**eAppendix 1**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

EAPPENDIX1.SAS

Program:

Creates a simulated dataset of 10,000 study participants.

Randomly misclassifies exposure.

Adjusts for exposure misclassification given inputs.

Programmer:

Candice Johnson

17 MAY 2014

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

CREATE SIMULATED STUDY POPULATION BASED ON CONTINGENCY TABLE FROM MANUSCRIPT TABLE 1

Numbers included are for the OR = 1.29 example

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** participants;

do id = **1** to **125**;

exposure = **1**; disease = **1**;

output;

end;

do id = **126** to **1000**;

exposure = **0**; disease = **1**;

output;

end;

do id = **1001** to **1900**;

exposure = **1**; disease = **0**;

output;

end;

do id = **1901** to **10000**;

exposure = **0**; disease = **0**;

output;

end;

run;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

MISCLASSIFY EXPOSURE

Run the section of code corresponding to the misclassification type wanted to be used as "truth"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

EXACTLY NONDIFFERENTIAL MISCLASSIFICATION

Misclassify the cases, apply the resultant Se and Sp to the non-cases

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*Misclassifies the cases;

**data** cases\_misclassify;

set participants;

if (disease = **1**);

retain seed **88846574**;

call ranbin(seed,**1**,**0.1**,misc);

if (misc = **1**) then do;

if (exposure = **1**) then mexposure = **0**;

else if (exposure = **0**) then mexposure = **1**;

end;

else if (misc = **0**) then mexposure = exposure;

run;

\*Find out what the Se and Sp are;

**proc** **freq** data = cases\_misclassify;

tables mexposure\*exposure / missprint;

**run**;

\*Calculate what the exposure classification table should be if the non-cases had this Se and Sp;

\*This ensures that misclassification is exactly nondifferential;

**data** noncases;

disease = **0**;

do id = **1001** to **1821**;

exposure = **1**;

mexposure = **1**;

output;

end;

do id = **1822** to **2821**;

exposure = **0**;

mexposure = **1**;

output;

end;

do id = **2822** to **2900**;

exposure = **1**;

mexposure = **0**;

output;

end;

do id = **2901** to **10000**;

exposure = **0**;

mexposure = **0**;

output;

end;

run;

\*Check that the Se and Sp for non-cases are the same as for cases;

**proc** **freq** data = noncases;

tables mexposure\*exposure / missprint;

**run**;

\*Combine cases and non-cases into one dataset;

**data** participants\_misclassify;

set cases\_misclassify noncases;

run;

\*Check the final Se and Sp;

**proc** **freq** data = participants\_misclassify;

tables mexposure\*exposure / missprint;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

APPROXIMATELY NONDIFFERENTIAL MISCLASSIFICATION

Misclassify participants without regard to case status

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** participants\_misclassify;

set participants;

retain seed **88846574**;

call ranbin(seed,**1**,**0.1**,misc);

if (misc = **1**) then do;

if (exposure = **1**) then mexposure = **0**;

else if (exposure = **0**) then mexposure = **1**;

end;

else if (misc = **0**) then mexposure = exposure;

run;

\*Check the Se and Sp;

**proc** **freq** data = participants\_misclassify;

tables mexposure\*exposure / missprint;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

DIFFERENTIAL A MISCLASSIFICATION

Misclassify cases and non-cases separately, according to assumptions

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** participants\_misclassify;

set participants;

retain seed **88846574**;

if (disease = **0**) then do;

call ranbin(seed,**1**,**0.1**,misc);

end;

else if (disease = **1**) then do;

call ranbin(seed,**1**,**0.05**,misc);

end;

if (misc = **1**) then do;

if (exposure = **1**) then mexposure = **0**;

else if (exposure = **0**) then mexposure = **1**;

end;

else if (misc = **0**) then mexposure = exposure;

run;

\*Check final Se and Sp;

**proc** **freq** data = participants\_misclassify;

tables disease\*mexposure\*exposure / missprint;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

DIFFERENTIAL B MISCLASSIFICATION

Misclassify cases and non-cases separately, according to assumptions

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** participants\_misclassify;

set participants;

retain seed **88846574**;

if (disease = **1**) then do;

call ranbin(seed,**1**,**0.15**,misc);

end;

else if (disease = **0**) then do;

call ranbin(seed,**1**,**0.1**,misc);

end;

if (misc = **1**) then do;

if (exposure = **1**) then mexposure = **0**;

else if (exposure = **0**) then mexposure = **1**;

end;

else if (misc = **0**) then mexposure = exposure;

run;

\*Check final Se and Sp;

**proc** **freq** data = participants\_misclassify;

tables disease\*mexposure\*exposure / missprint;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

