

## Using SAS to Model the Spread of Infectious Disease

Karl Mink, Mike Zdeb, University@Albany School of Public Health, Rensselaer, NY

### ABSTRACT

An epidemic is the rapid and extensive spread of disease. Environmental factors, human factors, and the properties of the infectious agent in question all play a role in the spread of that biological agent. It is this combination of factors that combine to form "a perfect epidemic." A program was created to enable educators to access a simple disease-modeling presentation to aid them in teaching students about the spread of disease. It takes into consideration only population density and assumes immunity occurs after infection. However, it is enough to visually demonstrate concepts of herd immunity, vaccination, and disease spread.

The program is a prototype and ideally would be hosted on a web site. Hopefully teachers could log on and change variables such as the infectiousness of the agent, the population density, and the number of people vaccinated against the agent.

### INTRODUCTION

A typical infectious cycle is characterized by: infection of a susceptible individual by an infected individual, an incubation period where the pathogen replicates, a symptomatic stage, and hopefully recovery. In some cases recovery provides immunity to future infections by the pathogen. Although not relevant to this disease model, it is important to note that transmission can occur prior to the symptomatic stage for some pathogens.

In the early days of epidemiology, before the advent of computers, epidemiologists used marbles to model the spread of infectious disease. A WHITE marble represented a susceptible individual. A RED marble represented an infected individual that is also contagious. And a BLACK marble represented a recovered immune survivor. Typically a RED marble was randomly placed into a group of WHITE marbles. The infectious RED marble infected the adjacent WHITE susceptibles (the white marbles were exchanged with RED marbles), and the infectious RED marble recovered (it was changed with a BLACK marble). The marbles were then again randomly distributed and the epidemic was allowed to run its course.

An epidemic is the rapid and extensive spread of disease. Environmental factors, human factors, and the properties of the infectious agent in question all play a role in the spread of that biological agent. It is this combination of factors that combine to influence the epidemic.

### Environmental Factors

Viruses are obligate intercellular organisms, meaning they need a host to survive. Bacteria can survive without a host. Some bacteria, such as the bacterium that causes anthrax, can encapsulate allowing it to remain resistant to environmental conditions that would normally kill it. These dormant bacteria can survive for years. Finally, some diseases rely upon vectors (mosquitoes, ticks, etc) for transmission. Humans can venture into new environments exposing them to new species that are infected with unknown pathogens. Environmental changes can also allow species to migrate to areas not previously suitable to the infected species. Either way, the end result is new human exposure to infected species. For the purposes of this disease-modeling program environmental changes are not included in the model.

### Human factors

People group together in sickness and in health. Population density is a primary variable in the spread of disease. Quite obviously a sick person must contact a susceptible person if an epidemic is to occur. This model allows adjustments of population density.

Human behavior is the other variable that influences transmission. Human behaviors like; safe sex, sharing needles when injecting drugs, hand washing, drinking clean water, etc all substantially influence disease transmission. Human behavior is not directly incorporated into this model but can indirectly be incorporated via the likelihood of infection. However, this is applied uniformly over the entire modeled population.

### Immunity

An individual's state of healthiness influences their likelihood of becoming infected and re-infected. Stress, sleep deprivation, infection with other pathogens, etc influence the body's ability to fight an infection. Some disease agents, like the virus that causes chicken pox, will result in long-term immunity. Long-term immunity against infectious agents is the logic behind vaccination. A vaccine actually puts some dead or weakened organism into your body to prime your immune system in the event of exposure to a pathogenic strain of that organism. This disease model assumes long-term immunity occurs after the initial exposure preventing re-infection.

### Properties of the pathogen

Some pathogens are very infectious (such as the bacterium that causes tularemia). Some pathogens cause high mortality because there is no treatment (such as the virus that causes Ebola hemorrhagic fever), or because the organism has developed ways to resist the treatment (such as a drug resistant form of the bacterium that causes Tuberculosis). This disease-modeling program allows for the level of infectiousness to be changed but requires the user to know what the appropriate level of infectiousness for the agent they are interested in studying.

