Population Dynamics of Infectious Disease

**Objective:** To model the progression of epidemics for diseases of different infectivities and to determine the effects of isolation and vaccination.

**Discussion:**

Population dynamics models are based on the same principle as accounting. If you want to know your daily cash total, you could count it every day. Or you could keep track of your income and expenditures and use them to update your daily totals. This is the principle used for epidemic models. We classify people according to their current status, such as Infected or Recovered, and use update formulas to calculate daily changes in the class counts. In disease models, as in most population models, the class count need not be an integer. This sounds horrible, but it isn’t when you realize that models are not meant to give “correct” answers. Class counts in the real world depend on chance; for example, the number of new infections might be 7 one day and 5 on each of the next 2 days. If the model assumes that the number is the same every day, we would use 5.67, which is an average of 7, 5, and 5. When studying a model, focus on the general trends, not the specific numbers.

The most important parameter in epidemic modeling is the *basic reproductive number*, denoted as $R\_{0}$ (and usually read as “R zero”). This quantity is the average number of secondary infections caused in a fully susceptible population by one infected person over the duration of that person’s illness. If the basic reproductive number is less than 1, the disease cannot propagate because the average person recovers before infecting anyone else. This is what happened with many strains of influenza, which has a seasonal basic reproductive number that is low in the summer. The most infectious human disease is measles, with a basic reproductive number estimated at 12-18.

**Notation and Assumptions:**

Disease models can vary greatly in complexity. In this module, we consider the prototype epidemic model, called the SIR model because the population is divided into three classes: Susceptible, Infected, and Removed. A more complicated model (called SEIR) separates infected individuals into a class of people who are not yet infectious (usually called Exposed) and a class of people who are infectious (I). Neither the SIR nor SEIR models can be used to model diseases for which recovery does not confer immunity, like the common cold. The SEIR model has been used for Covid-19, but it is too simple to capture all of the most important features of this disease. The Covid-19 model in module 403 is of type SEAIHRD (Susceptible, Exposed, Asymptomatic infective, symptomatic Infective, Hospitalized infective, Recovered, Deceased). More complicated models such as this can better capture details, but simpler models are much better for developing conceptual understanding. This is why it is important to understand the SIR model before studying more complicated ones. It is a reasonable first approximation for many less complicated diseases, such as influenza.

The basic SIR model has only two processes that change the class counts: infection and recovery.

A schematic diagram is helpful to show how the new infections and recoveries change the class counts. The new infections increase I and decrease S, while the recoveries increase R and decrease I.

recoveries

new infections

R

I

S

1. The infection process changes individuals from Susceptible to Infected. We assume that each infected individual would create an average of *b* new cases per day in a population where all *N* individuals are susceptible. Then *I* infected individuals would combine to create *bI* new cases per day. If the population is not wholly susceptible, then the expected number of new cases would be *bI\*S/N*, where *S/N* is the fraction of encounters that are with susceptibles.

Total new infections = average new cases per infective if everyone else is susceptible (b) \* number of infectives (I) \* fraction of contacts who are susceptible (S/N)

We define a parameter *B*=*b/N* so that we can write the new infection count per day more simply as *BSI*.

1. The recovery process changes individuals from Infected to Removed. Suppose *T* is the average amount of time an individual is infectious. We can reasonably assume a fraction 1/*T* of infected individuals recover in any given day. The number of individuals who remain infected from one day to the next is I-(1/T)\*I. We define *p*=1-1/*T*, so that number of infected individuals that are carried over each day is *pI* .

We will include one additional process: isolation of infected individuals. This is one of several tools that can be used to modify the course of an epidemic.

Other processes could be included. If our goal is to track a disease for several years, then we need to include births and deaths from other causes. We might also need to include an immunity loss process that changes individuals from Removed to Susceptible. Both of these processes operate over a much longer time window than a typical epidemic, so we will omit them here. This may seem like a ruinous error, but in modeling we must always keep in mind the amount of uncertainty in parameter values. A look at the well-documented Wikipedia page for Basic Reproductive Number shows that this important value is seldom known even to the nearest integer of precision. Uncertainty in this value makes the model results only rough approximations.