PROBABILISTIC ADJUSTMENTS

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*Get the counts for the misclassified contingency table;

**proc** **freq** data = participants\_misclassify;

tables disease\*mexposure / missprint;

**run**;

\*Get the true OR;

**proc** **freq** data = participants\_misclassify;

tables disease\*exposure / missprint relrisk;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

ADJUST ASSUMING EXACTLY NONDIFFERENTIAL MISCLASSIFICATION

Choose the same values of Se and Sp for cases and non-cases

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** exact;

retain seed **13579**;

\*Enter the contingency table cells for the misclassified exposure and the outcome;

a = **264**; \*exposure = 1, disease = 1;

b = **736**; \*exposure = 0, disease = 1;

c = **1640**; \*exposure = 1, disease = 0;

d = **7360**; \*exposure = 0, disease = 0;

\*Enter the true (non-misclassified) odds ratio to calculate the ROR;

trueor = **1.29**;

\*Enter the Se and Sp values (cases and non-cases get same value);

se = **0.9022**;

sp = **0.8978**;

\*Define minimum and maximum of the triangular distributions, here +/- 0.05 of the value;

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

semin = se - **0.05**;

semax = se + **0.05**;

spmin = sp - **0.05**;

spmax = sp + **0.05**;

\*10,000 simulations;

do i = **1** to **10000**;

\*Choose an Se from the triangular distribution for Se;

call rantri(seed, (se-semin)/(semax-semin), tempse);

setri = (semax-semin)\*tempse + semin;

\*Choose an Sp from the triangular distribution for Sp;

call rantri(seed, (sp-spmin)/(spmax-spmin), tempsp);

sptri = (spmax-spmin)\*tempsp + spmin;

\*Adjust cells of contingency table using Se and Sp values chosen from distributions - same values for cases and noncases;

newa = (a - (**1**-sptri)\*(a+b)) / (setri+sptri-**1**);

newc = (c - (**1**-sptri)\*(c+d)) / (setri+sptri-**1**);

newb = (a+b)-newa;

newd = (c+d)-newc;

\*Calculate the adjusted OR;

adjor = (newa\*newd)/(newb\*newc);

\*Calculate the ratio of odds ratios;

ror = adjor/trueor;

\*Check if the ROR is negative, these will be excluded from the results;

if (ror lt **0**) then countneg = **1**;

else if (ror ge **0**) then countneg = **0**;

output;

end;

run;

\*Check how many values were negative and will be excluded;

**proc** **freq** data = exact;

tables countneg / missprint;

**run**;

\*Find the median adjusted OR and ROR and the 95% simulation interval;

**proc** **univariate** data = exact noprint;

where countneg ne **1**;

var adjor ror;

output out = exactresults pctlpts = **2.5** **50** **97.5** pctlpre = adjor\_ ror\_;

**run**;

\*Results;

**proc** **print** data = exactresults;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

ADUST ASSUMING APPROXIMATELY NONDIFFERENTIAL MISCLASSIFICATION

Choose Se and Sp for cases and non-cases from the same distribution

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** approx;

retain seed **13579**;

\*Enter the contingency table cells for the misclassified exposure and the outcome;

a = **264**; \*exposure = 1, disease = 1;

b = **736**; \*exposure = 0, disease = 1;

c = **1640**; \*exposure = 1, disease = 0;

d = **7360**; \*exposure = 0, disease = 0;

\*Enter the true (non-misclassified) odds ratio to calculate the ROR;

trueor = **1.29**;

\*Assign Se and Sp for non-cases;

se0 = **0.9022**;

sp0 = **0.8978**;

\*Define minimum and maximum of the triangular distributions, here +/- 0.05 of the value;

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

semin0 = se0 - **0.05**;

semax0 = se0 + **0.05**;

spmin0 = sp0 - **0.05**;

spmax0 = sp0 + **0.05**;

\*Assign Se and Sp for cases;

se1 = **0.9022**;

sp1 = **0.8978**;

\*Define minimum and maximum of the triangular distributions, here +/- 0.05 of the value;

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

semin1 = se1 - **0.05**;

semax1 = se1 + **0.05**;

spmin1 = sp1 - **0.05**;

spmax1 = sp1 + **0.05**;

\*10,000 simulations;

do i = **1** to **10000**;

\*Choose Se, Sp from triangular distribution for non-cases;

call rantri(seed, (se0-semin0)/(semax0-semin0), tempse0);

setri0 = (semax0-semin0)\*tempse0 + semin0;

call rantri(seed, (sp0-spmin0)/(spmax0-spmin0), tempsp0);

sptri0 = (spmax0-spmin0)\*tempsp0 + spmin0;

