

Package ‘rprev’

September 2, 2016

Type Package

Title Estimating Disease Prevalence from Registry Data

Version 0.2.2

Date 2016-09-02

Description Estimates disease prevalence for a given index date, using existing registry data extended with Monte Carlo simulations.

License GPL-2

Depends survival

Imports doParallel, dplyr, foreach, ggplot2, lazyeval, magrittr, rms, tidy

RoxygenNote 5.0.1

Suggests testthat, knitr

URL <https://cran.r-project.org/package=rprev>

VignetteBuilder knitr

NeedsCompilation no

Author Simon Crouch [aut],
Stephanie Lax [aut],
Stuart Lacy [cre, aut]

Maintainer Stuart Lacy <stuart.lacy@york.ac.uk>

Repository CRAN

Date/Publication 2016-09-02 22:45:26

R topics documented:

determine_registry_years	2
functional_form_age	3
incidence	4
incidence_age_distribution	5
mean_incidence_rate	6
plot.incidence	7

plot.survfit.prev	8
population_survival_rate	9
prevalence	10
prevalence_counted	12
prevalence_simulated	14
prevsim	16
raw_incidence	17
rprev	18
summary.survfit.prev	18
survfit.prevalence	19
test_incidence_fit	20
test_prevalence_fit	21
UKmortality	22
Index	23

determine_registry_years
Calculate yearly dates.

Description

A helper function to calculate dates a year apart, starting from a given date and running for a set number of years.

Usage

```
determine_registry_years(start, num_reg_years)
```

Arguments

start	Date from which incident cases are included in the format YYYY-MM-DD. Defaults to the earliest entry date.
num_reg_years	The number of years of the registry for which incidence is to be calculated. Defaults to using all available complete years.

Value

A vector of dates delineating complete years of registry data.

functional_form_age *View the suitability of a linear effect of age on hazard.*

Description

Inspects whether a non-linear transform of age is more appropriate for survival modeling than the linear method which is currently used in the calculations in [prevalence](#).

Usage

```
functional_form_age(form, data, df = 4, plot_fit = TRUE, num_points = 200)
```

Arguments

form	Formula where the LHS is represented by a standard Surv object, and the RHS provides the name of the variable containing age in the supplied dataset.
data	A data frame with the corresponding column names provided in form.
df	The desired degrees of freedom for the cubic spline, must be ≥ 3 .
plot_fit	Whether to plot the relationship between age and relative hazard.
num_points	The number of values of age to calculate the hazard ratio at.

Value

An `rms::cph` object containing the Cox model fit between a cubic spline transform of age and the survival outcomes.

Examples

```
data(prevsim)

functional_form_age(Surv(time, status) ~ age, prevsim)

functional_form_age(Surv(time, status) ~ age, prevsim, df=3)

functional_form_age(Surv(time, status) ~ age, prevsim, df=3, plot_fit=FALSE)
```

incidence *Summarise disease incidence.*

Description

Calculates incidence by year of the registry data, along with mean incidence with confidence intervals. A smoothed cumulative incidence function is fit to the data for inspecting deviations in the registry data from a homogeneous Poisson process.

Usage

```
incidence(entry, population_size, start = NULL, num_reg_years = NULL,
          df = 6, precision = 2, level = 0.95)
```

Arguments

entry	Vector of diagnosis dates for each patient in the registry in the format YYYY-MM-DD.
population_size	Integer corresponding to the size of the population at risk.
start	Date from which incident cases are included in the format YYYY-MM-DD. Defaults to the earliest entry date.
num_reg_years	The number of years of the registry for which incidence is to be calculated. Defaults to using all available complete years.
df	The desired degrees of freedom of the smooth.
precision	The number of decimal places required.
level	The desired confidence interval width.

Value

An S3 object of class `incidence` with the following attributes:

raw_incidence	Vector of absolute incidence values for each included year of the registry, as generated using raw_incidence .
ordered_diagnoses	Vector of times (days) between diagnosis date and the earliest date of inclusion in the registry, ordered shortest to longest.
smooth	Smooth fitted to the cumulative incidence data.
index_dates	Dates delimiting the years in which incidence is calculated.
mean	List containing mean incidence per 100K with confidence intervals. See mean_incidence_rate .
dof	Degrees of freedom of the smooth.

