# Package 'SimHaz'

December 20, 2015

Title Simulated Survival and Hazard Analysis for Time-Dependent Exposure

**Version** 0.2 **Date** 2015-12-09

Author Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee	
Maintainer Nusrat Rabbee <rabbee@berkeley.edu></rabbee@berkeley.edu>	
<b>Depends</b> R (>= 3.1.1)	
Imports survival	
<ul> <li>Description Generate power for the Cox proportional hazards model by simulating survival events data with time dependent exposure status for subjects. A dichotomous exposure vable is considered with a single transition from unexposed to exposed status during the subject's time on study.</li> <li>License GPL (&gt;= 2)</li> </ul>	ari-
<pre>URL http://www.stat.berkeley.edu/~rabbee/research_webpage.htm</pre>	
<pre>BugReports https://github.com/rabbeelab/SimHaz/issues</pre>	
Suggests knitr, xtable, MASS, powerSurvEpi, vcd	
VignetteBuilder knitr	
R topics documented:	
SimHaz-package getpower.exp.matching getpower.exp.matching.opt getpower.method1 getpower.method2 getpower.multicenter plot_power plot_simuData tdSim.exp.matching tdSim.method1 tdSim.method2 tdSim.multicenter	
Index	26
1	

SimHaz-package

Simulated Survival and Hazard Analysis for time-dependent

### **Description**

This package generates power for the Cox proportional hazards model by simulating survival events data with time dependent exposure status for subjects. A dichotomous exposure variable is considered with a single transition from unexposed to exposed status during the subject time's in the study.

#### **Details**

Package: SimHaz
Type: Package
Version: 0.2

Date: 2015-12-09 License: GPL-2

Depends: R (>= 3.1.1) Imports: survival

#### Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

### **Examples**

```
# Simulate a dataset of 600 subjects with time-dependent exposure without
# considering minimum follow-up time or minimum post-exposure follow-up time.
# Specifically, set the duration of the study to be 24 months; the median time to
# event for control group to be 24 months; exposure effect to be 0.3; median time
# to censoring to be 14 months; and exposure proportion to be 20%.

df <- tdSim.method1(N = 600, duration = 24, lambda = log(2)/24, rho = 1,
    beta = 0.3, rateC = log(2)/14, exp.prop = 0.2,
    prop.fullexp = 0, maxrelexptime = 1, min.futime = 0,
    min.postexp.futime = 0)

# We recommend setting nSim to at least 500. It is set to 10 in the example to
# reduce run time for CRAN submission.

ret <- getpower.method1(nSim = 10, N = 600, b = 0.3, exp.prop = 0.2,
    type = "td", scenario = "scenario 1", maxrelexptime = 1/6, min.futime = 4,
    min.postexp.futime = 4, output.fn = "output.csv")</pre>
```

getpower.exp.matching Calculate power and betahat for Cox proportional Hazard model in the case of exposure matching.

#### **Description**

This functions runs nSim (Number of simulations, specified by the user) Monte Carlo simulations, each time it calling tdSim.method1 internally. This function is used in the case of exposure matching where exposed subject is matched with non-exposed subject according to the user-input ratio. Users specify the which Cox model approach to analyze the data with clusters(matching sets in this case).

The function returns a data frame of scenario-specific parameters (including statistical power) and appends the output to a file with file name specified in the input parameters list. The user also have an option whether to plot an incidence plot or not.

#### Usage

```
getpower.exp.matching(nSim, N_match, duration = 24, med.TTE.Control = 24,
rho = 1, med.TimeToCensor = 14, beta, matching.ratio, type, scenario,
method, prop.fullexp = 0, maxrelexptime = 1, min.futime = 0,
    min.postexp.futime = 0, output.fn, simu.plot = FALSE)
```

### **Arguments**

nSim Number of simulations.

N\_match Number of subjects to be screened.

duration Length of the study in months; the default value is 24 (months).

med.TTE.Control

Median time to event for control group; the default value is 24 (months).

rho Shape parameter of the Weibull distribution. Default is 1, which will generate

survival times by using the exponential distribution.

med.TimeToCensor

Median time to censoring for all subjects. The default value is 14 (months).

A numeric value that represents the exposure effect, which is the regression

coefficient (log hazard ratio) that represent the magnitude of the relationship

between the exposure covariate and the risk of an event.

matching.ratio Matching ratio used in exposure matching.For 1:1, the user should specify 1.in-

put value 3 corresponds to 1:3 (exposed: unexposed). input value 0.25 corre-

sponds to 4:1 (ie. 1:0.25)

type A text string indicating the what type of dataset is of interest. Either one of

"fixed" or "td" should be inputted.

method Specified which Cox model approach to analyze the data with clusters. Can be

one of the following: 'frailty', 'fixed effects', 'strata', 'Model with Indepen-

dence Assumption'.

scenario Any text string inputted by the user as an option to name a scenario that is being

simulated. The use can simply put " " if he/she decide to not name the scenario.

prop.fullexp A numeric value in interval [0, 1) that represents the proportion of exposed

subjects that are fully exposed from the beginning to the end of the study. The default value is 0, which means all exposed subjects have an exposure status transition at some point during the study. Only applies when type is "td". When

type is "td", the value is automatically 1.

maxrelexptime A numeric value in interval (0, 1] that represents the maximum relative expo-

sure time. Suppose this value is p, the exposure time for each subject is then uniformly distributed from 0 to p\*subject's time in the study. The default value is 1, which means all exposed subjects have an exposure status transition at any

point during the time in study.

min.futime

A numeric value that represents minimum follow-up time (in months). The default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study.

min.postexp.futime

A numeric value that represents minimum post-exposure follow-up time (in months). The defaul value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure.

