Package 'BioPET'

July 21, 2025

Title Biomarker Prognostic Enrichment Tool

Version 0.2.2

Date 2018-07-17

Description

Prognostic Enrichment is a clinical trial strategy of evaluating an intervention in a patient population with a higher rate of the unwanted event than the broader patient population (R. Temple (2010) <DOI:10.1038/clpt.2010.233>). A higher event rate translates to a lower sample size for the clinical trial, which can have both practical and ethical advantages. This package is a tool to help evaluate biomarkers for prognostic enrichment of clinical trials.

Imports ggplot2, gridExtra, pROC, VGAM

Depends R (>= 3.4.0)

License GPL (>= 2)

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Author Jeremy Roth [cre, aut],
Kathleen F. Kerr [aut],
Kehao Zhu [aut]

Maintainer Jeremy Roth <jhroth@uw.edu>

Repository CRAN

Date/Publication 2018-07-17 19:10:06 UTC

Contents

1

---

dcaData   *Simulated dataset for package 'DecisionCurve'*

Description

Simulated cohort data containing demographic variables, marker values and cancer outcome.

Usage

```
dcaData
```

Format

A data frame with 500 rows and 6 variables:

- Age: Age in years.
- Female: Indicator for female gender.
- Smokes: Indicator for smoking status.
- Marker1: simulated biomarker.
- Marker2: simulated biomarker.
- Cancer: Indicator for cancer.

---

enrichment_analysis   *Prognostic Enrichment with Real Data*

Description

Evaluating biomarkers for prognostic enrichment of clinical trials using real data

Usage

```
enrichment_analysis(formula, data, family = binomial(link = logit),
  reduction.under.treatment, cost.screening = NULL, cost.keeping = NULL,
  do.bootstrap = TRUE, n.bootstrap = 1000, smooth.roc = FALSE,
  power = 0.9, alpha = 0.025, alternative = c("one.sided", "two.sided"),
  selected.biomarker.quantiles = seq(from = 0, to = 0.95, by = 0.05))
```

Arguments

formula	Object of class "formula", in the form "outcome ~ predictors", where the outcome is a binary indicator with a value of 1 in cases and a value of 0 in controls.
data	Data frame containing the outcome and predictors specified in the "formula" argument. Observations with a missing value of the outcome or of any predictor are dropped.
family	Character object or call to the family() function specifying the link function that is passed to 'glm' to estimate a risk score when more than one predictor is specified. Defaults to binomial(link = "logit"), which yields logistic regression.
reduction.under.treatment	
	Number between 0 and 1 indicating the percent reduction in event rate under treatment that the trial should be able to detect with the specified power
cost.screening	Number indicating the cost of screening a patient to determine trial eligibility, This argument is optional; if both the "cost.screening" and "cost.keeping" arguments are specified, then the total cost of the trial based on each screening threshold is estimated and returned.
cost.keeping	Number indicating the cost of retaining a patient in the trial after enrolling. This argument is optional; if both the "cost.screening" and "cost.keeping" arguments are specified, then the total cost of the trial based on each screening threshold is estimated and returned.
do.bootstrap	Logical indicating whether bootstrap 95% confidence intervals should be computed. Defaults to TRUE.
n.bootstrap	Number of bootstrap samples to draw, if "do.bootstrap" is set to TRUE. Defaults to 1000.
smooth.roc	Logical indicating the "smooth" argument passed to the roc() function from the 'pROC' package when a single biomarker is given. Defaults to FALSE.
power	Number between 0 and 1 giving the power the trial should have to reject the null hypothesis that there is no treatment effect. Defaults to 0.9.
alpha	Number between 0 and 1 giving the type I error rate for testing the null hypothesis that there is no treatment effect. Defaults to 0.025.
alternative	Character specifying whether the alternative hypothesis is one-sided with a higher event rate in the treatment group ("one.sided") or two-sided ("two.sided"). Defaults to "one.sided".
selected.biomarker.quantiles	
	Numeric vector specifying the quantiles of the biomarker measured in controls that will be used to screen trial participants. Defaults to 0, 0.05, ..., 0.95. All entries must be between at least 0 and less than 1.