See Also

Other incidence functions: [incidence_age_distribution](#), [mean_incidence_rate](#), [plot.incidence](#), [raw_incidence](#), [test_incidence_fit](#)

Examples

```
data(prevsim)

## Not run:
incidence(prevsim$entrydate, 1e6)

incidence(prevsim$entrydate, 1e6, start = "2004-01-30", num_reg_years = 9)

## End(Not run)
```

```
incidence_age_distribution
```

Plot the age distribution of incident cases.

Description

This function plots the age distribution of incident cases from the registry.

Usage

```
incidence_age_distribution(agedata, df = 10)
```

Arguments

agedata	Vector of ages at diagnosis for each patient in the registry.
df	Degrees of freedom for the smooth.

Value

Plot of the raw data and smoothed age distribution function.

See Also

Other incidence functions: [incidence](#), [mean_incidence_rate](#), [plot.incidence](#), [raw_incidence](#), [test_incidence_fit](#)

Examples

```
data(prevsim)

incidence_age_distribution(prevsim$age)

incidence_age_distribution(prevsim$age, df=5)
```

mean_incidence_rate *Mean disease incidence.*

Description

Calculates the average incidence rate per one hundred thousand with confidence intervals for the given registry data.

Usage

```
mean_incidence_rate(raw_inc, population_size, precision = 2, level = 0.95)
```

Arguments

raw_inc	Vector of incidence values by year, as returned by raw_incidence .
population_size	Integer corresponding to the size of the population at risk.
precision	The number of decimal places required.
level	The desired confidence interval width.

Value

A list with the following values:

absolute	Overall incidence for the period of interest as a single double
per100K	Incidence for the period of interest per one hundred thousand
per100K.lower	Lower bounds of the specified confidence level on the per one hundred thousand estimate
per100K.upper	Upper bounds of the specified confidence level on the per one hundred thousand estimate

See Also

Other incidence functions: [incidence_age_distribution](#), [incidence](#), [plot.incidence](#), [raw_incidence](#), [test_incidence_fit](#)

Examples

```
data(prevsim)

rawinc <- raw_incidence(prevsim$entrydate)
mean_incidence_rate(rawinc, population_size=3.5e6)

rawinc2 <- raw_incidence(prevsim$entrydate, start="2005-05-01", num_reg_years=5)
mean_incidence_rate(rawinc2, population_size=3.5e6)
```

plot.incidence	<i>Visualise disease incidence.</i>
----------------	-------------------------------------

Description

Plots a comparison between the smoothed daily incidence function and actual incidence.

Usage

```
## S3 method for class 'incidence'  
plot(x, level = 0.95, ...)
```

Arguments

x	An incidence object.
level	The desired confidence interval width.
...	Arguments passed to plot.

Details

This function generates a plot from the cumulative incidence object. The incidence rate per year of the registry is shown in red. Mean incidence rate is shown as a solid blue line, with the confidence interval shown in dashed blue lines. The smooth fitted to the cumulative incidence data is shown in green.

Value

None, plots a side effect of incidence rate, confidence interval and smoothed incidence function.

See Also

Other incidence functions: [incidence_age_distribution](#), [incidence](#), [mean_incidence_rate](#), [raw_incidence](#), [test_incidence_fit](#)

Examples

```
data(prevsim)  
  
## Not run:  
inc <- incidence(prevsim$entrydate, population_size=1e6, start = "2004-01-30", num_reg_years = 9)  
  
plot(inc)  
  
## End(Not run)
```

```
plot.survfit.prev      Plot bootstrapped survival curves.
```

Description

This method plots survival curves for a `survfit.prev` object.

Usage

```
## S3 method for class 'survfit.prev'
plot(x, pct_show = 0.9, ...)
```

Arguments

<code>x</code>	A <code>survfit.prev</code> object.
<code>pct_show</code>	A list or dataframe with the covariate values to calculate survival probabilities.
<code>...</code>	Arguments passed to <code>plot</code> .

Details

The survival curve for a model formed on all the data is displayed in orange, while the 95% as a grey ribbon. Outlying survival curves are displayed in full, where the `pct_show` argument details the proportion of points outside of the confidence interval for a survival curve to be deemed as an outlier.

