

ReadMe.fil
07/19/17

Version 4.1

Documentation and user guide for SAS macro to project for absolute risk based on the relative risk models for (whites, hispanic, other), asian-american, or african-american. 1-AR, composite breast cancer incidences, competing hazards handling of missing covariate values and covariate editing procedures follow NCI BrCa Risk Assessment Tool (NCI BCRAT Ver4).

In this release of the SAS macro, in addition to the abs risk projection for the women with under investigation, for each women, an associated race specific abs risk projection for an "average" women is also provided. This quantity is included to follow the NCI Breast Cancer Risk Assessment Tool which provides an "avg" women risk projection as well.

Life time risk for a women can be obtained by setting her "projection age" to 90.

A simple 3 step example program (BCRAM_example.sas) on the use of the SAS macro (Br)east (Ca)ncer (R)isk (A)ssessment (M)acro -- BrCa_RAM.

Step 1: the included sas program BCRAM_example.sas reads the supplied data file

"Sample.in", which contains the Gail BrCa risk covarites and projection age interval for 26 hypothetical women. It then saves a temporary SAS system file with name of "ExampleIn" to be used as input to the SAS macro BrCa_RAM:

```
data ExampleIn;          *** name of the sas system file which
the macro parameter      &In_File should point to upon
macro invocation;

infile 'Sample.in' firstobs=9;  *** "Sample.in" is the
RR covariate input file      firstobs=9 skips
first 8 header                records on input
file "Sample.in";
*** SAS variable names;

input  IDD

      InitalAge
```

```

ProjtnAge

NBiop
HP

AgeMenarchy
AgeFstLive
Num_Rels

Ethnicity;

run;

```

Step 2: sas program BCRAM_example.sas runs the SAS macro BrCa_RAM:

```

%include "BrCa_RAM";          *** include the sas MACRO
BrCa_RAM;

```

Involking the sas macro BrCa_RAM to perform the BrCa projections.

The temporary sas input file is set to "ExampleIn".
The temporary sas output file is set to "ExampleOut".

The macro parameters WID, T1, T2, N_Biop, HyperPlasia, AgeMen, Agelst, N_Rels, and Race point to their corresponding sas variables on the sas file "ExampleIn", namely
IDD, InitalAge, ProjtnAge, NBiop, HP, AgeMenarchy, AgeFstLive, Num_Rels and Ethnicity respectively.

The macro parameter AbsRsk points to the sas variable Abs_Risk which will be added to the output sas file "ExampleOut". The output sas file will also contain all the variables on the input sas file.

	Macro parameter	pointing to	SAS file name or SAS variable name;
%BrCa_RAM	(In_File	=	ExampleIn ,
	Out_File	=	ExampleOut ,
	WID	=	IDD ,
	T1	=	InitalAge ,
	T2	=	ProjtnAge ,
	N_Biop	=	NBiop ,
	HyperPlasia	=	HP ,
	AgeMen	=	AgeMenarchy ,
	Agelst	=	AgeFstLive ,
	N_Rels	=	Num_Rels ,
	Race	=	Ethnicity ,
	CharRace	=	CharRace ,

```

RR_Star1      =      RR_Star1      ,
RR_Star2      =      RR_Star2      ,

AbsRsk        =      Absolute_Risk);

```

Step 3 It then lists the contents of the temporary output SAS system file "ExampleOut" which contains the projected absolute risk as well as the relative risk covariate values. Note that any further processing requiring the projected absolute risk, must be performed on the output SAS system file "ExampleOut" named in this sample program;

```

data ExampleOut;          *** output file from macro,
defined by pointing the;
set ExampleOut;          *** macro parameter &Out_File
to "ExampleOut";

file print;

if (_N_ eq 1) then do;
  put "
";
  put "          # Hypr   HP Age Age #
"
  "          "      RR      RR      Abs";
  put "          ID   T1   T2  Biop plas   RR Men 1st Rel
Race"
          " Age<50 Age>50   Risk(%)" ;
  put " ";
end;

*** all variables below take on their SAS variable names,
not their macro names;
*** see SAS variable names defined in Step 1;

if (_n_ le 100) then

  put IDD              7.0
  InitalAge           6.1
  ProjtnAge           6.1

  NBiop               6.0
  HP                  6.0
  R_Hyp               6.2

  AgeMenarchy         5.0
  AgeFstLive           5.0
  Num_Rels             5.0

  " "
  Ethnicity            2.0
  "="
  CharRace             $char2.

