Introduction to mathematical modelling: Modelling the dynamics of immunizing infections using Excel
Computer Practical

Emilia Vynnycky (HPA)

Introduction and Objectives

This practical is based on practical that we run as part of our modelling shortcourse "An Introduction to Infectious Disease Modelling and its Applications", which is organized jointly between the London School of Hygiene & Tropical Medicine and the Health Protection Agency Centre for Infections. The practical is divided in 2 parts.

In part I you will set up a simple model of the transmission dynamics of measles using difference equations in Excel. Part II is optional and explores the relationship between the basic and net reproduction numbers and the herd immunity threshold. In your spare time, you may like to try the supplementary questions which are designed to make you think about the factors which influence the regular cycles in immunizing infections observed in many populations.

By the end of the practical you should understand:
- the basic methods for setting up models
- the relationship between key input parameters in simple models of the transmission dynamics of immunizing infections such as measles
- the relationship between the basic and net reproduction numbers, the herd immunity threshold and trends in disease incidence.

PART I: Setting up models using difference equations

In the lecture, you saw that the following model of the transmission dynamics of measles in a closed population could be written using the difference equations below:
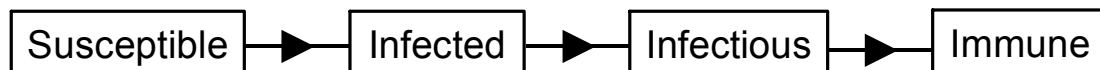
$$S_{t+1} = S_t - \beta S_t I_t \qquad (1)$$
$$E_{t+1} = E_t + \beta S_t I_t - f E_t \qquad (2)$$
$$I_{t+1} = I_t + f E_t - r I_t \qquad (3)$$
$$R_{t+1} = R_t + r I_t \qquad (4)$$

Here

S_t, E_t, I_t and R_t are the number of individuals in the population who are susceptible, infected (but not infectious), infectious and immune (recovered) respectively at time t,

$ß$ is the probability that a specific infectious and susceptible individual come into effective contact per unit time,
f is the rate at which infected individuals become infectious per unit time, and
r is the rate at which infectious individuals recover (become immune) per unit time.

We will set up this model in Excel by following the steps below.

Specifying the input parameters

1. Open up the spreadsheet measles1.xls.

The yellow cells show the key characteristics of the infection and the population in the model, namely the total population size, R_0, the average latent and infectious periods – these have been assigned the names in column G.

The blue cells contain the formulae for the parameters which we will actually be using when setting up the model, namely:
      a) the number of individuals "effectively" contacted by each person per day,
      b) the probability of an "effective" contact between 2 individuals per day
      c) the rate of infectious disease onset (equivalent to f in the the above equations)
      d) the rate at which infectious individuals recover become immune (equivalent to r in the above equations

The values in these (blue) cells have been assigned names in the adjacent column (ecr, beta, dis_rate, rec_rate) and are discussed below.

The parameter values are calculated as follows (do not worry about the methods for now):
1. The effective contact rate (ecr) is calculated as R_0/(average duration of infectiousness), where an "effective contact" is defined as one sufficient to lead to transmission between an infectious and a susceptible individual.   Recall that R_0 describes the average number of secondary infectious persons resulting from each infectious individual in a totally susceptible population and therefore reflects the total number of successful (i.e. effective) contacts during the infectious period.
2. beta is defined as ecr/total population. The logic behind this expression is that if each person effectively contacts 10 others per day, and the population comprises 100 individuals, then the probability that he/she comes into contact with another specific susceptible is 10/100.
3. The rate at which infected individuals develop infectious disease (dis_rate) is given by 1/(average latent period). Similarly the rate at which individuals recover from infectious disease (rec_rate) is given by 1/(average infectious period). These expressions stem from the fact that it can be shown (mathematically) that if an event occurs at a rate x, then the average time until that event is given by 1/x (do not worry about this for now)

Setting up the Excel equations

The lilac cells will contain the number of individuals in the susceptible, infected, infectious and immune categories at each of the times t=0, t=1, t=2 etc. At the start, there are 99999 susceptible, 1 infectious, 0 infected and 0 immune individuals.