Value

An S3 object of class `survfit.prev` with the following attributes:

Examples

```
data(prevsim)

## Not run:
prev_obj <- prevalence(Surv(time, status) ~ age(age) + sex(sex) +
  entry(entrydate) + event(eventdate),
  data=prevsim, num_years_to_estimate = c(5, 10),
  population_size=1e6, start = "2005-09-01",
  num_reg_years = 8, cure = 5)

survobj <- survfit(prev_obj, newdata=list(age=65, sex=0))

plot(survobj)

plot(survobj, pct_show=0) # Display curves with any outlying points
plot(survobj, pct_show=0.5) # Display curves with half outlying points
plot(survobj, pct_show=0.99) # Display curves with nearly all outlying points
```

```
## End(Not run)
```

```
population_survival_rate
```

```
Transform yearly mortality rates to daily.
```

Description

Calculates the daily mortality probability for a given population stratified by age, based on their yearly mortality rates.

Usage

```
population_survival_rate(form, data, max_age = 100)
```

Arguments

form	Formula where the LHS indicates the name of the mortality rate column, and the RHS is the column where age is located. This function assumes that the population data frame has already been stratified by sex, or any other categorical variable of interest.
data	Data frame of population mortality stratified by sex and age. The following columns must be present: sex, age, and rate.
max_age	Maximum age to calculate mortality for.

Value

An estimate of the survival rate by age in days, with $\text{max_age} * 365$ values.

Examples

```
data(UKmortality)

population_survival_rate(rate ~ age, UKmortality)
population_survival_rate(rate ~ age, subset(UKmortality, sex==0))
```

prevalence

*Estimate point prevalence at an index date.***Description**

Point prevalence at a specific index date is estimated using contributions to prevalence from both available registry data, and from Monte Carlo simulations of the incidence and survival process, as outlined by Crouch et al (2004) (see References).

Usage

```
prevalence(form, data, num_years_to_estimate, population_size,
  index_date = NULL, num_reg_years = NULL, cure = 10, N_boot = 1000,
  level = 0.95, precision = 2, proportion = 1e+05,
  population_data = NULL, n_cores = 1, start = NULL)
```

Arguments

form	Formula where the LHS is represented by a standard Surv object, and the RHS has three special function arguments: age, the column where age is located; sex, the column where sex is located; entry, the column where dates of entry to the registry are located; and event, the column where event dates are located. This formula is used in the following way: $\text{Surv}(\text{time}, \text{status}) \sim \text{age}(\text{age_column_name}) + \text{sex}(\text{sex_column_name}) + \text{entry}(\text{entry_column_name}) + \text{event}(\text{event_column_name})$ Using the supplied prevsim dataset, it is therefore called with: $\text{Surv}(\text{time}, \text{status}) \sim \text{age}(\text{age}) + \text{sex}(\text{sex}) + \text{entry}(\text{entrydate}) + \text{event}(\text{eventdate})$
data	A data frame with the corresponding column names provided in form.
num_years_to_estimate	Number of years of data to consider when estimating point prevalence; multiple values can be specified in a vector. If any values are greater than the number of years of registry data available before index_date, incident cases for the difference will be simulated.
population_size	Integer corresponding to the size of the population at risk.
index_date	The date at which to estimate point prevalence. Defaults to the latest registry entry date.
num_reg_years	The number of years of the registry for which incidence is to be calculated. Defaults to using all available complete years. Note that if more registry years are supplied than the number of years to estimate prevalence for, the survival data from the surplus registry years are still involved in the survival model fitting.
cure	Integer defining cure model assumption for the calculation (in years). A patient who has survived beyond the cure time has a probability of surviving derived from the mortality rate of the general population.

N_boot	Number of bootstrapped calculations to perform.
level	Double representing the desired confidence interval width.
precision	Integer representing the number of decimal places required.
proportion	The population ratio to estimate prevalence for.
population_data	A dataframe that must contain the columns age, rate, and sex, where each row is the mortality rate for a person of that age and sex. Ideally, age ranges from [0, 100]. Defaults to the supplied data; see UKmortality for the format required for custom datasets.
n_cores	Number of CPU cores to run the fitting of the bootstrapped survival models. Defaults to 1; multi-core functionality is provided by the doParallel package.
start	Deprecated: Use index_date instead and specify the number of years of registry data to use with num_reg_years. Date from which incident cases are included in the format YYYY-MM-DD. Defaults to the earliest entry date. This value is now inferred by counting back num_reg_years years of registry data from the index_date. and

Details

The most important parameter is `num_years_to_estimate`, which governs the number of previous years of data to use when estimating the prevalence at the index date. If this parameter is greater than the number of years of known incident cases available in the supplied registry data (specified with argument `num_registry_years`), then the remaining `num_years_to_estimate - num_registry_years` years of incident data will be simulated using Monte Carlo simulation.