 $\hbox{output. fn} \qquad \quad A. csv \ filename \ to \ write \ in \ the \ output. \ If \ the \ filename \ does \ not \ exist, \ the \ function$ 

will create a new .csv file for the output.

simu.plot A logical value indicating whether or not to output an incidence plot. The default

value is FALSE.

#### **Details**

The function calculates power based on the Cox regression model, which calls the coxph function from the survival library using the the simulated data from tdSim.method1.

#### Value

A data.frame object with columns corresponding to

i\_scenarioi\_typeScenario name specified by the userDataset type specified by the user

i\_N\_match Number of matching set specified by the user

i\_matching.ratio

Matching ratio used in exposure matching specified by the user

i\_min.futime Minimum follow-up time to be considered, specified by the user

i\_min.postexp.futime

Minimum post-exposure follow-up time to be considered, specified by the user

i\_lambda Value of the scale parameter of the Weibull distribution to generate survival

times. Calculated from median time to event for control group, which is speci-

fied by the user.

i\_rho User-specified value of the shape parameter of the Weibull distribution to gen-

erate survival times

i\_rateC Rate of the exponential distribution to generate censoring times. Calculated

from median time to censoring, which is specified by the user. i\_beta Input

value of regression coefficient (log hazard ratio).

i\_beta Value of the input beta

N\_eff Simulated number of evaluable subjects, which is the resulting number of sub-

jects with or without considering minimum follow-up time and/or minimum

post-exposure follow-up time.

N\_effexp\_p Simulated proportion of exposed subjects with or without considering minimum

follow-up time and/or minimum post-exposure follow-up time.

bhat Simulated value of regression coefficient (log hazard ratio)

HR Simulated value of hazard ratio
d Simulated number of events in total

with time-

d_c	Simulated number of events in control group
d_exp	Simulated number of events in exposed group
mst_c	Simulated median survival time in control group
mst_exp	Simulated median survival time in exposed group
pow	Simulated statistical power from the Cox regression model on data of dependent exposure

Variance of the betahat from the simulations

### Author(s)

variance

Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

#### References

Therneau T (2015). A Package for Survival Analysis in S. version 2.38, http://CRAN.R-project.org/package=survival

#### **Examples**

```
# We recommend setting nSim to at least 500. It is set to 10 in the example to
# reduce run time for CRAN submission.

# Run 10 simulations. Each time simulate a dataset of 100 matches
# time-dependent exposure with both minimum follow-up time (4 months) and
# minimum post-exposure follow-up time (4 months) imposed. Also consider a
# quick exposure after entering the study for each exposed subject. Set the
# maximum relative exposure time to be 1/6.

ret = getpower.exp.matching(nSim=10, N_match=100,duration=24,
    med.TTE.Control=24,rho=1, med.TimeToCensor=14, beta=0.7,
    matching.ratio=3,type="td", scenario="exposure_matching",
    method="marginal", prop.fullexp=0,maxrelexptime=1,
    min.futime=0, min.postexp.futime=0, output.fn="result_matching",
    simu.plot=FALSE)
```

```
getpower.exp.matching.opt
```

Calculate betahat bias and variance for different matching ratios in the case of exposure matching in the Cox proportional Hazard model

### **Description**

This function is used to explore what matching ratios should be used in the case of exposure matching. It simulate a large population (100000 subjects) in order to estimate a betahat value of that population. Then in repeated simulation, a subset of that population is drawn and exposure matching is done according to a user-specified list of ratios in order to compare the bias in the betahat estimate from the betahat value estimated from the large populations well as the variance of the betahat estimate for each ratio.

### Usage

```
getpower.exp.matching.opt(nSim, N, ratios=c(1,0.25,0.333,0.5,2,3,4,5),
    duration=24, med.TTE.Control=24,rho=1,med.TimeToCensor=14,beta,
exp.prop,type,scenario, method, prop.fullexp=0,maxrelexptime=1,
min.futime=0,min.postexp.futime=0, output.fn=NULL,simu.plot=FALSE)
```

#### **Arguments**

nSim Number of simulations.

N Number of subjects to be screened.

ratios Specification for matching ratio as a list of numbers. For 1:1, the user should

specify 1.input value 3 corresponds to 1:3 (exposed: unexposed). input value 0.25 corresponds to 4:1 (ie. 1:0.25). A list of c(1,0.25, 2) corresponds to the

following matching ratios: 1:1, 4:1, 1:2

duration Length of the study in months; the default value is 24 (months).

med.TTE.Control

Median time to event for control group; the default value is 24 (months).

rho Shape parameter of the Weibull distribution. Default is 1, which will generate

survival times by using the exponential distribution.

med.TimeToCensor

Median time to censoring for all subjects. The default value is 14 (months).