### THE VISION BEHIND CREATING THIS PROGRAM

This program was created to allow educators to access a simple disease-modeling program to aid them in teaching students about the spread of disease. It only takes into consideration population density and assumes immunity occurs after infection, but the program visually demonstrates concepts of herd immunity, vaccination, and disease spread. It is a proto-type and ideally would be hosted on a website.

Hopefully, teachers could log on and change variables like the infectiousness of the agent, the population density, and the number of people vaccinated against the agent. As you use this model of infectious disease, think about these questions:

*How does population density influence the number infected?*

*How low does the population density have to get to stop the outbreak from occurring, or burning out very quickly?*

*Does the entire population become infected? Why? or Why not?*

*If the agent were not as infectious, would less people become infected?*

*If the agent were not as infectious, would the epidemic be longer or shorter?*

*How many people need to get vaccinated to prevent an outbreak?*

### THE PROGRAM

SAS/GRAPH®, specifically PROC GMAP, is used to display the course of an epidemic given several choices made by a user of the program: the size of a square grid that represents the geographic area in which the epidemic takes place; the population size (the number of squares in the grid occupied by a person); the number of people in the population that are infected at the start of the epidemic. The user also chooses a seed for the SAS RANUNI function. This choice determines both the location of the population within the grid and the location of infected individuals.

Subsequent versions of the program allow a user to alter the probability that a non-infected individual will become infected once in contact with an infected individual. In the original version, contact of an infected with an uninfected (non-immune) has a 100% chance of transmission. The probability of infection should be governed by both the infectivity of the infected person and the susceptibility of an uninfected. It is also assumed the after an uninfected person become infected (or starts out at time 0 as infected), they will then become immune (resistant) to subsequent infection in the next cycle of the epidemic. Further program modifications should allow for a number of other possibilities: given a particularly virulent infect, that person might die and not take up a place in the grid; given that not all infections produce life-long immunity, immunity should not be exactly 100%, contact with an infected person could possibly cause another infection.

The original program allows free movement of all persons in the population to any point within the grid. All persons move at random. Another feature added in subsequent versions is the ability to limit the movement of the population within the grid. Not all individuals are allowed to move at random to any new place within the grid --- most individuals stay 'close to home' (home being their initial grid location) while a smaller portion can move to any location.