The larger `num_years_to_estimate`, the more accurate the prevalence estimate will be, provided an adequate survival model can be fitted to the registry data. It is therefore important to provide as much clean registry data as possible.

Simulated cases are marked with age and sex to enable agreement with population survival data where a cure model is used, and calculation of the posterior distributions of each.

Value

An S3 object of class `prevalence` with the following attributes:

estimates	Estimated prevalence at the index date for each of the years in <code>num_years_to_estimate</code> .
simulated	A list containing items related to the simulation of prevalence contributions, see prevalence_simulated
.	.
counted	Contributions to prevalence from each of the supplied registry years, see prevalence_counted .
start_date	The starting date of the registry data included in the estimation.
index_date	The index date at which the point prevalence was calculated for.
known_inc_rate	The known incidence rate for years included in the registry.
nregyears	Number of years of registry data that were used.
nbootstraps	The number of bootstrapped survival models fitted during the calculation.

pval	The p-value resulting from the chi-square test between the simulated and counted prevalent cases for the years of registry data available.
y	The Surv object used as the response in the survival modeling.
means	The covariate means from the data.

References

Crouch, Simon, et al. "Determining disease prevalence from incidence and survival using simulation techniques." *Cancer epidemiology* 38.2 (2014): 193-199.

See Also

Other prevalence functions: [prevalence_counted](#), [prevalence_simulated](#), [test_prevalence_fit](#)

Examples

```
data(prevsim)

## Not run:
prevalence(Surv(time, status) ~ age(age) + sex(sex) + entry(entrydate) + event(eventdate),
           data=prevsim, num_years_to_estimate = c(5, 10), population_size=1e6,
           index_date = '2013-09-01', num_reg_years = 8,
           cure = 5)

prevalence(Surv(time, status) ~ age(age) + sex(sex) + entry(entrydate) + event(eventdate),
           data=prevsim, num_years_to_estimate = 5, population_size=1e6)

# Run on multiple cores
prevalence(Surv(time, status) ~ age(age) + sex(sex) + entry(entrydate) + event(eventdate),
           data=prevsim, num_years_to_estimate = c(3,5,7), population_size=1e6, n_cores=4)

## End(Not run)
```

prevalence_counted *Count prevalence from registry data.*

Description

Counts contribution to prevalence at a specific index from each year of a registry. A person is included as contributing to disease prevalence if they are incident within the specified time-span, and are either alive or censored at the index date. The rationale for including censored cases in prevalence estimation is that such cases have typically been lost to follow-up, and are often more likely to have been alive at the index date than not.

Usage

```
prevalence_counted(entry, eventdate, status, index_date = NULL,
                  num_reg_years = NULL, start = NULL)
```

Arguments

entry	Vector of diagnosis dates for each patient in the registry in the format YYYY-MM-DD.
eventdate	Vector of dates corresponding to the indicator variable in the format YYYY-MM-DD.
status	Vector of binary values indicating if an event has occurred for each patient in the registry. entry, eventdate, and status must all have the same length.
index_date	The date at which to estimate point prevalence. Defaults to the latest registry entry date.
num_reg_years	The number of years of the registry for which incidence is to be calculated. Defaults to using all available complete years. Note that if more registry years are supplied than the number of years to estimate prevalence for, the survival data from the surplus registry years are still involved in the survival model fitting.
start	Deprecated: Use index_date instead and specify the number of years of registry data to use with num_reg_years. Date from which incident cases are included in the format YYYY-MM-DD. Defaults to the earliest entry date. This value is now inferred by counting back num_reg_years years of registry data from the index_date. and

Value

A vector of length equal to the number of complete years of registry data before index_date (given by num_reg_years), representing the number of incident cases in the corresponding year that contribute to the prevalence at the index date.

See Also

Other prevalence functions: [prevalence_simulated](#), [prevalence](#), [test_prevalence_fit](#)

Examples

```
data(prevsim)

prevalence_counted(prevsim$entrydate,
                  prevsim$eventdate,
                  prevsim$status)

prevalence_counted(prevsim$entrydate,
                  prevsim$eventdate,
                  prevsim$status,
                  index_date="2012-01-30")
```

```
prevalence_simulated
```

Estimate prevalence using Monte Carlo simulation.