\*Choose Se, Sp from triangular distribution for cases;

call rantri(seed, (se1-semin1)/(semax1-semin1), tempse1);

setri1 = (semax1-semin1)\*tempse1 + semin1;

call rantri(seed, (sp1-spmin1)/(spmax1-spmin1), tempsp1);

sptri1 = (spmax1-spmin1)\*tempsp1 + spmin1;

\*Adjust OR based on selected Se and Sp values;

newa = (a - (**1**-sptri1)\*(a+b)) / (setri1+sptri1-**1**);

newc = (c - (**1**-sptri0)\*(c+d)) / (setri0+sptri0-**1**);

newb = (a+b)-newa;

newd = (c+d)-newc;

\*Calculate adjusted OR;

adjor = (newa\*newd)/(newb\*newc);

\*Calculate ratio of odds ratios;

ror = adjor/trueor;

\*Identify negative RORs;

if (ror lt **0**) then countneg = **1**;

else if (ror ge **0**) then countneg = **0**;

output;

end;

run;

\*Check how many RORs are negative and will be excluded;

**proc** **freq** data = approx;

tables countneg / missprint;

**run**;

\*Median adjusted OR, ROR and 95% simulation intervals;

**proc** **univariate** data = approx noprint;

where countneg ne **1**;

var adjor ror;

output out = approxresults pctlpts = **2.5** **50** **97.5** pctlpre = adjor\_ ror\_;

**run**;

\*Results;

**proc** **print** data = approxresults;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

ADJUST ASSUMING DIFFERENTIAL A MISCLASSIFICATION

Assign Se and Sp to cases and non-cases separately

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** diffa;

retain seed **13579**;

\*Enter the contingency table cells for the misclassified exposure and the outcome;

a = **264**; \*exposure = 1, disease = 1;

b = **736**; \*exposure = 0, disease = 1;

c = **1640**; \*exposure = 1, disease = 0;

d = **7360**; \*exposure = 0, disease = 0;

\*Enter the true (non-misclassified) odds ratio to calculate the ROR;

trueor = **1.29**;

\*Assign Se and Sp to non-cases;

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

se0 = **0.9022**;

sp0 = **0.8978**;

semin0 = se0 - **0.05**;

semax0 = se0 + **0.05**;

spmin0 = sp0 - **0.05**;

spmax0 = sp0 + **0.05**;

\*Assign Se and Sp to cases (0.05 higher than non-cases, or correct values if making correct assumption);

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

se1 = **0.9522**;

sp1 = **0.9478**;

semin1 = se1 - **0.05**;

semax1 = **1**;

spmin1 = sp1 - **0.05**;

spmax1 = sp1 + **0.05**;

\*10,000 simulations;

do i = **1** to **10000**;

\*Find Se and Sp for non-cases from triangular distributions;

call rantri(seed, (se0-semin0)/(semax0-semin0), tempse0);

setri0 = (semax0-semin0)\*tempse0 + semin0;

call rantri(seed, (sp0-spmin0)/(spmax0-spmin0), tempsp0);

sptri0 = (spmax0-spmin0)\*tempsp0 + spmin0;

\*Find Se and Sp for cases from triangular distributions;

call rantri(seed, (se1-semin1)/(semax1-semin1), tempse1);

setri1 = (semax1-semin1)\*tempse1 + semin1;

call rantri(seed, (sp1-spmin1)/(spmax1-spmin1), tempsp1);

sptri1 = (spmax1-spmin1)\*tempsp1 + spmin1;

\*Adjust cells based on Se and Sp values selected;

newa = (a - (**1**-sptri1)\*(a+b)) / (setri1+sptri1-**1**);

newc = (c - (**1**-sptri0)\*(c+d)) / (setri0+sptri0-**1**);

newb = (a+b)-newa;

newd = (c+d)-newc;

\*Calculate adjusted OR;

adjor = (newa\*newd)/(newb\*newc);

\*Calculate ratio of odds ratios;

ror = adjor/trueor;

\*Check if the ROR is negative, to later exclude it;

if (ror lt **0**) then countneg = **1**;

else if (ror ge **0**) then countneg = **0**;

output;

end;

run;

\*Determine how many RORs are negative;

**proc** **freq** data = diffa;

tables countneg / missprint;

**run**;

\*Median adjusted OR, ROR and 95% simulation interval;

**proc** **univariate** data = diffa noprint;

where countneg ne **1**;

var adjor ror;

output out = diffaresults pctlpts = **2.5** **50** **97.5** pctlpre = adjor\_ ror\_;

**run**;

\*Results;

**proc** **print** data = diffaresults;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

ADJUST ASSUMING DIFFERENTIAL B MISCLASSIFICATION

Set Se and Sp separately for cases and non-cases

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** diffb;

retain seed **13579**;

