Package 'xhaz'

June 30, 2024

Title Excess Hazard Modelling Considering Inappropriate Mortality Rates

Version 2.0.2

Author Juste Goungounga [aut, cre] (<https://orcid.org/0000-0002-9039-2639>), Hadrien Charvat [aut] (<https://orcid.org/0000-0003-3624-1394>), Darlin Mba [aut] (<https://orcid.org/0000-0001-9768-0230>), Nathalie Graffeo [aut] (<https://orcid.org/0000-0001-7227-7525>), Roch Giorgi [aut] (<https://orcid.org/0000-0001-6135-3078>)

Maintainer Juste Goungounga <juste.goungounga@ehesp.fr>

Description Fits relative survival regression models with or without proportional excess hazards and with the additional possibility to correct for background mortality by one or more parameter(s). These models are relevant when the observed mortality in the studied group is not comparable to that of the general population or in population-based studies where the available life tables used for net survival estimation are insufficiently stratified. In the latter case, the proposed model by Touraine et al. (2020) <doi:10.1177/0962280218823234> can be used. The user can also fit a model that relaxes the proportional expected hazards assumption considered in the Touraine et al. excess hazard model. This extension was proposed by Mba et al. (2020) <doi:10.1186/s12874-020-01139-z> to allow non-proportional effects of the additional variable on the general population mortality. In non-population-based studies, researchers can identify non-comparability source of bias in terms of expected mortality of selected individuals. An excess hazard model correcting this selection bias is presented in Goungounga et al. (2019) <doi:10.1186/s12874-019-0747-3>. This class of model with a random effect at the cluster level on excess hazard is presented in Goungounga et al. (2023) <doi:10.1002/bimj.202100210>.

License AGPL (>= 3)

Depends R (>= 4.3.0), statmod (>= 1.5.0), stats, survival (>= 3.5.7)

Imports gtools, mexhaz (>= 2.6), numDeriv, optimParallel, splines, stringr, survexp.fr

Suggests knitr, rmarkdown, spelling, testthat (>= 2.1.0)

VignetteBuilder knitr

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

Language en-US

NeedsCompilation no

Repository CRAN

Date/Publication 2024-06-29 22:40:02 UTC

Contents

anova.bsplines . . . 2
anova.constant . . . 4
anova.mexhazLT . . . 6
breast . . . 8
ccr.mevents . . . 9
dataCancer . . . 10
duplicate . . . 11
exphaz . . . 13
mexhazLT . . . 15
plot.bsplines . . . 20
plot.predxhaz . . . 23
predict.bsplines . . . 25
predict.constant . . . 27
print.bsplines . . . 29
print.constant . . . 30
print.predxhaz . . . 31
qbs . . . 32
rescaledData . . . 33
simuData . . . 34
summary.bsplines . . . 35
summary.constant . . . 36
xhaz . . . 37

Index **44**

`anova.bsplines` *anova.bsplines function used for likelihood-ratio Test of two models from xhaz function*

Description

This function compute an analysis of deviance table for two excess hazard models fitted using xhaz R package.

Usage

```
## S3 method for class 'bsplines'
anova(object, ..., test = "LRT")
```

Arguments

`object`	an object of class bsplines
`...`	an object of class bsplines
`test`	a character string. The appropriate test is a likelihood-ratio test, all other choices result in Not yet implemented test.

Value

An object of class `anova` inheriting from class `matrix`. The different columns contain respectively the degrees of freedom and the log-likelihood values of the two nested models, the degree of freedom of the chi-square statistic, the chi-square statistic and the p-value of the likelihood ratio test.

Note

As expected, the comparison between two or more models by anova or more excess hazard models will only be valid if they are fitted to the same dataset, and if the compared models are nested. This may be a problem if there are missing values.

Author(s)

Juste Goungounga, Robert Darlin Mba, Nathalie Graff\'eo and Roch Giorgi

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res. 2020 Jan;29(1):122-136. doi: 10.1177/0962280218823234. Epub 2019 Jan 23. PMID: 30674229. (PubMed)

Mba RD, Goungounga JA, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting inaccurate background mortality in excess hazard models through breakpoints. BMC Med Res Methodol. 2020 Oct 29;20(1):268. doi: 10.1186/s12874-020-01139-z. PMID: 33121436; PMCID: PMC7596976. (PubMed)

See Also

`xhaz`, `summary.bsplines`, `print.constant`

Examples

```
# load the data set in the package

library("survival")
library("numDeriv")
library("survexp.fr")
```

```
library("statmod")

data("dataCancer", package = "xhaz")   # load the data set in the package

fit.phBS <- xhaz(
      formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
      data = dataCancer,
      ratetable = survexp.fr::survexp.fr,
      interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
      rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
      baseline = "bsplines", pophaz  = "classic")



fit.nphBS <- xhaz(
      formula = Surv(obs_time_year, event) ~ ageCentre + qbs(immuno_trt),
      data = dataCancer,
      ratetable = survexp.fr::survexp.fr,
      interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
      rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
      baseline = "bsplines", pophaz  = "classic")

anova(fit.phBS, fit.nphBS)
```

`anova.constant` *anova.constant function used for likelihood-ratio Test of two models from xhaz function*

Description

This function compute an analysis of deviance table for two excess hazard models fitted using xhaz R package.

Usage

```
## S3 method for class 'constant'
anova(object, ..., test = "LRT")
```

Arguments

`object`	an object of class constant
`...`	an object of class constant
`test`	a character string. The appropriate test is a likelihood-ratio test, all other choices result in Not yet implemented test.

Value

An object of class `anova` inheriting from class `matrix`. The different columns contain respectively the degrees of freedom and the log-likelihood values of the two nested models, the degree of freedom of the chi-square statistic, the chi-square statistic and the p-value of the likelihood ratio test.

Note

As expected, the comparison between two or more models by anova or more excess hazard models will only be valid if they are fitted to the same dataset, and if the compared models are nested. This may be a problem if there are missing values.

Author(s)

Juste Goungounga, Robert Darlin Mba, Nathalie Graff\'eo and Roch Giorgi

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res. 2020 Jan;29(1):122-136. doi: 10.1177/0962280218823234. Epub 2019 Jan 23. PMID: 30674229. (PubMed)

Mba RD, Goungounga JA, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting inaccurate background mortality in excess hazard models through breakpoints. BMC Med Res Methodol. 2020 Oct 29;20(1):268. doi: 10.1186/s12874-020-01139-z. PMID: 33121436; PMCID: PMC7596976. (PubMed)

Giorgi R, Abrahamowicz M, Quantin C, Bolard P, Esteve J, Gouvernet J, Faivre J. A relative survival regression model using B-spline functions to model non-proportional hazards. Statistics in Medicine 2003; 22: 2767-84. (PubMed)

See Also

`xhaz`, `summary.bsplines`, `print.constant`

Examples

```
# load the data set in the package
library("survival")
library("numDeriv")
library("survexp.fr")


data("dataCancer")   # load the data set in the package

fit.ph <- xhaz(
      formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
```

```
      data = dataCancer,
      ratetable = survexp.fr::survexp.fr,
      interval = c(0, NA, NA, NA, max(dataCancer$obs_time_year)),
      rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
      baseline = "constant", pophaz  = "classic")



fit.ph2 <- xhaz(
      formula = Surv(obs_time_year, event) ~ ageCentre ,
      data = dataCancer,
      ratetable = survexp.fr::survexp.fr,
      interval = c(0, NA, NA, NA, max(dataCancer$obs_time_year)),
      rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
      baseline = "constant", pophaz  = "classic")

  anova(fit.ph2, fit.ph)
```

`anova.mexhazLT` *anova.mexhazLT function used for likelihood-ratio Test of two models from mexhaz function*

Description

This function compute an analysis of deviance table for two excess hazard models fitted using xhaz R package.

Usage

```
## S3 method for class 'mexhazLT'
anova(object, ..., test = "LRT")
```

Arguments

`object`	an object of class mexhazLT
`...`	an object of class mexhazLT
`test`	a character string. The appropriate test is a likelihood-ratio test, all other choices result in Not yet implemented test.

