**Program 14.2 SAS Code for Propensity Score Bin Bootstrapping Analysis**

TITLE 'PSBB ANALYSIS';

PROC PRINTTO LOG='D:\TEMP\PSBBLog.log' NEW; RUN;

\*\* Assign summary statistics utilized in analysis code below \*\*;

\*\* Values not computed here to focus code on cost analysis steps \*\*;

%let nc1 = 210; \* Sample size for group B with cost data \*;

%let nc2 = 223; \* Sample size for group A with cost data \*;

%let ne1 = 200; \* Sample size for group B with effectiveness data \*;

%let ne2 = 218; \* Sample size for group A with effectiveness data \*;

%let mc1 = 21227; \* Mean of cost variable for group B \*;

%let mc2 = 20864; \* Mean of cost variable for group A \*;

%let me1 = 123.74; \* Mean of effectiveness variable for group B \*;

%let me2 = 129.23; \* Mean of effectiveness variable for group A \*;

\*\* Compute statistic used later in bootstrap CI calculations \*\*;

data icer;

set icer;

\* Compute score statistic \*;

if therapy = 'A' then do;

mnc1 = (&nc2\*&mc2 - totcost) / (&nc2 - **1**);

\* Mean of group without current obs \*;

mne1 = (&ne2\*&me2 - respdays) / (&ne2 - 1);

\* Mean of group without current obs \*;

diffc1 = mnc1 - &mc2;

\* Mean diff between groups without current obs \*;

diffe1 = mne1 - &me2;

\* Mean diff between groups without current obs \*;

uc = (&mc2 - &mc1 - diffc1);

ue = (&me2 - &me1 - diffe1);

\* Diff between overall and estimate without curren obs \*;

end;

if therapy = 'B' then do;

mnc1 = (&nc1\*&mc1 - totcost) / (&nc1 - 1);

\* Mean of group without current obs \*;

mne1 = (&ne1\*&me1 - respdays) / (&ne1 - 1);

\* Mean of group without current obs \*;

diffc1 = &mc1 - mnc1;

\* Mean diff between groups without current obs \*;

diffe1 = &me1 - mne1;

\* Mean diff between groups without current obs \*;

uc = (&mc2 - &mc1 - diffc1);

ue = (&me2 - &me1 - diffe1);

\* Diff between overall and estimate without current obs \*;

end;

if therapy = 'A' then ther = 1;

if therapy = 'B' then ther = 0;

run;

/\*\* Compute acceleration constant for later BCa CI calculations \*\*/

data accel;

set icer;

dm = 1;

uc\_cub + uc\*\*3;

ue\_cub + ue\*\*3;

uc\_sqr + uc\*\*2;

ue\_sqr + ue\*\*2;

keep patient dm uc ue uc\_cub ue\_cub uc\_sqr ue\_sqr;

run;

proc sort data = accel; by dm;

run;

data accel2;

set accel;

by dm;

if last.dm;

c\_aconst = uc\_cub / (((uc\_sqr\*\*1.5))\*6);

e\_aconst = ue\_cub / (((ue\_sqr\*\*1.5))\*6);

run;

\*\*Assign the c\_aconst and e\_aconst to macro variables for BCa calculation in macro PSBB \*\*;

data \_null\_; set accel2;

call symput('c\_aconst', trim(left(c\_aconst)));

call symput('e\_aconst', trim(left(e\_aconst)));

run;

%put c\_aconst=&c\_aconst e\_aconst=&e\_aconst;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\* Compute propensity score strata \*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

option spool;

ods listing close;

proc genmod data = icer;

class inv inptatst subsabdx bs\_bprsc insured;

model therapy = inv bs\_bprsc age inptatst subsabdx psycdur hospestmo

insured / dist = bin link = logit type3 obstats;

output out=pred6 pred = prdct;

run;

ods listing;

data premab;

set pred6;

predmo = 1-prdct;

predmc = prdct;

keep patient predmo predmc therapy ther totcost respdays age

gender inptatst subsabdx insured hospestmo psycdur bs\_bprsc;

run;

proc rank data = premab groups = 5 out = rankmab;

ranks rnkm\_ab;

var predmo;

run;

data rankmab;

set rankmab;

bin\_ps = rnkm\_ab + 1;

run;

proc sort data = rankmab;

by therapy;

run;

proc univariate data = rankmab;

by therapy;

var predmo;

title 'Distribution of propensity scores: oc';

run;

proc freq data = rankmab;

tables bin\_ps\*therapy;

title 'Therapy distribution among bins';

run;

proc tabulate data = rankmab;

class therapy bin\_ps;

var respdays totcost;

tables bin\_ps\*therapy,(respdays totcost)\*(n\*format=3. mean std);

title 'Summary of costs/responder days by bin and therapy';

run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Macro PSBB is for a propensity score bin bootstrap analysis

Inputs:

REP = Number of bootstrap samples

AVARC = Variable for Total Costs

AVARE = Variable for Effectiveness (response days calculated for BPRS)