```

```

RR_Star1      8.4
RR_Star2      8.4
Absolute_Risk 10.4;
run;

```

Detailed description of the operation and output items from the SAS macro BrCa_RAM:

Input data:

In_File= should "point" to a SAS data set containing all the required input data items needed to perform risk projections, such as initial age, projection age, BrCa relative risk covariates and race. See the paragraph "Input data items ..." below, for a detailed description of all required data items.

Output data:

Out_File= should "point" to a SAS output data set which will contain the projected absolute risk of BrCa as well as the original input data items.

Macro structure:

Macro name	Macro parameters	"points" to SAS names
%macro BrCa_RAM	(In_File =, Out_File =, WID =, T1 =, T2 =, N_Biop =, HyperPlasia =, AgeMen =,	name of input sas data set name of output sas data set ID # 1,2,3 ... positive initial age, age at projection projection age, age at end projection # biopsies performed did biopsy exhibit atypical age at menarchy
integers		
beginning of interval		
of interval		
hyperplasia?		

```

brca
    Agelst      =,      age at 1st live birth
    N_Rels      =,      # 1st degree relatives with

race
    Race        =,      race
    CharRace    =,      2 character abbreviation for

RR_Star1      =,      rr for ages lt 50
RR_Star2      =,      rr for ages ge 50

AbsRsk        =);     projected absolute risk of
brca (%)

```

appropriate sas file/sas variable names must be associated with all macro parameters on the invocation of the sas macro "BrCa_RAM".

For example by coding "In_File = AARPin" tells the macro that the user created sas file "AARPin" is to be used for input of variables. Similarly coding "N_Biop = Num_Biops", lets the macro know that the sas variable "Num_Biops" in the sas input file "AARPin" contains the count of the # of biopsies performed.

To involve the sas macro in your sas program, an %include statement must be coded in your sas program, which points to the sas macro "BrCa_RAM".

For example:

```

the statement: %include "BrCa_RAM";           points to the sas macro
BrCa_RAM                                           stored in your current
directory

```

```

the statement: %include "c:\sas.macro\BrCa_RAM"; points to the sas macro
BrCa_RAM                                           stored in the directory
c:\sas.macro

```

Input data items needed to project for BrCa absolute risk and consistency requirements:

Macro parameter	Definition	Valid values
WID	ID # for each woman	postive integers 1,2,3....
T1	Initial age	all real numbers T1 in [20,90)
T2	BrCa projection age	all real numbers T2 such that T2 > T1
CONSTRAINT on T1 and T2:		20 <= T1 < T2 <= 90

N_Biop recoded to 0)	# of biopsies	0,1,2 ...	99=unk (99
HyperPlasia biopsy	Did biopsy display atypical hyperplasia?	0=no, 1=yes,	99=unk or no
AgeMen or equal to T1,	Age at menarchy 99=unk	positive integer age less than	
Age1st age at menarchy initial age.	Age at first live birth	integer age greater or equal to and less than or equal to	
			98=nulliparous (no live birth), 99=unk
N_Rels	# 1st degree relatives with BrCa	0,1,2 ...	99=unk
Race (native americans and unknown race)	Race	1=Wh non-hispanic white 2=AA african-american, 3=HU US born hispanic, 4=NA other	
 (asian-pacific islanders)		5=HF Foreign born hispanic 6=Ch chinese	
		7=Ja japanese 8=Fi filipino 9=Hw hawaiian 10=oP other pacific islander 11=oA other asian	
log relative cancer composite based on white		note that other ethnic women risks are based on white women risks. other women breast incidence and competing risk are women seer rates.	

Recoding and checking of relative risk covariate values performed by
"BrCa_RAM":

	raw value	recoded
to		
N_Biop: # biopsies	0 or 99	
0		

		1
1		2,3,4 ... and not 99
2		
AgeMen: age at menarchy		14,15,16 ... 99
0		12,13
1		11 and younger
2		
Agelst: age at 1st live birth		19 and younger or 99
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: # 1st degree rel with BrCa		0 or 99
0		1
1		2,3,4 ... and not 99
2		

Consistency patterns for # of Biopsies and Hyperplasia:

