Package ‘BCRA’

April 25, 2018

Type Package
Title Breast Cancer Risk Assessment
Version 2.0
Date 2018-04-25
Author Fanni Zhang
Maintainer Fanni Zhang <zhangf@imsweb.com>

Description Functions provide risk projections of invasive breast cancer based on Gail model according to National Cancer Institute’s Breast Cancer Risk Assessment Tool algorithm for specified race/ethnic groups and age intervals.

License GPL (>= 2)
NeedsCompilation no

R topics documented:

BCRA-package ............................................................... 2
absolute.risk .............................................................. 3
BrCa_1_AR ................................................................. 4
BrCa_beta ................................................................. 4
BrCa_lambda1 ............................................................ 5
BrCa_lambda2 ............................................................ 5
check.summary .......................................................... 6
error.table ............................................................... 7
error.table.all ........................................................... 8
exampledata .............................................................. 8
list.constants ............................................................ 9
recode.check ........................................................... 10
relative.risk ............................................................ 12
risk.summary ........................................................... 13

Index 14
Description

This package is to project absolute risk of invasive breast cancer according to NCI’s Breast Cancer Risk Assessment Tool (BCRAT) algorithm for specified race/ethnic groups and age intervals. The updated version 2.0 includes the new Hispanic model.

Details

This package can be used to estimate the risk of developing breast cancer over a predetermined time interval with risk factors. As the same as Breast Cancer Risk Assessment SAS Macro, the users can specify the time interval as appropriate, not only limited to the 5 years risk prediction available with BCRAT.

The main function in this package is `absoluteNrisk`, which is defined based on a statistical model known as the "Gail model". Parameters and constants needed in this function include initial and projection age, recoded covariates using function `recodeNcheck`, relative risks of BrCa at age "<50" and ">=50" obtained from function `relativeNrisk` as well as other known constants listed from function `listNconstants` like BrCa composite incidences, competing hazards, 1-attributable risk using in NCI BrCa Risk Assessment Tool (NCI BCRAT). With risk factors and projection interval ages for a group of women, the function `absoluteNrisk` will return the corresponding absolute risk projections. If the function returns any missing values, the function `errorNtable` or `errorNtableNall` is used to find where the errors occurred. The function `checkNsummary` can give a quick check for errors of input file and missing values of risks.

For further analysis, a data frame is created from the function `riskNsummary`, which includes age, duration of the projection time interval, covariates and the projected risk.

The version 2.0 includes absolute risk projections for Hispanic women (US born and Foreign born) based on race specific RR risk models developed on the San Francisco Bay Area Breast Cancer Study (SFBCS). Race specific attributable risks, breast cancer composite incidences and competing hazards are added to the updated package.

Author(s)

Fanni Zhang <zhangf@imsweb.com>

References


absolute.risk

### Description

A function to estimate absolute risks of developing breast cancer

### Usage

```r
absolute.risk(data, Raw_Ind=1, Avg_White=0)
```

### Arguments

- **data**: A data set containing all the required input data needed to perform risk projections, such as initial age, projection age, BrCa relative risk covariates and race. See `exampledata` for details.
- **Raw_Ind**: The raw file indicator with default value 1. `Raw_Ind`=1 means RR covariates are in raw/original format. `Raw_Ind`=0 means RR covariates have already been re-coded to 0, 1, 2 or 3.
- **Avg_White**: Calculation indicator. `Avg_White`=0, calculate absolute risks; `Avg_White`=1, calculate average absolute risks based on the rates for average non-hispanic white women and average other (native american) women. The default value is 0.

### Details

For the projection of absolute risks, this function is defined based on Gail Model. Parameters and constants needed in this function include initial and projection age, recoded covariates from function `recode.check`, relative risks of BrCa at age "<50" and ">=50" from function `relative.risk` as well as other known constants like BrCa composite incidences, competing hazards, 1-attributable risk using in NCI BrCa Risk Assessment Tool (NCI BCRAT).

### Value

A vector which returns absolute risk values when `Avg_White`=0 or average absolute risk values when `Avg_White`=1.