INDAT = Data set to be analyzed

GRPVAR = Variable for therapy group number

FSEED0 = Starting randomization seed for therapy group 0

FSEED1 = Starting randomization seed for therapy group 1

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

%MACRO

PSBB(rep=,avarc=,avare=,indat=,grpvar=,fseed0=887583,fseed1=566126);

data temp; set &indat;

run;

proc freq data=temp noprint;

tables &grpvar / out=freqnums;

where not(&grpvar=.);

run;

data \_null\_; set freqnums;

call symput('val'||compress(put(\_n\_, 4.)), trim(left(&grpvar)));

call symput('ssize'||compress(put(\_n\_, 4.)), trim(left(count)));

run;

%\* Create data sets for each treatment \*;

data trt0 trt1;

set temp;

if (&grpvar=&val1) then output trt0;

else if (&grpvar=&val2) then output trt1;

keep &grpvar &avarc &avare bin\_ps;

run;

data bssumm; %\* Empty data set to add to later \*;

if \_n\_ eq 1 then stop;

run;

proc sort data=trt0; by &grpvar bin\_ps;

run;

proc sort data=trt1; by &grpvar bin\_ps;

run;

%do i=1 %to &rep;

%\*\* Generate random bootstrap sample data set for therapy0\*\*;

%\* Perform bootstrap resampling \*;

%let btnum=%qsysfunc(round(&ssize1/5,1));

%let rseed=%qsysfunc(round(&i + &fseed0, 1));

proc surveyselect data=trt0 method=urs outhits rep=1

n=&btnum. seed=&rseed. noprint out=trt0out;

strata &grpvar bin\_ps;

run;

%\*\* Generate random bootstrap sample data set for therapy1\*\*;

%let btnum=%qsysfunc(round(&ssize2/5,1));

%let rseed=%qsysfunc(round(&i + &fseed1, 1));

proc surveyselect data=trt1 method=urs outhits rep=1

n=&btnum. seed=&rseed. noprint out=trt1out;

strata &grpvar bin\_ps;

run;

data bothgrps;

set trt0out trt1out;

run;

%\*\* Compute overall statistics for the sample \*\*;

proc means data=bothgrps noprint;

class &grpvar;

var &avarc &avare;

output out = mn mean = out\_avgc out\_avge;

data mn; set mn end=eof;

label &avarc.\_avg1 = "Average for &AVARC, Group=&VAL1"

&avare.\_avg1 = "Average for &AVARE, Group=&VAL1"

&avarc.\_avg2 = "Average for &AVARC, Group=&VAL2"

&avare.\_avg2 = "Average for &AVARE, Group=&VAL2";

dumm= 1;

retain &avarc.\_avg1 &avare.\_avg1 &avarc.\_avg2 &avare.\_avg2;

if &grpvar=0 then do;

&avarc.\_avg1=out\_avgc;

&avare.\_avg1=out\_avge;

end;

if &grpvar=1 then do;

&avarc.\_avg2=out\_avgc;

&avare.\_avg2=out\_avge;

end;

keep dumm &avarc.\_avg1 &avare.\_avg1 &avarc.\_avg2 &avare.\_avg2;

if eof then output;

run;

%\*\* Update data set with statistics from this sample \*\*;

data bssumm;

set bssumm mn;

run;

%\*\*Clean work library\*\*;

proc datasets library=work memtype=data nolist;

delete trt0out trt01ut bothgrps mn;

run;

quit;

%end; %\* End of %do loop \*;

%\*\* Compute differences and test statistics \*\*;

data bssumm;

set bssumm;

&avare.\_diff = &avare.\_avg2 - &avare.\_avg1;

&avarc.\_diff = &avarc.\_avg2 - &avarc.\_avg1;

if &avarc.\_diff ne . and &avare.\_diff ne . then do;

if &avarc.\_diff ge 0 and &avare.\_diff ge 0 then ce\_quad = '++';

if &avarc.\_diff ge 0 and &avare.\_diff lt 0 then ce\_quad = '+-';

if &avarc.\_diff lt 0 and &avare.\_diff ge 0 then ce\_quad = '-+';

if &avarc.\_diff lt 0 and &avare.\_diff lt 0 then ce\_quad = '--';

end;

if &avarc.\_diff lt (&mc2 - &mc1) then zzeroctc + 1;

if &avare.\_diff lt (&me2 - &me1) then zzerocte + 1;

label &avare.\_diff="Average for &AVARE Diff: Grp2-Grp1"

&avarc.\_diff="Average for &AVARC Diff: Grp2-Grp1";

run;

\*Calculate quadrants percentage and assign macro variable for graph \*\*;

ods output OneWayFreqs=quadrt(keep=ce\_quad percent);

proc freq data = bssumm;

tables ce\_quad;

title2 "Quadrants distribution for cost effectiveness";

title3 "Variables &avarc and avare";

run;

data \_null\_; set quadrt;

if ce\_quad='++' then call symput('pospos', compress(percent));

if ce\_quad='+-' then call symput('posneg', compress(percent));

if ce\_quad='-+' then call symput('negpos', compress(percent));

if ce\_quad='--' then call symput('negneg', compress(percent));

run;

proc univariate data=bssumm freq noprint;

var &avarc.\_diff &avare.\_diff;

output out=pctls pctlpts=2.5 97.5 pctlpre = &avarc &avare

pctlname=\_lcl \_ucl;

run;

proc print data=pctls;

title2 "Bootstrap Percentile 95% confidence limits for &avarc and

&avare"; run;