1. Click on cell B44.

You should see that the expression for the number of susceptible individuals on day 1 is similar to equation 1 on page 1 of this handout.

Similarly, you should see that the expression for the number of infected individuals on day 1 is analogous to equation 2 on page 1 of this handout.

2. Use equations (3) and (4) on page 1 of this handout to set up appropriate expressions for:
  a) the number of infectious individuals on day 1, in terms of *dis_rate, rec_rate*, and the number infected and infectious on day 0 (in cell D44).
  b) the number of immune individuals on day 1, in terms of *rec_rate* and the number infectious and immune on day 0 (in cell E44).
If you are unfamiliar with Excel, you may wish to refer to the sheet on Excel tips at the end of this handout.

3. Copy the formulae for the number susceptible, infected, infectious and immune down until the 200th day. Select columns H and P together, click with the right mouse button and select the unhide option. You should see a graph of the number of individuals who are susceptible, infectious and immune over time.

Q1.1 What do you notice? Why do no further cases occur in this population after a certain time? How long does it take until there are no measles cases in the population?

Q1.2 Given the key inputs in the model (listed in the yellow cells), which would you change to make the measles epidemic finish sooner? later? Test your hypotheses by changing the values of the corresponding parameters.

Q1.3 Which assumptions might you alter or add to the model to describe the infection process of measles in
  a) a town
  b) a country,
over a period of weeks? years? (NB you are not expected to alter any assumptions now!)

We will now change the model to incorporate births and deaths in the population.

Incorporating births and deaths

1. Select rows 12 and 16 together, click on your right mouse button and select the unhide option. You should now see cells for the average mortality and birth rates and the number of births in this population per day.

Q1.4 What is the current value for the daily mortality rate? What is currently assumed about the birth rate? How should the total population size change over time in this situation?

Q1.5 What is the expression for the number of births which occur in the population each day?

For now, we will assume that all individuals are born susceptible in this population.

Q1.6 Is this realistic? Is this a reasonable assumption to make? Why? What alternative assumptions might be appropriate?

To incorporate births into the population, you would only need to modify equation (1) on page 1 as follows:

$$S_{t+1} = S_t - \text{ß}S_t I_t + num\_births \qquad (1')$$

Q1.7 Why don't we need to modify equations 2-4 on page 1 to incorporate births into the population?

To incorporate deaths from the susceptible population, equation (1') would need to be modified as follows:

$$S_{t+1} = S_t - \text{ß}S_t I_t + num\_births - m\_rate \times S_t \qquad (1'')$$

Q1.8 How would you modify equation 2 on page 1 to deal with deaths from the infected population? Similarly, how would you modify equations 3 and 4 on page 1 to deal with deaths from the infectious and immune populations respectively?

2. Open up the spreadsheet measles2.xls, where the formulae have been set up for you for day 1. Check that these formulae are consistent with your answer to the last question.

3. Copy the formulae for the number who are susceptible, infected, infectious and immune on day 1 down to time t=200.

Q1.9 How does this change your answer to question Q1.1? How do you think your answer would change if you were to simulate the dynamics for 10 years? 50 years?

4. Copy the formulae in the cells for t=200 until t=18250 (ie 365*50 days = 50 years). Select rows 18 and 36 together, click with the right mouse button and select the unhide option. You should see a graph of the number susceptible and immune for the whole 50 year period.

Q1.10 What do you notice? Is this consistent with your answer to the last question?

To extend the dynamics for an even longer time period (eg t=36500 days = 100 years), it would be possible to copy the formulae down still further. However, this would make the spreadsheet large (and possibly unmanageable, depending on the power of your computer).

One alternative is to take larger eg 2, 3 day or longer time steps in the simulations. You may wish to try this at the end of the practical. To do this, you need to convert the parameters currently used in the model into those appropriate for 2, 3 etc time steps. Another alternative is to reformulate the equations using *differential equations* and to use a package specially designed to deal with this problem. This is beyond the scope of this practical!