A numeric value that represents the exposure effect, which is the regression

coefficient (log hazard ratio) that represent the magnitude of the relationship

between the exposure covariate and the risk of an event.

exp.prop test

type A text string indicating the what type of dataset is of interest. Either one of

"fixed" or "td" should be inputted.

scenario Any text string inputted by the user as an option to name a scenario that is being

simulated. The use can simply put " " if he/she decide to not name the scenario.

method Specified which Cox model approach to analyze the data with clusters(matching

sets in this case). Can be one of the following: 'frailty', 'fixed effects', 'strata',

'Model with Independence Assumption'.

prop.fullexp A numeric value in interval [0, 1) that represents the proportion of exposed

subjects that are fully exposed from the beginning to the end of the study. The default value is 0, which means all exposed subjects have an exposure status transition at some point during the study. Only applies when type is "td". When

type is "td", the value is automatically 1.

maxrelexptime A numeric value in interval (0, 1] that represents the maximum relative expo-

sure time. Suppose this value is p, the exposure time for each subject is then uniformly distributed from 0 to p\*subject's time in the study. The default value is 1, which means all exposed subjects have an exposure status transition at any

point during the time in study.

min.futime A numeric value that represents minimum follow-up time (in months). The

default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a

short amount of time in the study.

min.postexp.futime

A numeric value that represents minimum post-exposure follow-up time (in months). The defaul value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure.

output. fn A .csv filename to write in the output. If the filename does not exist, the function

will create a new .csv file for the output.

simu.plot A logical value indicating whether or not to output an incidence plot. The default

value is FALSE.

#### **Details**

The function calculates power based on the Cox regression model, which calls the coxph function from the survival library using the the simulated data from tdSim.method1.

#### Value

A data frame object with columns corresponding to

i\_scenarioi\_typeScenario name specified by the userDataset type specified by the user

i\_min.futime Minimum follow-up time to be considered, specified by the user

i\_min.postexp.futime

Minimum post-exposure follow-up time to be considered, specified by the user

i\_lambda Value of the scale parameter of the Weibull distribution to generate survival

times. Calculated from median time to event for control group, which is speci-

fied by the user.

i\_rho User-specified value of the shape parameter of the Weibull distribution to gen-

erate survival times

i\_rateC Rate of the exponential distribution to generate censoring times. Calculated

from median time to censoring, which is specified by the user. i\_beta Input

value of regression coefficient (log hazard ratio).

i\_beta value of the input beta
 N\_match Number of matching sets
 N\_exposed Number of exposed subjects
 N\_unexposed Number of unexposed subjects

bhat Simulated value of regression coefficient (log hazard ratio)

pow Simulated statistical power from the Cox regression model on data with time-

dependent exposure

actual\_beta Value of betahat based on the population of 100000 subjects.

bias Value of betahat - actual\_betahat(approximated from the big population)

variance Variance of the betahat from the simulations

RE1 The variance of betahat of the ratio in that row divided by variance of betahat

for the 1:1 matching

RE2 The variance of betahat of the ratio in that row divided by variance of the betahat

for the closest ratio based on the exposure proportion in the population

8 getpower.method1

#### Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

#### References

Therneau T (2015). A Package for Survival Analysis in S. version 2.38, http://CRAN.R-project.org/package=survival

#### **Examples**

```
# We recommend setting nSim to at least 500. It is set to 5 in the example to
# reduce run time for CRAN submission.

# Run 5 simulations. Each time simulate a dataset of 400 subjects with
# time-dependent exposure with both minimum follow-up time (4 months) and
# minimum post-exposure follow-up time (4 months) imposed. Also consider a
# quick exposure after entering the study for each exposed subject. Set the
# maximum relative exposure time to be 1/6.

ret = getpower.exp.matching.opt(nSim=5, N=400, ratios=c(1), duration=24,
    med.TTE.Control=24, rho=1,med.TimeToCensor=14,beta=0.5, exp.prop=0.3,
    type="td",scenario="opt_exp_matching", method="marginal",prop.fullexp=0,
    maxrelexptime=1,min.futime=0,min.postexp.futime=0, output.fn="opt_matching",
    simu.plot=FALSE)
```

getpower.method1

Calculate power for the Cox proportional hazard model with timedependent exposure using method 1

### Description

This functions runs nSim (Number of simulations, specified by the user) Monte Carlo simulations, each time it calling tdSim.method1 internally. The function returns a data frame of scenario-specific parameters (including statistical power) and appends the output to a file with file name specified in the input parameters list. The user also have an option whether to plot an incidence plot or not.

### Usage

```
getpower.method1(nSim, N, duration = 24, med.TTE.Control = 24, rho = 1,
    med.TimeToCensor = 14, beta, exp.prop, type, scenario, prop.fullexp = 0,
    maxrelexptime = 1, min.futime = 0, min.postexp.futime = 0, output.fn,
    simu.plot = FALSE)
```

#### **Arguments**

nSim Number of simulations.

N Number of subjects to be screened.

duration Length of the study in months; the default value is 24 (months).

getpower.method1 9

med.TTE.Control

Median time to event for control group; the default value is 24 (months).

rho Shape parameter of the Weibull distribution. Default is 1, which will generate

survival times by using the exponential distribution.

med.TimeToCensor

Median time to censoring for all subjects. The default value is 14 (months).

A numeric value that represents the exposure effect, which is the regression

coefficient (log hazard ratio) that represent the magnitude of the relationship

between the exposure covariate and the risk of an event.

exp.prop A numeric value between 0 and 1 (not include 0 and 1) that represents the pro-

portion of subjects that are assigned with an exposure.

type A text string indicating the what type of dataset is of interest. Either one of

"fixed" or "td" should be inputted.

scenario Any text string inputted by the user as an option to name a scenario that is being

simulated. The use can simply put " " if he/she decide to not name the scenario.

prop.fullexp A numeric value in interval [0, 1) that represents the proportion of exposed

subjects that are fully exposed from the beginning to the end of the study. The default value is 0, which means all exposed subjects have an exposure status transition at some point during the study. Only applies when type is "td". When

type is "td", the value is automatically 1.

maxrelexptime A numeric value in interval (0, 1] that represents the maximum relative expo-

sure time. Suppose this value is p, the exposure time for each subject is then uniformly distributed from 0 to p\*subject's time in the study. The default value is 1, which means all exposed subjects have an exposure status transition at any

point during the time in study.

min.futime A numeric value that represents minimum follow-up time (in months). The

default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a

short amount of time in the study.

min.postexp.futime

A numeric value that represents minimum post-exposure follow-up time (in months). The defaul value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure.

output. fn A .csv filename to write in the output. If the filename does not exist, the function

will create a new .csv file for the output.

simu. plot A logical value indicating whether or not to output an incidence plot. The default

value is FALSE.

