Package 'RCTRecruit'

May 7, 2026

Type Package

Title Non-Parametric Recruitment Prediction for Randomized Clinical Trials

Version 0.2.0

Description Accurate prediction of subject recruitment for Randomized Clinical Trials (RCT) remains an ongoing challenge. Many previous prediction models rely on parametric assumptions. We present functions for non-parametric RCT recruitment prediction under several scenarios.

License MIT + file LICENSE

Depends R (>= 4.2.0)

Imports lubridate, methods, Rcpp

LinkingTo Rcpp

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

URL https://github.com/imalagaris/RCTRecruit

BugReports https://github.com/imalagaris/RCTRecruit/issues

Suggests knitr, magrittr, testthat (>= 3.0.0), withr

Config/testthat/edition 3

NeedsCompilation yes

Author Ioannis Malagaris [aut, cre, cph] (ORCID: <https://orcid.org/0000-0001-5126-2068>), Alejandro Villasante-Tezanos [aut], Christopher Kurinec [aut], Xiaoying Yu [aut]

Maintainer Ioannis Malagaris <iomalaga@utmb.edu>

Repository CRAN

Date/Publication 2025-04-21 18:30:02 UTC

Contents

`GetDistance` *Euclidean distance between predicted and actual recruitment*

Description

Euclidean distance between predicted and actual recruitment

Usage

```
GetDistance(
  target,
  nSim = 10000L,
  fillGaps = FALSE,
  cauchyWt = FALSE,
  efficiencyFactor = 1
)
```

Arguments

`target`	A vector with the actual recruitment by week
`nSim`	Number of simulations to run (default = 1e4L). Accepted values are in the range of 1 to 10,000.
`fillGaps`	Whether to fill recruitment gaps in the data (default = FALSE). Recruitment gaps are defined as any full week (Monday through Sunday) with no dates recorded in the loaded data. If at least one date is present within a given week, that week will not be considered a gap in recruitment.
`cauchyWt`	Whether to use Cauchy weights for sampling. If FALSE (default), binomial weights are used.
`efficiencyFactor`	An efficiency coefficient to apply to the recruitment rate (default = 1). If the efficiency of the recruitment process is expected to match the provided data, this value should be set to 1. If the recruitment process is expected to be slower, this value should less than 1. Finally, if the recruitment process is expected to proceed faster, this value should be greater than 1. Accepted values range from 0.1 to 2:

- 0.1: Indicates that the recruitment rate is expected to be 10% of the original rate.
- 2.0: Indicates that the recruitment rate is expected to be double the original rate.

Value

An object of `RCTDist` class with four elements.

1. `dist`: A numeric vector with length equal to `nSim` containing the simulated Euclidean distance.
2. `CI`: A numeric vector with the median and the 95% CI Euclidean distance.
3. `call.`: The call (deparsed) that created the object
4. `cargs`: A list with the arguments of the call that created the object including the default arguments

See Also

Other Links: `GetWeekPredCI()`, `LoadData()`, `RCTRecruit-package`, `Time2Nsubjects()`, `gripsYR1`, `gripsYR2`, `gripsYR2Weekly`, `plot.RCTPredCI()`

Examples

```
LoadData(gripsYR1, ScreenDt, Enrolled)
(res <- GetDistance(gripsYR2Weekly$enrolled))
str(res)
```

`GetWeekPredCI` *Calculate median recruitment with 95% CI for the next 104 weeks (two years)*

Description

Calculate median recruitment with 95% CI for the next 104 weeks (two years)

Usage

```
GetWeekPredCI(
  nSim = 10000L,
  fillGaps = FALSE,
  cauchyWt = FALSE,
  efficiencyFactor = 1
)
```

Arguments

`nSim`	Number of simulations to run (default = 1e4L). Accepted values are in the range of 1 to 10,000.
`fillGaps`	Whether to fill recruitment gaps in the data (default = FALSE). Recruitment gaps are defined as any full week (Monday through Sunday) with no dates recorded in the loaded data. If at least one date is present within a given week, that week will not be considered a gap in recruitment.
`cauchyWt`	Whether to use Cauchy weights for sampling. If FALSE (default), binomial weights are used.
`efficiencyFactor`	An efficiency coefficient to apply to the recruitment rate (default = 1). If the efficiency of the recruitment process is expected to match the provided data, this value should be set to 1. If the recruitment process is expected to be slower, this value should less than 1. Finally, if the recruitment process is expected to proceed faster, this value should be greater than 1. Accepted values range from 0.1 to 2: • 0.1: Indicates that the recruitment rate is expected to be 10% of the original rate. • 2.0: Indicates that the recruitment rate is expected to be double the original rate.

Value

An object of `RCTPredCI` class with 5 elements.