PART II (optional): The relationship between the basic and net reproduction numbers and the herd immunity threshold

We focus on how the net reproduction number changes during cycles in disease incidence for a simple immunizing infection. You should remember from the lecture that the net reproduction number, R_n is given by the following expression:

$$R_n = R_0 * proportion \ susceptible$$

1. Open up the file measles3.xls. It is closely related to the measles2.xls file you worked with in part 1. The orange cells will contain the disease incidence, proportion susceptible, proportion immune and R_n in the population over time.

2. Before continuing, check that you understand the expressions for the disease incidence and the proportions susceptible and immune set up for day 1.

3. Type in an appropriate expression for the net reproduction number for day 1 in cell I80 in terms of the proportion susceptible and the R_0, and copy the formulae for the disease incidence, proportions susceptible and immune and the R_n down until the 18250th day.

We now focus on the relationship between R_n and the other variables over short (10 year) intervals.

4. Select the columns I and V together, click with your right mouse button and select the unhide option.

You should now see Figure 1 which plots the R_n on the left hand y-axis and the disease incidence on the right hand y-axis in the population. The x-axis goes from the 14600th day (ie the 40th year) to the 18250th day (ie the 50th year). The gridlines on the x-axis occur at 2 year intervals.

Q2.1 How does the R_n change over time? What is the value of the R_n when the disease incidence reaches a peak? What is its value when the disease incidence reaches a trough?

Q2.2 What is the trend in the disease incidence when $R_n<1$? What is its trend when a) $R_n>1$ and b) $R_n=1$? Is this reasonable?

Q2.3 Given your answer to Q2.2, and given that

$$R_n=R_0*(proportion \ of \ individuals \ who \ are \ "susceptible")$$

how would you calculate the R_0 for your population if you had estimates of the proportion of individuals who are susceptible in the population at the same time as the disease incidence had reached a peak?

5. Select rows 22 and 43 together, click with the right mouse button and select the unhide option. You should now see Figure 2, which plots the proportion susceptible on the left hand y-axis and the disease incidence on the right hand y-axis.

Q2.4 What proportion of the population is susceptible to infection when the disease incidence peaks or troughs? Given your answer to Q2.3, check that this gives you the appropriate value for the R_0 in this population.

You should remember from the lecture that for a disease to "die out" in a population, the proportion of the population which is immune to infection has to maintained above the "Herd immunity threshold" (*HIT*), which is given by:

$$HIT = 1 - 1/R_0$$

Q2.5 Using this expression, calculate the herd immunity threshold in this population.

6. Select rows 49 and 70 together click with the right mouse button and select the unhide option. You should now see Figure 3 which plots the proportion immune in the population together with the disease incidence.

Q2.6 What proportion of the population is immune when the disease incidence peaks or troughs? What do you notice about the proportion who are immune when the disease incidence is declining? What is its value when the disease incidence is increasing. How does this relate to your estimate of the herd immunity threshold?

In your spare time, you may like to try the supplementary questions, which are designed to make you think about the factors which influence the cycles in immunizing infections.

Supplementary questions – analyses of factors influencing the cycles in disease incidence

This part of the practical is designed to make you think about the factors which influence the regular cycles in incidence of immunizing infections.

1. Open up the spreadsheet measles4.xls. This is very similar to the previous measles models, except that the transition rates are expressed in terms of a variable time step ("t_step"), which can be changed, rather than daily transition rates. In general, to convert a **daily** transition rate into that for a different time step, you just need to multiply it by the size of the time step. For example, if individuals develop disease at a rate of 5% per day, then you would expect roughly 3X5 =15% of individuals to have developed disease by the third day.

Q3.1. Complete the following Table by changing the R_0 accordingly and counting the number of complete cycles in disease that occur in each decade.

Time period	Number of cycles in disease incidence in each decade in a population with an R_0 of:		
	5	13	18
0-10 yrs			
11-20 yrs			
21-30 yrs			
31-40 yrs			
41-50 yrs			
51-60 yrs			
61-70 yrs			
71-80 yrs			
81-90 yrs			
91-100 yrs			
101-110 yrs		--	--
121-130 yrs		--	--
141-150 yrs		--	--
151-160 yrs		--	--

To put these in perspective, the R_0 estimated for Kansas (USA) was about 5-6 in 1918-9, 13-14 in Cirencester (UK) in 1947-50 and 18 in England and Wales in 1950-68.