Description

Estimates prevalent cases at a specific index date by use of Monte Carlo simulation. Simulated cases are marked with age and sex to enable agreement with population survival data where a cure model is used, and calculation of the posterior distributions of each.

Usage

```
prevalence_simulated(survobj, age, sex, entry, num_years_to_estimate,
  index_date, num_reg_years, cure = 10, start = NULL, N_boot = 1000,
  population_data = NULL, n_cores = 1)
```

Arguments

survobj	Surv object from survival package. Currently only right censoring is supported.
age	A vector of ages from the registry.
sex	A vector of sex, encoded as 0 and 1 for males and females respectively.
entry	A vector of entry dates into the registry, in the format YYYY-MM-DD.
num_years_to_estimate	Number of years of data to consider when estimating point prevalence; multiple values can be specified in a vector. If any values are greater than the number of years of registry data available before index_date, incident cases for the difference will be simulated.
index_date	The date at which to estimate point prevalence. Defaults to the latest registry entry date.
num_reg_years	The number of years of the registry for which incidence is to be calculated. Defaults to using all available complete years. Note that if more registry years are supplied than the number of years to estimate prevalence for, the survival data from the surplus registry years are still involved in the survival model fitting.
cure	Integer defining cure model assumption for the calculation (in years). A patient who has survived beyond the cure time has a probability of surviving derived from the mortality rate of the general population.
start	Deprecated: Use index_date instead and specify the number of years of registry data to use with num_reg_years. Date from which incident cases are included in the format YYYY-MM-DD. Defaults to the earliest entry date. This value is now inferred by counting back num_reg_years years of registry data from the index_date. and
N_boot	Number of bootstrapped calculations to perform.

population_data	A dataframe that must contain the columns age, rate, and sex, where each row is the mortality rate for a person of that age and sex. Ideally, age ranges from [0, 100]. Defaults to the supplied data; see UKmortality for the format required for custom datasets.
n_cores	Number of CPU cores to run the fitting of the bootstrapped survival models. Defaults to 1; multi-core functionality is provided by the <code>doParallel</code> package.

Value

A list with the following attributes:

mean_yearly_contributions	A vector of length <code>num_years_to_estimate</code> , representing the average number of prevalent cases subdivided by year of diagnosis across each bootstrap iteration.
posterior_age	Posterior distributions of age, sampled at every bootstrap iteration.
yearly_contributions	Total simulated prevalent cases from every bootstrapped sample.
pop_mortality	Population survival rates in the format of a list, stratified by sex.
nbootstraps	Number of bootstrapped samples used in the prevalence estimation.
coefs	The bootstrapped Weibull coefficients used by the survival models.
full_coefs	The Weibull coefficients from a model fitted to the full dataset.

See Also

Other prevalence functions: [prevalence_counted](#), [prevalence](#), [test_prevalence_fit](#)

Examples

```
data(prevsim)

## Not run:
prevalence_simulated(Surv(prevsim$time, prevsim$status), prevsim$age,
                     prevsim$sex, prevsim$entrydate,
                     num_years_to_estimate = 10,
                     index_date = "2013-09-01",
                     num_reg_years = 8, cure = 5)

prevalence_simulated(Surv(prevsim$time, prevsim$status), prevsim$age,
                     prevsim$sex, prevsim$entrydate,
                     num_years_to_estimate = 5,
                     index_date="2009-01-01",
                     num_reg_years=5)

# The program can be run using parallel processing.
prevalence_simulated(Surv(prevsim$time, prevsim$status), prevsim$age,
                     prevsim$sex, prevsim$entrydate,
                     num_years_to_estimate = 10,
```

```

index_date="2013-01-01",
num_reg_years=8, n_cores=4)

## End(Not run)

```

```
prevsim          Simulated patient dataset.
```

Description

A dataset in the format of a disease registry, where the outcome being modelled is death due to the simulated disease. The registry began in January 2003, with 1000 incident cases being recorded over a period of nearly exactly ten years. The patients are followed up for a further two years until 17.03.2015, at which point any subjects alive are marked as right censored.