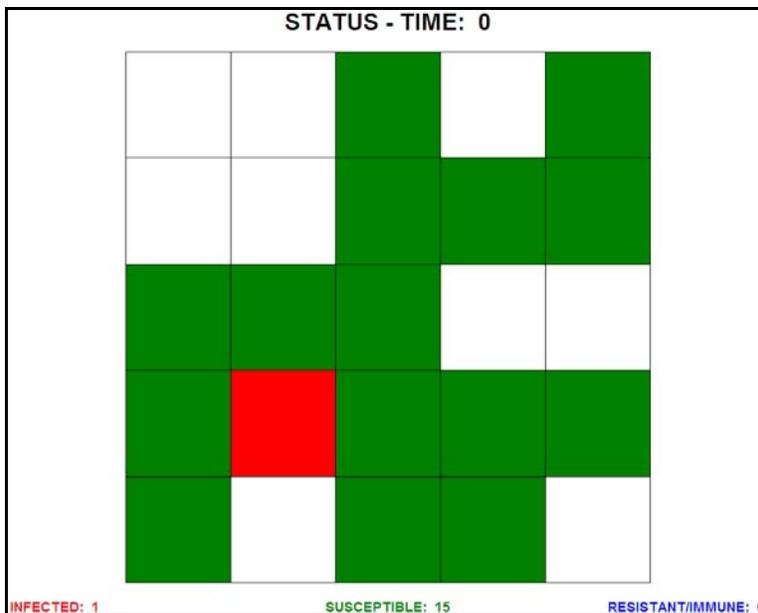
### VERSION 1

The original version of the program begins with the assignment of values to four macro variables:

```
* PARAMETERS;

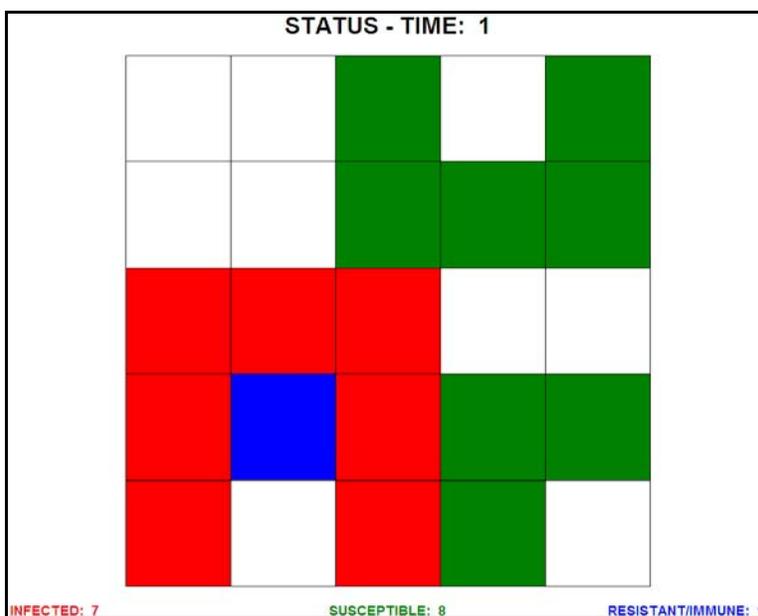
* grid size;
%let nsp = 5;
* number initially infected;
%let inf = 1;
* population size;
%let pop = 16;
* random number for initial assignment of 'infecteds' to grid;
%let seed = 1234;
```

The values of the macro variables determine parameters at the outbreak of an epidemic:  $\&nsp$ , the number of cells in a square grid (a value of 5 results in a square grid with 25 cells, 5 on a side);  $\&inf$ , the number of people placed in the grid who are initially infected;  $\&pop$ , the number of people who populate the grid, 1 infected and 15 susceptible (uninfected) given an initial value of 16;  $\&seed$ , a seed number for the RANUNI function, used to place people in the grid at random and to determine subsequent movement of people within the grid. This set of choices results in the following depiction of the epidemic at TIME 0:

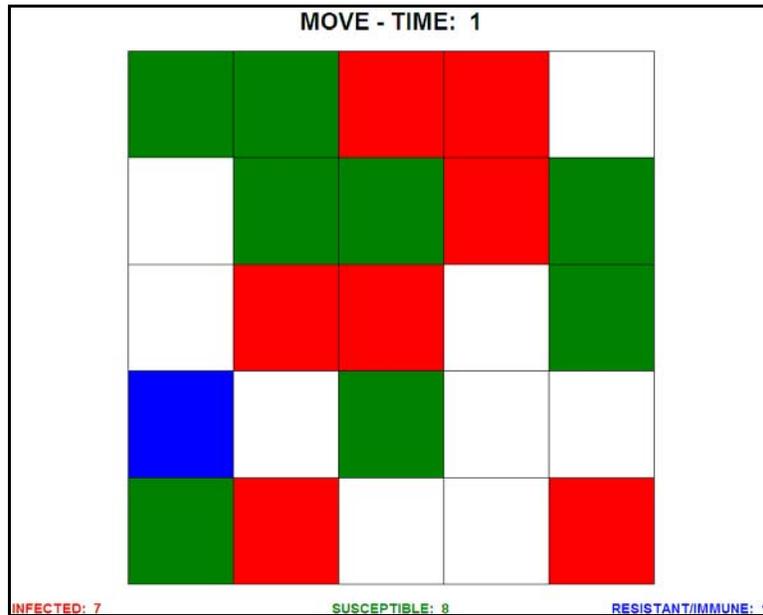


The 5x5 grid is produced with PROC GMAP. The 16 squares are filled at random. A TITLE statement is used to display the STATUS, the iteration in the cycle of the epidemic. A FOOTNOTE statement is used to display the number of infected, number of susceptible, and the number of resistant/immune. A person becomes immune in the cycle after they have become infected (they get 'well'). The original version of the program does not assign any of the initial population to the resistant/immune group. This is another modification to be made in subsequent versions, i.e. there may be some people who are resistant/immune to an infection at time 0.