Value

An object of class `anova` inheriting from class `matrix`. The different columns contain respectively the degrees of freedom and the log-likelihood values of the two nested models, the degree of freedom of the chi-square statistic, the chi-square statistic and the p-value of the likelihood ratio test.

Note

As expected, the comparison between two or more models by anova or more excess hazard models will only be valid if they are fitted to the same dataset, and if the compared models are nested. This may be a problem if there are missing values.

Author(s)

Juste Goungounga, Hadrien Charvat, Robert Darlin Mba, Nathalie Graff\'eo and Roch Giorgi

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Goungounga, JA, Graff\'eo N, Charvat H, Giorgi R. "Correcting for heterogeneity and non-comparability bias in multicenter clinical trials with a rescaled random-effect excess hazard model." Biometrical journal. Biometrische Zeitschrift vol. 65,4 (2023): e2100210. doi:10.1002/bimj.202100210.PMID: 36890623; (PubMed)

See Also

xhaz, mexhazLT, AIC.mexhazLT

Examples

```
# load the data set in the package
library("survival")
library("numDeriv")
library("survexp.fr")


breast$sexe <- "female"

fit.haz <- exphaz(
                  formula = Surv(temps, statut) ~ 1,
                  data = breast, ratetable = survexp.us,
                  only_ehazard = FALSE,
                  rmap = list(age = 'age', sex = 'sexe', year = 'date'))

breast$expected <- fit.haz$ehazard
breast$expectedCum <- fit.haz$ehazardInt

mod.bs3 <- mexhazLT(formula = Surv(temps, statut) ~ agecr + armt,
                  data = breast,
                  ratetable = survexp.us, degree = 3,
                  knots=quantile(breast[breast$statut==1,]$temps, probs=c(1:2/3)),
                  expected = "expected",expectedCum = "expectedCum",
                  base = "exp.bs", pophaz = "classic", random ="hosp")

mod.bs3
```

```
mod.bs4 <- mexhazLT(formula = Surv(temps, statut) ~ agecr + armt,
                  data = breast,
                  ratetable = survexp.us, degree = 3,
                  knots=quantile(breast[breast$statut==1,]$temps, probs=c(1:2/3)),
                  expected = "expected",expectedCum = "expectedCum",
                  base = "exp.bs", pophaz = "rescaled", random = "hosp")

mod.bs4

  anova(mod.bs3, mod.bs4)
```

`breast` *Simulated clinical trial data with non comparability bias in term of individuals expected hazard*

Description

Simulated data

Usage

```
data(breast)
```

Format

This dataset contains the following variables:

temps Follow-up time (years)

statut Vital status

age Age at diagnosis

agecr Centered and scaled age

date date of diagnosis.

SEX 2 for female

armt 2 arms of treatment (0,1)

hosp clinical centers

dept department of residence

References

Goungounga, JA, Graff\'eo N, Charvat H, Giorgi R. "Correcting for heterogeneity and non-comparability bias in multicenter clinical trials with a rescaled random-effect excess hazard model." Biometrical journal. Biometrische Zeitschrift vol. 65,4 (2023): e2100210. doi:10.1002/bimj.202100210.PMID: 36890623; (PubMed)

Examples

```
data(breast)
summary(breast)
```

`ccr.mevents` *colorectum cancer data with multiple events*

Description

multiple events data

Usage

```
data(dataCancer)
```

Format

This dataset contains the following variables:

id patient IDs.

sex gender with 1 for male and 2 for female.

sexe gender male and female.

age Age at diagnosis

stage lower to higher stage 1, 2, 3

time time-to-events (local or distant recurrence or death)

status 0 : no event; 1: local or distant recurrence or death

event 1: local recurrence; 2: distant recurrence; 3:death

date_diag date of diagnosis.

References

Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res. 2020 Jan;29(1):122-136. doi: 10.1177/0962280218823234. Epub 2019 Jan 23. PMID: 30674229. (PubMed)

Examples

```
data(ccr.mevents)
summary(ccr.mevents)
```

`dataCancer` *Simulated data with cause death information with non comparability bias in term of individuals expected hazard*

Description

Simulated data

Usage

```
data(dataCancer)
```

Format

This dataset contains the following variables:

obs_time Follow-up time (months)

obs_time_year Follow-up time (years)

event Vital status

age Age at diagnosis

agegrp "<30" , "30_60" and ">=60" age groups

ageCentre centered age at diagnosis

sexx Sex(Female,Male).

immuno_trt Treatment group

year_date date of diagnosis.

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res. 2020 Jan;29(1):122-136. doi: 10.1177/0962280218823234. Epub 2019 Jan 23. PMID: 30674229. (PubMed)

Mba RD, Goungounga JA, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting inaccurate background mortality in excess hazard models through breakpoints. BMC Med Res Methodol. 2020 Oct 29;20(1):268. doi: 10.1186/s12874-020-01139-z. PMID: 33121436; PMCID: PMC7596976. (PubMed)

Examples

```
data(dataCancer)
summary(dataCancer)
```

duplicate *duplicate function*

Description

Duplicate data for survival analysis in the context of competing risks, where an individual can experience only one of alternative events, using the Lunn & McNeil (Biometrics, 1995) approaches. Duplication of data proceeds as follows: Suppose that we study `J` distinct types of events. Each observation concerning a given subject is duplicated `J` times, with one row for each type of event. In addition, `(J-1)` dummy variables are created, each indicating the type of event in relation with that observation (`delta.j=1` if the event of type `j` is the observed one and `0` otherwise). Since, for a given subject, only the first occurring event is considered, the status indicator equals `1` for that event and `0` for all the others. In the case of a censored observation (dropout or administrative censoring), the same principle applies also: duplication of each subject's data is made `J` times with `(J-1)` dummy variables and a status indicator equal to `0` for all observations.

Usage

```
duplicate(status, event, data)
```

Arguments

`status`	the censoring status indicator (numeric vector), 0=alive, 1=dead.
`event`	the indicator of the event type (numeric vector). By default, the event==0 acts as the censoring indicator.
`data`	a data frame containing the data to duplicate.

Value

A data.frame containing the duplicated data with the new dummy variables, named `delta.number_of_the_event`, indicating the type of event.

Author(s)

Roch Giorgi

References

Lunn M and McNeil D. Applying Cox regression to competing risks. Biometrics 1995;51:524-532 (PubMed)

Examples

```
## Create the simplest test data set
data1 <- data.frame(futime     = c(1, 2, 5, 2, 1, 7, 3, 4, 8, 8),
                    fustat     = c(0, 1, 1, 1, 0, 0, 1, 0, 1, 1),
                    firstevent = c(0, 2, 1, 2, 0, 0, 1, 0, 2, 2),
                    sex        = c(1, 0, 0, 1, 0, 1, 1, 1, 0, 0))
```

```
## Duplicate data1 with firstevent == 0 as the censoring indicator.
library(xhaz)
dupli.data <- duplicate(status=fustat, event=firstevent, data=data1)


data2 <- data.frame(futime = c(10, 2, 7, 3, 4, 9, 13, 2, 5, 9),
                    fustat = c(0, 1, 1, 1, 0, 0, 1, 0, 1, 1),
                    firstevent = c(3, 2, 1, 2, 3, 3, 1, 3, 2, 2),
                    sex = c(1, 0, 0, 1, 0, 1, 1, 1, 0, 0))


## Duplicate data1 with firstevent == 3 as the censoring indicator.

dupli.data <- duplicate(status = fustat,
                        event = firstevent == 3,
                        data = data2)


# Joint modeling
coxph(Surv(futime, fustat) ~ delta.2 + sex + delta.2:(sex), data = dupli.data)

coxph(Surv(futime, fustat) ~ delta.1 + sex + delta.1:(sex), data = dupli.data)

# exemple using ccr.mevents data


ccr.mevents$loc.rec <- as.numeric(ccr.mevents$event == 1)
ccr.mevents$dist.rec <- as.numeric(ccr.mevents$event == 2)
ccr.mevents$death <- as.numeric(ccr.mevents$event == 3)
# Age centered to mean and scaled
ccr.mevents$agecr <-  scale(ccr.mevents$age, TRUE, TRUE)

## Duplication of the data with local recurrence as the reference
dupli.ccr.mevents <- duplicate(status = status,
                               event = event, data = ccr.mevents)
head(dupli.ccr.mevents)
# joint model including overall mortality modelling
fit <- coxph(Surv(time, status) ~ agecr + sexe + stage + delta.2 + delta.3,
             data = dupli.ccr.mevents)

fit

# add expected mortality from french life table to the data


library(survexp.fr)
fit.haz <- exphaz(formula = Surv(time, death) ~ 1,
                  data = dupli.ccr.mevents,
                  ratetable = survexp.fr, only_ehazard = TRUE,
                  rmap = list(age = 'age', sex = 'sexe', year = 'date_diag'))

 dupli.ccr.mevents$mua <- fit.haz$ehazard * dupli.ccr.mevents$delta.3
```

```
# joint model including excess hazard modelling
library(mexhaz)
fit.mort <- mexhaz(
    Surv(time, status) ~ delta.2 + delta.3,
    data = dupli.ccr.mevents, base = "exp.bs", degree = 3, knots = c(1),
    expected = "mua")

fit.mort
```

---

exphaz *exphaz function*

Description

Calculate the expected hazard and survival.

