**C2** 2nd confounder, 1 = present and 0 = absent

**C3** 3rd confounder, 1 = present and 0 = absent

**U** uncontrolled confounder, 1 = present and 0 = absent

---

evans

*Evans County dataset*

---

**Description**

Data from a cohort study in which white males in Evans County were followed for 7 years, with coronary heart disease as the outcome of interest.

**Usage**

evans

**Format**

A dataframe with 609 rows and 9 columns:

**ID** subject identification

**CHD** outcome variable; 1 = coronary heart disease

**AGE** age (in years)

**CHL** cholesterol, mg/dl

**SMK** 1 = subject has ever smoked

**ECG** 1 = presence of electrocardiogram abnormality

**DBP** diastolic blood pressure, mmHg

**SBP** systolic blood pressure, mmHg

**HPT** 1 = SBP greater than or equal to 160 or DBP greater than or equal to 95

**Source**

<<http://web1.sph.emory.edu/dkleinb/logreg3.htm#data>>

# Index

## \* datasets

df\_emc, 25  
df\_emc\_omc, 25  
df\_emc\_omc\_source, 26  
df\_emc\_sel, 27  
df\_emc\_sel\_source, 27  
df\_emc\_source, 28  
df\_omc, 29  
df\_omc\_sel, 29  
df\_omc\_sel\_source, 30  
df\_omc\_source, 30  
df\_sel, 31  
df\_sel\_source, 32  
df\_uc, 32  
df\_uc\_emc, 33  
df\_uc\_emc\_sel, 34  
df\_uc\_emc\_sel\_source, 34  
df\_uc\_emc\_source, 35  
df\_uc\_omc, 36  
df\_uc\_omc\_source, 36  
df\_uc\_sel, 37  
df\_uc\_sel\_source, 37  
df\_uc\_source, 38  
evans, 39

adjust\_emc, 3  
adjust\_emc\_omc, 4  
adjust\_emc\_sel, 5  
adjust\_multinom\_emc\_omc, 7  
adjust\_multinom\_uc\_emc, 8  
adjust\_multinom\_uc\_emc\_sel, 10  
adjust\_multinom\_uc\_omc, 12  
adjust\_omc, 14  
adjust\_omc\_sel, 15  
adjust\_sel, 16  
adjust\_uc, 18  
adjust\_uc\_emc, 19  
adjust\_uc\_emc\_sel, 20  
adjust\_uc\_omc, 22  
adjust\_uc\_sel, 23

df\_emc, 25  
df\_emc\_omc, 25  
df\_emc\_omc\_source, 26  
df\_emc\_sel, 27  
df\_emc\_sel\_source, 27  
df\_emc\_source, 28  
df\_omc, 29  
df\_omc\_sel, 29  
df\_omc\_sel\_source, 30  
df\_omc\_source, 30  
df\_sel, 31  
df\_sel\_source, 32  
df\_uc, 32  
df\_uc\_emc, 33  
df\_uc\_emc\_sel, 34  
df\_uc\_emc\_sel\_source, 34  
df\_uc\_emc\_source, 35  
df\_uc\_omc, 36  
df\_uc\_omc\_source, 36  
df\_uc\_sel, 37  
df\_uc\_sel\_source, 37  
df\_uc\_source, 38

evans, 39