### See Also

`recode.check`, `relative.risk`
**Examples**

```r
data(exampledata)
# calculate absolute risk
absolute.risk(exampledata)
# calculate average absolute risk
Avg_White <- 1
absolute.risk(exampledata, Raw_Ind=1, Avg_White)
```

---

**BrCa_1_AR** | *Breast cancer 1-Attributable Risk*
---

**Description**

1-Attributable Risk

**Usage**

```r
data("BrCa_1_AR")
```

**Format**

A data frame with 2 observations on the following 5 variables.

- *Wh.Gail* White
- *AA.CARE* African-American
- *HU.Gail* Hispanic-American (US born)
- *NA.Gail* Other (Native American and unknown race)
- *HF.Gail* Hispanic-American (Foreign born)
- *Asian.AABCS* Asian-American

---

**BrCa_beta** | *Breast cancer beta*
---

**Description**

The logistic regression coefficients derived from the Gail model.

**Usage**

```r
data("BrCa_beta")
```

**Format**

A data frame with 6 observations on the following 5 variables.

- *Wh.Gail* White, Gail model
- *AA.CARE* African-American, Care model
- *HU.Gail* Hispanic-American (US born), Gail model
- *NA.Gail* Other (Native American and unknown race), Gail model
- *HF.Gail* Hispanic-American (Foreign born), Gail model
- *Asian.AABCS* Asian-American, AABCS model
Description
Breast cancer composite incidences for different races and age groups from 20 to 90 by 5 years.

Usage
data("BrCa_lambda1")

Format
A data frame with 14 age groups on the following 12 variables.

- HF.1995_04  Hispanic-American (Foreign born) 1995:2004
- Ch.1998_02  Chinese-American SEER 18 1998:2002
- Fi.1998_02  Filipino-American SEER 18 1998:2002
- oP.1998_02  Other Pacific Islander SEER 18 1998:2002

Description
Breast cancer competing mortality for different races and age groups from 20 to 90 by 5 years.

Usage
data("BrCa_lambda2")
Format

A data frame with 14 age groups on the following 12 variables.

- HF.1995_04 Hispanic-American (Foreign born) 1995:2004
- Ch.1998_02 Chinese-American SEER 18 1998:2002
- Fi.1998_02 Filipino-American SEER 18 1998:2002
- oP.1998_02 Other Pacific Islander SEER 18 1998:2002

---

check.summary

**Summarize the error indicators, relative risks and absolute risks**

Description

A function to show descriptive statistics by applying function **mean** and **sd** to the quantities Error_Ind, AbsRisk, RR_Star1 and RR_Star2.

Usage

`check.summary(data, Raw_Ind=1, Avg_White=0)`

Arguments

- **data**
  A data set containing all the required input data needed to perform risk projections, such as initial age, projection age, BrCa relative risk covariates and race. See exampledata for details.

- **Raw_Ind**
  The raw file indicator with default value 1. Raw_Ind=1 means RR covariates are in raw/original format. Raw_Ind=0 means RR covariates have already been re-coded to 0, 1, 2 or 3.

- **Avg_White**
  Calculation indicator. Avg_White=0, calculate absolute risks; Avg_White=1, calculate average absolute risks based on the rates for average non-hispanic white women and average other (native american) women. The default value is 0.

Details

When the mean and standard deviation for the variable Error_Ind is 0, implies that no errors have not been found. Otherwise when the mean and std for Error_Ind is not 0, implies that errors have been found. When errors are found, the number of records with errors is the count associated with AbsRisk listed under NMiss (number of missing).
error.table

Value

A summary table for error indicators, relative risks and absolute risks

See Also

recode.check, relative.risk, absolute.risk

---

error.table  List the records and errors for IDs with missing absolute risks

Description

A function to list the records and errors for IDs with missing absolute risks. For each of the records with error, the record is listed followed by a line which gives some indication as to where the error occurred. Relative risks and risk pattern numbers are also included.

Usage

error.table(data, Raw_Ind=1)

Arguments

data  A data set containing all the required input data needed to perform risk projections, such as initial age, projection age, BrCa relative risk covariates and race. See exampledata for details.

