### Population Simulation Macro

/\*The following macro simulates a TF dataset at the individual level for Amhara, Ethiopia

User inputs: seed -> randomizes data, can be any integer (default=983758932)

numgott -> randomized number of gotts (villages) per woreda, can be any positive integer (default=100)

high1-3,mid1-4,low1-7 -> average prevalence levels in each gott (randomized), 14 levels in total, can be any number in range (0,1)

   (defaults=0.4,0.35,0.3,0.25,0.2,0.15,0.1,0.08,0.05,0.04,0.03,0.02,0.01,0.005)

   A segment prevalence is randomized for each segment based on the gott prevalence.

   There are exactly 30 households per segment, with an average of 1.1 children aged 1-9 per household (# of children randomized).\*/

libname prac "H:\Thesis\_Practicum";

%macro tfdata(seed=983758932,/\*numgott=100\*/high1=0.4,high2=0.35,high3=0.3,mid1=0.25,mid2=0.2,mid3=0.15,mid4=0.10,low1=0.08,

low2=0.05,low3=0.04,low4=0.03,low5=0.02,low6=0.01,low7=0.005);

data prac.population;

call streaminit(&seed); /\*setting seed so data is reproduceable\*/

do woreda=1 to 30; /\*30 woredas\*/

do gott=1 to rand("negbinomial",0.01,1.51);  /\* distributed around average number (156) of specified gotts per woreda\*/

unique\_gott=(woreda\*1000)+gott; /\*creating unique id for each gott\*/

/\*high prevalence districts\*/

if 1<=woreda<=3 then do;

tf\_gott=rand("beta",20\*&high1,20\*(1-&high1)); end; /\*default mean of 0.4\*/

else if 4<=woreda=<6 then do;

tf\_gott=rand("beta",20\*&high2,20\*(1-&high2)); end; /\*default mean of 0.35\*/

else if 7<=woreda<=9 then do;

tf\_gott=rand("beta",20\*&high3,20\*(1-&high3)); end; /\*default mean of 0.3\*/

/\*middle prevalence districts\*/

else if 10<=woreda<=12 then do;

tf\_gott=rand("beta",25\*&mid1,25\*(1-&mid1)); end; /\*default mean of 0.25\*/

else if 13=<woreda=<15 then do;

tf\_gott=rand("beta",25\*&mid2,25\*(1-&mid2)); end; /\*default mean of 0.2\*/

else if 16<=woreda<=18 then do;

tf\_gott=rand("beta",25\*&mid3,25\*(1-&mid3)); end; /\*default mean of 0.15\*/

else if 19<=woreda<=21 then do;

tf\_gott=rand("beta",25\*&mid4,25\*(1-&mid4)); end; /\*default mean of 0.10\*/

/\*low prevalence districts\*/

else if 22<=woreda<=24 then do;

tf\_gott=rand("beta",25\*&low1,25\*(1-&low1)); end; /\*default mean of 0.08\*/

else if woreda=25 then do;

tf\_gott=rand("beta",25\*&low2,25\*(1-&low2)); end; /\*default mean of 0.05\*/

else if woreda=26 then do;

tf\_gott=rand("beta",25\*&low3,25\*(1-&low3)); end; /\*default mean of 0.04\*/

else if woreda=27 then do;

tf\_gott=rand("beta",25\*&low4,25\*(1-&low4)); end; /\*default mean of 0.03\*/

else if woreda=28 then do;

tf\_gott=rand("beta",25\*&low5,25\*(1-&low5)); end; /\*default mean of 0.02\*/

else if woreda=29 then do;

tf\_gott=rand("beta",25\*&low6,25\*(1-&low6)); end; /\*default mean of 0.01\*/

else if woreda=30 then do;

tf\_gott=rand("beta",25\*&low7,25\*(1-&low7)); end; /\*default mean of 0.005\*/

 do segment=1 to (rand("negbinomial",0.1,0.25)+1); /\*average number of segments per gott is 3.24, this gives expected value of about 3.24 with signficant variance\*/

 unique\_seg=(unique\_gott\*10)+segment; /\*creating unique value for each segment\*/

 tf\_segment=rand("beta?",(200\*tf\_gott),(200\*(1-tf\_gott))); /\*creating segment level prevalence to account for correlation within segments, want small variance so use 200 as multiplier\*/

 randnum=rand("uniform");

  do household=1 to 30; /\*30 houses per segment\*/

  unique\_house=(unique\_seg\*100)+household; /\*creating unique value for each household\*/

   do member\_1to9=0 to rand("Poisson",1.107); /\*Poisson distributed number of kids 1-9 per household, mean calculated as 1.107=(.27\*4.1)\*/

   unique\_mem=(unique\_house\*10)+member\_1to9;

   if member\_1to9>0 then do;

    tf\_ind=rand("Bernoulli",tf\_segment); /\*Indicator variable for TF based on prevalence of TF in the gott\*/

   end;

   else do;

    tf\_ind=.;

   end;

   output;

   end; end; end; end; end;

run;

%mend tfdata;