Example of BCRA

April 26, 2018

Example data

Following is a listing of the sample raw input data set "exampledata",

```
> library(BCRA)

> data(exampledata, package="BCRA")
> exampledata
```

	ID	T1	T2	N_Biop	HypPlas	AgeMen	Age1st	N_Rels	Race
1	1	45.2	53.3	99	99	10	20	1	0
2	2	45.2	53.3	99	1	10	20	1	1
3	3	45.2	53.3	99	0	10	20	1	2
4	4	45.2	53.3	0	99	10	20	1	3
5	5	45.2	53.3	1	99	10	20	1	4
6	6	45.2	53.3	1	99	14	19	1	5
7	7	45.2	53.3	99	99	99	19	1	6
8	8	45.2	53.3	1	1	14	19	1	7
9	9	45.2	53.3	99	1	14	99	1	8
10	10	45.2	53.3	1	0	14	19	1	9
11	11	45.2	53.3	99	0	99	99	1	10
12	12	45.2	53.3	0	0	14	19	1	11
13	13	45.2	53.3	0	99	10	20	1	12
14	14	45.2	53.3	0	1	10	20	1	0
15	15	45.2	53.3	0	0	10	20	1	1
16	16	45.2	53.3	1	0	10	20	1	2
17	17	35.0	40.0	4	99	11	25	0	3
18	18	35.0	40.0	4	99	11	98	0	4
19	19	35.0	40.0	4	99	11	10	0	5
20	20	35.0	40.0	4	99	36	25	0	6
21	21	27.0	90.0	99	99	13	22	0	7
22	22	27.0	90.0	99	99	13	22	99	8
23	23	18.0	26.0	99	99	13	22	99	9
24	24	27.0	26.0	99	99	13	22	99	10
25	25	85.0	91.0	99	99	13	22	99	11
26	26	86.0	90.0	99	99	13	22	99	12

Quick check

A quick check for errors of input file and missing values of risks using the function "check.summary".

```
> check.summary(exampledata)
```

```
   Variable                                          Label              Mean
1 Error_Ind        If mean not 0, implies ERROR in file 0.576923076923077
2   AbsRisk Abs risk(%) of BrCa in age interval [T1,T2)  3.76765716742822
3  RR_Star1                     Relative risk age lt 50  3.43948187580407
4  RR_Star2                     Relative risk age ge 50  2.86656310613911
              StdDev  N NMiss
1 0.503831473655779 26     0
2   2.5784420169801 11    15
3  1.92321120759021 13    13
4  1.54840080626899 13    13
```

The mean and std for "Error_Ind" are not 0, which implies that errors have been found. The number of records with errors is the count asscociated with "AbsRisk" listed under NMiss (the number of missing).

Error checking

Since NMiss=15 for Absolute Risk, we note that the error table lists the below 15 records.

```
> error.table(exampledata)
```

	ID	T1	T2	N_Biop	HypPlas	R_Hyp	AgeMen	Age1st	N_Rels	Race	RR_Star1
1	1	45.2	53.3	99	99	1	10	20	1	0	<NA>
2	1	45.2	53.3	0	99	1	2	1	1	??	
3	2	45.2	53.3	99	1	<NA>	10	20	1	1	<NA>
4	2	45.2	53.3	A	A	A	2	1	1	Wh	
5	3	45.2	53.3	99	0	<NA>	10	20	1	2	<NA>
6	3	45.2	53.3	A	A	A	1	0	1	AA	
7	9	45.2	53.3	99	1	<NA>	14	99	1	8	<NA>
8	9	45.2	53.3	A	A	A	0	0	1	Fi	
9	11	45.2	53.3	99	0	<NA>	99	99	1	10	<NA>
10	11	45.2	53.3	A	A	A	0	0	1	oP	
11	12	45.2	53.3	0	0	<NA>	14	19	1	11	<NA>
12	12	45.2	53.3	A	A	A	0	0	1	oA	
13	13	45.2	53.3	0	99	1	10	20	1	12	<NA>
14	13	45.2	53.3	0	99	1	2	1	1	??	
15	14	45.2	53.3	0	1	<NA>	10	20	1	0	<NA>
16	14	45.2	53.3	A	A	A	2	1	1	??	
17	15	45.2	53.3	0	0	<NA>	10	20	1	1	<NA>
18	15	45.2	53.3	A	A	A	2	1	1	Wh	
19	19	35	40	4	99	1	11	10	0	5	<NA>
20	19	35	40	2	99	1	2	<NA>	0	HF	
21	20	35	40	4	99	1	36	25	0	6	<NA>

```
22 20   35   40      2      99     1   <NA>   <NA>      0   Ch
23 23   18   26     99      99     1     13     22     99    9              <NA>
24 23 <NA>   26      0      99     1      1   <NA>      0   Hw
25 24   27   26     99      99     1     13     22     99   10 1.42102047641751
26 24 <NA> <NA>      0      99     1      1      1      0   oP
27 25   85   91     99      99     1     13     22     99   11 1.42102047641751
28 25   85 <NA>      0      99     1      1      1      0   oA
29 26   86   90     99      99     1     13     22     99   12              <NA>
30 26   86   90      0      99     1      1      1      0   ??
           RR_Star2 AbsRisk PatternNum
1              <NA>    <NA>         29
2
3              <NA>    <NA>       <NA>
4
5              <NA>    <NA>       <NA>
6
7              <NA>    <NA>       <NA>
8
9              <NA>    <NA>       <NA>
10
11             <NA>    <NA>       <NA>
12
13             <NA>    <NA>         29
14
15             <NA>    <NA>       <NA>
16
17             <NA>    <NA>       <NA>
18
19             <NA>    <NA>       <NA>
20
21             <NA>    <NA>       <NA>
22
23             <NA>    <NA>       <NA>
24
25 1.42102047641751    <NA>         16
26
27 1.42102047641751    <NA>         16
28
29             <NA>    <NA>         16
30
```