\*Enter the contingency table cells for the misclassified exposure and the outcome;

a = **264**; \*exposure = 1, disease = 1;

b = **736**; \*exposure = 0, disease = 1;

c = **1640**; \*exposure = 1, disease = 0;

d = **7360**; \*exposure = 0, disease = 0;

\*Enter the true (non-misclassified) odds ratio to calculate the ROR;

trueor = **1.29**;

\*Assign Se and Sp to non-cases;

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

se0 = **0.9022**;

sp0 = **0.8978**;

semin0 = se0 - **0.05**;

semax0 = se0 + **0.05**;

spmin0 = sp0 - **0.05**;

spmax0 = sp0 + **0.05**;

\*Assign Se and Sp to cases, 0.05 lower than for non-cases (or enter correct values when making correct assumption);

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

se1 = **0.8720**;

sp1 = **0.8229**;

semin1 = se1 - **0.05**;

semax1 = se1 + **0.05**;

spmin1 = sp1 - **0.05**;

spmax1 = sp1 + **0.05**;

\*10,000 simulations;

do i = **1** to **10000**;

\*Select Se and Sp for non-cases from triangular distribution;

call rantri(seed, (se0-semin0)/(semax0-semin0), tempse0);

setri0 = (semax0-semin0)\*tempse0 + semin0;

call rantri(seed, (sp0-spmin0)/(spmax0-spmin0), tempsp0);

sptri0 = (spmax0-spmin0)\*tempsp0 + spmin0;

\*Select Se and Sp for cases from triangular distribution;

call rantri(seed, (se1-semin1)/(semax1-semin1), tempse1);

setri1 = (semax1-semin1)\*tempse1 + semin1;

call rantri(seed, (sp1-spmin1)/(spmax1-spmin1), tempsp1);

sptri1 = (spmax1-spmin1)\*tempsp1 + spmin1;

\*Adjust cells based on selected Se and Sp values;

newa = (a - (**1**-sptri1)\*(a+b)) / (setri1+sptri1-**1**);

newc = (c - (**1**-sptri0)\*(c+d)) / (setri0+sptri0-**1**);

newb = (a+b)-newa;

newd = (c+d)-newc;

\*Calculate adjusted OR;

adjor = (newa\*newd)/(newb\*newc);

\*Calculate ratio of odds ratios;

ror = adjor/trueor;

\*Determine if ROR is negative;

if (ror lt **0**) then countneg = **1**;

else if (ror ge **0**) then countneg = **0**;

output;

end;

run;

\*Find out how many RORs are negative;

**proc** **freq** data = diffb;

tables countneg / missprint;

**run**;

\*Median adjusted OR and ROR and 95% simulation intervals;

**proc** **univariate** data = diffb noprint;

where countneg ne **1**;

var adjor ror;

output out = diffbresults pctlpts = **2.5** **50** **97.5** pctlpre = adjor\_ ror\_;

**run**;

\*Results;

**proc** **print** data = diffbresults;

**run**;

**eAppendix 2**

**Literature review methods**

We searched Embase and Medline for studies published between 1985 and April 2011 reporting sensitivity and specificity, or data sufficient so these could be calculated, of obesity classification in American adult females. Search terms included self-report, height, weight, BMI (body mass index), sensitivity, and specificity. We identified 62 articles in the search and excluded articles for the following reasons: irrelevant topic (n = 18), commentary without original data (n = 1), included only children or the elderly (n = 12), included only males (n = 1), or not conducted in the United States (n = 23). Of the 7 remaining studies, we excluded 4 because they provided estimates from NHANES; this exclusion was made because we wanted to approximate scenarios in which internal validation data was unavailable.

To the 3 studies identified through the search strategy20,22,23 we added 2 studies that we were aware of, but that were not identified by the search strategy.19,21

**eAppendix 3**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

EAPPENDIX3.SAS

Program:

Clean NHANES data.

Adjust for exposure misclassification given assumptions.

