

Package ‘RecurRisk’

June 23, 2020

Type Package

Title Recurrence Risk Assessment Tool

Version 1.0.2

Date 2020-06-22

Author Fanni Zhang <fannizhang@gmail.com>

Maintainer Joe Zou <zouj@imsweb.com>

Description Functions to estimate the risk of recurrence using disease-specific survival data. Mariotto AB, Zou Z, Zhang F, et al (2018) <doi:10.1158/1055-9965.EPI-17-1129>.

License GPL (>= 2)

Imports SEER2R, flexsurvcure, stats, survival

Depends R (>= 3.6.0)

NeedsCompilation no

Repository CRAN

Date/Publication 2020-06-23 12:30:06 UTC

R topics documented:

| | |
|--------------------------------|---|
| RecurRisk-package | 2 |
| data.cansurv | 3 |
| data.group | 4 |
| data.individual | 5 |
| read.seerstat | 5 |
| recurrisk.group | 6 |
| recurrisk.individual | 8 |

| | |
|--------------|-----------|
| Index | 10 |
|--------------|-----------|

RecurRisk-package *Recurrence Risk Assessment Tool*

Description

This package is to project the risk of recurrence using disease-specific survival data.

Details

This package has been developed to estimate the risk of progressing to distant recurrence using disease-specific survival from cancer registry data. The disease-specific survival is assessed via cause-specific survival or relative survival using SEER*Stat software. The cause-specific survival or relative survival is assumed to follow a mixture-cure model and the risk of recurrence is inferred from the survival among the non-cured fraction. The cure fraction and parametric survival distribution among those not cured are estimated using CanSurv software for group data and R flexsurvcure package for individual data. The current version can handle Weibull and log-logistic distributions for the non-cured survival.

Author(s)

Fanni Zhang <fannizhang@gmail.com>

References

- Mariotto AB, Zou Z, Zhang F, Howlader N, Kurian AW, Etzioni R. Can We Use Survival Data from Cancer Registries to Learn about Disease Recurrence? The Case of Breast Cancer. *Cancer Epidemiol Biomarkers Prev.* 2018 Nov; 27(11):1332-1341.
- Yu B, Tiwari RC, Cronin KA, McDonald C, Feuer EJ. CANSURV: A Windows program for population-based cancer survival analysis. *Comput.Methods Programs Biomed.* 2005;80(3):195-203.
- Yu B, Tiwari RC, Cronin KA, Feuer EJ. Cure fraction estimation from the mixture cure models for grouped survival data. *Stat.Med.* 2004;23(11):1733-47.
- De Angelis R, Capocaccia R, Hakulinen T, Soderman B, Verdecchia A. Mixture models for cancer survival analysis: application to population-based data with covariates. *Stat. Med.* 1999;18(4):441-54.
- Capocaccia R. Relationships between incidence and mortality in non-reversible diseases. *Stat.Med.* 1993;12(24):2395-415.
- Li CS, Taylor JMG, Sy JP. Identifiability of cure models. *Statistics & Probability Letters.* 2001;54(4):389-95.
- Howlader N, Ries LA, Mariotto AB, Reichman ME, Ruhl J, Cronin KA. Improved estimates of cancer-specific survival rates from population-based data. *J Natl Cancer Inst.* 2010;102(20):1584-98.

| | |
|--------------|----------------------------------|
| data.cansurv | <i>CanSurv CSV format output</i> |
|--------------|----------------------------------|

Description

An example CSV format output from CanSurv software including parameters for the mixture cure survival model.

Usage

```
data("data.cansurv")
```

Format

A data frame with 99 observations on the following 15 variables. Two strata "SEER_historic_stage_LRD" and "Age_breast", and one covariate "Year_of_diagnosis_BC" are defined in the model.

Link The parametric survival distribution among those not cured specified in CanSurv.

Stratum_1__SEER_historic_stage_LRD The stratum SEER_historic_stage_LRD value.

Stratum_2__Age_breast The stratum Age_breast value.

estimate The estimated parameters for the mixture cure survival model.

c_int A numeric vector in covariance matrix wrt cure intercept.

c_1__Year_of_diagnosis_BC__0 A numeric vector in covariance matrix wrt cure parameter for Year_of_diagnosis_BC baseline.

c_1__Year_of_diagnosis_BC__1 A numeric vector in covariance matrix wrt cure parameter for Year_of_diagnosis_BC value 1.

c_1__Year_of_diagnosis_BC__2 A numeric vector in covariance matrix wrt cure parameter for Year_of_diagnosis_BC value 2.

c_1__Year_of_diagnosis_BC__3 A numeric vector in covariance matrix wrt cure parameter for Year_of_diagnosis_BC value 3.

u_int A numeric vector in covariance matrix wrt mu intercept.

u_1__Year_of_diagnosis_BC__0 A numeric vector in covariance matrix wrt mu parameter for Year_of_diagnosis_BC baseline.

u_1__Year_of_diagnosis_BC__1 A numeric vector in covariance matrix wrt mu parameter for Year_of_diagnosis_BC value 1.

u_1__Year_of_diagnosis_BC__2 A numeric vector in covariance matrix wrt mu parameter for Year_of_diagnosis_BC value 2.

u_1__Year_of_diagnosis_BC__3 A numeric vector in covariance matrix wrt mu parameter for Year_of_diagnosis_BC value 3.

s_int A numeric vector in covariance matrix wrt sigma intercept.