Usage

```
exphaz(
  formula = formula(data),
  data = sys.parent(),
  ratetable,
  rmap = list(age = NULL, sex = NULL, year = NULL),
  ratedata = sys.parent(),
  only_ehazard = TRUE,
  subset,
  na.action,
  scale = 365.2425
)
```

Arguments

formula	a formula object of the Surv function with the response on the left of a ~ operator and the terms on the right. The response must be a survival object as returned by the Surv function (time in first and status in second).
data	a data frame in which to interpret the variables named in the formula
ratetable	a rate table stratified by age, sex, year (if missing, ratedata is used)
rmap	a list that maps data set names to the ratetable names.
ratedata	a data frame of the hazards mortality in general population.
only_ehazard	a boolean argument (by default, only_ehazard=TRUE). If TRUE, the cumulative population hazard is not provided.
subset	an expression indicating which subset of the rows in data should be used in the fit. All observations are included by default

na.action	a missing data filter function. The default is na.fail, which returns an error if any missing values are found. An alternative is na.exclude, which deletes observations that contain one or more missing values.
scale	a numeric argument specifying by default scale = 365.2425 (or using the value corresponding to attributes(ratetable)$cutpoints[[1]][2], often equal to 365.25) if the user wants to extract a yearly hazard rate, or scale = 1 if he wants to extract a daily hazard rate from a ratetable containing daily hazard rates for a matched subject from the population, defined as -log(1-q)/365.25 where q is the 1-year probability of death.

Value

An object of class list containing the following components:

ehazard	expected hazard calculated from the matching ratetable.
ehazardInt	cumulative expected hazard calculated from the matching ratetable. if only_ehazard=TRUE, this quantity is not provided.
dateDiag	date of diagnosis

Note

Time is OBLIGATORY in YEARS.

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Therneau, T. M., Grambsch, P. M., Therneau, T. M., & Grambsch, P. M. (2000). Expected survival. Modeling survival data: extending the Cox model, 261-287.

Examples

```
library(survival)
library(survexp.fr)
library(xhaz)
fit.haz <- exphaz(
               formula = Surv(obs_time_year, event) ~ 1,
               data = dataCancer,
               ratetable = survexp.fr, only_ehazard = TRUE,
               rmap = list(age = 'age', sex = 'sexx', year = 'year_date')
)
```

mexhazLT *mexhazLT function*

Description

Extends excess hazard models from the mexhaz R-package to allow rescaling (Goungounga et al. (2019) doi:10.1186/s12874-019-0747-3) of the background mortality in the presence or absence of multilevel data (Goungounga et al. (2023) <doi: 10.1002/bimj.202100210>). It allows for different shapes of the baseline hazard, the ability to include time-dependent effects of variable(s), and a random effect at the cluster level.

Usage

```
mexhazLT(
  formula,
  data,
  expected = "expected",
  expectedCum = "expectedCum",
  pophaz = "classic",
  base = c("weibull", "exp.bs", "exp.ns", "pw.cst"),
  degree = 3,
  knots = NULL,
  bound = NULL,
  n.gleg = 20,
  init = NULL,
  random = NULL,
  n.aghq = 10,
  fnoptim = c("nlm", "optim"),
  verbose = 0,
  method = "Nelder-Mead",
  iterlim = 10000,
  numHess = FALSE,
  print.level = 1,
  exactGradHess = TRUE,
  gradtol = ifelse(exactGradHess, 1e-08, 1e-06),
  testInit = TRUE,
  keep.data = FALSE,
  ...
)
```

Arguments

formula	a formula object of the function with the response on the left of a ~ operator and the terms on the right. The response must be a survival object as returned by the Surv function (time in first and status in second).
data	a data frame in which the variables named in the formula are to be interpreted.

`expected` name of the variable (must be given in quotes) representing the population instantaneous hazard.

`expectedCum` name of the variable (must be given in quotes) representing the population cumulative hazard.

`pophaz` specifies two possible arguments in character: classic and rescaled. If `pophaz = "classic"` is chosen, it fits the models that do not require the background mortality to be rescaled and assumes that the comparability assumption holds; if `pophaz = "rescaled"` is chosen, it fits the models that require that require the background mortality to be rescaled.

`base` functional form that should be used to model the baseline hazard. Selection can be made between the following options: `"weibull"` for a Weibull hazard, `"exp.bs"` for a hazard described by the exponential of a B-spline (only B-splines of degree 1, 2 or 3 are accepted), `"exp.ns"` for a hazard described by the exponential of a restricted cubic spline (also called 'natural spline'), `"pw.cst"` for a piecewise constant hazard. By default, base="weibull" as in mexhaz R-package.

`degree` if `base="exp.bs"`, degree represents the degree of the B-spline used. Only integer values between 1 and 3 are accepted, and 3 is the default.

`knots` if `base="exp.bs"` or `"exp.ns"`, knots is the vector of interior knots of the spline. If `base="pw.cst"`, knots is the vector defining the endpoints of the time intervals on which the hazard is assumed to be constant. By default, `knots=NULL` (that is, it produces a B-spline with no interior knots if base="exp.bs", a linear B-spline with no interior knots if base="exp.ns", or a constant hazard over the whole follow-up period if `base="pw.cst"`).

`bound` a vector of two numerical values corresponding to the boundary knots of the spline functions. If `base="exp.bs"` or `base="exp.ns"`, computation of the B-spline basis requires that boundary knots be given. The bound argument allows the user to specify these boundary knots. If `base="exp.bs"`, the interval defined by the boundary knots must at least include the interval `c(0,max(time))` (otherwise, there could be problems with ill-conditioned bases). If `base="exp.ns"`,

`n.gleg` corresponds to the number of quadrature nodes to be specified as in `mexhaz`.

`init` vector of initial values as in `mexhaz`.

`random` name of the variable to be entered as a random effect (must be given between quotes), representing the cluster membership. As in `mexhaz` `random=NULL` means that the function fits a fixed effects model.

`n.aghq` corresponds to the number of quadrature points to be specified as in `mexhaz` for the estimation of the cluster-specific marginal likelihoods by adaptative Gauss-Hermite quadrature.

`fnoptim` name of the R optimisation procedure used to maximise the likelihood. Selection can be made between "nlm" (by default) and "optim". Note: if `exactGradHess=TRUE`, this argument will be ignored (fnoptim will be set automatically to `"nlm"`).

`verbose` integer parameter representing the frequency at which the current state of the optimisation process is displayed. If verbose=0 (default), nothing is displayed.