Programmer:

Candice Johnson

17 MAY 2014

NHANES datasets can be downloaded from:

http://www.cdc.gov/nchs/nhanes/nhanes\_questionnaires.htm

NHANES datasets used:

Demographics (DEMO)

Body measures (BMX)

Weight History (WHQ)

Diabetes (DIQ)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

CREATE NHANES DATASETS

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

libname bmx99 xport "FILEPATH\BMX.xpt";

libname bmx01 xport "FILEPATH\BMX\_B.xpt";

libname bmx03 xport "FILEPATH\BMX\_C.xpt";

libname bmx05 xport "FILEPATH\BMX\_D.xpt";

libname bmx07 xport "FILEPATH\BMX\_E.xpt";

libname bmx09 xport "FILEPATH\BMX\_F.xpt";

libname whq99 xport "FILEPATH\WHQ.xpt";

libname whq01 xport "FILEPATH\WHQ\_B.xpt";

libname whq03 xport "FILEPATH\WHQ\_C.xpt";

libname whq05 xport "FILEPATH\WHQ\_D.xpt";

libname whq07 xport "FILEPATH\WHQ\_E.xpt";

libname whq09 xport "FILEPATH\WHQ\_F.xpt";

libname diq99 xport "FILEPATH\DIQ.xpt";

libname diq01 xport "FILEPATH\DIQ\_B.xpt";

libname diq03 xport "FILEPATH\DIQ\_C.xpt";

libname diq05 xport "FILEPATH\DIQ\_D.xpt";

libname diq07 xport "FILEPATH\DIQ\_E.xpt";

libname diq09 xport "FILEPATH\DIQ\_F.xpt";

libname demo99 xport "FILEPATH\DEMO.xpt";

libname demo01 xport "FILEPATH\DEMO\_B.xpt";

libname demo03 xport "FILEPATH\DEMO\_C.xpt";

libname demo05 xport "FILEPATH\DEMO\_D.xpt";

libname demo07 xport "FILEPATH\DEMO\_E.xpt";

libname demo09 xport "FILEPATH\DEMO\_F.xpt";

\*Macro to create datasets, keeping only essential variables;

**%macro** createds(xpt, filen, keepvar);

data &xpt (keep = seqn &keepvar);

set &xpt.**.**&filen;

run;

**%mend**;

\*Measured BMI, 1999-2010;

%***createds***(bmx99,bmx,bmxbmi);

%***createds***(bmx01,bmx\_b,bmxbmi);

%***createds***(bmx03,bmx\_c,bmxbmi);

%***createds***(bmx05,bmx\_d,bmxbmi);

%***createds***(bmx07,bmx\_e,bmxbmi);

%***createds***(bmx09,bmx\_f,bmxbmi);;

\*Combine measured BMI datasets and dichotomize BMI at 30 kg/m2;

**data** measbmi (keep = seqn mobesity);

set bmx99 bmx01 bmx03 bmx05 bmx07 bmx09;

\*Dichotomize BMI;

if (bmxbmi gt **0** & bmxbmi lt **30**) then mobesity = **0**;

else if (bmxbmi ge **30** & bmxbmi lt **200**) then mobesity = **1**;

run;

\*Self-reported BMI, 1999-2010;

%***createds***(whq99,whq,whd010 whd020);

%***createds***(whq01,whq\_b,whd010 whd020);

%***createds***(whq03,whq\_c,whd010 whd020);

%***createds***(whq05,whq\_d,whd010 whd020);

%***createds***(whq07,whq\_e,whd010 whd020);

%***createds***(whq09,whq\_f,whd010 whd020);

\*Combine self-reported BMI datasets, calculate BMI, dichotomize;

**data** srbmi (keep = seqn srobesity);

set whq99 whq01 whq03 whq05 whq07 whq09;

\*Clean height variable;

srheight = whd010;

if (srheight gt **7000**) then srheight = **.**;

\*Clean weight variable;

srweight = whd020;

if (srweight gt **7000**) then srweight = **.**;

\*Calculate BMI from inches/pounds;

srbmi = **703**\*srweight/(srheight\*\***2**);

\*Dichotomize BMI;

if (srbmi gt **0** & srbmi lt **30**) then srobesity = **0**;

else if (srbmi ge **30** & srbmi lt **200**) then srobesity = **1**;

run;

\*Self-reported diabetes, 1999-2010;

%***createds***(diq99,diq,diq010);

%***createds***(diq01,diq\_b,diq010);

%***createds***(diq03,diq\_c,diq010);

%***createds***(diq05,diq\_d,diq010);

%***createds***(diq07,diq\_e,diq010);

%***createds***(diq09,diq\_f,diq010);

\*Combine diabetes datasets, dichotomize self-reported diabetes;

**data** diabetes (keep = seqn diabetes);

set diq99 diq01 diq03 diq05 diq07 diq09;

\*Dichotomize diabetes (if diabetes = "borderline" then diabetes = "no");

if (diq010 = **1**) then diabetes = **1**;

else if (diq010 in (**2**,**3**)) then diabetes = **0**;

run;

\*Demographics, 1999-2010;

%***createds***(demo99,demo,riagendr ridageyr ridexprg);

%***createds***(demo01,demo\_b,riagendr ridageyr ridexprg);

%***createds***(demo03,demo\_c,riagendr ridageyr ridexprg);