Raw_Ind  The raw file indicator with default value 1. Raw_Ind=1 means RR covariates are in raw/original format. Raw_Ind=0 means RR covariates have already been re-coded to 0, 1, 2 or 3.

Value

A data frame listing the raw records, errors, relative risks and pattern numbers for IDs with missing absolute risks. If there is nothing wrong with the input data, the function will return "NO ERROR!".

See Also

recode.check, error.table.all
**error.table.all**  
*List all records and errors*

**Description**

A function to list all records with both raw values and recoded values (or indications for errors). For each of the records, the record is listed followed by a line which gives some indication as to where the error occurred.

**Usage**

```r
error.table.all(data, Raw_Ind=1)
```

**Arguments**

- `data`  
  A data set containing all the required input data needed to perform risk projections, such as initial age, projection age, BrCa relative risk covariates and race. See `exampledata` for details.

- `Raw_Ind`  
  The raw file indicator with default value 1. `Raw_Ind=1` means RR covariates are in raw/original format. `Raw_Ind=0` means RR covariates have already been re-coded to 0, 1, 2 or 3.

**Value**

A data frame listing all records and errors. If there is nothing wrong with the input data, the function will return "NO ERROR!".

**See Also**

- `recode.check`
- `error.table`

---

**exampledata**  
*Example data set*

**Description**

A data set containing all the required input data needed to perform risk projections, such as initial age, projection age, BrCa relative risk covariates and race.

**Usage**

```r
data("exampledata")
```
list.constants

Format

A data frame with 26 observations on the following 9 variables.

ID  Woman’s ID, positive integer 1, 2, 3,...
T1  Initial age, all real numbers T1 in [20, 90).
T2  BrCa projection age, all real numbers T2 in [20, 90] such that T1<T2.
N_Biop The number of biopsies, 0, 1, 2,..., 99=unk (99 recoded to 0).
HypPlas Did biopsy display atypical hyperplasia? 0=no, 1=yes, 99=unk or not applicable.
AgeMen Age at menarchy, less than or equal to initial age, 99=unk.
Age1st Age at first live birth, greater or equal to age at menarchy and less than or equal to initial age, 98=nulliparous, 99=unk.
N_Re1s The number of 1st degree relatives with BrCa, 0, 1, 2,... 99=unk.
Race  Race, positive integer 1, 2, 3,...11. See details.

Details

1=Wh  White 1983-87 SEER rates (rates used in NCI BCRAT)
2=AA  African-American
3=HU  Hispanic-American (US born) 1995-04
4=NA  Other (Native American and unknown race)
5=HF  Hispanic-American (Foreign born) 1995-04
6=Ch  Chinese-American
7=Ja  Japanese-American
8=Fi  Filipino-American
9=Hw  Hawaiian-American
10=oP  Other Pacific Islander
11=oA  Other Asian

Description

A function to create a text file under user’s working directory which contains all constants required for BrCa absolute risk projections.

Usage

list.constants(BrCa_lambda1, BrCa_lambda2, BrCa_beta, BrCa_1_AR)

Arguments

BrCa_lambda1  Breast Cancer Composite Incidences
BrCa_lambda2  Breast Cancer Competing Mortality
BrCa_beta     The logistic regression coefficients (beta) derived from the Gail model
BrCa_1_AR     1-Attributable Risk
Details

See "BrCa_lambda1.rda", "BrCa_lambda2.rda", "BrCa_beta.rda", "BrCa_1_AR.rda" in package data folder.

Value

A text file "list_all_constants.txt" exported under user’s working directory for reading convenience.

| recode.check | Recode and check the relative risk covariate values |

Description

A function to recode the relative risk covariates and check errors.

Usage

recode.check(data, Raw_Ind=1)

Arguments

data A data set containing all the required input data needed to perform risk projections, such as initial age, projection age, BrCa relative risk covariates and race. See exampledata for details.

Raw_Ind The raw file indicator with default value 1. Raw_Ind=1 means RR covariates are in raw/original format. Raw_Ind=0 means RR covariates have already been re-coded to 0, 1, 2 or 3.