\*\* Compute BCa confidence intervals \*\*;

data zerodat;

set bssumm;

by dumm;

if last.dumm;

keep zzeroctc zzerocte;

run;

data bcacalc;

set zerodat;

zzeroc = probit( zzeroctc / &rep );

zzeroe = probit( zzerocte / &rep );

zzl = probit(.025);

zzh = probit(.975);

bcaclo = zzeroc + ((zzeroc + zzl) / (1 - &c\_aconst.\*(zzeroc + zzl)));

bcachi = zzeroc + ((zzeroc + zzh) / (1 - &c\_aconst.\*(zzeroc + zzh)));

bcaelo = zzeroe + ((zzeroe + zzl) / (1 - &e\_aconst.\*(zzeroe + zzl)));

bcaehi = zzeroe + ((zzeroe + zzh) / (1 - &e\_aconst.\*(zzeroe + zzh)));

bcacl = probnorm(bcaclo);

bcach = probnorm(bcachi);

bcael = probnorm(bcaelo);

bcaeh = probnorm(bcaehi);

run;

data \_null\_; set bcacalc;

call symput('bcacl', trim(left(bcacl\*100)));

call symput('bcach', trim(left(bcach\*100)));

call symput('bcael', trim(left(bcael\*100)));

call symput('bcaeh', trim(left(bcaeh\*100)));

run;

%put bcacl=&bcacl bcach=&bcach bcael=&bcael baceh=&bcaeh;

proc univariate data=bssumm freq noprint;

var &avarc.\_diff;

output out=pctls2 pctlpts=&bcacl. &bcach. pctlpre = &avarc

pctlname=\_lcl \_ucl;

run;

proc print data=pctls2;

title2 "BCa bootstrap 95% confidence limits for &avarc";

run;

proc univariate data=bssumm freq noprint;

var &avare.\_diff;

output out=pctls3 pctlpts=&bcael. &bcaeh. pctlpre = &avare

pctlname=\_lcl \_ucl;

run;

proc print data=pctls3;

title2 "BCa bootstrap 95% confidence limits for &avare";

run;

%\*\* Create graph of bootstrap ce \*\*;

axis1 label=(h=1.5 c=black a=90 "Effectiveness Difference: A - B"

J=CENTER) value=(h=1.5 c=black) ;

axis2 label=(h=1.5 c=black "Cost Difference: A - B" J=CENTER)

value=(h=1.5 c=black) ;

proc gplot data=bssumm;

plot &avare.\_diff\*&avarc.\_diff = '\*'/nolegend haxis=axis2

vaxis=axis1

href=0

vref=0 ;

%\*\*Add quadrant frequency percentage \*\*;

note height=1.75 m=(80pct,80pct) "&pospos.%";

note height=1.75 m=(80pct,30pct) "&posneg.%";

note height=1.75 m=(20pct,30pct) "&negneg.%";

note height=1.75 m=(20pct,80pct) "&negpos.%";

title1 h=2.5 lspace=1 "Quadrant distribution for cost effectiveness";

run; quit;

%\*\*Clean work library\*\*;

proc datasets library=work memtype=data nolist;

delete temp freqnums trt0 trt1 pctls pctls2 pctls3 zerodat;

run; quit;

goptions reset=all;

%MEND PSBB; /\* End of macro psbb \*/

/\* Call the bootstrap macro \*/

filename myfile1 "D:\Temp\ICER\_TOTCOST\_RDBPRS\_DIFF2.gif";

goptions reset=all device=gif gsfname=MYFILE1 gsfmode=replace htext=1 ftext=swiss rotate=landscape noborder;

%*PSBB*(rep=10000,avarc=totcost,avare=respdays,indat=rankmab,grpvar=ther);

quit;

goptions reset=all;

\*\*\*Draw histogram of mean difference in costs\*\*\*;

data bssumm; set bssumm;

totcost\_diff1=totcost\_diff/1000;

\*\*resize the cost to show in the graph x-axis\*\*;

run;

filename myfile2 "D:\Temp\ICER\_TOTCOST\_DIFF2.gif";

goptions reset=all device=gif gsfname=MYFILE2 gsfmode=replace htext=1.25 ftext='arial/bo' rotate=landscape noborder;

title ' ';

footnote ' ';

pattern1 v=solid c=black;

footnote1 h=1.5 "Mean Difference in Costs: A - B (thousand $)";

AXIS1 LABEL=(H=2 C=BLACK angle=90 "Frequency" J=CENTER) value=(H=1.5

C=BLACK) order=(0 to 1600 by 200);

AXIS2 LABEL=(H=2 C=BLACK J=CENTER ' ') ;

PROC GCHART data=bssumm;

VBAR totcost\_diff1/ref=(0 to 1600 by 200) midpoints=(-12 to 12 by 1) raxis=axis1 maxis=axis2 space=5 width=2;

run; quit;

goptions reset=all;

PROC PRINTTO; RUN;