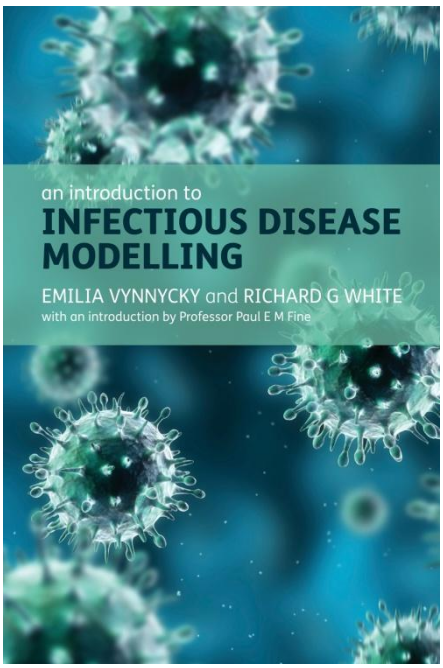
An Introduction to Infectious Disease Modelling

Guide to the online material



Version 4, released 03/11/10

Table of Contents

Page number

+ remaining files to become available in November/December 2010

Software requirements

The model files accompanying the book are written for Microsoft Excel 2007 and Berkeley Madonna. The Berkeley Madonna files have been tested using version 8.3.11 of the software.

Berkeley Madonna can be downloaded and purchased from the website http://www.berkeleymadonna.com/

Copyright

.

Using the Berkeley Madonna files

1.1 Introduction

Berkeley Madonna is a specialist model-building package that is used to set up models using difference or (most commonly) differential equations.

For all of the online exercises accompanying the book you will just need to run the models that have already been set up in Berkeley Madonna, or to change the input parameters. This document provides an overview of the key features of the package and the model files that you need to know in order to do this.

Further detailed instructions about using Berkeley Madonna can be found in the online guide available at http://www.berkeleymadonna.com/

1.2 The basics of running models in Berkeley Madonna

1.2.1 The flowchart and equation editors

Models can be set up in Berkeley Madonna using either the flowchart or equation editor.

The flowchart approach involves drawing a flowchart of the model, which provides a visual summary of all the equations and how the components are linked together. The equation editor requires users to type in the differential equations directly and, with this approach, Berkeley Madonna does not provide a flow diagram.

Wherever possible, the models accompanying this book are set up using both methods and you can choose which approach you work with.

The method for running models in Berkeley Madonna is identical irrespective of whether the model is set up using the flowchart or equation editor. We outline the main differences between models set up using the two approaches below.