For each of the records with error, the record is listed followed by a line which gives some indication as to where the error occured. For example, the record with ID=2 has an "A" listed under the 3 variables associated with Biopy i.e. N_Biop, Hyperplasia and Hypr_RR. This means that ID=2 has violated consistency defined by Requirement (A) above. Similarly for IDs 3,9,11,12,14 and 15, violations are observed for Requirements (A). For IDs 19 and 20, they have violated AgeMen and/or Age1st consistency. Note the missing value "NA" listed under AgeMen and/or Age1st. For IDs 23, 24 and 25, violations of T1 and/or T2 consistency requirements are observed. So there are NA values listed

under T1 and/or T2. This small sample data set "exampledata" in no way exhausts all the possible ways in which the data can be in error, but it should give a guide and indication on how to check and correct errors when they do occur.

Risk results

List the absolute risks as well as all the original input variables in a data frame,

```
> risk.summary(exampledata)
```

```
   ID   T1   T2 Proj_Intvl N_Biop HypPlas AgeMen Age1st N_Rels Race CharRace
1   1 45.2 53.3        8.1     99      99     10     20      1    0       ??
2   2 45.2 53.3        8.1     99       1     10     20      1    1       Wh
3   3 45.2 53.3        8.1     99       0     10     20      1    2       AA
4   4 45.2 53.3        8.1      0      99     10     20      1    3       HU
5   5 45.2 53.3        8.1      1      99     10     20      1    4       NA
6   6 45.2 53.3        8.1      1      99     14     19      1    5       HF
7   7 45.2 53.3        8.1     99      99     99     19      1    6       Ch
8   8 45.2 53.3        8.1      1       1     14     19      1    7       Ja
9   9 45.2 53.3        8.1     99       1     14     99      1    8       Fi
10 10 45.2 53.3        8.1      1       0     14     19      1    9       Hw
11 11 45.2 53.3        8.1     99       0     99     99      1   10       oP
12 12 45.2 53.3        8.1      0       0     14     19      1   11       oA
13 13 45.2 53.3        8.1      0      99     10     20      1   12       ??
14 14 45.2 53.3        8.1      0       1     10     20      1    0       ??
15 15 45.2 53.3        8.1      0       0     10     20      1    1       Wh
16 16 45.2 53.3        8.1      1       0     10     20      1    2       AA
17 17   35   40          5      4      99     11     25      0    3       HU
18 18   35   40          5      4      99     11     98      0    4       NA
19 19   35   40          5      4      99     11     10      0    5       HF
20 20   35   40          5      4      99     36     25      0    6       Ch
21 21   27   90         63     99      99     13     22      0    7       Ja
22 22   27   90         63     99      99     13     22     99    8       Fi
23 23   18   26          8     99      99     13     22     99    9       Hw
24 24   27   26         -1     99      99     13     22     99   10       oP
25 25   85   91          6     99      99     13     22     99   11       oA
26 26   86   90          4     99      99     13     22     99   12       ??
              AbsRisk
1               <NA>
2               <NA>
3               <NA>
4   2.10808457698158
5   4.44129446296421
6   3.97624470498605
7     1.249550474885
8   5.77567590434044
9               <NA>
10  3.90606305710754
11              <NA>
```

```
12              <NA>
13              <NA>
14              <NA>
15              <NA>
16  2.68990404389925
17 0.678914509377765
18  1.02298019251626
19              <NA>
20              <NA>
21  8.82766887411449
22  6.76784804053789
23              <NA>
24              <NA>
25              <NA>
26              <NA>
```