### Details

The function calculates power based on the Cox regression model, which calls the coxph function from the survival library using the the simulated data from tdSim.method1.

#### Value

A data frame object with columns corresponding to

i\_scenarioi\_typeScenario name specified by the userDataset type specified by the user

10 getpower.method1

i_N	Number of subjects to be screened, specified by the user
i_min.futime	Minimum follow-up time to be considered, specified by the user
i_min.postexp.	futime
	Minimum post-exposure follow-up time to be considered, specified by the user
i_exp.prop	Exposure rate specified by the user
i_lambda	Value of the scale parameter of the Weibull distribution to generate survival times. Calculated from median time to event for control group, which is specified by the user.
i_rho	User-specified value of the shape parameter of the Weibull distribution to generate survival times
i_rateC	Rate of the exponential distribution to generate censoring times. Calculated from median time to censoring, which is specified by the user. i_beta Input value of regression coefficient (log hazard ratio).
N_eff	Simulated number of evaluable subjects, which is the resulting number of subjects with or without considering minimum follow-up time and/or minimum post-exposure follow-up time.
N_effexp_p	Simulated proportion of exposed subjects with or without considering minimum follow-up time and/or minimum post-exposure follow-up time.
bhat	Simulated value of regression coefficient (log hazard ratio)
HR	Simulated value of hazard ratio
d	Simulated number of events in total
d_c	Simulated number of events in control group
d_exp	Simulated number of events in exposed group
mst_c	Simulated median survival time in control group
mst_exp	Simulated median survival time in exposed group
pow	Simulated statistical power from the Cox regression model on data with time-dependent exposure

### Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

#### References

Therneau T (2015). A Package for Survival Analysis in S. version 2.38, http://CRAN.R-project.org/package=survival

### **Examples**

```
\# We recommend setting nSim to at least 500. It is set to 10 in the example to \# reduce run time for CRAN submission.
```

```
# Run 10 simulations. Each time simulate a dataset of 600 subjects with # time-dependent exposure with both minimum follow-up time (4 months) and # minimum post-exposure follow-up time (4 months) imposed. Also consider a # quick exposure after entering the study for each exposed subject. Set the # maximum relative exposure time to be 1/6.
```

getpower.method2 11

```
# Set the duration of the study to be 24 months; the median time to event for
# control group to be 24 months; exposure effect to be 0.3; median time to
# censoring to be 14 months; and exposure proportion to be 20%.

ret <- getpower.method1(nSim = 10, N = 600, b = 0.3, exp.prop = 0.2,
    type = "td", scenario = " ", maxrelexptime = 1/6, min.futime = 4,
    min.postexp.futime = 4, output.fn = "output.csv")</pre>
```

getpower.method2

Calculate power for the Cox proportional hazard model with timedependent exposure using method 2

### **Description**

This functions runs nSim (Number of simulations, specified by the user) Monte Carlo simulations, each time calling tdSim.method2 internally. The function returns a data frame of scenario-specific input parameters- and also output statistical power. The user has the option to append the output to a file with file name specified in the input parameters list.

### Usage

```
getpower.method2(nSim = 500, N, duration = 24, scenario, lambda12,
lambda23 = NULL, lambda13, HR = NULL, exp.prop, rateC, min.futime,
min.postexp.futime, output.fn, simu.plot = FALSE)
```

### **Arguments**

nSim	Number of simulations.
N	Number of subjects to be screened.
duration	Length of the study in months; the default value is 24 (months).
scenario	Any text string inputted by the user as an option to name a scenario that is being simulated. The use can simply put " " if he/she decides to not name the scenario.
lambda12	Lambda12 parameter to control time to exposure.
lambda23	Lambda23 parameter to control time to event after exposure.
lambda13	Lambda13 parameter to control time to event in the control group.
HR	Hazard Ratio. This input is optional. If HR is set and lambda23 is not set, $lambda23 = lambda13*HR$ .
exp.prop	A numeric value between 0 and 1 (not include 0 and 1) that represents the proportion of subjects that are assigned with an exposure.
rateC	Rate of the exponential distribution to generate censoring times.
min.futime	A numeric value that represents minimum follow-up time (in months). The default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study.

12 getpower.method2

min.postexp.futime

A numeric value that represents minimum post-exposure follow-up time (in months). The defaul value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure.

output. fn A .csv filename to write in the output. If the filename does not exist, the function

will create a new .csv file for the output.

simu. plot A logical value indicating whether or not to output an incidence plot. The default

value is FALSE.

#### **Details**

The function calculates power based on the Cox regression model, which calls the coxph function from the survival library using the the simulated data from tdSim.method2.