Details

This function is to recode the following relative risk covariates. Recoded RR covariates are named as NB_Cat, AM_Cat, AF_Cat and NR_Cat for N_Biop, AgeMen, Age1st and N_Rels, respectively.

N_Biop: The number of biopsies.
AgeMen: Age at menarchy.
Age1st: Age at first live birth.
N_Rels: The number of first degree relatives with BrCa.

See the following table for recoding details.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Raw Value</th>
<th>Recoded to</th>
</tr>
</thead>
<tbody>
<tr>
<td>N_Biop</td>
<td>0 or 99 (unk or not applicable)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2,3,4 ... and not 99</td>
<td>2</td>
</tr>
<tr>
<td>AgeMen</td>
<td>14,15,16 ... or 99 (unk)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>12,13</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>11 and younger</td>
<td>2</td>
</tr>
<tr>
<td>Age1st</td>
<td>19 and younger or 99 (unk)</td>
<td>0</td>
</tr>
</tbody>
</table>
This function is also used to check consistency and errors of input data. Let \( \text{set\_T1\_missing} \) and \( \text{set\_T2\_missing} \) be the checking variables for \( T_1 \) and \( T_2 \). The constraint on \( T_1 \) and \( T_2 \) is \( 20 \leq T_1 < T_2 \leq 90 \). If it is violated, \( \text{set\_T1\_missing} \) and \( \text{set\_T2\_missing} \) and the absolute risk will be set to the missing value \( \text{NA} \).

Let \( \text{RacCat} \) be the checking variable for \( \text{Race} \). If the \( \text{Race} \) value is not included in the 11 races defined, the absolute risk will be set to the missing value \( \text{NA} \) and \( \text{RacCat} \) will be set to "U" (undefined). The corresponding character of Race \( \text{CharRace} \) will be set to "??".

Let \( \text{set\_HyperP\_missing} \) and \( \text{set\_R\_Hyp\_missing} \) be the checking variables for \( \text{HyperP} \) and \( \text{R\_Hyp} \). Consistency patterns for the number of Biopsies and Hyperplasia are:

\[
\begin{align*}
\text{Requirement (A)} & \quad \text{N\_Biops} = 0 \text{ or } 99, \text{ then } \text{HyperP} = 99 \text{ (not applicable).} \\
\text{Requirement (B)} & \quad \text{N\_Biops} > 0 \text{ and } < 99, \text{ then } \text{HyperP} = 0, 1 \text{ or } 99.
\end{align*}
\]

If ANY of the above 2 REQUIREMENTS is violated, \( \text{NB\_Cat}, \text{set\_HyperP\_missing} \) and \( \text{set\_R\_Hyp\_missing} \) will be set to the corresponding character "A" or "B" and the absolute risk will be set to the missing value \( \text{NA} \). The consequences to the relative risk (RR) for the above two requirements are:

\[
\begin{align*}
\text{(A)} & \quad \text{N\_Biops} = 0 \text{ or } 99, \text{ HyperP} = 99 \text{ (not applicable) inflates RR by } 1.00. \\
\text{(B)} & \quad \text{N\_Biops} > 0 \text{ and } < 99, \text{ HyperP} = 0 \text{ (no) inflates RR by } 0.93; \\
& \quad \text{N\_Biops} > 0 \text{ and } < 99, \text{ HyperP} = 1 \text{ (yes) inflates RR by } 1.82; \\
& \quad \text{N\_Biops} > 0 \text{ and } < 99, \text{ HyperP} = 99 \text{ (unk) inflates RR by } 1.00.
\end{align*}
\]

For remaining relative risk covariates, \( \text{AgeMen}, \text{Age1st} \) and \( \text{N\_Rel} \):

\begin{itemize}
  \item **AgeMen** Age at menarchy must be postive integer less than or equal to initial age \( T_1 \).
  \item **Age1st** Age at 1st live birth must be postive integer greater than equal to \( \text{AgeMen} \) and less than or equal to initial age \( T_1 \).
\end{itemize}