1. `predCI`: An 104x3 matrix with the 2.5%, 50% and 97.5% weekly percentiles
2. `pargs`: An environment which contains objects and functions used to construct the plot with `base::plot()`. For internal use only.
3. `call.`: The call (deparsed) that created the object
4. `cargs`: A list with the arguments of the call that created the object including the default arguments

See Also

Other Links: `GetDistance()`, `LoadData()`, `RCTRecruit-package`, `Time2Nsubjects()`, `gripsYR1`, `gripsYR2`, `gripsYR2Weekly`, `plot.RCTPredCI()`

Examples

```
LoadData(gripsYR1, ScreenDt, Enrolled)
(res <- GetWeekPredCI(fillGaps = TRUE, efficiencyFactor = 1.5))
scenarios <- list(
  sc1 = GetWeekPredCI(),
  sc2 = GetWeekPredCI(cauchyWt = TRUE),
  sc3 = GetWeekPredCI(fillGaps = TRUE),
  sc4 = GetWeekPredCI(fillGaps = TRUE, efficiencyFactor = 1.2)
)
```

```
maxY <- sapply(scenarios, \(x) x$pargs$maxY) |> max()
defaultGraphicParams <- par(no.readonly = TRUE)
graphics::par(mfrow = c(2, 2), cex.main = 1)
for (x in scenarios) plot(x, yMax = maxY, main = x$call.)
do.call(par, defaultGraphicParams)
```

`gripsYR1` *Daily recruitment data for the 1st year of the GRIPS study*

Description

Modified recruitment data from the first year of the GRIPS study. This dataset includes the number or participants recruited for each calendar date during active recruitment periods.

Usage

```
gripsYR1
```

Format

A data frame with **159** observations of **2** variables

[,1]	ScreenDt	`character`	Calendar date of the screening process for recruitment in the study
[,2]	Enrolled	`integer`	Number of new subjects enrolled in the study on that date

Source

Villasante-Tezanos A, Kuo Y, Kurinec C, Li Y, Yu X (2024). "A non-parametric approach to predict the recruitment for randomized clinical trials: an example in elderly inpatient settings." *BMC medical research methodology*, *24*, 189. ISSN 1471-2288, https://pubmed.ncbi.nlm.nih.gov/39210285/.

See Also

Other Links: `GetDistance()`, `GetWeekPredCI()`, `LoadData()`, `RCTRecruit-package`, `Time2Nsubjects()`, `gripsYR2`, `gripsYR2Weekly`, `plot.RCTPredCI()`

gripsYR2 *Daily recruitment data for the 2nd year of the GRIPS study*

Description

Modified recruitment data from the second year of the GRIPS study. This dataset includes the number or participants recruited for each calendar date during active recruitment periods.

Usage

```
gripsYR2
```

Format

A data frame with **292** observations of **2** variables

[,1]	ScreenDt	`character`	Calendar date of the screening process for recruitment in the study
[,2]	Enrolled	`integer`	Number of new subjects enrolled in the study on that date

Source

Villasante-Tezanos A, Kuo Y, Kurinec C, Li Y, Yu X (2024). "A non-parametric approach to predict the recruitment for randomized clinical trials: an example in elderly inpatient settings." *BMC medical research methodology*, *24*, 189. ISSN 1471-2288, https://pubmed.ncbi.nlm.nih.gov/39210285/.

See Also

Other Links: `GetDistance()`, `GetWeekPredCI()`, `LoadData()`, `RCTRecruit-package`, `Time2Nsubjects()`, `gripsYR1`, `gripsYR2Weekly`, `plot.RCTPredCI()`

gripsYR2Weekly *Weekly recruitment data for the 2nd year of the GRIPS study*

Description

Modified recruitment data from the second year of the GRIPS study, aggregated by calendar week.

Usage

```
gripsYR2Weekly
```

Format

A data frame with **52** observations of **4** variables

[,1]	week	`double`	Calendar week
[,2]	year	`double`	Calendar year
[,3]	enrolled	`integer`	Number of people enrolled that week
[,4]	activeDays	`integer`	Number of days in that week when recruitment was active

Source

Villasante-Tezanos A, Kuo Y, Kurinec C, Li Y, Yu X (2024). "A non-parametric approach to predict the recruitment for randomized clinical trials: an example in elderly inpatient settings." *BMC medical research methodology*, *24*, 189. ISSN 1471-2288, https://pubmed.ncbi.nlm.nih.gov/39210285/.

See Also

Other Links: `GetDistance()`, `GetWeekPredCI()`, `LoadData()`, `RCTRecruit-package`, `Time2Nsubjects()`, `gripsYR1`, `gripsYR2`, `plot.RCTPredCI()`

`LoadData` *Load recruitment data.*

Description

This function must be called before any other function in this package. LoadData checks the input data and stores the results internally for the session. Calling this function more than once in the same session will overwrite the previously created internal data.