Requirment: (A) N_Biops = 0 or 99 then Hyperplasia MUST = 99 (not applicable)
 (B) N_Biops > 0 and < 99 then Hyperplasia = 0, 1 or 99 (unk)

if ANY of the above 2 REQUIREMENTS are violated, the absolute risk will be set to the sas missing value ".". The consequences to the relative risk (RR) for the above two requirements is:

(A) # biopsies = 0 or 99	&	Hyperplasia =99 (not applicable)	inflates RR
by 1.00			
(B) # biopsies > 0 and <99	&	Hyperplasia = 0 (no hyprplasia)	inflates RR
by 0.93			
		= 1 (yes hyprplasia)	inflates RR
by 1.82			
		=99 (unk hyprplasia)	inflates RR
by 1.00			

Edit checking for remaining relative risk covariates, AgeMen, Age1st and N_Rels:

AgeMen: age at menarchy must be positive integer less than 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 positive 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: # of 1st degree relatives with BrCa must be 0,1,2...

Following is a listing of the sample raw input data set "Sample.in" (column heading included for clarity):

IDD	T1	T2	Num Biop	Hyp Plas	Age Men	Age 1st	Num Rel	Race
1	45.2	53.3	99	99	10	20	1	0
2	45.2	53.3	99	1	10	20	1	1
3	45.2	53.3	99	0	10	20	1	2
4	45.2	53.3	0	99	10	20	1	3
5	45.2	53.3	1	99	10	20	1	4
6	45.2	53.3	1	99	14	19	1	5
7	45.2	53.3	99	99	99	19	1	6
8	45.2	53.3	1	1	14	19	1	7
9	45.2	53.3	99	1	14	99	1	8
10	45.2	53.3	1	0	14	19	1	9
11	45.2	53.3	99	0	99	99	1	10
12	45.2	53.3	0	0	14	19	1	11
13	45.2	53.3	0	99	10	20	1	12
14	45.2	53.3	0	1	10	20	1	0
15	45.2	53.3	0	0	10	20	1	1
16	45.2	53.3	1	0	10	20	1	2
17	35.0	40.0	4	99	11	25	0	3
18	35.0	40.0	4	99	11	98	0	4
19	35.0	40.0	4	99	11	10	0	5
20	35.0	40.0	4	99	36	25	0	6
21	27.0	90.0	99	99	13	22	0	7
22	27.0	90.0	99	99	13	22	99	8
23	18.0	26.0	99	99	13	22	99	9
24	27.0	26.0	99	99	13	22	99	10
25	85.0	91.0	99	99	13	22	99	11
26	86.0	90.0	99	99	13	22	99	12

1	45.2	53.3	99	99	1.00	10	20	1	0	.	.
.	29										
	45.2	53.3	0	99	1.00	2	1	1	??		
2	45.2	53.3	99	1	.	10	20	1	1	.	.
.	.										
	45.2	53.3	A	A	A	2	1	1	Wh		
3	45.2	53.3	99	0	.	10	20	1	2	.	.
.	.										
	45.2	53.3	A	A	A	1	0	1	AA		
9	45.2	53.3	99	1	.	14	99	1	8	.	.
.	.										
	45.2	53.3	A	A	A	0	0	1	Fi		
11	45.2	53.3	99	0	.	99	99	1	10	.	.
.	.										
	45.2	53.3	A	A	A	0	0	1	oP		
12	45.2	53.3	0	0	.	14	19	1	11	.	.
.	.										
	45.2	53.3	A	A	A	0	0	1	oA		
13	45.2	53.3	0	99	1.00	10	20	1	12	.	.
.	29										
	45.2	53.3	0	99	1.00	2	1	1	??		
14	45.2	53.3	0	1	.	10	20	1	0	.	.
.	.										
	45.2	53.3	A	A	A	2	1	1	??		
15	45.2	53.3	0	0	.	10	20	1	1	.	.
.	.										
	45.2	53.3	A	A	A	2	1	1	Wh		
19	35.0	40.0	4	99	1.00	11	10	0	5	.	.
.	.										
	35.0	40.0	2	99	1.00	2	.	0	Wo		
20	35.0	40.0	4	99	1.00	36	25	0	6	.	.
.	.										
	35.0	40.0	2	99	1.00	.	.	0	Ch		
23	18.0	26.0	99	99	1.00	13	22	99	9	.	.
.	.										
	.	26.0	0	99	1.00	1	.	0	Hw		
24	27.0	26.0	99	99	1.00	13	22	99	10	1.42	1.42
.	16										
	.	.	0	99	1.00	1	1	0	oP		
25	85.0	91.0	99	99	1.00	13	22	99	11	1.42	1.42
.	16										
	85.0	.	0	99	1.00	1	1	0	oA		

```

26  86.0  90.0   99   99  1.00  13  22  99   12   .   .
.   16
.   86.0  90.0   0   99  1.00   1   1   0   ??