Mathematical models have a lot of parameters, making it difficult to decide how to study their results. One way to reduce the number of options is to give some of the parameters fixed values. The total population is not particularly important, as it is best to report results in terms of percentages. We’ll take a total population of 10000, which is just slightly larger than the number of people on the largest cruise ship. The infectivity parameter is very important, but it is difficult to determine from data. Instead, we will consider $R\_{0}$ to be the principal measure of infectiousness and use it to calculate the infectivity.

**Governing Equations:**

**Eq. 1: Infectivity**

$$b=\frac{R\_{0}}{T}$$

**Eq. 2: New Infections**

$$I\_{+}=b\frac{S}{N}I=BSI$$

where $S$ and $I$ are the populations of the Susceptible and Infected classes, *N* is the total population, $b$ is the infectivity (the average number of new infections per day caused by one infected individual in a fully susceptible population), and $B=b/N$ is defined for convenience.

**Eq. 3: Recovery / Death**

$$I\_{-}=\frac{1}{T}I$$

where $T$ is the average time an individual spends in the infected class. With *T*=10, one tenth of infected individuals will move to the removed class each day.

**Eq. 4: Isolation**

$$I\_{q}=p^{3}qI$$

where *p*=1-1/*T* represents the fraction of today’s infected class who are still contagious tomorrow and *q* is the fraction of symptomatic individuals who are put into isolation. The formula assumes that symptoms appear in the third day of infectivity: the factor $p^{3}I$ is the number of individuals who have been infected for at least three days, and therefore symptomatic, and a fraction *q* of these are isolated.

**Eq. 5, 6, 7: Population Updates**

$$S\_{n+1}=S\_{n}-I\_{+}$$

$I\_{n+1}=I\_{n}+I\_{+}-I\_{-}-I\_{q}$=$ pI\_{n}+I\_{+}-I\_{q}$

$$R\_{n+1}=N-S\_{n+1}-I\_{n+1}$$

where *n* is the current time, from 0 on up, and *p*=1-1/*T* is defined so as to combine $I\_{n}$ and $I\_{-}$ into a single term.

**Eq. 8, 9, 10: Initial Conditions**

$$I\_{0}=fN$$

$$R\_{0}=vN$$

$$S\_{0}=\left(1-f-v\right)N$$

where *f* is the fraction of the population that is initially infected and *v* is the fraction of the population that is initially immune (this provides a way of introducing vaccination into the model).

**Setting Up:**

1. Open Microsoft Excel®
2. Set up your spreadsheet exactly as it appears below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|   | A | B | C | D | E | F |
| 1 | Total population (N) | Isolation (q) | Inf. frac. (f) | Rem. frac. (v)  | Duration (T) | R0 |
| 2 | 10000  |  0 | 0.001 | 0 | 5 | 2 |
| 3 |   |   |  |  |   |  |
| 4 |   |  | Carryover (p)  |  Infectivity (b) | B |  |
| 5 |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |
| 7 | Time (days) | Susceptible  | Infected  |  Removed | New Infections  |  |

Note that only 4 data values are needed to define a simulation with the SIR model in the absence of isolation or vaccination. N=10000 is large enough that rounding off the class counts to integers changes the model only slightly without making the numbers unreadably large. The population size is just a little bigger than the number of people on a large cruise ship and roughly that of a small rural city. The initial choices for T and R0 are estimates for the Spanish flu, which has been estimated to have killed at least 1% of the world’s population in the last pandemic before Covid-19. We are using yellow highlighting to call attention to the cells that contain the model data.