`method` if fnoptim="optim", method represents the optimisation method to be used by optim. By default, `method="Nelder-Mead"`. This parameter is not used if `fnoptim="nlm"`.

iterlim	if fnoptim="nlm", iterlim represents the maximum number of iterations before the nlm optimisation procedure is terminated. By default, iterlim is set to 10000. This parameter is not used if fnoptim="optim" (in this case, the maximum number of iterations must be given as part of a list of control parameters via the control argument: see the help page of optim for further details).
numHess	logical value allowing the user to choose between the Hessian returned by the optimization algorithm (default) or the Hessian estimated by the hessian function from the numDeriv package.
print.level	his argument is only used if fnoptim="nlm". It determines the level of printing during the optimisation process. The default value (for the mexhaz function) is set to '1' which means that details on the initial and final step of the optimisation procedure are printed (see the help page of nlm for further details).
exactGradHess	logical value allowing the user to decide whether maximisation of the likelihood should be based on the analytic gradient and Hessian computed internally (default, corresponding to exactGradHess=TRUE).
gradtol	this argument is only used if fnoptim="nlm". It corresponds to the tolerance at which the scaled gradient is considered close enough to zero to terminate the algorithm. The default value depends on the value of the argument exactGradHess.
testInit	this argument is used only when exactGradHess=TRUE and when the model is not an excess hazard random effect model. It instructs the mexhaz function to try several vectors of initial values in case optimization was not successful with the default (or user-defined) initial values. Because optimization based on the analytical gradient and Hessian is usually fast, this simple and empirical procedure proves useful to increase the probability of convergence in cases when it is difficult to specify appropriate initial values.
keep.data	logical argument determining whether the dataset should be kept in the object returned by the function: this can be useful in certain contexts (e.g., to calculate cluster-specific posterior predictions from a random intercept model) but might create unnecessarily voluminous objects. The default value is set to FALSE.
...	other parameters used with the mexhazLT function

Value

An object of class mexhaz, xhaz or mexhazLT. This object is a list containing the following components:

dataset	name of the dataset used to fit the model.
call	function call on which the model is based.
formula	formula part of the call.
withAlpha	logical value indicating whether the model corresponds to a class of models correcting for life tables.
expected	name of the variable corresponding to the population hazard.
expectedCum	name of the variable corresponding to the cumulative population hazard.
xlevels	information concerning the levels of the categorical variables used in the model.
n.obs.tot	total number of observations in the dataset.

n.obs	number of observations used to fit the model (after exclusion of missing values).
n.events	number of events (after exclusion of missing values).
n.clust	number of clusters.
n.time.0	number of observations for which the observed follow-up time was equal to 0 (only for right censored type data).
base	function used to model the baseline hazard.
max.time	maximal observed time in the dataset.
boundary.knots	vector of boundary values used to define the B-spline (or natural spline) bases.
degree	degree of the B-spline used to model the logarithm of the baseline hazard.
knots	vector of interior knots used to define the B-spline (or natural spline) bases.
names.ph	names of the covariables with a proportional effect.
random	name of the variable defining cluster membership (set to NA in the case of a purely fixed effects model).
init	a vector containing the initial values of the parameters.
coefficients	a vector containing the parameter estimates.
std.errors	a vector containing the standard errors of the parameter estimates.
vcov	the variance-covariance matrix of the estimated parameters.
gradient	the gradient of the log-likelihood function evaluated at the estimated parameters.
hessian	the Hessian of the log-likelihood function evaluated at the estimated parameters.
mu.hat	a data.frame containing the estimated cluster-specific random effects (shrinkage estimators).
var.mu.hat	the covariance matrix of the cluster-specific shrinkage estimators.
vcov.fix.mu.hat	a matrix containing the covariances between the fixed effect and the cluster-specific shrinkage estimators. More specifically, the i-th line of the matrix represents the covariances between the shrinkage estimator of the i-th cluster and the fixed effect estimates. This matrix is used by the function predict.mexhaz to make cluster-specific predictions.
data	original dataset used to fit the model (if keep.data was set to TRUE).
n.par	number of estimated parameters.
n.gleg	number of Gauss-Legendre quadrature points used to calculate the cumulative (excess) hazard (only relevant if a B-spline of degree 2 or 3 or a cubic restricted spline was used to model the logarithm of the baseline hazard).
n.aghq	number of adaptive Gauss-Hermite quadrature points used to calculate the cluster-specific marginal likelihoods (only relevant if a multi-level model is fitted).
fnoptim	name of the R optimisation procedure used to maximise the likelihood.
method	optimisation method used by optim.
code	code (integer) indicating the status of the optimisation process (this code has a different meaning for nlm and for optim: see their respective help page for details).

loglik	value of the log-likelihood at the end of the optimisation procedure. Note that this is different to that calculated in mexhaz as the cumulative expected hazard cannot be removed from the log-likelihood.
iter	number of iterations used in the optimisation process.
eval	number of evaluations used in the optimisation process.
time.elapsed	total time required to reach convergence.

Note

time is OBLIGATORY in YEARS.

Author(s)

Juste Goungounga, Hadrien Charvat, Nathalie Graffeo, Roch Giorgi

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Goungounga, JA, Graff\'eo N, Charvat H, Giorgi R. "Correcting for heterogeneity and non-comparability bias in multicenter clinical trials with a rescaled random-effect excess hazard model." Biometrical journal. Biometrische Zeitschrift vol. 65,4 (2023): e2100210. doi:10.1002/bimj.202100210.PMID: 36890623; (PubMed)

Examples

```
library("numDeriv")
library("survexp.fr")
library("splines")
library("statmod")
data("breast")
# load the data sets 'breast'.

 # Flexible mexhaz model: baseline excess hazard with cubic B-splines
 # assumption on the life table available :
 # other cause mortality in the cohort is comparable to the mortality
 # observed in the general population with the same characteristics.

# The life table to be used is survexp.us. Note that SEX is coded 2 instead of female in survexp.us.
breast$sexe <- "female"

fit.haz <- exphaz(
                  formula = Surv(temps, statut) ~ 1,
                  data = breast, ratetable = survexp.us,
                  only_ehazard = FALSE,
                  rmap = list(age = 'age', sex = 'sexe', year = 'date'))

breast$expected <- fit.haz$ehazard
```

```
breast$expectedCum <- fit.haz$ehazardInt

mod.bs <- mexhazLT(formula = Surv(temps, statut) ~ agecr + armt,
                   data = breast,
                   ratetable = survexp.us, degree = 3,
                   knots=quantile(breast[breast$statut==1,]$temps, probs=c(1:2/3)),
                   expected = "expected",expectedCum = "expectedCum",
                   base = "exp.bs", pophaz = "classic")

mod.bs


 # Flexible mexhaz model: baseline excess hazard with cubic B-splines
 # assumption on the life table available :
 # other cause mortality in the cohort is different to the mortality
 # observed in the general population with the same characteristics.

mod.bs2 <- mexhazLT(formula = Surv(temps, statut) ~ agecr + armt,
                   data = breast, degree = 3,
                   knots=quantile(breast[breast$statut==1,]$temps, probs=c(1:2/3)),
                   expected = "expected",expectedCum = "expectedCum",
                   base = "exp.bs", pophaz = "rescaled")

mod.bs2


 # Flexible mexhaz model with a random effects at cluster level:
 # baseline excess hazard with cubic B-splines
 # assumption on the life table used :
 # other cause mortality in the cohort is different to the mortality
 # observed in the general population with the same characteristics.

mod.bs3 <- mexhazLT(formula = Surv(temps, statut) ~ agecr + armt,
                   data = breast, degree = 3,
                   knots=quantile(breast[breast$statut==1,]$temps, probs=c(1:2/3)),
                   expected = "expected",expectedCum = "expectedCum",
                   base = "exp.bs", pophaz = "rescaled", random = "hosp")

mod.bs3
```

---

plot.bsplines *plot.bsplines*

Description

to plot the log hazard ratio functions for non-proportional hazards model

Usage

```
## S3 method for class 'bsplines'
plot(
  x,
  cov,
  conf.int = TRUE,
  baseline = FALSE,
  xrange,
  yrange,
  xlegend,
  ylegend,
  glegend,
  xaxs = NULL,
  add = FALSE,
  col = 1,
  lty = 1,
  lwd = 1,
  ...
)
```