%***createds***(demo05,demo\_d,riagendr ridageyr ridexprg);

%***createds***(demo07,demo\_e,riagendr ridageyr ridexprg);

%***createds***(demo09,demo\_f,riagendr ridageyr ridexprg);

\*Combine demographics datasets;

**data** demo;

set demo99 demo01 demo03 demo05 demo07 demo09;

run;

\*Merge all NHANES datasets together;

**%macro** sortit(dsname);

proc sort data = &dsname;

by seqn;

run;

**%mend**;

%***sortit***(measbmi);

%***sortit***(srbmi);

%***sortit***(diabetes);

%***sortit***(demo);

**data** merged;

merge demo measbmi srbmi diabetes;

by seqn;

\*Exclude if male, age not 18-49, pregnant, missing diabetes, missing srbmi, missing measbmi;

if (riagendr ne **2**) then delete;

if (ridageyr lt **18** | ridageyr gt **49**) then delete;

if (ridexprg = **1**) then delete;

if (diabetes = **.**) then delete;

if (srobesity = **.**) then delete;

if (mobesity = **.**) then delete;

run;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Crude associations

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*Measured obesity and diabetes;

**proc** **freq** data = merged;

tables mobesity\*diabetes / missprint relrisk;

**run**;

\*Self-reported obesity and diabetes;

**proc** **freq** data = merged;

tables srobesity\*diabetes / missprint relrisk;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Classification tables

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*Classification table: whole population;

**proc** **freq** data = merged;

tables srobesity\*mobesity / missprint;

**run**;

\*Classification table: diabetes;

**proc** **freq** data = merged;

where diabetes = **1**;

tables srobesity\*mobesity / missprint;

**run**;

\*Classification table: no diabetes;

**proc** **freq** data = merged;

where diabetes = **0**;

tables srobesity\*mobesity / missprint;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

PROBABILISTIC ADJUSTMENTS

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

ADJUST ASSUMING EXACTLY NONDIFFERENTIAL MISCLASSIFICATION

Choose the same values of Se and Sp for cases and non-cases

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** exact;

retain seed **13579**;

\*Enter the contingency table cells for the misclassified exposure and the outcome;

a = **197**; \*exposure = 1, disease = 1;

b = **89**; \*exposure = 0, disease = 1;

c = **2240**; \*exposure = 1, disease = 0;

d = **5597**; \*exposure = 0, disease = 0;

\*Enter the true (non-misclassified) odds ratio to calculate the ROR;

trueor = **6.06**;

\*Enter the Se and Sp values (from literature review);

se = **0.90**;

sp = **0.97**;

\*Define minimum and maximum of the triangular distributions, here +/- 0.05 of the value;

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

semin = se - **0.05**;

semax = se + **0.05**;

spmin = sp - **0.05**;

spmax = **1**;

\*10,000 simulations;

do i = **1** to **10000**;

\*Choose an Se from the triangular distribution for Se;

call rantri(seed, (se-semin)/(semax-semin), tempse);

setri = (semax-semin)\*tempse + semin;

\*Choose an Sp from the triangular distribution for Sp;

call rantri(seed, (sp-spmin)/(spmax-spmin), tempsp);

sptri = (spmax-spmin)\*tempsp + spmin;

\*Adjust cells of contingency table using Se and Sp values chosen from distributions - same values for cases and noncases;

newa = (a - (**1**-sptri)\*(a+b)) / (setri+sptri-**1**);

newc = (c - (**1**-sptri)\*(c+d)) / (setri+sptri-**1**);

newb = (a+b)-newa;

newd = (c+d)-newc;

\*Calculate the adjusted OR;

adjor = (newa\*newd)/(newb\*newc);

\*Calculate the ratio of odds ratios;

ror = adjor/trueor;

\*Check if the ROR is negative, these will be excluded from the results;

if (ror lt **0**) then countneg = **1**;

else if (ror ge **0**) then countneg = **0**;

output;

end;

run;

\*Check how many values were negative and will be excluded;

**proc** **freq** data = exact;

tables countneg / missprint;

**run**;

\*Find the median adjusted OR and ROR and the 95% simulation interval;

**proc** **univariate** data = exact noprint;

where countneg ne **1**;

var adjor ror;

output out = exactresults pctlpts = **2.5** **50** **97.5** pctlpre = adjor\_ ror\_;

**run**;

\*Results;

**proc** **print** data = exactresults;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

ADUST ASSUMING APPROXIMATELY NONDIFFERENTIAL MISCLASSIFICATION

Choose Se and Sp for cases and non-cases from the same distribution

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** approx;

retain seed **13579**;

