

Package ‘BCRA’

April 27, 2015

Type Package

Title Breast Cancer Risk Assessment in R

Version 1.0

Date 2015-03-24

Author Fanni Zhang

Maintainer Fanni Zhang <zhangf@imsweb.com>

Description An R package for Breast Cancer Absolute Risk Projection

License GPL (>= 2)

R topics documented:

BCRA-package	1
absolute.risk	2
BrCa_1_AR	3
BrCa_beta	4
BrCa_lambda1	4
BrCa_lambda2	5
check.summary	5
error.table	6
error.table.all	7
exampledata	7
list.constants	8
recode.check	9
relative.risk	11
risk.summary	12

Description

BCRA is an R package that projects 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.

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 `absolute.risk`, 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 `recode.check`, relative risks of BrCa at age "<50" and "≥50" obtained from function `relative.risk` as well as other known constants listed from function `list.constants` 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 `absolute.risk` will return the corresponding absolute risk projections. If the function returns any missing values, the function `error.table` or `error.table.all` is used to find where the errors occurred. The function `check.summary` 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 `risk.summary`, which includes age, duration of the projection time interval, covariates and the projected risk. This function can also export a CSV file "risk_summary.csv" to save the data frame under user's working directory for reading convenience.

Any changes/modifications to the package would be at the user's own discretion and risk.

Author(s)

Fanni Zhang <zhangf@imsweb.com>

References

Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Shairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *JNCI* 81(24): 1879-86, 1989.

Costantino J, Gail MH, Pee D, Anderson S, Redmond CK, Benichou J, Wieand HS. Validation studies for models to project the risk of invasive and total breast cancer. *JNCI* 91(18):1541-48, 1999.

Gail MH, Costantino JP, Pee D, Bondy M, Newman L, Selvan M, Anderson GL, Malone KE, Marchbanks PA, McCaskill-Stevens W, Norman SA, Simon MS, Spirtas R, Ursin G, Berstein L. Projecting individualized absolute invasive breast cancer risk in African American women. *JNCI* 99(23):1782-92, 2007.

Matsuno RK, Costantino JP, Ziegler RG, Anderson GL, Li H, Pee D, Gail MH. Projecting individualized absolute invasive breast cancer risk in asian and pacific islander american women. *JNCI* 103(12):951-61, 2011.

absolute.risk	<i>Estimate absolute risks</i>
---------------	--------------------------------

Description

A function to estimate absolute risks of developing breast cancer

Usage

```
absolute.risk(dat, iloop = 1, Raw_Ind = 1)
```

Arguments

dat	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 <code>exampledata</code> for details.
iloop	Loop indicator. <code>iloop=1</code> , calculate absolute risks; <code>iloop=2</code> , calculate average absolute risks. The default value is 1.
Raw_Ind	The raw file indicator with default value 1. <code>Raw_Ind=1</code> means RR covariates are in raw/original format. <code>Raw_Ind=0</code> means RR covariates have already been re-coded to 0, 1, 2 or 3.

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 `iloop=1` or average absolute risk values when `iloop=2`.

See Also

`recode.check`, `relative.risk`

BrCa_1_AR	<i>Breast cancer 1-Attributable Risk</i>
-----------	--

Description

1-Attributable Risk

Usage

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

Format

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

Wh.Gail White
 AA.CARE African-American
 Hi.Gail Hispanic-American
 NA.Gail Native American and unknown race
 Asian.AABCS Asian-American

BrCa_beta	<i>Breast cancer beta</i>
-----------	---------------------------

Description

The logistic regression coefficients derived from the Gail model.

Usage

```
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
 Hi.Gail Hispanic-American, Gail model
 NA.Gail Native American and unknown race, Gail model
 Asian.AABCS Asian-American, AABCS model

BrCa_lambda1	<i>Breast cancer composite incidences</i>
--------------	---

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.

Wh.1983_87 White SEER 1983:1987
 AA.1994_98 African-American SEER 1994:1998
 Hi.1990_96 Hispanic-American 1990:1996
 NA.1983_87 Native American and unknown race 1983:1987
 Wo.1995_03 White SEER 1995:2003
 Ch.1998_02 Chinese-American SEER 18 1998:2002
 Ja.1998_02 Japanese-American SEER 18 1998:2002
 Fi.1998_02 Filipino-American SEER 18 1998:2002
 Hw.1998_02 Hawaiian SEER 18 1998:2002
 oP.1998_02 Other Pacific Islander SEER 18 1998:2002
 oA.1998_02 Other Asian SEER 1998:2002
 Wh_Avg.1992_96 Average White SEER 1992:1996

 BrCa_lambda2

Breast cancer competing mortality

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.

Wh.1983_87 White SEER 1983:1987
 AA.1994_98 African-American SEER 1994:1998
 Hi.1990_96 Hispanic-American 1990:1996
 NA.1983_87 Native American and unknown race 1983:1987
 Wo.1995_03 White SEER 1995:2003
 Ch.1998_02 Chinese-American SEER 18 1998:2002
 Ja.1998_02 Japanese-American SEER 18 1998:2002
 Fi.1998_02 Filipino-American SEER 18 1998:2002
 Hw.1998_02 Hawaiian SEER 18 1998:2002
 oP.1998_02 Other Pacific Islander SEER 18 1998:2002
 oA.1998_02 Other Asian SEER 1998:2002
 Wh_Avg.1992_96 Average White SEER 1992:1996

check.summary	<i>Summarize the error indicators, relative risks and absolute risks</i>
---------------	--

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(dat, iloop = 1, Raw_Ind = 1)
```

Arguments

<code>dat</code>	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 <code>exampledata</code> for details.
<code>iloop</code>	Loop indicator. <code>iloop=1</code> , calculate absolute risks; <code>iloop=2</code> , calculate average absolute risks. The default value is 1.
<code>Raw_Ind</code>	The raw file indicator with default value 1. <code>Raw_Ind=1</code> means RR covariates are in raw/original format. <code>Raw_Ind=0</code> means RR covariates have already been re-coded to 0, 1, 2 or 3.