Arguments

x	An object of class xhaz
cov	specify covariates for which a plot is required.
conf.int	a vector of logical values indicating whether (if TRUE) confidence intervals will be plotted. The default is to do so if the plot concerns only one curve.
baseline	a vector of logical values indicating whether (if baseline = TRUE) to plot the curve for the baseline group. Default is FALSE, except if cov is unspecified.
xrange	vector indicating the minimum and the maximum values of the x axis. By default, these values are automatically calculated for the first plot (i.e before the use of add argument).
yrange	vector indicating the minimum and the maximum values of the y axis. By default, these values are automatically calculated for the first plot (i.e before the use of add argument).
xlegend	value indicating the location of the legend over x axis. By default, location at the left of the plot.
ylegend	value indicating the location of the legend over y axis. By default, location at the top of the plot
glegend	vectors of names attributed to each lines of the excess hazard to be displayed in the plot. If (baseline = TRUE), glegend is "baseline".
xaxs	the x axis style, as listed in 'par'. Survival curves are traditionally drawn with the curve touching the bounding box on the left edge, but not touching it on the right edge. This corresponds to neither of the two standard S axis styles of "e" (neither touches) or "i" (both touch). If xaxis is missing or NULL the internal axis style is used (xaxs= i) but only after the right endpoint has been extended.

add	a logical value indicating whether to add the survival curves to the current plot (if add = TRUE). Default is FALSE.
col	a vector of integers specifying colors for each curve. The default value is 1.
lty	a vector of integers specifying line types for each curve. The default value is fixed by the number of covariates (plus 1 if baseline = TRUE).
lwd	a vector of numeric values for line widths. The default value is 1.
...	additional arguments affecting the plot function

Value

The return of this function produce graphics of log hazard ratio functions for non-proportional hazards model

Author(s)

Juste Goungounga, Robert Darlin Mba, Nathalie Graff\'eo and Roch Giorgi

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res. 2020 Jan;29(1):122-136. doi: 10.1177/0962280218823234. Epub 2019 Jan 23. PMID: 30674229. (PubMed)

Mba RD, Goungounga JA, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting inaccurate background mortality in excess hazard models through breakpoints. BMC Med Res Methodol. 2020 Oct 29;20(1):268. doi: 10.1186/s12874-020-01139-z. PMID: 33121436; PMCID: PMC7596976. (PubMed)

Giorgi R, Abrahamowicz M, Quantin C, Bolard P, Esteve J, Gouvernet J, Faivre J. A relative survival regression model using B-spline functions to model non-proportional hazards. Statistics in Medicine 2003; 22: 2767-84. (PubMed)

Examples

```
# load the data set in the package
library("xhaz")
library("survexp.fr")

data("dataCancer", package = "xhaz")   # load the data set in the package

fit.nphBS <- xhaz(
      formula = Surv(obs_time_year, event) ~ ageCentre + qbs(immuno_trt),
      data = dataCancer,
      ratetable = survexp.fr,
      interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
```

```
      rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
      baseline = "bsplines", pophaz  = "classic")

plot(fit.nphBS, cov = "immuno_trt", col = "blue", baseline = FALSE)
```

`plot.predxhaz` *plots of excess hazard and net Survival from an* `predxhaz` *object*

Description

Function to plot excess hazard or net survival

Usage

```
## S3 method for class 'predxhaz'
plot(x, what = "survival", ...)
```

Arguments

`x`	An object of class predxhaz
`what`	allow to choose between excess hazard (`what="hazard"`) or net survival (`what="survival"`).
`...`	additional arguments affecting the plot function

Value

The return of this function produce graphics of excess hazard or net survival, or time-dependent effects, when times.pts argument is provided in prediction call.

Author(s)

Juste Goungounga, Robert Darlin Mba, Nathalie Graff\'eo and Roch Giorgi

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res. 2020 Jan;29(1):122-136. doi: 10.1177/0962280218823234. Epub 2019 Jan 23. PMID: 30674229. (PubMed)

Mba RD, Goungounga JA, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting inaccurate background mortality in excess hazard models through breakpoints. BMC Med Res Methodol. 2020 Oct 29;20(1):268. doi: 10.1186/s12874-020-01139-z. PMID: 33121436; PMCID: PMC7596976. (PubMed)

Examples

```
data("dataCancer")
# load the data set in the package
library("survival")
library("numDeriv")
library("survexp.fr")
data("simuData", package = "xhaz") # load the data sets 'simuData'

#define the levels of variable sex

# Esteve et al. model


fit.estv1 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
                  data = simuData, ratetable = survexp.us,
                  interval = c(0, NA, NA, NA, NA, NA, max(simuData$time_year)),
                  rmap = list(age = 'age', sex = 'sex', year = 'date'),
                  baseline = "constant", pophaz = "classic")


predict_est <- predict(object = fit.estv1,
                       new.data = simuData,
                       times.pts = c(seq(0, 4, 0.1)),
                       baseline = TRUE)

plot(predict_est, what = "survival",
     xlab = "time since diagnosis (year)",
     ylab = "net survival", ylim = c(0, 1))
data("dataCancer", package = "xhaz")   # load the data set in the package

fit.phBS <- xhaz(
      formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
      data = dataCancer, ratetable = survexp.fr::survexp.fr,
      interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
      rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
      baseline = "bsplines", pophaz  = "classic")


predict_mod1 <- predict(object = fit.phBS, new.data = dataCancer,
                        times.pts = c(seq(0, 10, 0.1)), baseline = FALSE)

old.par <- par(no.readonly = TRUE)
par(mfrow = c(2, 1))


plot(predict_mod1, what = "survival",
     xlab = "time since diagnosis (year)",
     ylab = "net survival", ylim = c(0, 1))

plot(predict_mod1, what = "hazard",
     xlab = "time since diagnosis (year)",
     ylab = "excess hazard")
```

```
par(old.par)
```

`predict.bsplines` *Predictions of excess hazard and net Survival from a* `bsplines` *object*

Description

Function to predict excess hazard and net survival based on an object of class `bsplines`. The function allows the predictions at several time points but not exceeding the maximum time of follow-up from the baseline model.

Usage

```
## S3 method for class 'bsplines'
predict(object, new.data = NULL, times.pts = NULL, baseline = TRUE, ...)
```

Arguments

`object`	an object of class `bsplines`
`new.data`	new.data where is covariates
`times.pts`	time in year scale to calculate the excess hazard. The default value is NULL. In this case, time variable must be provided in the new.data
`baseline`	default is survival baseline; put `baseline = FALSE` to estimate the net survival with covariates
`...`	additional arguments affecting the predictions of excess hazard and net survival

Value

An object of class predxhaz, which is a list of data.frame. Each element of the list contains the estimates of hazard and survival at a fixed time point. The return of this function can be used to produce graphics of excess hazard or net survival, when times.pts argument is provided. This object contains:

`times.pts`	the times value in year at which the excess hazard and or the net survival have been estimated
`hazard`	the excess hazard values based on the model of interest
`survival`	the net survival values based on the model of interest

Author(s)

Juste Goungounga, Robert Darlin Mba, Nathalie Graff\'eo and Roch Giorgi

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res. 2020 Jan;29(1):122-136. doi: 10.1177/0962280218823234. Epub 2019 Jan 23. PMID: 30674229. (PubMed)

Mba RD, Goungounga JA, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting inaccurate background mortality in excess hazard models through breakpoints. BMC Med Res Methodol. 2020 Oct 29;20(1):268. doi: 10.1186/s12874-020-01139-z. PMID: 33121436; PMCID: PMC7596976. (PubMed)

See Also

xhaz, print.bsplines, print.constant

Examples

```
library("survival")
library("numDeriv")
library("survexp.fr")
library("splines")
data("dataCancer", package = "xhaz")   # load the data set in the package

fit.phBS <- xhaz(
        formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
        data = dataCancer, ratetable = survexp.fr,
        interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
        rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
        baseline = "bsplines", pophaz = "classic")


print(fit.phBS)


predicted <- predict(object = fit.phBS,
                     new.data = dataCancer[1:10,],
                     times.pts = c(seq(0,10,1)),
                     baseline = TRUE)


#a list of predicted hazard and survival at different time points
print(predicted)


#predicted hazard and survival at time points 10 years
print(predicted[[10]])
```

predict.constant *Predictions of excess hazard and net Survival from an* constant *object*

Description

Function to predict excess hazard and net survival based on an object of class constant. The function allows the predictions at several time points but not exceeding the maximum time of follow-up from the baseline model.