\*Enter the contingency table cells for the misclassified exposure and the outcome;

a = **197**; \*exposure = 1, disease = 1;

b = **89**; \*exposure = 0, disease = 1;

c = **2240**; \*exposure = 1, disease = 0;

d = **5597**; \*exposure = 0, disease = 0;

\*Enter the true (non-misclassified) odds ratio to calculate the ROR;

trueor = **6.06**;

\*Assign Se and Sp for non-cases (from literature review);

se0 = **0.90**;

sp0 = **0.97**;

\*Define minimum and maximum of the triangular distributions, here +/- 0.05 of the value;

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

semin0 = se0 - **0.05**;

semax0 = se0 + **0.05**;

spmin0 = sp0 - **0.05**;

spmax0 = **1**;

\*Assign Se and Sp for cases (from literature review);

se1 = **0.90**;

sp1 = **0.97**;

\*Define minimum and maximum of the triangular distributions, here +/- 0.05 of the value;

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

semin1 = se1 - **0.05**;

semax1 = se1 + **0.05**;

spmin1 = sp1 - **0.05**;

spmax1 = **1**;

\*10,000 simulations;

do i = **1** to **10000**;

\*Choose Se, Sp from triangular distribution for non-cases;

call rantri(seed, (se0-semin0)/(semax0-semin0), tempse0);

setri0 = (semax0-semin0)\*tempse0 + semin0;

call rantri(seed, (sp0-spmin0)/(spmax0-spmin0), tempsp0);

sptri0 = (spmax0-spmin0)\*tempsp0 + spmin0;

\*Choose Se, Sp from triangular distribution for cases;

call rantri(seed, (se1-semin1)/(semax1-semin1), tempse1);

setri1 = (semax1-semin1)\*tempse1 + semin1;

call rantri(seed, (sp1-spmin1)/(spmax1-spmin1), tempsp1);

sptri1 = (spmax1-spmin1)\*tempsp1 + spmin1;

\*Adjust OR based on selected Se and Sp values;

newa = (a - (**1**-sptri1)\*(a+b)) / (setri1+sptri1-**1**);

newc = (c - (**1**-sptri0)\*(c+d)) / (setri0+sptri0-**1**);

newb = (a+b)-newa;

newd = (c+d)-newc;

\*Calculate adjusted OR;

adjor = (newa\*newd)/(newb\*newc);

\*Calculate ratio of odds ratios;

ror = adjor/trueor;

\*Identify negative RORs;

if (ror lt **0**) then countneg = **1**;

else if (ror ge **0**) then countneg = **0**;

output;

end;

run;

\*Check how many RORs are negative and will be excluded;

**proc** **freq** data = approx;

tables countneg / missprint;

**run**;

\*Median adjusted OR, ROR and 95% simulation intervals;

**proc** **univariate** data = approx noprint;

where countneg ne **1**;

var adjor ror;

output out = approxresults pctlpts = **2.5** **50** **97.5** pctlpre = adjor\_ ror\_;

**run**;

\*Results;

**proc** **print** data = approxresults;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

ADJUST ASSUMING DIFFERENTIAL A MISCLASSIFICATION

Assign Se and Sp to cases and non-cases separately

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** diffa;

retain seed **13579**;

\*Enter the contingency table cells for the misclassified exposure and the outcome;

a = **197**; \*exposure = 1, disease = 1;

b = **89**; \*exposure = 0, disease = 1;

c = **2240**; \*exposure = 1, disease = 0;

d = **5597**; \*exposure = 0, disease = 0;

\*Enter the true (non-misclassified) odds ratio to calculate the ROR;

trueor = **6.06**;

\*Assign Se and Sp to non-cases (from literature review);

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

se0 = **0.90**;

sp0 = **0.97**;

semin0 = se0 - **0.05**;

semax0 = se0 + **0.05**;

spmin0 = sp0 - **0.05**;

spmax0 = **1**;

\*Assign Se and Sp to cases (0.05 higher than non-cases);

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

se1 = **0.95**;

sp1 = **1**;

semin1 = se1 - **0.05**;

semax1 = **1**;

spmin1 = **0.97**;

spmax1 = **1**;

\*10,000 simulations;

do i = **1** to **10000**;

\*Find Se and Sp for non-cases from triangular distributions;

call rantri(seed, (se0-semin0)/(semax0-semin0), tempse0);

setri0 = (semax0-semin0)\*tempse0 + semin0;

call rantri(seed, (sp0-spmin0)/(spmax0-spmin0), tempsp0);

sptri0 = (spmax0-spmin0)\*tempsp0 + spmin0;

\*Find Se and Sp for cases from triangular distributions;

call rantri(seed, (se1-semin1)/(semax1-semin1), tempse1);

setri1 = (semax1-semin1)\*tempse1 + semin1;