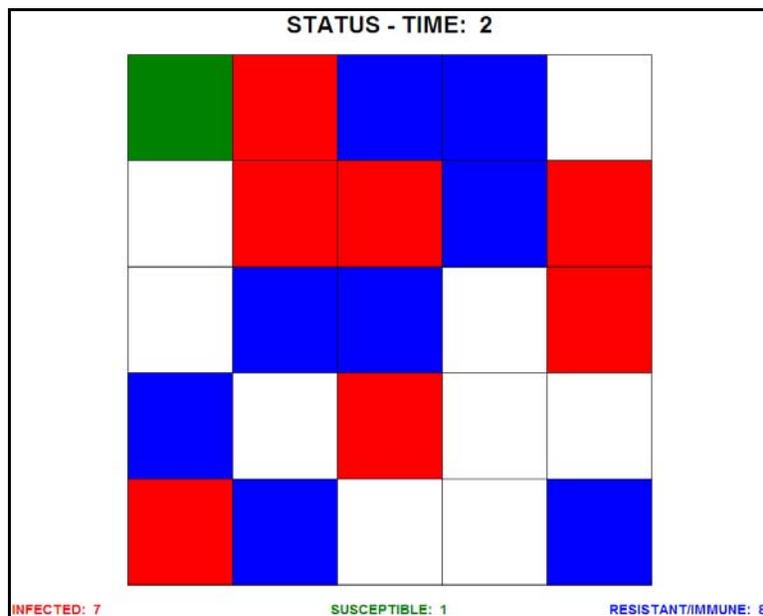
The original program makes an assumption that the infected person infects all the susceptible individuals in adjoining squares of the grid, leading to the epidemic at TIME 1:



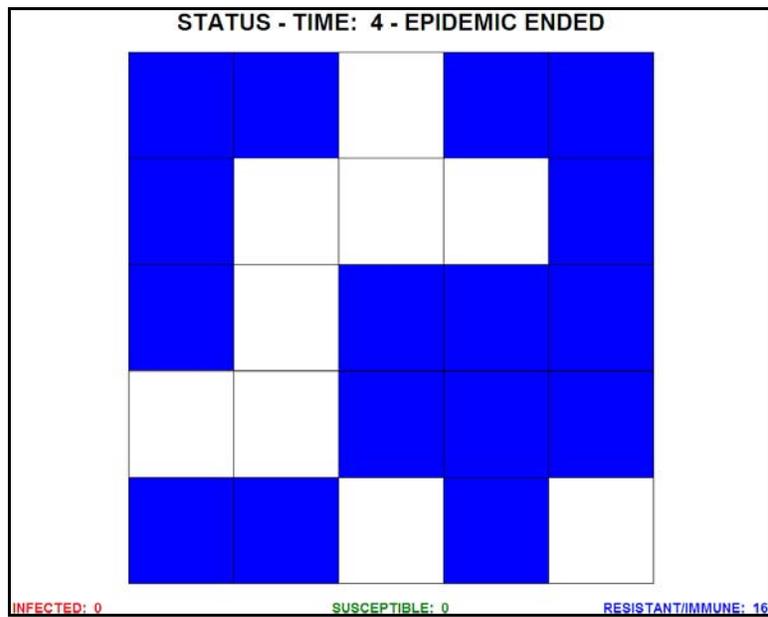
The first infected person is now resistant to further infection and all those in adjoining cells are infected. Another assumption in the original program is that individuals can move to any other position in the grid at random. This results in the following:



Now, the newly infected individual can infect those that are susceptible, resulting in:



This cycle (status, move, status, move...) continues until there are no infected individuals left in the grid. Given the original set of parameters, the values of the macro variables, this occurs at time 4:



The epidemic ends in this scenario since all individuals at some point have become infected are all now resistant/immune to further infection.