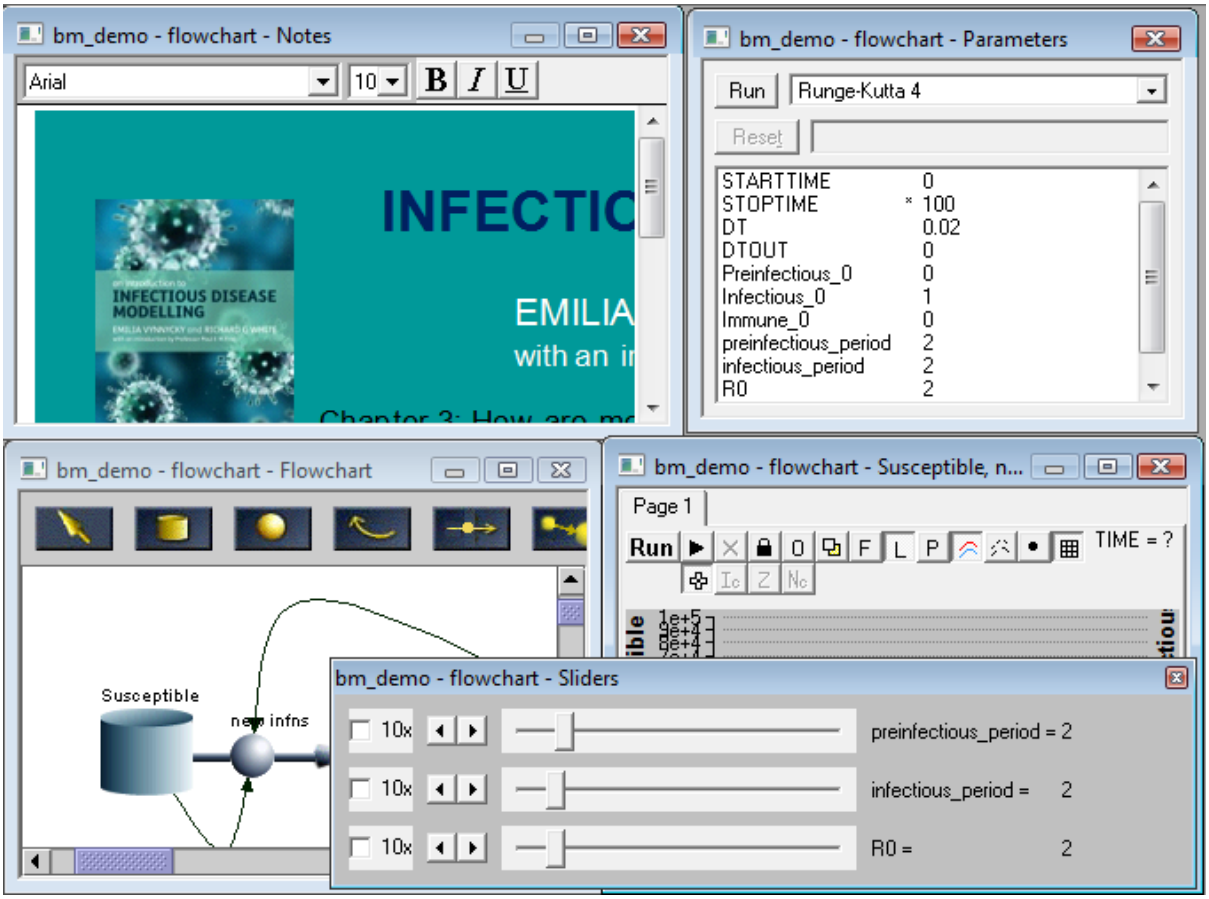
1.2.2 Layout of the model files

1. Start up Berkeley Madonna. If error messages appear, relating to chemrxnl.dll failing to load, and Chemical Reaction Server registration failing, click on OK to continue. These are

minor errors which are usually present when you install Berkeley Madonna and do not affect the running of the program.

2. Open the file bm_demo – flowchart.mmd.

You should now see four windows resembling the following:

