Package 'RegCalib'

August 7, 2026

Title Regression Calibration for Measurement Error Correction

Version 0.1.0

Description Corrects for measurement error in continuous exposures and covariates using regression calibration methods. Provides corrected coefficients, standard errors, p-values, confidence intervals, and variance-covariance matrices for linear and generalized linear outcome models under an external validation study design. Implements the deattenuation factor method (RegCalibDF) and the substitution approach (RegCalibSub). Supports single and multiple error-prone exposures.

License GPL-3

Encoding UTF-8

Imports stats, dplyr, Matrix, matrixcalc

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

URL https://github.com/JingyuCui639/RegCalib

Config/roxygen2/version 8.0.0

Depends R (>= 3.5)

LazyData true

VignetteBuilder knitr

BugReports https://github.com/JingyuCui639/RegCalib/issues

LazyDataCompression xz

NeedsCompilation no

Author Jingyu Cui [aut, cre] (Affiliation: Department of Biostatistics, Yale School of Public Health, ORCID: <https://orcid.org/0000-0002-8073-1422>),
Wenze Tang [aut] (Affiliation: Vertex Pharmaceuticals),
Molin Wang [aut] (Affiliation: Department of Biostatistics, Department of Epidemiology, Harvard T.H. Chan School of Public Health)

Maintainer Jingyu Cui <jingyu.cui@yale.edu>

Repository CRAN

Date/Publication 2026-08-07 16:30:20 UTC

Contents

`main_data_sim` *Simulated Main-Study Data*

Description

A simulated main-study dataset for demonstrating regression calibration with error-prone dietary exposures.

Usage

```
main_data_sim
```

Format

A data frame with 89,538 observations and the following variables:

id Participant identifier.

fqcal Food-frequency questionnaire estimate of caloric intake.

fqcalinc Transformed caloric intake.

fqtfat Food-frequency questionnaire estimate of total fat intake.

fqtfatinc Transformed total fat intake.

fqalc Food-frequency questionnaire estimate of alcohol intake.

fqalcinc Transformed alcohol intake.

age Participant age.

agec Categorized participant age.

case Binary outcome indicator.

Source

Simulated data created for the RegCalib package.

RegCalibDF *Regression Calibration Using the Deattenuation Factor Method*

Description

Corrects for measurement error in continuous exposures (and covariates) and returns corrected coefficients, standard errors, p-values, and variance-covariance matrices. Users may supply their own uncorrected estimates (e.g., from logistic regression model) instead of using the built-in outcome model. Only externalvalidation study design is supported. Based on Rosner, Spiegelman & Willett (1989, 1990) and Spiegelman, McDermott & Rosner (1997).

Usage

```
RegCalibDF(
  supplyEstimates = FALSE,
  ms,
  vs,
  sur,
  exp,
  covCalib = NULL,
  covOutcomePlus = NULL,
  outcome = NA,
  method = "lm",
  family = NA,
  link = NA,
  external = TRUE,
  pointEstimates = NA,
  vcovEstimates = NA
)
```

Arguments

supplyEstimates	Logical. If TRUE, uncorrected estimates are supplied by the user via pointEstimates and vcovEstimates, and ms is optional. Standard regression results are not returned in this case. Default FALSE.
ms	Main study data frame. Must contain all variables specified in sur, covCalib, and covOutcomePlus. Required when supplyEstimates = FALSE.
vs	Internal or external validation study data frame. Must contain all variables specified in exp, sur, and covCalib.
sur	Character vector of mismeasured exposure(s)/covariate(s) (surrogates) in the main study dataset.
exp	Character vector of correctly-measured exposure(s)/covariate(s) in the validation dataset. Must correspond one-to-one with sur and have the same length.
covCalib	Character vector of correctly-measured covariates to adjust for in both the calibration model and the outcome model. Default NULL.

`covOutcomePlus` Character vector of correctly-measured risk factors for the outcome that are *not* associated with the exposure or surrogate. These are included in the outcome model only and must not overlap with `covCalib`. Default `NULL`.

`outcome` Character. Name of the outcome variable. Required when `method` is `"lm"` or `"glm"`.

`method` Character. Outcome modelling method: `"lm"`, `"glm"`. Default `"lm"`.

`family` Family function for `glm` (e.g., `binomial`). Not a character string. Required when `method = "glm"`.

`link` Character. Link function for `glm` (e.g., `"logit"`). Required when `method = "glm"`.

`external` Logical. `TRUE` (default) for an external validation study; `FALSE` for an internal validation study. When `external = FALSE`, `vs` must contain the outcome variable and `supplyEstimates` must be `FALSE`.

`pointEstimates` Named numeric vector of uncorrected point estimates from standard regression (intercept excluded). Names must match the (expanded) covariate names from `covCalib` followed by `covOutcomePlus`. Required when `supplyEstimates = TRUE`.

`vcovEstimates` Named square matrix of uncorrected variance-covariance estimates (intercept excluded). Column names must match those of `pointEstimates`. Required when `supplyEstimates = TRUE`.

Value

A named list containing:

correctedCoefTable Data frame of corrected estimates, standard errors, Z-values, p-values, and 95% confidence intervals.

correctedVCOV Variance-covariance matrix of the corrected estimates.

standardCoefTable (When `supplyEstimates = FALSE`) Results from the uncorrected standard regression.

standardVCOV (When `supplyEstimates = FALSE`) Variance- covariance matrix from the uncorrected standard regression.

calibrationModelCoefTable Calibration model slope estimates.

calibrationModelVCOV Residual variance-covariance matrix from the calibration model.