#### Value

A data frame object with columns corresponding to

i\_scenario
 i\_N
 Number of subjects needs to be screened, specified by the user
 i\_min.futime
 Minimum follow-up time to be considered, specified by the user

i\_min.postexp.futime

Minimum post-exposure follow-up time to be considered, specified by the user

i\_exp.prop Exposure rate specified by the user

i\_lambda12 Lambda12 parameter to control time to exposure

i\_lambda23 Lambda23 parameter to control time to event after exposure

i\_lambda13 Lambda13 parameter to control time to event in the control group

i\_rateC Rate of the exponential distribution to generate censoring times. Calculated

from median time to censoring, which is specified by the user. i\_beta Input

value of regression coefficient (log hazard ratio)

N\_eff Simulated number of evaluable subjects, which is the resulting number of sub-

jects with or without considering minimum follow-up time and/or minimum

post-exposure follow-up time

N\_effexp\_p Simulated proportion of exposed subjects with or without considering minimum

follow-up time and/or minimum post-exposure follow-up time

bhat Simulated value of regression coefficient (log hazard ratio)

HR Simulated value of hazard ratio
d Simulated number of events in total

d\_c
 d\_exp
 Simulated number of events in control group
 mst\_c
 Simulated median survival time in control group
 mst\_exp
 Simulated median survival time in exposed group

pow Simulated statistical power from the Cox regression model on data with time-

dependent exposure

getpower.multicenter 13

#### Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

#### **Examples**

```
# We recommend setting nSim to at least 500. It is set to 10 in the example to
# reduce run time for CRAN submission.

# Run 10 simulations. Each time simulate a dataset of 600 subjects

ret <- getpower.method2(nSim=10, N=600, duration=24, scenario="test",
    lambda12=1.3, lambda23=0.04, lambda13=0.03, HR=NULL,exp.prop=0.2, rateC=0.05,
    min.futime=4, min.postexp.futime=4,output.fn="database.csv", simu.plot=FALSE)</pre>
```

getpower.multicenter

Calculate power for the Cox proportional hazard model with timedependent exposure and multiple centers using method1

### **Description**

This functions runs nSim (Number of simulations, specified by the user) Monte Carlo simulations, each time calling tdSim.multicenter internally. The function returns a data frame of scenario-specific input parameters- and also output statistical power. The user has the option to append the output to a file with file name specified in the input parameters list.

#### Usage

### Arguments

nSim Number of simulations.

N Number of subjects to be screened.

duration Length of the study in months; the default value is 24 (months).

rho Shape parameter of the Weibull distribution. Default is 1, which will generate

survival times by using the exponential distribution.

beta A numeric value that represents the exposure effect, which is the regression

coefficient (log hazard ratio) that represent the magnitude of the relationship

between the exposure covariate and the risk of an event.

med.TimeToCensor

Median time to censoring for all subjects. The default value is 14 (months).

df

A user-specified n by 4 clustering data frame with columns corresponding to cat\_id (category id, which is the physician site id. It can be either text strings or integers), center.size (number of subjects within each center), cat\_exp.prop (proportion of exposed subjects in each center), and med.TTE.Control (median time to control event for each center). n rows corresponds to n different centers

14 getpower.multicenter

dist The distribution of the center effect across centers. Default is NULL. If dist= 'gamma', then a random frailty effect from a gamma distribution with scale 0.5

and shape 2 is used.

method Specified which Cox model approach to analyze the data with centers. Can be

one of the following: 'frailty', 'fixed effects', 'strata', 'Model with Indepen-

dence Assumption'.

type A text string indicating the what type of dataset is of interest. Either one of

"fixed" or "td" should be inputted.

scenario Any text string inputted by the user as an option to name a scenario that is being

simulated. The use can simply put " " if he/she decide to not name the scenario.

prop.fullexp A numeric value in interval [0, 1) that represents the proportion of exposed

subjects that are fully exposed from the beginning to the end of the study. The default value is 0, which means all exposed subjects have an exposure status transition at some point during the study. Only applies when type is "td". When

type is "td", the value is automatically 1.

maxrelexptime A numeric value in interval (0, 1] that represents the maximum relative expo-

sure time. Suppose this value is p, the exposure time for each subject is then uniformly distributed from 0 to p\*subject's time in the study. The default value is 1, which means all exposed subjects have an exposure status transition at any

point during the time in study.

min.futime A numeric value that represents minimum follow-up time (in months). The

default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a

short amount of time in the study.

min.postexp.futime

A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure.

output. fn A .csv filename to write in the output. If the filename does not exist, the function

will create a new .csv file for the output.

simu.plot A logical value indicating whether or not to output an incidence plot. The default

value is FALSE.

#### **Details**

The function calculates power based on the Cox regression model, which calls the coxph function from the survival library using the the simulated data from tdSim.multicenter.

### Value

A data.frame object with columns corresponding to

i\_scenarioi\_scenario name specified by the useri\_typeDataset type specified by the user

i\_N Number of subjects to be screened, specified by the user

i\_min.futime Minimum follow-up time to be considered, specified by the user

i\_min.postexp.futime

Minimum post-exposure follow-up time to be considered, specified by the user

getpower.multicenter 15

i_exp.prop	Exposure rate specified by the user
i_lambda	Value of the scale parameter of the Weibull distribution to generate survival times. Calculated from median time to event for control group, which is specified by the user.
i_rho	User-specified value of the shape parameter of the Weibull distribution to generate survival times
i_rateC	Rate of the exponential distribution to generate censoring times. Calculated from median time to censoring, which is specified by the user. i_beta Input value of regression coefficient (log hazard ratio).
N_eff	Simulated number of evaluable subjects, which is the resulting number of subjects with or without considering minimum follow-up time and/or minimum post-exposure follow-up time.
N_effexp_p	Simulated proportion of exposed subjects with or without considering minimum follow-up time and/or minimum post-exposure follow-up time.
bhat	Simulated value of regression coefficient (log hazard ratio)
HR	Simulated value of hazard ratio
d	Simulated number of events in total
d_c	Simulated number of events in control group
d_exp	Simulated number of events in exposed group
mst_c	Simulated median survival time in control group
mst_exp	Simulated median survival time in exposed group
pow	Simulated statistical power from the Cox regression model on data with time-dependent exposure

### Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

### **Examples**

16 plot\_power

plot_power	Plot power curves for survival analysis with time-dependent exposure

### **Description**

This function plots a power curve at each time and returns a subsetted data frame that match the list of input parameters.