Value

A list with components

- estimates: A data frame with the following summary measures for each biomarker threshold that is used to screen trial participants: 'selected.biomarker.quantiles': quantiles of observed biomarker values used for screening. 'biomarker.screening.thresholds': the values of

the biomarker corresponding to the quantiles, 'event.rate': post-screening event rate, 'NNS': The estimated number of patients needed to screen to identify one patient eligible for the trial, 'SS': The sample size in a clinical trial enrolling only patients whose biomarker-based disease risk is above the level used for screening, 'N.screen': The total number of individuals whose biomarker values are screened to determine whether they should be enrolled in the trial, 'N.screen.increase.percentage': Percentage in N.screen relative to a trail that does not based on the biomarker. 'total.cost': The estimated total cost of running the trial if the biomarker were used for prognostic enrichment (if cost.screening and cost.keeping are specified), 'cost.reduction.percentage': The reduction in total cost relative to a trial that does not screen based on the biomarker.

- estimates.min.total.cost: The row of the estimates data frame corresponding the screening strategy that results in the lowest total cost.

- bootstrap.CIs: 95% bootstrap-based CIs for reported summary measures (if do.bootstrap=TRUE).

- simulation: A logical indicating whether data were simulated.

- bootstrap.CIs: 95% bootstrap-based CIs for reported summary measures (if do.bootstrap=TRUE).

- biomarker: Biomarker from the given dataset, either the single biomarker specified or the predicted values from logistic regression if multiple biomarkers are specified).

- response: Response variable specified in the dataset.

See Also

[enrichment_simulation](), [plot_enrichment_summaries]()

Examples

```
data(dcaData)

## using a single biomarker in the dataset
analysis.single.marker <- enrichment_analysis(Cancer ~ Marker1,
data=dcaData,
reduction.under.treatment=0.3,
cost.screening=100,
cost.keeping=1000)
head(analysis.single.marker$estimates)
head(analysis.single.marker$bootstrap.CIs)

## combining two biomarkers in the dataset
analysis.two.markers <- enrichment_analysis(Cancer ~ Marker1 + Marker2,
data=dcaData,
reduction.under.treatment=0.3,
cost.screening=100,
cost.keeping=1000)
head(analysis.two.markers$estimates)
head(analysis.two.markers$bootstrap.CIs)
```

enrichment_simulation	*Prognostic Enrichment with Simulated Data*

Description

Evaluating biomarkers for prognostic enrichment of clinical trials using simulated data

Usage

```
enrichment_simulation(baseline.event.rate, reduction.under.treatment,
  estimated.auc, roc.type = NULL, cost.screening = NULL,
  cost.keeping = NULL, simulation.sample.size = 5e+05,
  alternative = c("one.sided", "two.sided"), power = 0.9, alpha = 0.025,
  selected.biomarker.quantiles = seq(from = 0, to = 95, by = 5))
```

Arguments

baseline.event.rate

A number between 0 and 1 indicating the prevalence of the event in the study population.

reduction.under.treatment

A number between 0 and 1 indicating the percent reduction in event rate under treatment that the trial should be able to detect with the specified power.

estimated.auc   A numeric vector, with each entry between 0.5 and 1, that specifies the AUC for each biomarker to use in simulations.

roc.type   A character vector with the same length as the estimated.auc argument. Each entry must be one of "symmetric", "right.shifted", or "left.shifted", which describes the general shape of the ROC curve to use for simulated data. Defaults to "symmetric" for each biomarker.

cost.screening A positive number indicating the cost of screening a patient to determine trial eligibility, This argument is optional; if both cost.screening and cost.keeping are specified, then then the total cost of the trial based on each screening threshold is estimated and returned.

cost.keeping   A positive number indicating the cost of retaining a patient in the trial after enrolling. This argument is optional; if both cost.screening and cost.keeping are specified, then then the total cost of the trial based on each screening threshold is estimated and returned.

simulation.sample.size

A positive number giving the sample size to use for simulated data. Defaults to 500,000 (to help see trends).

alternative   Character specifying whether the alternative hypothesis is one-sided ("one.sided") with a higher outcome probability in the treatment group or two-sided ("two.sided"). Defaults to "one.sided".

power   Number between 0 and 1 giving the power the trial should have to reject the null hypothesis that there is no treatment effect. Defaults to 0.9.

alpha   Number between 0 and 1 giving the type I error rate for testing the null hypoth-
                         esis that there is no treatment effect. Defaults to 0.025.

selected.biomarker.quantiles
                         Numeric vector specifying the quantiles of the biomarker measured in controls
                         that will be used to screen trial participants. Defaults to 0, 5, ..., 95. All entries
                         must be between at least 0 and less than 001.