References

Rosner B, Willett WC, Spiegelman D (1989). Correction of logistic relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine* 8:1051–1069.

Rosner B, Spiegelman D, Willett WC (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American Journal of Epidemiology* 132:734–745.

Spiegelman D, McDermott A, Rosner B (1997). The many uses of the 'regression calibration' method for measurement error bias correction in nutritional epidemiology. *American Journal of Clinical Nutrition* 65:1179S–1186S.

Spiegelman D, Carroll RJ, Kipnis V (2001). Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument. *Statistics in Medicine* 20:139–160.

Examples

```
data("main_data_sim", package = "RegCalib")
data("valid_data_sim", package = "RegCalib")

set.seed(123)
selected_rows <- sample(
  seq_len(nrow(main_data_sim)),
  1000
)

main_example <- main_data_sim[
  selected_rows,
  ,
  drop = FALSE
]

result <- RegCalibDF(
  ms = main_example,
  vs = valid_data_sim,
  sur = c("fqtfatinc", "fqcalinc", "fqalcinc"),
  exp = c("drtfatinc", "drcalinc", "dralcinc"),
  covCalib = "agec",
  outcome = "case",
  method = "glm",
  family = binomial,
  link = "logit",
  external = TRUE
)

result$correctedCoefTable
```

RegCalibSub *Regression Calibration Using Substitution Method*

Description

Corrects for measurement error in continuous exposures (and covariates) and returns corrected coefficients, standard errors, p-values, and variance-covariance matrices using the Carroll-Ruppert-Stefanski-Crainiceanu (CRS) sandwich variance estimator. Supports linear and generalized linear outcome models under external validation study design. Standard errors are derived analytically via the sandwich estimator (not bootstrap). Non-linear terms such as interaction terms should be pre-computed as permanent columns in the input datasets rather than specified in the formula.

Usage

```
RegCalibSub(
  ms,
  vs,
  sur,
  exp,
  vsIndicator,
  covCalib = NULL,
  covOutcome = NULL,
  outcome = NA,
  method = "lm",
  family = NA,
  link = NA,
  external = TRUE
)
```

Arguments

ms	Main study data frame. Must contain all variables specified in sur, outcome, and covOutcome (if any).
vs	External validation study data frame. Must contain all variables specified in exp, sur, and covCalib (if any). Required when external = TRUE.
sur	Character vector of mismeasured exposure(s)/covariate(s) (surrogates) in the main study dataset.
exp	Character vector of correctly-measured exposure(s)/covariate(s) in the validation dataset. Must correspond one-to-one with sur and have the same length.
vsIndicator	Character. Name of the indicator variable in the main study dataset that identifies subjects with a validation record (1 = has validation record). Required when external = FALSE.
covCalib	Character vector of correctly-measured covariates to adjust for in the calibration model, including any non-linear terms. Default NULL.
covOutcome	Character vector of correctly-measured covariates to adjust for in the outcome model (with corrected exposure), including any non-linear terms. Default NULL.
outcome	Character. Name of the outcome variable. Required.
method	Character. Outcome modelling method: "lm" or "glm". Default "lm".
family	Family function for glm (e.g., binomial). Not a character string. Required when method = "glm".
link	Character. Link function for glm (e.g., "logit" or "log"). Required when method = "glm".
external	Logical. TRUE (default) for an external validation study; FALSE for an internal validation study embedded in ms.

Value

A named list containing:

correctedCoefTable Matrix of corrected estimates, standard errors, Z-values, p-values, and 95% confidence intervals.

correctedVCOV Variance-covariance matrix of the corrected estimates.

References

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006). *Measurement Error in Nonlinear Models*, 2nd ed. Chapman & Hall/CRC.

Examples

```
data("main_data_sim", package = "RegCalib")
data("valid_data_sim", package = "RegCalib")

set.seed(123)
selected_rows <- sample(
  seq_len(nrow(main_data_sim)),
  1000
)

main_example <- main_data_sim[
  selected_rows,
  ,
  drop = FALSE
]

result <- RegCalibSub(
  ms = main_example,
  vs = valid_data_sim,
  sur = c("fqtfatinc", "fqcalinc", "fqalcinc"),
  exp = c("drtfatinc", "drcalinc", "dralcinc"),
  covCalib = "agec",
  covOutcome = "agec",
  outcome = "case",
  method = "glm",
  family = binomial,
  link = "logit",
  external = TRUE
)

result$correctedCoefTable
result$correctedVCOV
```

---

valid_data_sim *Simulated External Validation Data*

Description

A simulated external validation dataset containing surrogate and reference measurements of dietary exposures.

Usage

```
valid_data_sim
```

Format

A data frame with 173 observations and variables containing food-frequency questionnaire measurements, reference measurements, and age.

id Participant identifier.

fqcal Surrogate caloric intake measurement.

fqcalinc Transformed surrogate caloric intake.

fqtfat Surrogate total-fat intake measurement.

fqtfatinc Transformed surrogate total-fat intake.

fqalc Surrogate alcohol intake measurement.

fqalcinc Transformed surrogate alcohol intake.

drcal Reference caloric intake measurement.

drcalinc Transformed reference caloric intake.

drtfat Reference total-fat intake measurement.

drtfatinc Transformed reference total-fat intake.

dralc Reference alcohol intake measurement.

dralcinc Transformed reference alcohol intake.

age Participant age.

agec Categorized participant age.

Source

Simulated data created for the RegCalib package.

Index