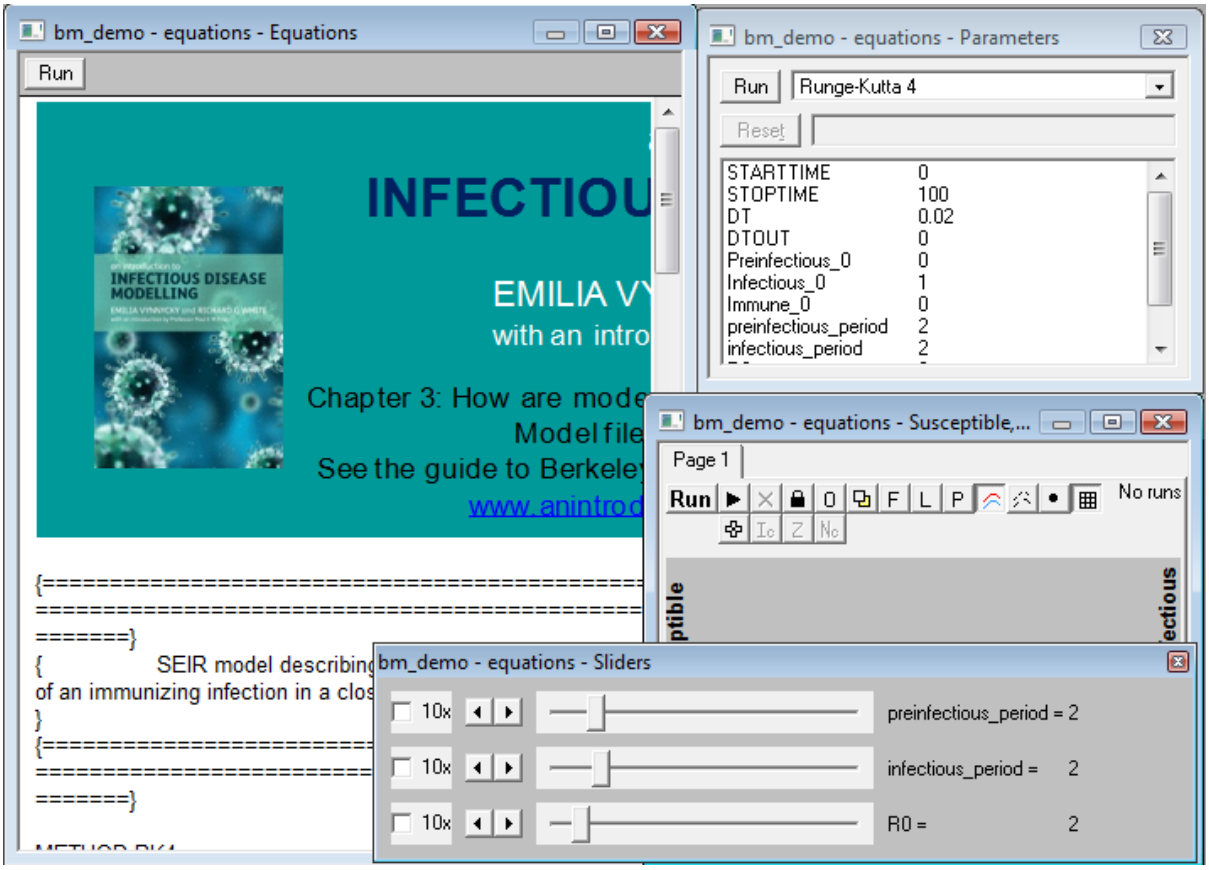


The four windows are as follows:

1. The Notes window, which holds the name and key features of the model.
2. The Parameters window, which holds the values for all the parameters in the model.
3. The Flowchart window, which holds the flow diagram of the model.
4. The “Figures” window, which, when you open up the file, holds an empty graph.

You should also see a sliders window – the sliders also allow you to change the values for the model’s input parameters.

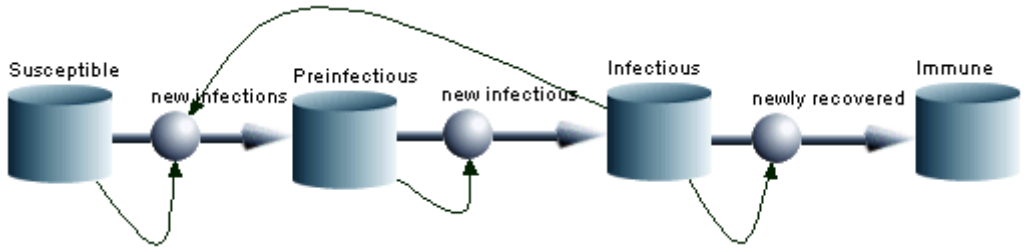
The next page shows the layout of the same model, which is set up using the equation editor (bm_demo – equations.mmd). It is identical to the layout for the model set up using the flowchart, except that the windows with the flow diagram and the notes are replaced by the “Equations” window. This window holds the name of the model, the differential equations and input parameters of the model:



Each of the windows is described below. Note that if any window is not visible at any time, it can usually be found in the list accessible through the Windows option in the main menu.