Usage

```
prevsim
```

Format

A data frame with 1000 rows and 6 columns:

time time between date of diagnosis and death or censorship in days

status event marker; 1 if patient is deceased and 0 if alive or censored

age age in years at point of entry into the registry

sex boolean variable; 0 for males and 1 for females

entrydate date of entry into the registry in YYYY-MM-DD format

eventdate date of death or censorship in YYYY-MM-DD format

Details

Demographic and disease-specific data required for prevalence estimations are included, such as sex, age, and dates of entry and event. `eventdate` marks the date of the last known follow-up with the patient, corresponding to death (`status = 1`) or censorship (`status = 0`).

raw_incidence	<i>Disease incidence.</i>
---------------	---------------------------

Description

Calculates yearly incidence for the available registry data.

Usage

```
raw_incidence(entry, start = NULL, num_reg_years = NULL)
```

Arguments

entry	Vector of diagnosis dates for each patient in the registry in the format YYYY-MM-DD.
start	Date from which incident cases are included in the format YYYY-MM-DD. Defaults to the earliest entry date.
num_reg_years	The number of years of the registry for which incidence is to be calculated. Defaults to using all available complete years.

Value

Vector of length `num_reg_years` of integers, representing the number of absolute incidence values for each included year of the registry.

See Also

Other incidence functions: [incidence_age_distribution](#), [incidence](#), [mean_incidence_rate](#), [plot.incidence](#), [test_incidence_fit](#)

Examples

```
data(prevsim)

raw_incidence(prevsim$entrydate, start="2004-01-01", num_reg_years=8)
raw_incidence(prevsim$entrydate)
raw_incidence(prevsim$entrydate, start="2005-05-01", num_reg_years=5)
raw_incidence(prevsim$entrydate, start="2005-05-01")
```

rprev	<i>rprev: Estimate disease point prevalence using a combination of registry data and Monte Carlo simulations.</i>
-------	---

Description

The rprev package uses available registry data to estimate point prevalence at a specified index date. This is done by calculating yearly incident cases and estimating survival probabilities of these cases at the index date, to establish yearly contributions of incidence to the point prevalence estimate.

Details

This process relies upon accurate modeling of both the incidence and survival process, requiring that two assumptions are met:

- That the disease incidence is a homogeneous Poisson process
- That survival can be modeled by a Weibull model, incorporating both age and sex

Prevalence is estimated using incident cases from a set number of years, where the larger this values the more accurate the prevalence estimates are. However, if the user asks to use more years of incident cases than are available in the registry data set, then the remaining years of incidence are simulated.

The primary function in this package is thereby [prevalence](#), which performs the combination of counted incidence from the registry data, and the simulated cases, along with the calculation of their survival probabilities at the index date.

[incidence](#) provides a summary of the incident cases in the registry data set, allowing for inspection of whether the homogeneous Poisson process assumption holds for the disease in question.

summary.survfit.prev	<i>Obtain N-year survival probability estimates.</i>
----------------------	--

Description

Summarises survival information at pre-specified years of interest on a survfit.prev object.

Usage

```
## S3 method for class 'survfit.prev'
summary(object, years = c(1, 3, 5), ...)
```

Arguments

object	A survfit.prev object.
years	A vector of years for which to estimate survival probability from the bootstrapped survival curves.
...	Arguments passed to main summary function.

Details

Survival probability is estimated as the mean of the bootstrapped survival curves at a specific time-point, with 2.5 confidence intervals. Survival probability can only be estimated at time points less than the maximum survival time in the original dataset that the prevalence object was fitted to.

Value

None, displays the survival probabilities to screen as a side-effect.

Examples

```
data(prevsim)

## Not run:
prev_obj <- prevalence(Surv(time, status) ~ age(age) + sex(sex) +
  entry(entrydate) + event(eventdate),
  data=prevsim, num_years_to_estimate = c(5, 10),
  population_size=1e6, start = "2005-09-01",
  num_reg_years = 8, cure = 5)

survobj <- survfit(prev_obj, newdata=list(age=65, sex=0))

summary(survobj)

summary(survobj, years=c(1, 3, 5, 7))

## End(Not run)
```

survfit.prevalence *Form bootstrapped survival curves.*

Description

Calculates bootstrapped survival probabilities from the Weibull models fitted to the prevalence object.

Usage

```
## S3 method for class 'prevalence'
survfit(formula, newdata = NULL, ...)
```

Arguments

formula	A prevalence object.
newdata	A list or dataframe with the covariate values to calculate survival probabilities for. Defaults to using the mean values from the the original dataset that the model was fit to.
...	Other arguments to survfit.