1. Enter the time 0 in cell A8.
2. The time corresponding to each increment needs to be entered in column A. Rather than enter the values individually, we can use an Excel formula. Enter the formula “=A8+1” into cell A9, which will set the next time step to be 1 day. Now select cell A9. Hover the mouse pointer over the lower right-hand corner so that the cursor appears as a plus sign. Click and drag the cell down to A128. You should see a column of times from 0 to 120.
3. The parameters p, b, and B need to be calculated from the original parameters. Enter “=1-1/E2” in cell C5, “F2/E2” in cell D5, and “=D5/A2” in cell E5. With the given total population, disease duration, and basic reproductive number, the results are p=0.8, b=0.4, and B=0.00004.
4. Calculate the initial populations from Equations 8, 9, and 10 by entering “=C2\*A2” in cell C8, “=D2\*A2” in cell D8, and “=A2-C8-D8” in cell B8. The simulation will start with 9990 susceptible, 10 infected, and none removed.
5. Calculate the initial number of new infections from Equation 2 by entering “=MIN($E$5\*B8\*C8, B8)” in cell E8. The first formula calculates the quantity B\*S\*I, and the MIN function uses the smaller of this result and the number of susceptibles. This is necessary to make sure that the number calculated as becoming infected does not exceed the number who could have become infected.
6. Calculate the numbers in each class for time 1 from Equations 4, 5, and 6 by entering “=B8-E8” in cell B9, “=$C$5\*C8+E8” in cell C9, and “=$A$2-B9-C9” in cell D9. Note that there is no isolation term in the formula for I because nobody has been sick long enough to have symptoms. The results should be approximately 9986, 12, and 2.
7. Complete the calculations for time 1 by dragging the formula from cell E8 down to E9.
8. Spreadsheets are often hard to read because the formulas result in answers with too many decimal places. To keep this from happening, highlight cells B8 through E9. Click on Format in the Cells menu and then Format Cells at the bottom of the drop-down menu. In the dialog box, choose Number and 0 decimal places.
9. Complete the calculations by highlighting cells B9 through E9, hovering the mouse over the lower right corner of cell E9 so that it becomes a plus sign, and then double clicking. This should fill in the table of data down to row 128.
10. The spreadsheet is now set up almost correctly, but the isolation term Iq has not been entered. Edit cell C11 so that it reads “=$C$5\*C10+E10-$C$5^3\*$B$2\*C10”, which matches Equation 5. Then double-click the fill handle of cell C10 so that the new formula is carried down through column C. The numbers you see in any cell should be unchanged by these modifications because q=0.

**Graphing the results:**

Follow the instructions below to create the graphs.

**Making an X – Y Scatter Plot on Excel 2013:** \*for Excel 2003 refer to the end of the document for steps 1-3.

1. Click on the Insert tab at the top of the screen.
2. Select the Scatter icon and choose the graph with straight lines and no data points on it.



1. ***Do not trust that the program will select the correct data, even if the graph looks correct.*** To select the correct data, right click the chart area and choose “Select Data” from the drop down menu. Remove any data series that is automatically selected. You will be adding a series for each model.
2. To add the data, start by clicking the "Add" icon. Then name the series “Susceptible” and click the icon to the right of the typing field for the X values (shown here).