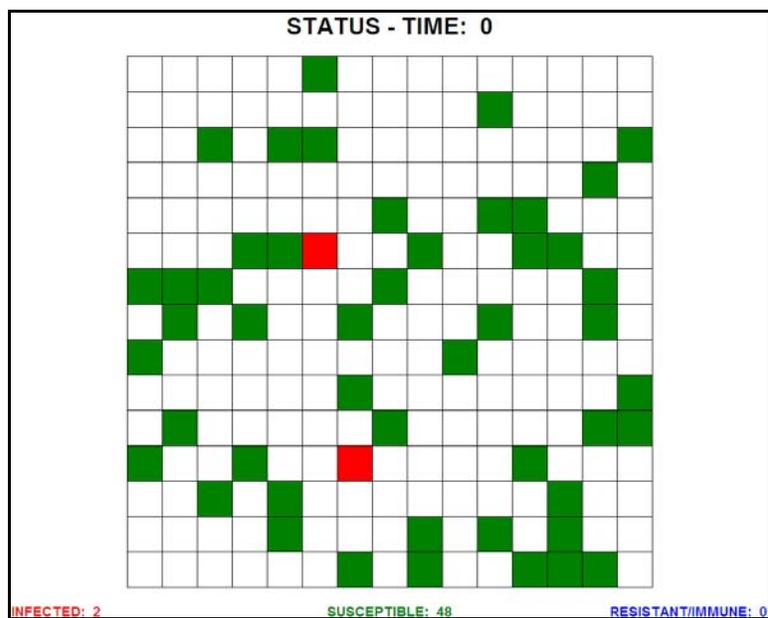
Another possible end point is when there are still susceptible individuals, but there are no longer any individuals with an active infection. For example, given the following set of parameters...

```

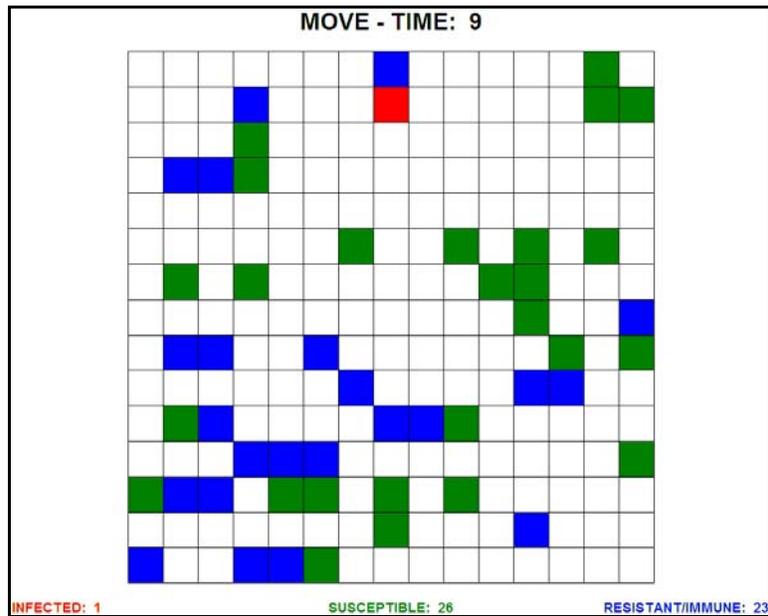
* PARAMETERS;
* grid size;
%let nsp = 15;
* number initially infected;
%let inf = 2;
* population size;
%let pop = 50;
* random number for initial assignment of 'infecteds' to grid;
%let seed = 3456;