\[
\begin{align*}
\text{AgeMen} & \quad \text{Age at menarchy must be postive integer less than or equal to initial age } T_1. \\
\text{NOTE: (1)} & \quad \text{For African-American women } \text{AgeMen} \leq 11 \text{ are grouped with } \text{AgeMen} = 12 \text{ or 13. (2) For US Born Hispanic women } \text{AgeMen} \text{ is not included in the RR model and all values for this variable are recoded to 0.}
\end{align*}
\]

\[
\begin{align*}
\text{Age1st} & \quad \text{Age at 1st live birth must be postive integer greater than equal to } \text{AgeMen} \\
\text{NOTE: (1)} & \quad \text{For African-American women, Age1st is not included in the RR model and all values for this variable are recoded to 0. (2) For US Born and Foreign Born Hispanic women, the recoding for this variable follows:}
\end{align*}
\]

\[
\begin{align*}
\text{Age1st} & \quad \text{19 and younger or 99 (unk) } 0 \\
& \quad \text{20 - 29 } 1 \\
& \quad \text{30+ or 98 (nulliparous) and not 99 } 2
\end{align*}
\]

\[
\begin{align*}
\text{N\_Rel} & \quad \text{The number of 1st degree relatives with BrCa must be 0,1,2,...}
\end{align*}
\]
NOTE: For Asian-Americans Race=6-11 and Hispanic-Americans (US and foreign born), the number of 1st degree relative coded value of 2 gets grouped with 1.

Value
A data frame containing the error indicators, recoded covariates as well as other checking variables defined for checking the consistency of the input data.

See Also
error.table.all, error.table

Examples
```
data(exampledata)
recode.check(exampledata)
```

### Description
A function to estimate relative risks for risk factor combinations

### Usage
```
relative.risk(data, Raw_Ind=1)
```

### Arguments
- **data**: A data set containing all the required input data needed to perform risk projections, such as initial age, projection age, BrCa relative risk covariates and race. See exampledata for details.
- **Raw_Ind**: The raw file indicator with default value 1. Raw_Ind=1 means RR covariates are in raw/original format. Raw_Ind=0 means RR covariates have already been re-coded to 0, 1, 2 or 3.

### Details
The age is dichotomized as "age less than 50 years" and "age 50 years or more". The relative risks can be obtained from Gail Model, an unconditional logistic regression that included main effects NB_Cat, AM_Cat, AF_Cat, NR_Cat as well as interactions between AF_Cat and NR_Cat and between the age category and NR_Cat.

### Value
- **RR_Star1**: Relative risk for woman of interest at ages < 50.
- **RR_Star2**: Relative risk for woman of interest at ages >= 50.
- **PatternNumber**: The sequence number of risk patterns. There are 3 levels for NB_Cat, 3 for AM_Cat, 4 for AF_Cat, 3 for NR_Cat, 3*3*4*3 = 108 patterns in total. Pattern Number=NB_Cat*3*3*4+AM_Cat*3*4+AF_Cat*3+NR_Cat*1+1.
See Also

recode.check

Examples

data(exampledata)
relative.risk(exampledata)

risk.summary

List the records with relative risks and absolute risks

Description

A function to list all the records with relative risks and absolute risks.

Usage

risk.summary(data, Raw_Ind=1)

Arguments

data A data set containing all the required input data needed to perform risk projections, such as initial age, projection age, BrCa relative risk covariates and race. See exampledata for details.

Raw_Ind The raw file indicator with default value 1. Raw_Ind=1 means RR covariates are in raw/original format. Raw_Ind=0 means RR covariates have already been re-coded to 0, 1, 2 or 3.

Value

A data frame that includes age, duration of the projection time interval, covariates and the projected risk. A CSV file is created to save the data frame under user’s working directory for reading convenience.

See Also

relative.risk, absolute.risk

Examples

data(exampledata)
risk.summary(exampledata)
Index

absolute.risk, 3, 7, 13
BCRA (BCRA-package), 2
BCRA-package, 2
BrCa_1_AR, 4
BrCa_beta, 4
BrCa_lambda1, 5
BrCa_lambda2, 5
check.summary, 6
error.table, 7, 8, 12
error.table.all, 7, 8, 12
examedata, 8
list.constants, 9
recode.check, 3, 7, 8, 10, 13
relative.risk, 3, 7, 12, 13
risk.summary, 13