Usage

```
## S3 method for class 'constant'
predict(object, new.data = NULL, times.pts = NULL, baseline = TRUE, ...)
```

Arguments

object	An object of class constant
new.data	new.data where is covariates
times.pts	time in year scale to calculate the excess hazard. The default value is NULL. In this case, time variable must be provided in the new.data
baseline	default is survival baseline; put baseline = FALSE to estimate the net survival with covariates
...	additional arguments affecting the predictions of excess hazard and net survival

Value

An object of class predxhaz. The return of this fonction can be used to produce graphics of excess hazard or net survival, when times.pts argument is provided. This object contains:

times.pts	the times value in year at which the excess hazard and or the net survival have been estimated
hazard	the excess hazard values based on the model of interest
survival	the net survival values based on the model of interest

Author(s)

Juste Goungounga, Robert Darlin Mba, Nathalie Graff\'eo and Roch Giorgi

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res. 2020 Jan;29(1):122-136. doi: 10.1177/0962280218823234. Epub 2019 Jan 23. PMID: 30674229. (PubMed)

Mba RD, Goungounga JA, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting inaccurate background mortality in excess hazard models through breakpoints. BMC Med Res Methodol. 2020 Oct 29;20(1):268. doi: 10.1186/s12874-020-01139-z. PMID: 33121436; PMCID: PMC7596976. (PubMed)

See Also

xhaz, print.bsplines, print.constant

Examples

```
# load the data set in the package
library("xhaz")
library("numDeriv")

# load the data sets 'simuData

data("simuData", package = "xhaz")

#define the levels of variable sex
levels(simuData$sex) <- c("male", "female")

# Esteve et al. model

set.seed(1980)
simuData2 <- simuData[sample(nrow(simuData), size = 500), ]

fit.estv2 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
                  data = simuData2,
                  ratetable = survexp.us,
                  interval = c(0, NA, NA, NA, NA, NA, 6),
                  rmap = list(age = 'age', sex = 'sex', year = 'date'),
                  baseline = "constant", pophaz = "classic")


predict_est <- predict(object = fit.estv2,
                       new.data = simuData2,
                       times.pts = c(seq(0, 4, 1)),
                       baseline = TRUE)
predict_est
```

`print.bsplines` *A print.bsplines Function used to print a object of class* `bsplines`

Description

This function present the estimated coefficients for the excess hazard baseline coefficient and for the covariate effects

Usage

```
## S3 method for class 'bsplines'
print(x, digits = max(options()$digits - 4, 3), ...)
```

Arguments

`x`	an object of class `bsplines`
`digits`	minimal number of significant digits.
`...`	additionnal parameters which can be used in the `print` function

Value

Estimated parameters of the model in different scales for interpretation purposes.

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res. 2020 Jan;29(1):122-136. doi: 10.1177/0962280218823234. Epub 2019 Jan 23. PMID: 30674229. (PubMed)

Mba RD, Goungounga JA, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting inaccurate background mortality in excess hazard models through breakpoints. BMC Med Res Methodol. 2020 Oct 29;20(1):268. doi: 10.1186/s12874-020-01139-z. PMID: 33121436; PMCID: PMC7596976. (PubMed)

See Also

`xhaz`, `plot.predxhaz`, `print.constant`

Examples

```
library("xhaz")
library("survival")
library("numDeriv")
library("survexp.fr")
library("splines")
data("dataCancer", package = "xhaz")   # load the data set in the package

fit.phBS <- xhaz(
              formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
              data = dataCancer, ratetable = survexp.fr,
              interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
              rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
              baseline = "bsplines", pophaz = "classic")

print(fit.phBS)
```

`print.constant` *A print.constant Function used to print a object of class constant*

Description

This function present the estimated coefficients for the excess hazard baseline coefficient and for the covariate effects

Usage

```
## S3 method for class 'constant'
print(x, ci_type = "lognormal", digits = max(options()$digits - 4, 3), ...)
```

Arguments

`x`	an object of class xhaz.constant
`ci_type`	method for confidence intervals calculation
`digits`	minimal number of significant digits.
`...`	additionnal parameters which can be used in the `print` function

Value

Estimated parameters of the model in different scales for interpretation purposes.

See Also

`xhaz`, `summary.constant`, `print.bsplines`

Examples

```
library("numDeriv")
library("survexp.fr")

data("simuData","rescaledData", "dataCancer")
# load the data sets 'simuData', 'rescaledData' and 'dataCancer'.

# Esteve et al. model: baseline excess hazard is a piecewise function
#                      linear and proportional effects for the covariates on
#                      baseline excess hazard.

set.seed(1980)
simuData2 <- simuData[sample(nrow(simuData), size = 500), ]

fit.estv2 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
                  data = simuData2,
                  ratetable = survexp.us,
                  interval = c(0, NA, NA, NA, NA, NA, 6),
                  rmap = list(age = 'age', sex = 'sex', year = 'date'),
                  baseline = "constant", pophaz = "classic")


print(fit.estv2)
```

`print.predxhaz` *A print.predxhaz Function used to print a object of class predxhaz*

Description

This function present the print of the predict function

Usage

```
## S3 method for class 'predxhaz'
print(x, ...)
```

Arguments

`x`	an object of class predxhaz
`...`	other parameters used for print function

Value

an object of class data.frame containing the following components:

`times.pts`	The time at which the estimations of excess hazard and net survival are predicted
`hazard`	the predicted excess hazard at the fixed times
`survival`	the predicted net survival at the fixed times

Examples

```
library("xhaz")
library("survexp.fr")
library("splines")

data("dataCancer", package = "xhaz")   # load the data set in the package

fit.phBS <- xhaz(
        formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
        data = dataCancer, ratetable = survexp.fr,
        interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
        rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
        baseline = "bsplines", pophaz = "classic")


fit.phBS


predicted <- predict(object = fit.phBS,
                     new.data = dataCancer[1:10,],
                     times.pts = c(seq(0,10,1)),
                     baseline = TRUE)



#a list of predicted hazard and survival at different time points
print(predicted)


#predicted hazard and survival at time points 10 years
print(predicted[[10]])
```

---

qbs *qbs function*

Description

a function indicating which covariates have a time-dependent effect in the formula.

Usage

```
qbs(x)
```

Arguments

x	a covariate to be considered in the xhaz formula with a time-dependant effect. Quadratic B-splines with two interior knots are used.

Value

No return value, called for side effects.

Examples

```
library("xhaz")
library("numDeriv")
library("survexp.fr")
library("splines")

fit.tdphBS <- xhaz(
              formula = Surv(obs_time_year, event) ~ ageCentre + qbs(immuno_trt),
              data = dataCancer, ratetable = survexp.fr,
              interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
              rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
              baseline = "bsplines", pophaz = "classic")

print(fit.tdphBS)
```

---

rescaledData *Simulated data with cause death information with non comparability bias in term of individuals expected hazard*

Description

Simulated data

Usage

```
data(rescaledData)
```

Format

This dataset contains the following variables:

time Follow-up time (months)

status Vital status

age Age at diagnosis

age.c Centred age

sex Sex(Female,Male)

hormTh Treatment group variable

date date of diagnosis

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Examples

```
data(rescaledData)
summary(rescaledData)
```

`simuData` *Simulated data with cause death information in long term follow-up setting without non comparability bias in term of individuals expected hazard*

Description

Simulated data

Usage

```
data(simuData)
```

Format

This dataset contains the following variables:

age Age at diagnosis

agec Centered age

sex Sex(Female,Male)

race Race

date date of diagnosis.

time Follow-up time (months)

time_year Follow-up time (years)

status Vital status

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res. 2020 Jan;29(1):122-136. doi: 10.1177/0962280218823234. Epub 2019 Jan 23. PMID: 30674229. (PubMed)

Examples

```
data(simuData)
summary(simuData)
```

`summary.bsplines` *A summary.bsplines Function used to print a object of class* `bsplines`

Description

This function present the estimated coefficients for the excess hazard baseline coefficient and for the covariate effects

Usage

```
## S3 method for class 'bsplines'
summary(object, ...)
```

Arguments

`object`	an object of class `bsplines`
`...`	additionnal parameters which can be used in the `summary` function

Value

Estimated parameters of the model in different scales for interpretation purposes.