```

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. 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). Similarly for IDs 3,9,11,12,14 and 15 which display violations of Requirements (A). For IDs 19 and 20, violation of AgeMen and/or Age1st consistency are seen. Note the SAS missing value "." listed under AgeMen and/or Age1st. For IDs 23, 24 and 25 violation of T1 and/or T2 consistency requirements are seen. Again, note the "." listed under T1 and/or T2. This small sample data set "Sample.in" 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.

Finally, the listing from Step3:

Listing of the first 100 records in temporary output sas system file ExampleOut
Further analysis depending on the projected abs risk must be performed using the output sas system file which is invoked by the sas macro parameter 'Out_File'

RR	AbsRisk	AbsRisk	#	Hypr	HP	Age	Age	#		RR	
Age>=50	ID	T1	T2	Biop	plas	RR	Men	1st	Rel	Race	Age<50
			(%)	AvgWm(%)							
	1	45.2	53.3	99	99	1.00	10	20	1	0=??	. .
	2	45.2	53.3	99	1	.	10	20	1	1=Wh	. .
	3	45.2	53.3	99	0	.	10	20	1	2=AA	. .
	4	45.2	53.3	0	99	1.00	10	20	1	3=Hi	3.2354
3.2354		2.1081		1.1313							
	5	45.2	53.3	1	99	1.00	10	20	1	4=NA	5.4926
4.1180		4.4413		1.7673							
	6	45.2	53.3	1	99	1.00	14	19	1	5=Wo	4.4263
3.3185		3.9762		1.7673							
	7	45.2	53.3	99	99	1.00	99	19	1	6=Ch	2.2075
2.2075		1.2496		1.1644							

8	45.2	53.3	1	1	1.82	14	19	1	7=Ja	6.9820	
6.9820	5.7757		1.7279								
9	45.2	53.3	99	1	.	14	99	1	8=Fi	.	.
.	.										
10	45.2	53.3	1	0	0.93	14	19	1	9=Hw	3.5677	
3.5677	3.9061		2.2614								
11	45.2	53.3	99	0	.	99	99	1	10=oP	.	.
.	.										
12	45.2	53.3	0	0	.	14	19	1	11=oA	.	.
.	.										
13	45.2	53.3	0	99	1.00	10	20	1	12=??	.	.
.	.										
14	45.2	53.3	0	1	.	10	20	1	0=??	.	.
.	.										
15	45.2	53.3	0	0	.	10	20	1	1=Wh	.	.
.	.										
16	45.2	53.3	1	0	0.93	10	20	1	2=AA	2.3458	
2.0974	2.6899		1.6479								
17	35.0	40.0	4	99	1.00	11	25	0	3=Hi	5.3860	
3.0274	0.6789		0.2183								
18	35.0	40.0	4	99	1.00	11	98	0	4=NA	5.3860	
3.0274	1.0230		0.2814								
19	35.0	40.0	4	99	1.00	11	10	0	5=Wo	.	.
.	.										
20	35.0	40.0	4	99	1.00	36	25	0	6=Ch	.	.
.	.										
21	27.0	90.0	99	99	1.00	13	22	0	7=Ja	1.4210	
1.4210	8.8277		12.2076								
22	27.0	90.0	99	99	1.00	13	22	99	8=Fi	1.4210	
1.4210	6.7678		9.4245								
23	18.0	26.0	99	99	1.00	13	22	99	9=Hw	.	.
.	.										
24	27.0	26.0	99	99	1.00	13	22	99	10=oP	1.4210	
1.4210	.		.								
25	85.0	91.0	99	99	1.00	13	22	99	11=oA	1.4210	
1.4210	.		.								
26	86.0	90.0	99	99	1.00	13	22	99	12=??	.	.
.	.										

Statistical issues should be directed to: Dr. Mitchell Gail
gailm@exchange.nih.gov
Technical details should be directed to: Mr. David Pee
djkeno88@yahoo.com