### Usage

```
plot_power(table_df, N, type, exp.prop, min.futime, min.postexp.futime,
    show.plot = FALSE, newplot = FALSE, col = NULL, lty, lwd, pch)
```

### **Arguments**

table_df	A data frame read from a .csv file of a format output from the getpower.method1 function.
N	Number of subjects needs to be screened
type	A text string indicating the what type of dataset is of interest. Either one of "fixed" or "td" should be inputted
exp.prop	A numeric value between 0 and 1 (not include 0 and 1) that represents the proportion of subjects that are assigned with an exposure
min.futime	A numeric value that represents minimum follow-up time (in months). The default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study
min.postexp.futime	
	A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-

up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure A logical value indicating whether to output a power curve or not. The default

show.plot value is TRUE

newplot A logical value indicating whether to create a new plot or add to an existing plot

col, lty, lwd, pch

Graphical parameters as in the regular plot function in R

### **Details**

The gist of this function is that the user can check the plot with the values in the output data frame. Moreover, it is flexible that the user can choose to only output the data frame in order to plot their own graph (such as add titles or legends) based on the data. The user can also add as many lines as possible to an existing power curve plot so that he/she is allowed to compare different scenarios.

### Value

A data frame object with columns corresponding to

Number of subjects needs to be screened, specified by the user i\_N

plot\_simuData 17

N_eff	Simulated number of evaluable subjects, which is the resulting number of sub-
	jects with or without considering minimum follow-up time and/or minimum
	post-exposure follow-up time
i_beta	Input value of regression coefficient (log hazard ratio)
pow	Simulated statistical power from the Cox regression model on data with time-
	dependent exposure

#### Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

#### **Examples**

```
# We recommend setting nSim to at least 500. It is set to 10 in the example to
# reduce run time for CRAN submission.
ret \leftarrow getpower.method1(nSim = 10, N = 600, b = 0.3, exp.prop = 0.2,
    type = "td", scenario = " ", maxrelexptime = 1/6, min.futime = 4,
    min.postexp.futime = 4, output.fn = "output.csv")
ret2 <- getpower.method1(nSim = 10, N = 600, b = 0.3, exp.prop = 0.2,
    type = "td", scenario = " ", maxrelexptime = 1/6, min.futime = 4,
    min.postexp.futime = 0, output.fn ="output.csv")
# Read in .csv file as a data frame
tb <- read.csv("output.csv", header = TRUE, sep = ",")
# Visualize the subsetted data frame of interest and create a new plot
visualize1 <- plot_power(table_df = tb, N = 600, type = "td", exp.prop = 0.2,</pre>
    min.futime = 4, min.postexp.futime = 4, show.plot = TRUE, newplot = TRUE,
    col = "red", lty = 1, lwd = 2, pch = 16)
# Add a different power curve to the previously created plot
visualize2 <- plot_power(table_df = tb, N = 600, type = "td", exp.prop=0.2,</pre>
    min.futime = 4, min.postexp.futime = 0, show.plot = TRUE, newplot = FALSE,
    col = "blue", lty = 1, lwd = 2, pch = 16)
```

plot\_simuData

Make an incidence plot from simulated data.

### **Description**

Create an incidence plot ordered by follow-up time from a survival data simulated.

### Usage

```
plot_simuData(data, title="Sample Survival Data")
```

18 tdSim.exp.matching

### **Arguments**

data A dataframe of survival data containing the following columns: id, start, stop,

status, x

title Title of the graph

#### **Details**

This makes the incidence plot of the survival data based on the input dataframe from the td-Sim.method1 or tdSim.method2 function. More generally, this function would also works with a dataframe containing survival data with the same columns name as indicated above.

#### Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

### **Examples**

tdSim.exp.matching

Simulate 1 dataframe (1 simulation) of time-dep exposure in the case of exposure matching

### **Description**

This function simulate Survival Data. It generates a simulated dataset with time-dependent exposure with a user-specified list of parameters as input as well as matching id in order to do exposure matching.

### Usage

```
tdSim.exp.matching(N_match, duration=24,lambda, rho=1, beta,
rateC,matching.ratio=3,prop.fullexp=0,maxrelexptime=1,min.futime=0,
min.postexp.futime=0)
```

### **Arguments**

N_match	Number of matching sets
duration	Length of the study in Months. The default value is 24 (months)
lambda	Scale parameter of the Weibull distribution, which is calculated as $\log(2)$ / median time to event for control group
rho	Shape parameter of the Weibull distribution, which is defaulted as 1, as we generate survival times by using the exponential distribution
beta	A numeric value that represents the exposure effect, which is the regression coefficient (log hazard ratio) that represent the magnitude of the relationship between the exposure covariate and the risk of an event

tdSim.exp.matching 19

rateC Rate of the exponential distribution to generate censoring times, which is calcu-

lated as log(2) / median time to censoring

matching.ratio Matching ratio used in exposure matching.For 1:1, the user should specify 1.in-

put value 3 corresponds to 1:3 (exposed: unexposed). input value 0.25 corre-

sponds to 4:1 (ie. 1:0.25).

