### Drawing Samples of Varying Numbers of Households (Ignoring Segment Structure) Macro

proc sql;

create table t12

as select \*, max(gott) as numgott

from prac.population

group by woreda;

proc sort data=t12;

by woreda gott segment household;

run;

/\*creating variable for number of segment in each gott\*/

proc sql;

create table t22

as select \*, max(segment) as numseg

from t12

group by unique\_gott;

/\*creating max houses variable per gott\*/

data house;

set t22;

maxh=30\*numseg;

run;

/\*tabulating woreda by gott by segment\*/

proc freq data=house noprint;

   tables woreda\*gott\*maxh / out=x;

run;

proc sort data=x;

   by woreda gott maxh;

run;

/\*creating data set with one observation for each gott, with variable that is total number of segments per gott\*/

data frame;

   set x;

   by woreda gott maxh;

   drop count percent;

   if last.gott;

run;

/\*macro to draw a bunch of samples\*/

/\*The following macro draws 1000 samples of a user input number of clusters and households. The segment structure is ignored.\*/

%macro sampleh(cluster=30,house=10);

/\*selecting user input number of clusters (gotts) within each of the 30 woredas\*/

proc surveyselect data=frame method=sys N=&cluster out=sample\_&cluster&&house seed=6212018 rep=1000;

samplingunit gott;

strata woreda;

run;

proc sort data=sample\_&cluster&&house;

   by woreda gott maxh;

run;

/\*expanding data set to the household level so that we can randomly select a number of households\*/

data samph\_&cluster&&house;

   set sample\_&cluster&&house;

   do house=1 to maxh; output;

   end;

   /\*adding this line to try to fix issues, may need to remove\*/

   rename house=household;

run;

/\*sorting by random number within replicate, woreda, and house\*/

proc sort data=samph\_&cluster&&house;

   by replicate woreda gott;

run;

/\*selecting households\*/

proc surveyselect data=samph\_&cluster&&house method=srsN=&house selectall

out=samp\_&cluster&&house seed=6212018;

samplingunit household;

strata replicate woreda gott;

run;

proc sort data=samp\_&cluster&&house;

   by woreda gott household;/\*this was changed trying to fix the error\*/

run;

proc sort data=t2;

   by woreda gott household;

run;

/\*merging population clinical data to sample data from above\*/

proc sql;

create table sql\_sample\_&cluster&&house as

select f.replicate, f.woreda, p.woreda, f.gott, p.gott, f.household, p.household,

p.member\_1to9, p.tf\_ind, f.selectionprob, p.numgott, f.maxh

from work.samp\_&cluster&&house f

left join work.t2 p

on f.woreda=p.woreda AND f.gott=p.gott AND f.household=p.household

;

quit;

proc sort data=sql\_sample\_&cluster&&house;

by replicate woreda gott household member\_1to9;

run;

/\*calculating true weights by multiplying selection prob from both stages\*/

data prac.sample\_&cluster&&house;

set sql\_sample\_&cluster&&house;

sprob1=&house/maxh;

if sprob1>1 then sprob2=1;

else sprob2=sprob1;

finalprob=sprob2\*selectionprob;

final\_weight=1/finalprob;

run;

/\*getting bootstrap confidence intervals for each woreda\*/

/\*calculating prevalence for each replicate and each woreda weighted based on sampling weights\*/

proc means data=prac.sample\_&cluster&&house noprint;

weight final\_weight;

class replicate woreda;

var tf\_ind;

output out=means\_&cluster&&house;

run;

proc sort data=prac.sample\_&cluster&&house;

by gott;

run;

/\*manipulating data set of prevalences\*/

data mn\_&cluster&&house;

set means\_&cluster&&house;

if \_STAT\_ ne "MEAN" then delete;

drop \_TYPE\_ \_FREQ\_ \_STAT\_;

rename tf\_ind=Mean\_prev;

if woreda=. then delete;

if replicate=. then delete;

run;

/\*calculating percentiles within replicates for 95% confidence intervals\*/

proc univariate data=mn\_&cluster&&house noprint;

class woreda;

var mean\_prev;

output out=percentile\_&cluster&&house mean=avg\_est\_mean pctlpre=replicate pctlpts=2.5,97.5;

run;

/\*creating dataset with lower and upper bounds, and length on confidence interval\*/

data prac.ci\_&cluster&&house;

set percentile\_&cluster&&house ;

rename replicate2\_5=lower replicate97\_5=upper;

label replicate2\_5=" " replicate97\_5=" ";

clusters=&cluster;

households=&house;

length=replicate97\_5-replicate2\_5;

run;

%mend sampleh;