Value

An S3 object of class `survfit.prev` with the following attributes:

<code>time</code>	A vector of time points at which survival probability has been calculated.
<code>surv</code>	A matrix of survival probabilities, where the rows represent a different bootstrapped Weibull model, and the columns are each timepoint.
<code>fullsurv</code>	A vector of survival probabilities for the predictors provided in <code>newdata</code> .

Examples

```
data(prevsim)

## Not run:
prev_obj <- prevalence(Surv(time, status) ~ age(age) + sex(sex) +
  entry(entrydate) + event(eventdate),
  data=prevsim, num_years_to_estimate = c(5, 10),
  population_size=1e6, start = "2005-09-01",
  num_reg_years = 8, cure = 5)

survobj <- survfit(prev_obj)

survobj <- survfit(prev_obj, newdata=list(age=65, sex=0))

## End(Not run)
```

`test_incidence_fit` *Test homogeneous Poisson assumption of disease incidence.*

Description

This function compares the actual variance of the yearly incidence rates with rates simulated from a Poisson process with overall rate equal to the overall mean rate.

Usage

```
test_incidence_fit(inc, N_sim = 1e+05)
```

Arguments

<code>inc</code>	Vector of absolute incidence values for each included year of the registry as generated by raw_incidence .
<code>N_sim</code>	Number of simulations to perform.

Value

Vector of p-values for over- and under-dispersion based on the position of the observed sequence variance in the distribution.

See Also

Other incidence functions: [incidence_age_distribution](#), [incidence](#), [mean_incidence_rate](#), [plot.incidence](#), [raw_incidence](#)

Examples

```
data(prevsim)

inc <- raw_incidence(prevsim$entrydate)

test_incidence_fit(inc)
```

test_prevalence_fit *Test simulated prevalence fit.*

Description

Calculates a Chi squared test between predicted yearly contributions to prevalence, and the observed values obtained from the registry, indicating whether the simulated prevalence values are accurate.

Usage

```
test_prevalence_fit(object)
```

Arguments

object A prevalence object.

Value

P-value from a chi-squared test of difference between prevalence prediction and counted prevalence at the index date.

See Also

Other prevalence functions: [prevalence_counted](#), [prevalence_simulated](#), [prevalence](#)

Examples

```
data(prevsim)

## Not run:

obj <- prevalence(Surv(time, status) ~ age(age) + sex(sex) + entry(entrydate) + event(eventdate),
                 data=prevsim, num_years_to_estimate = c(5, 10), population_size=1e6,
                 start = "2005-09-01",
                 num_reg_years = 8, cure = 5)

test_prevalence_fit(obj)
```

```
## End(Not run)
```

UKmortality

General population survival data.

Description

A dataset containing basic mortality rates stratified by sex and age (0 to 100 year) from the UK population, yearly, obtained from: <http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Life+Tables#tab-data-tables>. Adapted from public sector information licensed under the Open Government Licence v3.0. Data were relabelled according to the mean year of the three-year birth window.

Usage

UKmortality

Format

A data frame with 6666 rows and 4 columns:

calendar_year year

sex boolean variable; 0 for males and 1 for females

age age in years

rate death rate

Index

*Topic **datasets**

- prevsim, [16](#)
- UKmortality, [22](#)

determine_registry_years, [2](#)

functional_form_age, [3](#)

incidence, [4](#), [5–7](#), [17](#), [18](#), [21](#)

incidence_age_distribution, [4](#), [5](#), [6](#), [7](#), [17](#), [21](#)

mean_incidence_rate, [4](#), [5](#), [6](#), [7](#), [17](#), [21](#)

plot.incidence, [4–6](#), [7](#), [17](#), [21](#)

plot.survfit.prev, [8](#)

population_survival_rate, [9](#)

prevalence, [3](#), [10](#), [13](#), [15](#), [18](#), [21](#)

prevalence_counted, [11](#), [12](#), [12](#), [15](#), [21](#)

prevalence_simulated, [11–13](#), [14](#), [21](#)

prevsim, [16](#)

raw_incidence, [4–7](#), [17](#), [20](#), [21](#)

rprev, [18](#)

rprev-package (rprev), [18](#)

summary.survfit.prev, [18](#)

survfit.prevalence, [19](#)

test_incidence_fit, [4–7](#), [17](#), [20](#)

test_prevalence_fit, [12](#), [13](#), [15](#), [21](#)

UKmortality, [11](#), [15](#), [22](#)