\*Sp1=Spmax1 so let the triangular distribution be symmetric to allow the program to run;

call rantri(seed, **0.5**, tempsp1);

sptri1 = (spmax1-spmin1)\*tempsp1 + spmin1;

\*Adjust cells based on Se and Sp values selected;

newa = (a - (**1**-sptri1)\*(a+b)) / (setri1+sptri1-**1**);

newc = (c - (**1**-sptri0)\*(c+d)) / (setri0+sptri0-**1**);

newb = (a+b)-newa;

newd = (c+d)-newc;

\*Calculate adjusted OR;

adjor = (newa\*newd)/(newb\*newc);

\*Calculate ratio of odds ratios;

ror = adjor/trueor;

\*Check if the ROR is negative, to later exclude it;

if (ror lt **0**) then countneg = **1**;

else if (ror ge **0**) then countneg = **0**;

output;

end;

run;

\*Determine how many RORs are negative;

**proc** **freq** data = diffa;

tables countneg / missprint;

**run**;

\*Median adjusted OR, ROR and 95% simulation interval;

**proc** **univariate** data = diffa noprint;

where countneg ne **1**;

var adjor ror;

output out = diffaresults pctlpts = **2.5** **50** **97.5** pctlpre = adjor\_ ror\_;

**run**;

\*Results;

**proc** **print** data = diffaresults;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

ADJUST ASSUMING DIFFERENTIAL B MISCLASSIFICATION

Set Se and Sp separately for cases and non-cases

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** diffb;

retain seed **13579**;

\*Enter the contingency table cells for the misclassified exposure and the outcome;

a = **197**; \*exposure = 1, disease = 1;

b = **89**; \*exposure = 0, disease = 1;

c = **2240**; \*exposure = 1, disease = 0;

d = **5597**; \*exposure = 0, disease = 0;

\*Enter the true (non-misclassified) odds ratio to calculate the ROR;

trueor = **6.06**;

\*Assign Se and Sp to non-cases (from literature review);

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

se0 = **0.90**;

sp0 = **0.97**;

semin0 = se0 - **0.05**;

semax0 = se0 + **0.05**;

spmin0 = sp0 - **0.05**;

spmax0 = **1**;

\*Assign Se and Sp to cases, 0.05 lower than for non-cases;

\*If Se or Sp + 0.05 is greater than 1, limit it to 1. If Se or Sp - 0.05 is less than 0.5, limit it to 0.5;

se1 = **0.85**;

sp1 = **0.92**;

semin1 = se1 - **0.05**;

semax1 = se1 + **0.05**;

spmin1 = sp1 - **0.05**;

spmax1 = sp1 + **0.05**;

\*10,000 simulations;

do i = **1** to **10000**;

\*Select Se and Sp for non-cases from triangular distribution;

call rantri(seed, (se0-semin0)/(semax0-semin0), tempse0);

setri0 = (semax0-semin0)\*tempse0 + semin0;

call rantri(seed, (sp0-spmin0)/(spmax0-spmin0), tempsp0);

sptri0 = (spmax0-spmin0)\*tempsp0 + spmin0;

\*Select Se and Sp for cases from triangular distribution;

call rantri(seed, (se1-semin1)/(semax1-semin1), tempse1);

setri1 = (semax1-semin1)\*tempse1 + semin1;

call rantri(seed, (sp1-spmin1)/(spmax1-spmin1), tempsp1);

sptri1 = (spmax1-spmin1)\*tempsp1 + spmin1;

\*Adjust cells based on selected Se and Sp values;

newa = (a - (**1**-sptri1)\*(a+b)) / (setri1+sptri1-**1**);

newc = (c - (**1**-sptri0)\*(c+d)) / (setri0+sptri0-**1**);

newb = (a+b)-newa;

newd = (c+d)-newc;

\*Calculate adjusted OR;

adjor = (newa\*newd)/(newb\*newc);

\*Calculate ratio of odds ratios;

ror = adjor/trueor;

\*Determine if ROR is negative;

if (ror lt **0**) then countneg = **1**;

else if (ror ge **0**) then countneg = **0**;

output;

end;

run;

\*Find out how many RORs are negative;

**proc** **freq** data = diffb;

tables countneg / missprint;

**run**;

\*Median adjusted OR and ROR and 95% simulation intervals;

**proc** **univariate** data = diffb noprint;

where countneg ne **1**;

var adjor ror;

output out = diffbresults pctlpts = **2.5** **50** **97.5** pctlpre = adjor\_ ror\_;

**run**;

\*Results;

**proc** **print** data = diffbresults;

**run**;