```

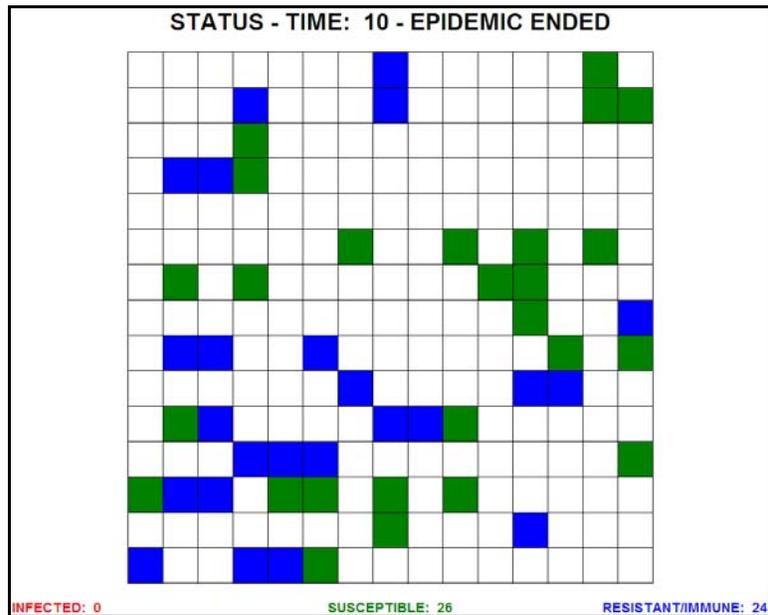
a population of 50 persons is spread out within a grid of 225 cells.



Given the sparsely populated the grid, the chances of an infected person coming into contact with a susceptible person are reduced. It takes more cycles for the epidemic to end. By the time 9 is reached, there is only one infected person.



and the status at time 10 looks as follows:



The epidemic ends not because everyone is resistant/immune as in the first example, but because there are no longer any infected people in the population.

**VERSION 1 SAS CODE**

The following is a listing of the SAS code used to produce the output shown on the previous pages. There are a number of comments within the code explaining the function of each section. Work is underway on adding options (infection not 100% probable but varying across population, restricted movement of a portion of the population).

```

* PARAMETERS;
* grid size;
%let nsp=5;
* number initially infected;
%let inf=1;
* population size;
%let pop=16;
* random number for initial assignment of 'infecteds' to grid;
%let seed=1234;

*****;

* calculate the number uninfected;
%let uni=%eval(&pop - &inf);
* calculate the number of cells in the grid;
%let dim=%eval(&nsp * &nsp);
%let mul=1;
* set the total number infected at start of the epidemic (will accumulate);
%let tinf=&inf;
* set the total number uninfected at start of the epidemic (will decrease);
%let tuni=&uni;
* set the number resistant/immune at start of the epidemic to 0 (will accumulate);
%let timm=0;

* reset all graphics options to default values - choose a font;
options reset=all ftext='Helvetica/bo' htext=2.5 gunits=pct;

* select colors to represent three state: red-infected; green-uninfected; blue - resistant/immune;
pattern1 v=s c=red;
pattern2 v=s c=green;
pattern3 v=s c=blue;

* a macro that creates x-y coordinates for the grid (the map data set for PROC GMAP);
%macro grid;
*** create a grid;
data space (keep=sp x y);
do j = 1 to &nsp;
do i = 1 to &nsp;
  sp+1;
  x=i; y=j; output;
  x=i+1 ; output;
  y=j+1 ; output;
  x=i ; output;
end;
end;

x1=1; y=1;
do i=1 to 3;
  sp+1;
  output;
end;
run;
%mend;

* a macro that creates the display using the map data set SPACE and the response data set EPI;
%macro makemap (x);
%if &tinf gt 0 %then title "&x - TIME: &j";
%else title "&x - TIME: &j - EPIDEMIC ENDED";
footnote h=1.5 j=1 c=red "INFECTED: %left(&tinf)"
          j=c c=green "SUSCEPTIBLE: %left(&tuni)"
          j=r c=blue "RESISTANT/IMMUNE: %left(&timm)";

proc gmap map=space data=epi all;
id sp;
choro status / discrete nolegend outline=black ;
run;
quit;
title;
footnote;
%mend;

* a macro that places people into spots on the grid based on random numbers from the RANUNI function;
%macro status;
*** first assignment of status;
data epi (keep=sp status);
array s(&dim) _temporary_;

status = 1;
do i=1 to &inf;
  sp = int(&dim*ranuni(&seed))+1;
  s(sp) = 1;
  output;
end;
end;