prop.fullexp A numeric value in interval [0, 1) that represents the proportion of exposed

subjects that are fully exposed from the beginning to the end of the study. The default value is 0, which means all exposed subjects have an exposure status

transition at some point during the study

maxrelexptime A numeric value in interval (0, 1] that represents the maximum relative expo-

sure time. Suppose this value is p, the exposure time for each subject is then uniformly distributed from 0 to p\*subject's time in the study. The default value is 1, which means all exposed subjects have an exposure status transition at any

point during the time in study.

min.futime A numeric value that represents minimum follow-up time (in months). The

default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a

short amount of time in the study

min.postexp.futime

A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure

#### **Details**

Simulata a Survival dataset using a modified version of illness-death model controlled by lambda12, lambda13

### Value

A data.frame object with columns corresponding to

id Integer that represents a subject's identification number

For counting process formulation. Represents the start of each time interval For counting process formulation. Represents the end of each time interval

status Indicator of event. status = 1 when event occurs and 0 otherwise x Indicator of exposure. x = 1 when exposed and 0 otherwise

match\_id Integer that represents a subject's matching set.

#### Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

#### **Examples**

```
df = tdSim.exp.matching<-function(N_match, duration=24,lambda, rho=1,
    beta, rateC,matching.ratio=3, prop.fullexp=0,maxrelexptime=1,min.futime=0,
    min.postexp.futime=0)</pre>
```

20 tdSim.method1

tdSim.method1	Simulate 1 dataframe (1 simulation) of time-dep exposure under method 1
---------------	---

### Description

This function generates a simulated dataset with time-dependent exposure under method 1 with a user-specified list of parameters as input. Survival times and censoring times are generated from the exponential distribution.

### Usage

```
tdSim.method1(N, duration = 24, lambda, rho = 1, beta, rateC, exp.prop,
    prop.fullexp = 0, maxrelexptime = 1, min.futime = 0, min.postexp.futime = 0)
```

### Arguments

N	Number of subjects needs to be screened
duration	Length of the study in Months. The default value is 24 (months)
lambda	Scale parameter of the Weibull distribution, which is calculated as $\log(2)$ / median time to event for control group
rho	Shape parameter of the Weibull distribution, which is defaulted as 1, as we generate survival times by using the exponential distribution
beta	A numeric value that represents the exposure effect, which is the regression coefficient (log hazard ratio) that represent the magnitude of the relationship between the exposure covariate and the risk of an event
rateC	Rate of the exponential distribution to generate censoring times, which is calculated as $\log(2)$ / median time to censoring
exp.prop	A numeric value between 0 and 1 (not include 0 and 1) that represents the proportion of subjects that are assigned with an exposure
prop.fullexp	A numeric value in interval [0, 1) that represents the proportion of exposed subjects that are fully exposed from the beginning to the end of the study. The default value is 0, which means all exposed subjects have an exposure status transition at some point during the study
maxrelexptime	A numeric value in interval (0, 1] that represents the maximum relative exposure time. Suppose this value is p, the exposure time for each subject is then uniformly distributed from 0 to p*subject's time in the study. The default value is 1, which means all exposed subjects have an exposure status transition at any point during the time in study.
min.futime	A numeric value that represents minimum follow-up time (in months). The default value is 0, which means no minimum follow-up time is considered. If it

short amount of time in the study

min.postexp.futime

A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure

has a positive value, this argument will help exclude subjects that only spend a

tdSim.method1 21

#### **Details**

If no minimum follow-up time or minimum post-exposure follow-up time is considered (min.fut = 0 and min.postexp.fut = 0), then the output data frame will have N subjects. If we consider minimum follow-up time or both, then the output data frame will have at most N subjects.

#### Value

A data frame object with columns corresponding to

id	Integer that represents a subject's identification number
start	For counting process formulation. Represents the start of each time interval
stop	For counting process formulation. Represents the end of each time interval
status	Indicator of event. status = $1$ when event occurs and $0$ otherwise
X	Indicator of exposure. $x = 1$ when exposed and 0 otherwise

### Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

#### References

Therneau and C. Crowson (2015). Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model.

https://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf

#### **Examples**

```
# Simulate a dataset of 600 subjects with time-dependent exposure without
# considering minimum follow-up time or minimum post-exposure follow-up time.
# Specifically, set the duration of the study to be 24 months; the median time to
# event for control group to be 24 months; exposure effect to be 0.3; median time
# to censoring to be 14 months; and exposure proportion to be 20%.
df1 \leftarrow tdSim.method1(N = 600, duration = 24, lambda = log(2)/24, rho = 1,
   beta = 0.3, rateC = log(2)/14, exp.prop = 0.2, prop.fullexp = 0,
   maxrelexptime = 1, min.fut = 0, min.postexp.fut = 0)
# Simulate a dataset of 600 subjects with time-dependent exposure with
# both minimum follow-up time (4 months) and minimum post-exposure
# follow-up time (4 months) imposed. Other parameters remain the same as
# in the first case.
df2 \leftarrow tdSim.method1(N = 600, duration = 24, lambda = log(2)/24, rho = 1,
   beta = 0.3, rateC = log(2)/14, exp.prop = 0.2, prop.fullexp = 0,
   maxrelexptime = 1, min.fut = 4, min.postexp.fut = 4)
# Simulate a dataset of 600 subjects with time-dependent exposure with
# both minimum follow-up time (4 months) and minimum post-exposure
# follow-up time (4 months) imposed. Also consider a quick exposure after
# entering the study for each exposed subject. Set the maximum relative
# exposure time to be 1/6. Other parameters remain the same as in the first case.
df3 \leftarrow tdSim.method1(N = 600, duration = 24, lambda = log(2)/24, rho = 1,
```

22 tdSim.method2

```
beta = 0.3, rateC = log(2)/14, exp.prop = 0.2, prop.fullexp = 0,
maxrelexptime = 1/6, min.fut = 4,min.postexp.fut = 4)
```

tdSim.method2 Simulate 1 dataframe (1 simulation) of time-dep exposimethod 2	e under
--	---------

### Description

This function simulate Survival Data. It generates a simulated dataset with time-dependent exposure under method 2 with a user-specified list of parameters as input.