Usage

```
LoadData(data, date, enrolled)
```

Arguments

`data`	Main dataset containing at least two columns: • A `date` column with the calendar date of the screening • A `enrolled` column with the number of subjects enrolled If the entries cover a period longer than 1 year, only the entries within one year prior to the latest date will be retained.
`date`	The name (symbol or string) of the column in the dataset with the calendar dates of active screening. All active calendar dates should be included, even if the recruitment for that date is 0. Only dates with truly non-active recruitment should be omitted. The date column must be:

- an object inheriting from class the `Date` class
- or a character vector with a valid date format.

`enrolled` The name (symbol or string) of the column in the dataset with the number of subjects recruited on the corresponding calendar date. It must be a numeric vector.

Value

This function does not return any value. It runs several tests and proccesses the data and stores internally the results. It prints a message to the console if the data is successfully loaded or an error message if there is an issue with the input data. Once the dataset is loaded, the following functions can be used:

- `Time2Nsubjects()`: simulates the number of weeks needed to recruit a given number of subjects
- `GetDistance()`: calculates the Euclidean distance between the predicted and actual recruitment
- `GetWeekPredCI()`: calculates the median recruitment with 95% CI for up to the next 104 weeks (two years)

See Also

Other Links: `GetDistance()`, `GetWeekPredCI()`, `RCTRecruit-package`, `Time2Nsubjects()`, `gripsYR1`, `gripsYR2`, `gripsYR2Weekly`, `plot.RCTPredCI()`

Examples

```
# Load using names as symbols
LoadData(gripsYR1, ScreenDt, Enrolled)

# Load using names as strings
LoadData(gripsYR1, "ScreenDt", "Enrolled")

# Load using base pipe operator
gripsYR1 |> LoadData(ScreenDt, Enrolled)

# Load using magrittr pipe operator
if (base::requireNamespace("magrittr", quietly = TRUE)) {
  library(magrittr)
  gripsYR1 %>% LoadData(ScreenDt, Enrolled)
}
```

---

plot.RCTPredCI *Plots RCTPredCI object*

Description

Plots RCTPredCI object

Usage

```
## S3 method for class 'RCTPredCI'
plot(x, yMax = NULL, includeYR2 = FALSE, xGrid = TRUE, yGrid = TRUE, ...)
```

Arguments

x	RCTPredCI object
yMax	It sets upper value for the y-axis. It is useful when several figures are plotted together
includeYR2	Whether to plot predictions for 104 weeks. By default (default = FALSE) it only plots the first 52 weeks.
xGrid	Whether to plot vertical grid lines (default = TRUE)
yGrid	Whether to plot horizontal grid lines (default = TRUE)
...	Additional arguments to be passed as in base::plot() function

See Also

Other Links: GetDistance(), GetWeekPredCI(), LoadData(), RCTRecruit-package, Time2Nsubjects(), gripsYR1, gripsYR2, gripsYR2Weekly

---

Time2Nsubjects *Simulate the number of weeks needed to recruit a given number of subjects*

Description

Simulate the number of weeks needed to recruit a given number of subjects

Usage

```
Time2Nsubjects(
  nSub = 50L,
  nSim = 10000L,
  fillGaps = FALSE,
  cauchyWt = FALSE,
  efficiencyFactor = 1
)
```

Arguments

nSub	Number of subjects to recruit (default = 50L)
nSim	Number of simulations to run (default = 1e4L). Accepted values are in the range of 1 to 10,000.

`fillGaps` Whether to fill recruitment gaps in the data (default = FALSE). Recruitment gaps are defined as any full week (Monday through Sunday) with no dates recorded in the loaded data. If at least one date is present within a given week, that week will not be considered a gap in recruitment.

`cauchyWt` Whether to use Cauchy weights for sampling. If FALSE (default), binomial weights are used.

`efficiencyFactor`
An efficiency coefficient to apply to the recruitment rate (default = 1). If the efficiency of the recruitment process is expected to match the provided data, this value should be set to 1. If the recruitment process is expected to be slower, this value should less than 1. Finally, if the recruitment process is expected to proceed faster, this value should be greater than 1. Accepted values range from 0.1 to 2:

- 0.1: Indicates that the recruitment rate is expected to be 10% of the original rate.
- 2.0: Indicates that the recruitment rate is expected to be double the original rate.

Value

An object of `RCTNWeeks` class with four elements.

1. `weeks` is an integer vector with length equal to `nSim` containing the simulation results.
2. `CI` shows the median and the 95% CI.
3. `call.`: The call (deparsed) that created the object
4. `cargs`: A list with the arguments of the call that created the object including the default arguments

See Also

Other Links: `GetDistance()`, `GetWeekPredCI()`, `LoadData()`, `RCTRecruit-package`, `gripsYR1`, `gripsYR2`, `gripsYR2Weekly`, `plot.RCTPredCI()`

Examples

```
LoadData(gripsYR1, ScreenDt, Enrolled)
(res <- Time2Nsubjects())
str(res)
```

Index