Value

A list with components

- estimates: A data frame with the following summary measures for each biomarker thresh-
  old that is used to screen trial participants: 'selected.biomarker.quantiles': quantiles of ob-
  served biomarker values used for screening. 'biomarker.screening.thresholds': the values of
  the biomarker corresponding to the quantiles, 'event.rate': post-screening event rate, 'NNS':
  The estimated number of patients needed to screen to identify one patient eligible for the
  trial, 'SS': The sample size in a clinical trial enrolling only patients whose biomarker-based
  disease risk is above the level used for screening, 'N.screen': The total number of individ-
  uals whose biomarker values are screened to determine whether they should be enrolled in
  the trial, 'N.screen.increase.percentage': Percentage in N.screen relative to a trail that does
  not based on the biomarker. 'total.cost': The estimated total cost of running the trial if the
  biomarker were used for prognostic enrichment (if cost.screening and cost.keeping are spec-
  ified), 'cost.reduction.percentage': The reduction in total cost relative to a trial that does not
  screen based on the biomarker. 'Biomarker': label for the biomarker.

- simulation: Logical indicating whether data were simulated (always TRUE for the plot_enrichment_summaries
  function).

See Also

enrichment_analysis, plot_enrichment_summaries

Examples

```
## three biomarkers with symmetric ROC curves
simulation.three.markers <- enrichment_simulation(baseline.event.rate=0.2,
reduction.under.treatment=0.3,
estimated.auc=c(0.72, 0.82, 0.85),
roc.type=c("symmetric", "symmetric", "symmetric"),
cost.screening=1,
cost.keeping=10,
simulation.sample.size=1e+5)
head(simulation.three.markers$estimates)
```

plot_enrichment_summaries

Plotting Prognostic Enrichment Estimates

Description

Plot summaries of prognostic enrichment of clinical trials estimated by the enrichment_analysis
and enrichment_simulation functions.

Usage

```
plot_enrichment_summaries(x, text.size.x.axis = 10, text.size.y.axis = 10,
  text.size.plot.title = 10, text.size.axis.ticks = 10,
  annotate.no.screening.cost = FALSE, smooth.roc = TRUE)
```

Arguments

x	Object returned by either the enrichment_analysis or the enrichment_simulation function.
text.size.x.axis	
	Size of text for the x-axis of plots. Defaults to 10.
text.size.y.axis	
	Size of text for the y-axis of plots. Defaults to 10.
text.size.plot.title	
	Size of text for the plot titles. Defaults to 10.
text.size.axis.ticks	
	Size of axis tick marks for plots. Defaults to 10.
annotate.no.screening.cost	
	Logical indicating whether to annotate the relative to total cost curve at the point where no biomarker screening occurs. Defaults to FALSE
smooth.roc	Logical indicating whether the ROC curves (plotting with the roc() function in the 'pROC' package) should be smoothed. Defaults to TRUE.

Value

A grid of either 4 or 6 plots, summarizing the results of either the enrichment_analysis or the
enrichment_simulation function.

See Also

enrichment_analysis, enrichment_simulation

Examples

```
data(dcaData)
# one marker
analysis.single.marker <- enrichment_analysis(Cancer ~ Marker1,
data=dcaData,
reduction.under.treatment=0.3,
cost.screening=100, cost.keeping=1000)
plot_enrichment_summaries(analysis.single.marker)

# two markers
analysis.two.markers <- enrichment_analysis(Cancer ~ Marker1 + Marker2,
data=dcaData,
reduction.under.treatment=0.3,
cost.screening=100,
cost.keeping=1000)
plot_enrichment_summaries(analysis.two.markers)
```

Index