Package: RefBasedMI (via r-universe)

October 21, 2026

Title Reference-Based Imputation for Longitudinal Clinical Trials with Protocol Deviation

Version 0.2.0

Description Imputation of missing numerical outcomes for a longitudinal trial with protocol deviations. The package uses distinct treatment arm-based assumptions for the unobserved data, following the general algorithm of Carpenter, Roger, and Kenward (2013) <doi:10.1080/10543406.2013.834911>, and the causal model of White, Royes and Best (2020) <doi:10.1080/10543406.2019.1684308>. Sensitivity analyses to departures from these assumptions can be done by the Delta method of Roger. The program uses the same algorithm as the 'mimix' 'Stata' package written by Suzie Cro, with additional coding for the causal model and delta method. The reference-based methods are jump to reference (J2R), copy increments in reference (CIR), copy reference (CR), and the causal model, all of which must specify the reference treatment arm. Other methods are missing at random (MAR) and the last mean carried forward (LMCF). Individual-specific imputation methods (and their reference groups) can be specified.

URL https://github.com/UCL/RefBasedMI

License GPL-3

Encoding UTF-8

LazyData true

Imports data.table, Hmisc, mice, pastecs, assertthat

Depends R (>= 3.5.0)

NeedsCompilation yes

RoxygenNote 7.2.3

Author Kevin McGrath [aut], Matteo Quartagno [cre]

Maintainer Matteo Quartagno <m.quartagno@ucl.ac.uk>

Repository https://cran.r-universe.dev

Date/Publication 2024-09-10 02:55:08 UTC

RemoteUrl https://github.com/cran/RefBasedMI

RemoteRef HEAD

RemoteSha bcee397a450f5b252d1c17d2f09d50a5c6364cbf

Contents

acupuncture 2
antidepressant 3
asthma 3
RefBasedMI 4

Index **7**

`acupuncture` *Sample data: acupuncture trial*

Description

A data set containing results of a randomised, double-blind, parallel-group comparing active treatment with placebo The primary outcome is head, measured at time 3 and 12

Usage

```
acupuncture
```

Format

A data frame with 802 rows and 11 columns

id

time

age

sex

migraine

chronicity

practice_id

treat

head_base covariate

head outcome variable

withdrawal_reason

antidepressant *Sample data: antidepressant trial*

Description

A data set containing antidepressant trial data as described in paper by White,Royes,Best (2019) The primary outcome is HAMD17.TOTAL measured at visit number 4,5,6,7.

Usage

```
antidepressant
```

Format

dataframe containing 688 rows and 14 columns

PATIENT.NUMBER

HAMA.TOTAL

PGI_IMPROVEMENT

VISIT...VISIT.3.DATE

VISIT.NUMBER

TREATMENT.NAME

PATIENT.SEX

POOLED.INVESTIGATOR

basval

HAMD17.TOTAL outcome variable

change

miss_flag

methodcol individual-specific method

referencecol individual-specific reference arm

asthma *Sample data: asthma trial*

Description

A data set containing asthma trial data as used in the Stata mimix help file The primary outcome variable is fev, measured at 2,4,8,12 weeks

Usage

```
asthma
```

Format

A data frame containing 732 rows and 5 columns

id patient identifier

time

treat

base covariate

fev outcome variable

`RefBasedMI` *Reference-based multiple imputation of longitudinal data*

Description

Performs reference-based multiple imputation of longitudinal data where data are missing after treatment discontinuation. Methods available are missing at random, jump to reference, copy reference, copy increments in reference, last mean carried forward, the causal model, and delta-adjustment.

Usage

```
RefBasedMI(
  data,
  covar = NULL,
  depvar,
  treatvar,
  idvar,
  timevar,
  method = NULL,
  reference = NULL,
  methodvar = NULL,
  referencevar = NULL,
  K0 = NULL,
  K1 = NULL,
  delta = NULL,
  dlag = NULL,
  M = 1,
  seed = 101,
  prior = "jeffreys",
  burnin = 1000,
  bbetween = NULL,
  mle = FALSE
)
```

Arguments

`data`	Dataset in long format
`covar`	Baseline covariate(s): must be complete (no missing values)
`depvar`	Outcome variable
`treatvar`	Treatment group variable: can be numeric or character
`idvar`	Participant identifiervariable
`timevar`	Variable indicating time point for repeated measures
`method`	Reference-based imputation method: must be "J2R", "CR", "CIR", "MAR", "Causal" or "LMCF"
`reference`	Reference group for "J2R", "CIR", "CR" methods, or control group for causal method: can be numeric or string
`methodvar`	Variable in dataset specifying individual method
`referencevar`	Variable in dataset specifying reference group for individual method
`K0`	Causal constant for use with Causal method
`K1`	Exponential decaying causal constant for use with Causal method
`delta`	Optional vector of delta values to add onto imputed values (non-mandatory) (a's in Five_Macros user guide), length equal to number of time points
`dlag`	Optional vector of delta values to add onto imputed values (non-mandatory) (b's in Five_Macros user guide), length equal to number of time points
`M`	Number of imputations to be created
`seed`	Seed value: specify this so that a new run of the command will give the same imputed values
`prior`	Prior when fitting multivariate normal distributions: can be one of "jeffreys" (default), "uniform" or "ridge"
`burnin`	Number of burn-in iterations when fitting multivariate normal distributions
`bbetween`	Number of iterations between imputed data sets when fitting multivariate normal distributions
`mle`	Use with extreme caution: do improper imputation by drawing from the model using the maximum likelihood estimates. This does not allow for uncertainty in the MLEs and invalidates interval estimates from Rubin's rules.

Details

The program works through the following steps:

1. Set up a summary table based on treatment arm and missing data pattern (i.e. which timepoints are unobserved)
2. Fit a multivariate normal distribution to each treatment arm using MCMC methods in package norm2
3. Impute all interim missing values under a MAR assumption, looping over treatments and patterns

4. Impute all post-discontinuation missing values under the user-specified assumption, looping over treatments and patterns (and over methodvar and referencevar if specified)
5. Perform delta-adjustment if specified
6. Repeat steps 2-5 M times and form into a single data frame

The baseline value of the outcome could be handed as an outcome, but this would allow a treatment effect at baseline. We instead recommend handling it as a covariate.

The program is based on Suzie Cro's Stata program mimix

The user can use the as.mids() function in the mice package to convert the output data to mids data type and then perform analysis using Rubin's rules.

Value

A data frame containing the original data stacked above the M imputed data sets. The original ID variable (idvar) is renamed as .id. A new variable .imp indicates the original data (.imp=0) or the imputed data sets (.imp=1,...,M).

Examples

```
# Perform jump to reference imputation on asthma trial data, with reference arm 1
asthmaJ2R <- RefBasedMI(data=asthma, depvar=fev, treatvar=treat,
 idvar=id, timevar=time, method="J2R", reference=1, M=5, seed=54321)
# Fit regression model to each imputed data set by treating output data frame as mids object
library(mice)
fit <- with(data = as.mids(asthmaJ2R), lm(fev ~ factor(treat), subset=(time==12)))
# Find pooled treatment effects using Rubin's rules
summary(pool(fit))
```

Index

∗ **datasets**
acupuncture, 2
antidepressant, 3
asthma, 3

acupuncture, 2
antidepressant, 3
asthma, 3

RefBasedMI, 4