1. Once you do this, the spreadsheet will be active and you can highlight any column of cells you wish to make the X values for the graph. **Do not highlight column headings, just data points.** Highlight the first five time values, starting with cell A8. *\*The formula for the X values in the field for the ideal model should read* "=Sheet1!$A$8:$A$12". Now edit that formula by changing that final number to 128.
2. Do the same for the corresponding Y values, but using column B to get the susceptible population instead of column A. Note that you must highlight an equal number of X and Y values for a scatter plot, so the final number should again be 128. Hit OK until the dialog boxes go away.
3. Repeat steps 4 through 6 to add three more series to the graph, one for the Infected population, one for the Removed population, and one for the New Infections count. For each of these, the last row for both X and Y should be 128.
4. Note the range of values on the X and Y axes. Software cannot think through the problem of choosing meaningful variable ranges, so we often have to change the default values manually. Population graphs should always start at 0 and go to a reasonable value. Right-click on the numbers on the Y axis and choose Format Axis from the menu. Change the Y axis range so that the maximum is 10000. The minimum value is currently set to 0 using the default algorithm of the program. Type in the number 0 instead so that the program will not be able to choose a different minimum value. Similarly, set the range on the X axis to be 0 to 120.
5. Graphs should always have axis labels, legends, and a title (in professional scientific papers, a caption is used instead of a title). Left-click on the plot area and choose the + sign that appears to the right of the graph. Check the boxes for Axes, Axis Titles, Chart Title, Gridlines, and Legend.
6. If the legend appears to the right of the graph, change it so that it appears on the bottom instead. To do this, put the cursor where the legend is and double-click. Choose Legend Options in the Format Legend dialog box and click the radio button that puts the legend on the bottom. Close the Format Legend box.
7. The generic axis and chart labels need to be changed. Left-click on each of these and type “Days” for the X axis, “Populations” for the Y axis, and “Epidemic Progress” for the chart title.
8. Move the plot so that the upper left corner of the plot window is in cell F4. This positions it so that you can easily see how it is affected by changes in the parameters.

**Making an Area Plot on Excel 2013:**

Another way of visualizing data for an epidemic model is with a stacked area plot.

1. Select the block of cells from B7 to D128. This includes all the data for Susceptible, Infected, and Removed, along with the column heading.
2. Click on the Insert tab at the top of the screen.
3. Select the Line or Area Chart icon and choose the right-most icon in the 2D Area group. This will automatically create the area plot. You do not need to modify the data set. You will need to move it up so that it is just under or beside the scatter plot.



1. As with any plot, there should be axis labels. Repeat the procedure you used to get the axis labels on the scatter plot. A stacked area plot shows how the proportion of population groups changes over time. Any vertical line drawn at a particular time will be divided into colors, with the length of each colored segment matching the height of the corresponding population on the line graph. Make sure that the same color codes are being used for the area and line plots.

**Final Setup:**

Graphs give a good indication of the progress of a simulation, but there are also important properties that are useful in characterizing what happens. While these quantities can be found by visual examination of the data, this examination can be automated using formulas. This makes is much easier to quickly identify the effect of any change in the current scenario.

1. The maximum number of individuals in class I is a measure of the stress the epidemic causes on the health care and economic systems. Type “maximum %I” in cell G1 and “=100\*MAX(C8:C128)/A2” in cell G2. Use the Format Cells menu to specify that cell G2 is a number with one decimal place. In our starter scenario, the maximum number of infected individuals at any one time is 16.1%.
2. The number of susceptibles at the end of the scenario is a measure of the risk of a resurgence of the epidemic. Type “final %S” in cell H1 and “=100\*B128/A2” in cell H2. As before, set the format to number with one decimal place. In our starter scenario, 19.1% of the population remains susceptible at the end of the epidemic.
3. Highlight cells G1:H2 in light blue to make it easier to identify key results.