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res. 2020 Jan;29(1):122-136. doi: 10.1177/0962280218823234. Epub 2019 Jan 23. PMID: 30674229. (PubMed)

Mba RD, Goungounga JA, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting inaccurate background mortality in excess hazard models through breakpoints. BMC Med Res Methodol. 2020 Oct 29;20(1):268. doi: 10.1186/s12874-020-01139-z. PMID: 33121436; PMCID: PMC7596976. (PubMed)

See Also

`xhaz`, `summary.bsplines`, `plot.bsplines`

Examples

```
library("xhaz")
library("survival")
library("numDeriv")
library("survexp.fr")
library("splines")

data("dataCancer", package = "xhaz")   # load the data set in the package

fit.phBS <- xhaz(
              formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
              data = dataCancer, ratetable = survexp.fr,
              interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
              rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
              baseline = "bsplines", pophaz = "classic")

summary(fit.phBS)
```

`summary.constant` *A summary.constant Function used to print a object of class* `xhaz.constant`

Description

This function present the estimated coefficients for the excess hazard baseline coefficient and for the covariate effects

Usage

```
## S3 method for class 'constant'
summary(object, ci_type = "lognormal", ...)
```

Arguments

`object`	an object of class xhaz.constant
`ci_type`	method for confidence intervals calculation
`...`	additionnal parameters which can be used in the `print` function

Value

Estimated parameters of the model in different scales for interpretation purposes.

See Also

`xhaz`, `summary.constant`, `print.bsplines`

Examples

```
library("xhaz")
library("numDeriv")
data("simuData", package = "xhaz")    # load the data sets 'simuData'

# Esteve et al. model: baseline excess hazard is a piecewise function
#                      linear and proportional effects for the covariates on
#                      baseline excess hazard.


set.seed(1980)
simuData2 <- simuData[sample(nrow(simuData), size = 500), ]

fit.estv2 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
                  data = simuData2,
                  ratetable = survexp.us,
                  interval = c(0, NA, NA, NA, NA, NA, 6),
                  rmap = list(age = 'age', sex = 'sex', year = 'date'),
                  baseline = "constant", pophaz = "classic")




summary(fit.estv2)
```

---

xhaz *xhaz function*

Description

Fits the excess hazard models proposed by Esteve et al. (1990) doi:10.1002/sim.4780090506, with the possibility to account for time dependent covariates. Fits also the non-proportional excess hazard model proposed by Giorgi et al. (2005) doi:10.1002/sim.2400. In addition, fits excess hazard models with possibility to rescale (Goungounga et al. (2019) doi:10.1186/s12874-019-0747-3) or to correct the background mortality with a proportional (Touraine et al. (2020) doi:10.1177/0962280218823234) or non-proportional (Mba et al. (2020) doi:10.1186/s12874-020-01139-z) effect.

Usage

```
xhaz(
  formula = formula(data),
  data = sys.parent(),
  ratetable,
  rmap = list(age = NULL, sex = NULL, year = NULL),
  baseline = c("constant", "bsplines"),
```

```
  pophaz = c("classic", "rescaled", "corrected"),
  only_ehazard = FALSE,
  add.rmap = NULL,
 add.rmap.cut = list(breakpoint = FALSE, cut = NA, probs = NULL, criterion = "BIC",
    print_stepwise = FALSE),
  interval,
  ratedata = sys.parent(),
  subset,
  na.action,
  init,
  control = list(eps = 1e-04, iter.max = 800, level = 0.95),
  optim = TRUE,
  scale = 365.2425,
  trace = 0,
  speedy = FALSE,
  nghq = 12,
  method = "L-BFGS-B",
  ...
)
```

Arguments

formula	a formula object of the function with the response on the left of a ~ operator and the terms on the right. The response must be a survival object as returned by the Surv function (time in first and status in second).
data	a data frame in which to interpret the variables named in the formula
ratetable	a rate table stratified by age, sex, year (if missing, ratedata is used)
rmap	a list that maps data set names to the ratetable names.
baseline	an argument to specify the baseline hazard: if it follows a piecewise constant, baseline = "constant" is used and corresponds to the baseline in Esteve et al. model; if the baseline follows a quadratic b-splines, baseline = "bsplines" is used, corresponding to the baseline excess hazard in Giorgi et al model.
pophaz	indicates three possibles arguments in character: classic or rescaled and corrected. If pophaz = "classic" chosen, fits the model that do not require to rescale or to correct the background mortality (i.e. the Esteve et al. model or Giorgi et al. model); if pophaz = "rescaled" or pophaz = "corrected" chosen, fits the models that require to rescale or to correct the background mortality.
only_ehazard	a boolean argument (by default, only_ehazard=FALSE). If only_ehazard = TRUE, pophaz = "classic" must be provided and the total value of the log-likelihood will not account for the cumulative population hazard.
add.rmap	character that indicates the name in character of the additional demographic variable from data to be used for correction of the life table, in particular when one is in the presence of an insufficiently stratified life table (see Touraine et al. model). This argument is not used if pophaz = "classic" or pophaz = "rescaled".
add.rmap.cut	a list containing arguments to specify the modeling strategy for breakpoint positions, which allows a non-proportional effect of the correction term acting on

	the background mortality. By default `list(breakpoint = FALSE)`, i.e. a proportional effect of the correction term acting on the background mortality is needed; in this case, all the other argument of the list are not working for the model specification; if `list(breakpoint = TRUE, cut = c(70))`, the chosen cut-point(s) is (are) the numeric value(s) proposed. If `list(breakpoint = TRUE, cut = NA)`, there is the same number of breakpoints as the number of NA, with their possible positions specified as here by `probs`, i.e. `list(breakpoint = TRUE, cut = NA, probs = seq(0, 1, 0.25))`. That corresponds to a numeric vector of probabilities with values between 0 and 1 as in `quantile` function. `criterion` is used to choose the best model, using the AIC or the BIC (the default criterion). If needed, all the fitted models are printed by the user by adding in the list `print_stepwise = FALSE`.
`interval`	a vector indicating either the location of the year-scale time intervals for models with piecewise constant function, or the location of the knots for models with B-splines functions for their baseline hazard (see the appropriate specification in `baseline` argument). The first component of the vector is 0, and the last one corresponds to the maximum time fellow-up of the study.
`ratedata`	a data frame of the hazards mortality in general population.
`subset`	an expression indicating which subset of the data should be used in the modeling. All observations are included by default
`na.action`	as in the `coxph` function, a missing-data filter function.
`init`	a list of initial values for the parameters to estimate. For each elements of the list, give the name of the covariate followed by the vector of the fixed initials values
`control`	a list of control values used to control the optimization process. In this list, `eps`, is a convergence criteria (by default, `eps=10^-4`), `iter.max` is the maximum number of iteration (by default, `iter.max=15`), and `level`, is the level used for the confidence intervals (by default, `level=0.95`).
`optim`	a Boolean argument (by default, `optim = FALSE`). If `optim = TRUE`), the maximization algorithm uses the `optim` function
`scale`	a numeric argument to specify whether the life table contains death rates per day (default `scale = 365.2425`) or death rates per year (`scale = 1`).
`trace`	a Boolean argument, if `trace = TRUE`), tracing information on the progress of the optimization is produced
`speedy`	a Boolean argument, if `speedy = TRUE`, optimization is done in a parallel mode
`nghq`	number of nodes and weights for Gaussian quadrature
`method`	corresponds to `optim` function argument.
`...`	other parameters used with the `xhaz` function

Details

Use the `Surv(time_start, time_stop, status)` notation for time dependent covariate with the appropriate organization of the data set (see the help page of the `Surv` function)

Only two interior knots are possible for the model with B-splines functions to fit the baseline (excess) hazard. Determination of the intervals might be user's defined or automatically computed according to the quantile of the distribution of deaths. Use NA for an automatic determination (for example, `interval = c(0, NA, NA, 5)`).