Details

For each stratum, the variable order in the estimate column should be cure intercept, cure coefficients for each level of the covariate, mu intercept, mu coefficients for each level of the covariate, sigma intercept.

Examples

```
data("data.cansurv")
```

| | |
|------------|--------------------------------------|
| data.group | <i>SEER*Stat group survival data</i> |
|------------|--------------------------------------|

Description

An example SEER*Stat data set including the cause-specific group survival.

Usage

```
data("data.group")
```

Format

A data frame with 803 observations on the following 12 variables.

Page_type Page type with format 0=Life Page, 1=Summary Page, etc..

SEER_historic_stage_LRD Stage variable with format 0=Localized, 1=Regional, 2=Distant.

Year_of_diagnosis_BC Year variable with format 0=1985-89, 1=1990-94, 2=1995-99, 3=2000-13.

Age_breast Age variable with format 0=15-59, 1=60-69, 2=70-84.

Interval Follow-up years.

Alive_at_Start The number of people who are alive at the beginning of interval.

Died The number of cancer deaths in the interval.

Lost_to_Followup The number of people lost to follow-up or dying from other causes in the interval.

CauseSpecific_Survival_Interval The cause-specific survival for the interval.

CauseSpecific_Survival_Cum The cause-specific cumulative survival.

CauseSpecific_SE_Interval The standard error for cause-specific interval survival.

CauseSpecific_SE_Cum The standard error for cause-specific cumulative survival.

Examples

```
data("data.group")
```

| | |
|-----------------|---------------------------------|
| data.individual | <i>Individual survival data</i> |
|-----------------|---------------------------------|

Description

An example individual survival data including the least required variables for `recurrisk.individual` function, which is created using the case-listing unformatted SEER*Stat CSV data file.

Usage

```
data("data.individual")
```

Format

A data frame with 421198 observations on the following 5 variables.

`stage` Stage variable with format 0=Localized, 1=Regional, 2=Distant.

`yeargroup` Year variable with format 0=1985-89, 1=1990-94, 2=1995-99, 3=2000-13.

`agegroup` Age variable with format 0=15-59, 1=60-69, 2=70-84.

`time` Follow-up years.

`status` The status indicator, 0=alive, 1=dead due to cancer.

Details

Note that, dead events due to other causes should be defined as censoring events.

Examples

```
data("data.individual")
```

| | |
|---------------|--|
| read.seerstat | <i>Read data from SEER*Stat export files</i> |
|---------------|--|

Description

A function to read data and dictionary information from SEER*Stat export files. It should be called to read in group survival data before running function `recurrisk.group`.

Usage

```
read.seerstat(file.dic, file.txt)
```

Arguments

| | |
|----------|--|
| file.dic | The dictionary file exported from SEER*Stat software with .dic extension which contains the information describing the layout of the export data file. |
| file.txt | The cause-specific or relative survival data generated from SEER*Stat in .txt format. Note that, if the .txt file is not exported from SEER*Stat but manually converted, the user should make sure the delimiter used in the txt file is the same as the one specified in the dic file - field delimiter option. |

Details

This function makes use of function read.SeerStat in R package SEER2R.

Value

A data frame containing SEER*Stat survival data.

See Also

[recurrisk.group](#)

recurrisk.group

Estimate recurrence risk using grouped survival data

Description

A function to estimate the risk of recurrence using cancer registry disease-specific grouped survival data.

Usage

```
recurrisk.group(data, data.cansurv, stagevar, stage.dist.value, adj.r = 1)
```

Arguments

| | |
|------------------|--|
| data | The SEER*Stat cause-specific or relative group survival data frame returned by function read.seerstat. |
| data.cansurv | The data frame including parameter estimation and covariance matrix for the mixture cure survival model from Cansurv CSV format output. It also contains the information on strata and covariates. |
| stagevar | The stage variable containing the distant stage from SEER*Stat data. |
| stage.dist.value | The numeric value of distant stage. |
| adj.r | The adjustment factor used to adjust the registry-based survival curves for sensitivity analysis. The default value is 1. |

Value

A data frame containing the following items.