**Development Questions:**

1. Why did we not include the new infections column in the area plot?

**Analysis Questions:**

1. Compare the days when the maximum number of new infections occurred and when the maximum infected population size occurred. Explain the time lag between these events.
2. Make a record of the maximum number of infectives and the day on which this occurs. Then change the initial fraction of infectives to 0.01 and record the changes in these two quantities. Go back to an initial fraction of 0.001 and observe how long it takes for the population of infectives to increase by a factor of 10, so as to reach 0.01. Use the results to draw conclusions about how a different initial number of infectives changes the course of an epidemic.
3. Covid-19 has a basic reproductive number of approximately 5. Set R0 to 5 and describe the changes in the epidemic progress. Pay attention to the key output quantities and also the speed with which the epidemic develops and resolves. (Note: We are using the same basic reproductive number, but the SIR model with a 5-day disease duration is not a good match for Covid-19.)
4. The duration of Covid-19 infectivity is approximately 10 days rather than 5. Change this parameter and describe the effect this has on the progress of the epidemic. It is helpful to think of the end of the epidemic as roughly the point when the number of infectives is down to what it was initially. (Note: We still are not modeling Covid-19, because the SIR model fails to account for a long incubation period and asymptomatic transmission.)
5. The Incan Empire had a population of over one million when it was conquered by 168 Spanish Conquistadores in 1525. The Spanish had gunpowder weapons and horses, but these advantages would not have been sufficient to defeat the huge Incan army. (It took about 2 minutes to reload a single-shot arquebus.) They also benefitted by joining forces with peoples subjugated by the Incas, but that would not have happened on its own. In *Guns, Germs, and Steel*, author Jared Diamond argues that the key factor in the Incan defeat was the European diseases the Spanish brought with them. To test this theory, set the basic reproductive number at 5 and the duration at 20, values that roughly match smallpox. Describe the effect introduction of smallpox into Incan civilization would have had, even without considering the death toll of the disease.
6. The most contagious human disease is thought to be measles, with a basic reproductive number of approximately 15. It has an infective period of about 8 days. Describe the progress of a measles outbreak caused by 1 initial infective in a population of 10000 that had not previously been exposed. This event would have happened numerous times in human history.
7. Do a more thorough study of the effect of the basic reproductive number on the total number of people who get a disease having an infective period averaging 5 days.
	1. Set the duration to 5 days. Label two columns “R0” and “total I”. Enter the numbers 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 3.0, 3.5, 4.0, 4.5, and 5.0 in the R0 column. Fill in the total I column one value at a time using the corresponding value of R0. Use the percentage of people who are no longer susceptible at the end of the simulation as a measure of the total number infected.
	2. Get a graph of the results. The easiest way to do this is to select the two-column table of numbers, click Insert, and choose a scatter plot with a smooth curve. This combination of key strokes will be correctly interpreted by Excel, so you won’t have to change the data series.
	3. Describe and explain the results.
8. Isolation[[1]](#footnote-1) of the sick to combat infection dates back centuries, at least to the establishment of a leper colony in England in 1084. Set the initial fraction of infectives to 0.001 and the duration and basic reproductive number to 5.
	1. Compare the effects of isolation levels of 20%, 50%, and 80% with the base level of 0%.
	2. Make a table to record the total percentage of people who get infected for the isolation levels 0%, 10%, 20%, and up to 90%. Make a graph of the results. Explain what the graph tells us about the effectiveness of isolation (note that the results you see are specifically for a disease with a five-day duration and two days of infectivity before symptoms appear).
9. Vaccination grew out of a practice called “variolation,” the deliberate use of cowpox infection to confer immunity against smallpox, which was discovered independently in China in the 17th Century and by the English doctor Edward Jenner in 1796. Vaccination with killed pathogen began in 1798 and quickly replaced variolation. Set the initial fraction of infectives to 0.001 and the duration and basic reproductive number to 5. Make sure q is set at 0.
	1. Compare the effects of vaccination levels of 20%, 50%, and 80% with the base level of 0%.
	2. Make a table to record the total percentage of people who get infected for the vaccination levels 0%, 10%, 20%, and up to 90%. Make a graph of the results. Explain what the graph tells us about the effectiveness of vaccination for this disease.
	3. Do an internet search of the term “herd immunity.” Add what you learn through reading with what you learned by experiment to draw some conclusions about the effectiveness of vaccination at different levels.
	4. Compare the effectiveness of vaccination with that of isolation.
1. The word ``quarantine’’ is often used incorrectly for the practice of isolation. Quarantine is the practice of isolating individuals who are thought to have been exposed to a disease, not individuals who are showing symptoms. [↑](#footnote-ref-1)