```

```

end;

status = 2;
do i=1 to &uni;
  sp = int(&dim*ranuni(&seed*2))+1;
  if s(sp) ne 1 then do;
    s(sp) = 1;
    output;
  end;
  else i=i-1;
end;
run;
%mend;

* a macro that checks the status of people in the grid;
* used to 'infect' susceptibles who come into contact with infecteds;
%macro check;
data epi (keep=sp status);
array s(&dim) _temporary_;
array o(&nsp,&nsp) _temporary_;
array n(&nsp,&nsp) _temporary_;

do i=1 to &pop;
  set epi;
  s(sp) = status;
  row = int((sp-1)/&nsp) + 1;
  col = sp - (&nsp*(row-1));
  o(row,col) = status;
  n(row,col) = status;
end;

do row = 1 to &nsp;
do col = 1 to &nsp;

  if o(row,col) eq 1 then do;

    sr = row - 1;
    if sr lt 1 then sr = 1;
    er = row + 1;
    if er gt &nsp then er = &nsp;
    sc = col - 1;
    if sc lt 1 then sc = 1;
    ec = col + 1;
    if ec gt &nsp then ec = &nsp;

    do r = sr to er;
    do c = sc to ec;
      if o(r,c) eq 2 then n(r,c) = 1;
    end;
  end;
end;

end;
end;

inf=0;
uni=0;
imm=0;

do r=1 to &nsp;
do c=1 to &nsp;
  if o(r,c) eq 1 then n(r,c) = 3;
  sp = (&nsp*(r-1)) + c;
  status = n(r,c);
  if status ne . then output;
  if status eq 1 then inf+1;
  else
  if status eq 2 then uni+1;
  else
  if status eq 3 then imm+1;
end;
end;

call symput('tinf',put(inf,best.));
call symput('tuni',put(uni,best.));
call symput('timm',put(imm,best.));

sp=&dim;
do status=1 to 3;
  sp+1;
  output;
end;
stop;
run;
%mend;

* a macro that moves people around the grid based on random numbers from the RANUNI function;
%macro move;
%let mul=%eval(&mul + 1);

```

```

data epi (keep=sp status);
array o(&dim) _temporary_;
array n(&dim) _temporary_;

do i=1 to &pop;
  set epi;
  o(sp) = status;
end;

do i=1 to &dim;
  status = o(i);
if status ne . then do;
  sp = int(&dim*ranuni(&seed*&mul))+1;
  if n(sp) eq . then do;
    n(sp) = status;
  output;
  end;
  else i=i-1;
end;
end;

sp=&dim;
do status=1 to 3;
  sp+1;
  output;
end;
stop;
run;
%mend;

* combine the status, makemap, check, and move macros into one macro - epidemic;
%macro epidemic;
%let j=0;
%status;
%makemap(STATUS);

%do j=1 %to 100;
%check;
%makemap(STATUS);

%if &tinf eq 0 %then %goto exit;

%move;
%makemap(MOVE);
%end;
%exit: %mend;

* create the map data set that is used to draw the grid;
%grid;

* use the epididemic macro;
ods results off;
%epidemic;
ods results on;

```

## CONCLUSION

A combination of data steps, PROC GMAP, macro variables, and macro programs make it possible to simulate the progress of an epidemic. Though a number of assumptions have been made in the first version of this program, it does fulfill its aim in providing an aid for teaching students about the spread of an infectious disease given a set of parameters at the start of the spread. The parameters are easy to modify. Options added to future versions will allow modification of even more parameters, changing the program's current reliance on 100% probability of infection and unrestricted movement of the population.

## CONTACT INFORMATION

Questions about this paper can be addressed to Mike Zdeb using e-mail...msz03@albany.edu

by phone...518-402-6479

or by regular mail...

Mike Zdeb  
 U@Albany School of Public Health  
 1 University Place  
 Rensselaer, NY 12144-3456

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and  other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.