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).

Value

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

See Also

```
recode.check, relative.risk, absolute.risk
```

error.table	<i>List the records and errors for IDs with missing absolute risks</i>
-------------	--

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(dat, Raw_Ind = 1)
```

Arguments

dat	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 <code>exampledata</code> for details.
Raw_Ind	The raw file indicator with default value 1. <code>Raw_Ind=1</code> means RR covariates are in raw/original format. <code>Raw_Ind=0</code> 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

```
error.table.all(dat, Raw_Ind = 1)
```

Arguments

dat	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 <code>exampledata</code> for details.
Raw_Ind	The raw file indicator with default value 1. <code>Raw_Ind=1</code> means RR covariates are in raw/original format. <code>Raw_Ind=0</code> 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

```
data("exampledata")
```

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_Rels 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=Hi	Hispanic-American
4=NA	Other (Native American and unknown race)
5=Wo	White 1995-03 SEER rates (rates under further investigation)
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

Note that Hispanic and other ethnic women risks are based on white women log relative risks. Hispanic women risk are based on Hispanic seer rates while other women risk are based on white women seer rates. Average other women rates are also based on average white women seer rates as well.

`list.constants` *List all constants required for BrCa absolute risk projections*

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(dat, Raw_Ind=1)
```

Arguments

`dat` 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.

Covariate	Raw Value	Recoded to
N_Biop	0 or 99 (unk or not applicable)	0
	1	1
	2,3,4 ... and not 99	2
AgeMen	14,15,16 ... or 99 (unk)	0
	12,13	1
	11 and younger	2
Age1st	19 and younger or 99 (unk)	0
	20,21,22,23,24	1
	25,26,27,28,29 or 98 (nulliparous)	2
	30,31,32 ... and not 98 and not 99	3
N_Rels	0 or 99 (unk)	0
	1	1
	2,3,4 ... and not 99	2

This function is also used to check consistency and errors of input data. Let `set_T1_missing` and `set_T2_missing` be the checking variables for T1 and T2. The constraint on T1 and T2 is $20 \leq T1 < T2 \leq 90$. If it is violated, `set_T1_missing` and `set_T2_missing` and the absolute risk will be set to the missing value NA.

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

Let `set_HyperP_missing` and `set_R_Hyp_missing` be the checking variables for `HypPlas` and `R_Hyp`. Consistency patterns for the number of Biopsies and Hyperplasia are:

- Requirement (A) N_Biops=0 or 99, then HypPlas MUST = 99 (not applicable).
- Requirement (B) N_Biops>0 and <99, then HypPlas = 0, 1 or 99.

If ANY of the above 2 REQUIREMENTS is violated, NB_Cat, `set_HyperP_missing` and `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 NA. The consequences to the relative risk (RR) for the above two requirements are:

(A) `N_Biops=0` or `99`, `HypPlas=99` (not applicable) inflates RR by 1.00.

(B) `N_Biops>0` and `<99`, `HypPlas=0` (no) inflates RR by 0.93;

`N_Biops>0` and `<99`, `HypPlas=1` (yes) inflates RR by 1.82;

`N_Biops>0` and `<99`, `HypPlas=99` (unk) inflates RR by 1.00.

For remaining relative risk covariates, `AgeMen`, `Age1st` and `N_Rels`:

- `AgeMen` Age at menarchy must be postive integer less than or equal to initial age T1.
NOTE: For African-American women `AgeMen<=11` are grouped with `AgeMen=12` or 13.
- `Age1st` Age at 1st live birth must be postive integer greater than equal to `AgeMen` and less than or equal to initial age T1.
NOTE: For African-American women, `Age1st` is not included in the RR model and all values for this variable are recoded to 0.
- `N_Rels` The number of 1st degree relatives with BrCa must be 0,1,2....
NOTE: For Asian-Americans `Race=6-11`, the number of 1st degree relative coded value of 2 gets grouped with 1.

Value

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

See Also

`error.table.all`, `error.table`

<code>relative.risk</code>	<i>Estimate relative risks</i>
----------------------------	--------------------------------

Description

A function to estimate relative risks for risk factor combinations

Usage

```
relative.risk(dat, Raw_Ind = 1)
```

Arguments

- `dat` 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`

<code>risk.summary</code>	<i>List the records with relative risks and absolute risks</i>
---------------------------	--

Description

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

Usage

```
risk.summary(dat, iloop = 1, Raw_Ind = 1)
```

Arguments

<code>dat</code>	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 <code>exampledata</code> for details.
<code>iloop</code>	Loop indicator. <code>iloop=1</code> , calculate absolute risks; <code>iloop=2</code> , calculate average absolute risks. The default value is 1.
<code>Raw_Ind</code>	The raw file indicator with default value 1. <code>Raw_Ind=1</code> means RR covariates are in raw/original format. <code>Raw_Ind=0</code> 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`