| | |
|----------------------|---|
| link | The parametric survival distribution among those not cured specified in CanSurv. |
| cure | The cure fraction estimated from the mixture cure survival model. |
| lambda | The estimated scale parameter of the survival distribution for those not cured. |
| k | The estimated shape parameter of the survival distribution for those not cured. |
| theta | The exponential hazard of the time from recurrence to cancer death. |
| surv_curemodel | The survival estimated from the mixture cure survival model. |
| surv_notcure | The estimated survival for the non-cured fraction. |
| median_surv_notcured | The median survival time for the non-cured fraction. |
| s1_numerical | The numerical estimated survival to recurrence (recurrence-free survival) for the non-cured fraction. |
| G_numerical | The numerical estimated survival to recurrence. |
| CI_numerical | 1-G_numerical, the numerical estimated cumulative incidence of recurrence which is the probability of progressing to cancer recurrence. |
| s1_analytical | The analytical estimated survival to recurrence (recurrence-free survival) for the non-cured fraction. |
| G_analytical | The analytical estimated survival to recurrence. |
| CI_analytical | 1-G_analytical, the analytical estimated cumulative incidence of recurrence. |
| se_CI_analytical | The standard error of CI_analytical. |
| obs_surv | The observed survival from SEER*Stat. |
| obs_dist_surv | The observed survival for distant stage from SEER*Stat. |

See Also

[data.group](#), [read.seerstat](#).

Examples

```
data("data.group")
data("data.cansurv")
stagevar<-"SEER_historic_stage_LRD"
stage.dist.value<-2
adj.r<-1.5
out<-recurrisk.group(data.group, data.cansurv, stagevar, stage.dist.value, adj.r)
```

recurrisk.individual *Estimate recurrence risk using individual survival data*

Description

A function to estimate the risk of recurrence using cancer registry disease-specific individual survival data.

Usage

```
recurrisk.individual(data, stratum, covar, timevar, eventvar,
stagevar, stage.dist.value, link = "Log-logistic", adj.r = 1)
```

Arguments

| | |
|------------------|--|
| data | The individual survival data set created using the case-listing unformatted CSV data exported from SEER*Stat software or user-generated individual data by other statistical software. |
| stratum | The user-defined strata variables which should be categorical coded as integers. One or more strata can be defined. If there is no stratum, user should define it as NULL. |
| covar | The covariates defined in the mixture cure survival model which should be categorical coded as integers. One or more covariates can be defined. If there is no covariate, user should define it as NULL. |
| timevar | The variable specified as follow-up time in the survival model. |
| eventvar | The status indicator, 0=alive, 1=dead (due to cancer). Note that, dead events due to other causes should be defined as censoring events. |
| stagevar | The stage variable defined in SEER*Stat data. If there are more than 1 stage variable, the user will need to define the one which contains the distant stage. |
| stage.dist.value | The numeric value of distant stage from the values of stage variable. |
| link | The latency distribution for the cure model (non-cured survival). The current version can handle Weibull and log-logistic distributions. The default link is log-logistic. |
| adj.r | The adjustment factor used to adjust the registry-based survival curves for sensitivity analysis. The default value is 1. |

Value

A data frame containing the following items.

| | |
|------|--|
| link | The parametric survival distribution among those not cured specified in CanSurv. |
| cure | The cure fraction estimated from the mixture cure survival model. |

| | |
|----------------------|---|
| lambda | The estimated scale parameter of the survival distribution for those not cured. |
| k | The estimated shape parameter of the survival distribution for those not cured. |
| theta | The exponential hazard of the time from recurrence to cancer death. |
| surv_curemodel | The survival estimated from the mixture cure survival model. |
| surv_notcure | The estimated survival for the non-cured fraction. |
| median_surv_notcured | The median survival time for the non-cured fraction. |
| s1_numerical | The numerical estimated survival to recurrence (recurrence-free survival) for the non-cured fraction. |
| G_numerical | The numerical estimated survival to recurrence. |
| CI_numerical | 1-G_numerical, the numerical estimated cumulative incidence of recurrence which is the probability of progressing to cancer recurrence. |
| s1_analytical | The analytical estimated survival to recurrence (recurrence-free survival) for the non-cured fraction. |
| G_analytical | The analytical estimated survival to recurrence. |
| CI_analytical | 1-G_analytical, the analytical estimated cumulative incidence of recurrence. |
| se_CI_analytical | The standard error of CI_analytical. |
| obs_surv | The observed survival from SEER*Stat. |
| obs_dist_surv | The observed survival for distant stage from SEER*Stat. |

Note

Note that, the cure fraction and parametric survival distribution are estimated using flexsurvcure function in R package flexsurvcure for individual data.

See Also

[data.individual](#)

Examples

```
#data("data.individual")
#timevar<-"time"
#eventvar<-"status"
#stagevar<-"stage"
#stage.dist.value<-3
#stratum<-c("stage","agegroup")
#covar<-"yeargroup"
#link<-"Log-logistic"
#adj.r<-1.2
#out<-recurrisk.individual(data.individual, stratum, covar, timevar, eventvar,
#                           stagevar, stage.dist.value, link, adj.r)
```

Index

*Topic **datasets**

data.cansurv, 3

data.group, 4

data.individual, 5

data.cansurv, 3

data.group, 4, 7

data.individual, 5, 9

read.seerstat, 5, 7

RecurRisk (RecurRisk-package), 2

RecurRisk-package, 2

recurrisk.group, 6, 6

recurrisk.individual, 8