Value

An object of class `constant` or `bsplines`, according to the type of functions chosen to fit the baseline hazard of model (see details for argument `baseline`). This object is a list containing the following components:

`coefficients`	estimates found for the model
`varcov`	the variance-covariance matrix
`loglik`	for the Estève et al. model: the log-likelihood of the null model, i.e without covariate, and the log-likelihood of the full model, i.e with all the covariates declared in the formula; for the Giorgi et al. model: the log-likelihood of the full model
`cov.test`	for the Esteve et al.model: the log-likelihood of the null model, i.e without covariate, and the log-likelihood of the full model, i.e with all the covariates declared in the formula; for the Giorgi et al. model: the log-likelihood of the full model
`message`	a character string returned by the optimizer see details in `optim` help page
`convergence`	an integer code as in `optim` when `"L-BFGS-B"` method is used.
`n`	the number of individuals in the dataset
`n.events`	the number of events in the dataset. Event are considered as death whatever the cause
`level`	the confidence level used
`interval`	the intervals used to split time for piecewise baseline excess hazard, or knots positions for Bsplines baseline
`terms`	the representation of the terms in the model
`call`	the function `call` based on model
`pophaz`	the assumption considered for the life table used in the excess hazard model
`add.rmap`	the additional variable for which the life table is not stratified
`ehazardInt`	the cumulative hazard of each individuals calculated from the ratetable used in the model
`ehazard`	the individual expected hazard values from the ratetable used to fit the model
`data`	the dataset used to run the model
`time_elapsed`	the time to run the model

Note

`time` is OBLIGATORY in YEARS.

Author(s)

Juste Goungounga, Darlin Robert Mba, Nathalie Graffeo, Roch Giorgi

References

Goungounga JA, Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res Methodol. 2019 May 16;19(1):104. doi: 10.1186/s12874-019-0747-3. PMID: 31096911; PMCID: PMC6524224. (PubMed)

Touraine C, Graff\'eo N, Giorgi R; CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res. 2020 Jan;29(1):122-136. doi: 10.1177/0962280218823234. Epub 2019 Jan 23. PMID: 30674229. (PubMed)

Mba RD, Goungounga JA, Graff\'eo N, Giorgi R; CENSUR working survival group. Correcting inaccurate background mortality in excess hazard models through breakpoints. BMC Med Res Methodol. 2020 Oct 29;20(1):268. doi: 10.1186/s12874-020-01139-z. PMID: 33121436; PMCID: PMC7596976. (PubMed)

Giorgi R, Abrahamowicz M, Quantin C, Bolard P, Esteve J, Gouvernet J, Faivre J. A relative survival regression model using B-spline functions to model non-proportional hazards. Statistics in Medicine 2003; 22: 2767-84. (PubMed)

Examples

```
library("numDeriv")
library("survexp.fr")
library("splines")
library("statmod")
data("simuData","rescaledData", "dataCancer")
# load the data sets 'simuData', 'rescaledData' and 'dataCancer'.

# Esteve et al. model: baseline excess hazard is a piecewise function
#                      linear and proportional effects for the covariates on
#                      baseline excess hazard.


fit.estv1 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
                  data = simuData,
                  ratetable = survexp.us,
                  interval = c(0, NA, NA, NA, NA, NA, 6),
                  rmap = list(age = 'age', sex = 'sex', year = 'date'),
                  baseline = "constant", pophaz = "classic")


fit.estv1


# Touraine et al. model: baseline excess hazard is a piecewise function
#                        with a linear and proportional effects for the
#                        covariates on the baseline excess hazard.
# An additionnal cavariate (here race) missing in the life table is
```

```
# considered by the model.


fit.corrected1 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
                       data = simuData,
                       ratetable = survexp.us,
                       interval = c(0, NA, NA, NA, NA, NA, 6),
                       rmap = list(age = 'age', sex = 'sex', year = 'date'),
                       baseline = "constant", pophaz = "corrected",
                       add.rmap = "race")



fit.corrected1

# extension of Touraine et al model: baseline excess hazard is a piecewise
# constant function with a linear and proportional effects for the covariates
# on the baseline excess hazard.

# An additionnal cavariate (here race) missing in the life table is
# considered by the model with a breakpoint at 75 years

fit.corrected2 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
                       data = simuData,
                       ratetable = survexp.us,
                       interval = c(0, NA, NA, NA, NA, NA, 6),
                       rmap = list(age = 'age', sex = 'sex', year = 'date'),
                       baseline = "constant", pophaz = "corrected",
                       add.rmap = "race",
                        add.rmap.cut = list(breakpoint = TRUE, cut = 75))



fit.corrected2


#Giorgi et al model: baseline excess hazard is a quadratic Bsplines
#                    function with two interior knots and allow here a
#                    linear and proportional effects for the covariates on
#                    baseline excess hazard.


fitphBS <- xhaz(formula = Surv(time_year, status) ~ agec + race,
                data = simuData,
                ratetable = survexp.us,
                interval = c(0, NA, NA, 6),
                rmap = list(age = 'age', sex = 'sex', year = 'date'),
                baseline = "bsplines", pophaz = "classic")

fitphBS
```

```
# Application on `dataCancer`.
#Giorgi et al model: baseline excess hazard is a quadratic Bspline
#                    function with two interior knots and allow here a
#                    linear and proportional effect for the variable
#                    "immuno_trt" plus a non-proportional effect
#                    for the variable "ageCentre" on baseline excess hazard.


fittdphBS <- xhaz(formula = Surv(obs_time_year, event) ~ qbs(ageCentre) + immuno_trt,
                  data = dataCancer,
                  ratetable = survexp.fr,
                  interval = c(0, 0.5, 12, 15),
                  rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
                  baseline = "bsplines", pophaz = "classic")

fittdphBS




# Application on `rescaledData`.
# rescaled model: baseline excess hazard is a piecewise function with a
# linear and proportional effects for the covariates on baseline excess hazard.

# A scale parameter on the expected mortality of general population is
# considered to account for the non-comparability source of bias.

rescaledData$timeyear <- rescaledData$time/12
rescaledData$agecr <- scale(rescaledData$age, TRUE, TRUE)

fit.res <- xhaz(formula = Surv(timeyear, status) ~ agecr + hormTh,
                data = rescaledData,
                ratetable = survexp.fr,
                interval = c(0, NA, NA, NA, NA, NA, max(rescaledData$timeyear)),
                rmap = list(age = 'age', sex = 'sex', year = 'date'),
                baseline = "constant", pophaz = "rescaled")

 fit.res
```

Index

∗ **anova.bsplines**
anova.bsplines, 2
∗ **anova.constant**
anova.constant, 4
∗ **anova.mexhazLT**
anova.mexhazLT, 6
∗ **datasets**
breast, 8
ccr.mevents, 9
dataCancer, 10
rescaledData, 33
simuData, 34
∗ **duplicate**
duplicate, 11
∗ **mexhazLT**
mexhazLT, 15
∗ **plot.bsplines**
plot.bsplines, 20
∗ **plot.predxhaz**
plot.predxhaz, 23
∗ **predict.bsplines**
predict.bsplines, 25
∗ **predict.constant**
predict.constant, 27
∗ **print.bsplines**
print.bsplines, 29
∗ **print.constant**
print.constant, 30
∗ **print.predxhaz**
print.predxhaz, 31
∗ **qbs**
qbs, 32
∗ **summary.bsplines**
summary.bsplines, 35
∗ **summary.constant**
summary.constant, 36
∗ **xhaz**
xhaz, 37

AIC.mexhazLT, *7*
anova.bsplines, 2
anova.constant, 4
anova.mexhazLT, 6

breast, 8

ccr.mevents, 9

dataCancer, 10
duplicate, 11

exphaz, 13

mexhazAlpha *(*mexhazLT*)*, 15
mexhazLT, *7*, 15

plot.bsplines, 20, *35*
plot.predxhaz, 23, *29*
predict.bsplines, 25
predict.constant, 27
print.bsplines, *26*, *28*, 29, *30*, *36*
print.constant, *3*, *5*, *26*, *28*, *29*, 30
print.predxhaz, 31

qbs, 32

rescaledData, 33

simuData, 34
summary.bsplines, *3*, *5*, *35*, 35
summary.constant, *30*, *36*, 36

xhaz, *3*, *5*, *7*, *26*, *28*–*30*, *35*, *36*, 37