Q3.2 How long does it take before the measles epidemics start to occur at regular intervals in populations with these R_0 values? What is the average inter-epidemic period when this occurs?

Q3.2 Why is the inter-epidemic period shorter for populations with an R_0 of 18 than for populations with an R_0 of 5?

Anderson and May (1991) provide a formula for approximating the inter-epidemic period for infections which impart full protection against (re) infection:
$$T = 2\pi \, [L(D+D')/(R_0-1)]^{1/2}$$

where D' is the average duration of the latent period, D' is the average duration of infectiousness, and L is the life expectancy.

Q3.3 Do your results from Q3.2 fit this formula? (Note that the life expectancy is currently set to be 70 years – it can be calculated using the expression 1/mortality rate).

Q3.4 Would you expect the inter-epidemic period for measles to be shorter than that for chickenpox? mumps? rubella?   Complete the following table either by using the above formula or by running the model (and writing down your findings when the cycles stabilize) to verify your hypothesis.

Infection	R_0	Latent period (days)	Infectious period (days)	Inter-epidemic period (yrs)
Measles	13	8	7	
Mumps	8	15	6	
Rubella	7	10	11	
Chickenpox	7	10	10	

Summary of useful Excel commands

The following summarizes some basic Excel commands.

Entering formulae/equations

1. Click on the cell where you need the equation to be located.
2. Type the equals sign, = followed by the rest of the equation and press enter, eg

$$= B3*B6$$

An alternative (slightly easier) method to set up this formula would be to:
1. type the equals sign,
2. click on cell B3
3. type *
4. click on cell B6
5. press enter

If the contents of cells already have names, then you can replace the cell reference in the formula with its name. For example, if the contents of cells B3 and B6 were called num_mosquitoes and num_humans respectively, then you could have typed in

$$= num\_mosquitoes*num\_humans$$

Selecting cells
If the cells you need to select are adjacent you can either:
1. Hold the shift key down and move around the spreadsheet using either the arrow keys or page up, page down etc.
2. Hold the left mouse button down and move around the spreadsheet.

Copying cell contents
There are several methods for doing this,:
1) Select the cells you want copied, press Ctrl and the C key together, move to the destination of these cells and press enter (or Ctrl V).
2) Select the cells you want copied using the mouse, click on the copy button, move to the destination of these cells, and click on the paste button.
3) Select the cells you want copied using the mouse, position your mouse in the bottom right-hand corner of these cells so that you see a + sign, click on the + sign and, holding the mouse down, drag the mouse down to the bottom of the range where you want the cell contents to go. Then release the mouse.

Moving cell contents

There are several methods for doing this,:
1) Select the cells you want moved, press Ctrl and the X key together, move to the destination of these cells and press enter (or Ctrl V).

2) Select the cells you want moved using the mouse, click on the cut button, move to the destination of these cells, and click on the paste button.
3) Select the cells you want moved using the mouse, position your mouse just below the last cell so that you see an arrow sign, click and holding the mouse down, drag the mouse down to the bottom of the range where you want the cell contents to go. Then release the mouse.

Moving around spreadsheets

The following combinations of key strokes takes you to different positions in the spreadsheet:

	Destination
Ctrl-Home	A1
Ctrl End	Bottom right hand corner of the data in your spreadsheet
End →	End of the current row of completed data
End ←	Start of the current row of completed data
End ↓	Bottom of the current column of completed data
End ↑	Top of the current column of completed data

Holding the shift key down at the same time as you perform these key strokes will also select these blocks of data.

Mathematical expressions

The following are commands which you may find helpful:

Mathematical expression	Excel command
$\log_e(x)$	=ln(x)
e^c	= exp(c)
e^{-c}	=exp(-c)
a^c	=a^c
a^{-c}	=a^(-c)
a^{cd}	=a^(c*d)