### Usage

### **Arguments**

N	Number of subjects	
duration	Duration of the study. This is used in censoring	
lambda12	Lambda12 parameter to control time to exposure	
lambda23	Lambda23 parameter to control time to event after exposure	
lambda13	Lambda13 parameter to control time to event in the control group	
HR	Hazard Ratio. This input is optional. If HR is set and lambda23 is not set, lambda23 = lambda13*HR	
exp.prop	A numeric value between 0 and 1 (not include 0 and 1) that represents the proportion of subjects that are assigned with an exposure	
rateC	Rate of the exponential distribution to generate censoring times	
min.futime	A numeric value that represents minimum follow-up time (in months). The default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study	
min postovn futimo		

min.postexp.futime

A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure

### **Details**

Simulata a Survival dataset using a modified version of illness-death model controlled by lambda12, lambda23, lambda13

tdSim.multicenter 23

#### Value

A data.frame	object	with co	lumns	corres	ponding to

id	Integer that represents a subject's identification number
start	For counting process formulation. Represents the start of each time interval
stop	For counting process formulation. Represents the end of each time interval
status	Indicator of event. status = $1$ when event occurs and $0$ otherwise
X	Indicator of exposure. $x = 1$ when exposed and 0 otherwise

### Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

### **Examples**

tdSim.multicenter	Simulate 1 dataframe (1 simulation) of time-dependent exposure with
	multiple centers

### Description

This function allows the user to input a data frame with multi-center parameters and generates a simulated dataset with time-dependent exposure. In particular, the output dataset has a column corresponding to the center id, which will be used as a clustering variable in the Cox regression model in power calculation.

### Usage

```
tdSim.multicenter(N,duration=24,rho=1,beta,rateC,df,
    prop.fullexp=0,maxrelexptime=1,min.futime=0,
    min.postexp.futime=0, dist=NULL)
```

### **Arguments**

N	Number of subjects needs to be screened
duration	Length of the study in Months. The default value is 24 (months)
rho	Shape parameter of the Weibull distribution, which is defaulted as 1, as we generate survival times by using the exponential distribution
beta	A numeric value that represents the exposure effect, which is the regression coefficient (log hazard ratio) that represent the magnitude of the relationship between the exposure covariate and the risk of an event
rateC	Rate of the exponential distribution to generate censoring times, which is calculated as log(2) / median time to censoring

24 tdSim.multicenter

df A user-specified n by 4 clustering data frame with columns corresponding to

cat\_id (category id, which is the physician site id. It can be either text strings or integers), center.size (number of subjects within each center), cat\_exp.prop (proportion of exposed subjects in each center), and med.TTE.Control (median time to control event for each center). n rows corresponds to n different centers

prop.fullexp A numeric value in interval [0, 1) that represents the proportion of exposed

subjects that are fully exposed from the beginning to the end of the study. The default value is 0, which means all exposed subjects have an exposure status

transition at some point during the study

maxrelexptime A numeric value in interval (0, 1] that represents the maximum relative expo-

sure time. Suppose this value is p, the exposure time for each subject is then uniformly distributed from 0 to p\*subject's time in the study. The default value is 1, which means all exposed subjects have an exposure status transition at any

point during the time in study.

min.futime A numeric value that represents minimum follow-up time (in months). The

default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a

short amount of time in the study

min.postexp.futime

A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure

dist The distribution of the center effect across centers. Default is NULL. If dist=

'gamma', then a random frailty effect from a gamma distribution with scale  $0.5\,$ 

and shape 2 is used.

### Details

The current version of this function allows the user to input a data frame with at least 3 categories of physician sites, because the function uses a multinomial distribution to assign subjects into each category according to the corresponding category proportion

### Value

A data frame object with columns corresponding to

id Integer that represents a subject's identification number

start For counting process formulation. Represents the start of each time interval stop For counting process formulation. Represents the end of each time interval

status Indicator of event. status = 1 when event occurs and 0 otherwise

x Indicator of exposure. x = 1 when exposed and 0 otherwise

clst\_id For clustering in the Cox proportional hazard model. Represents label of each

subject's corresponding physician site

### Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Hiroto Udagawa, Nusrat Rabbee

Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

tdSim.multicenter 25

#### References

T. Therneau and C. Crowson (2015). Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model.

https://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf

### **Examples**

df = input\_df1, maxrelexptime = 1/6, min.futime = 4, min.postexp.futime = 4)

## **Index**

```
*Topic Plot
    plot_power, 16
    plot_simuData, 17
*Topic Power_Calculation
    getpower.exp.matching, 2
    getpower.exp.matching.opt, 5
    getpower.method1, 8
    getpower.method2, 11
    getpower.multicenter, 13
*Topic Simulation
    tdSim.exp.matching, 18
    tdSim.method1, 20
    tdSim.method2, 22
    tdSim.multicenter, 23
getpower.exp.matching, 2
getpower.exp.matching.opt, 5
getpower.method1, 8
getpower.method2, 11
getpower.multicenter, 13
plot_power, 16
plot_simuData, 17
SimHaz-package, 2
tdSim.exp.matching, 18
tdSim.method1, 20
tdSim.method2, 22
tdSim.multicenter, 23
```