1.2.3 The flowchart window (flowchart version only)

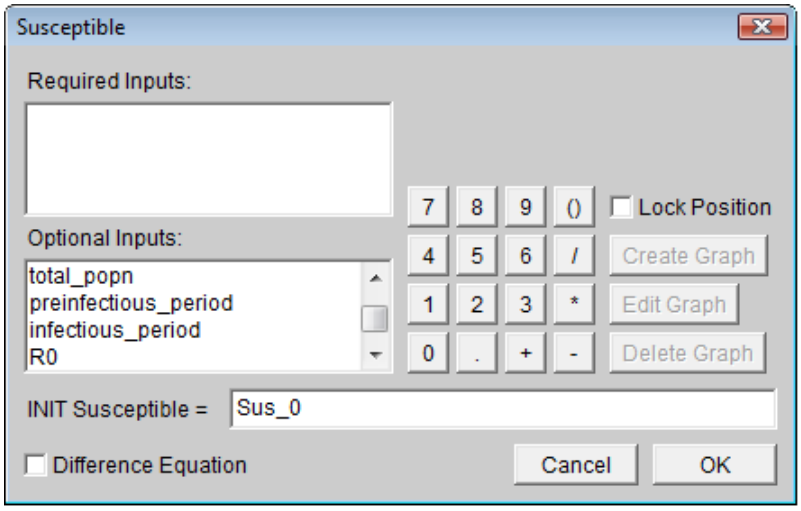
The flowchart window has a diagram which shows the general structure of the model. The following shows the flow diagram that you should see in the flowchart window for Bm_demo, after you've resized the window.



Berkeley Madonna refers to the blue cylinders as 'reservoirs'. These represent the number of individuals in each category (i.e. Susceptible, Pre-infectious, Infectious or Immune) at a given time *t*.

1. Double click on the Susceptible reservoir.

You should see now the following window, which specifies the number of individuals that are **susceptible at the start**. This currently equals Sus_0; we will show where Sus_0 is defined later. The initial values for the other compartments are set up in a similar way.



2. If you have not yet done so, click on the cancel button to exit this window.

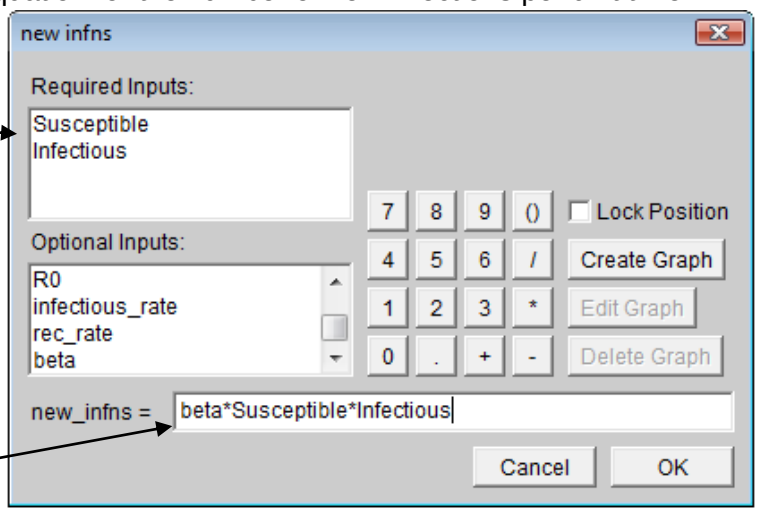
The flow diagram has two types of arrows. The **flow arrow with the circle in the middle**  contains the expression for the number of individuals who move from one category to the next. The **thin arc arrows**  tell Berkeley Madonna which components in the model are used in the expression in the component to which it points. For example, as we shall see below, the expression for the number of new infections per unit time is given by:

$$\beta * \text{Susceptible} * \text{Infectious}$$

Therefore, since the numbers of susceptible and infectious individuals are used in this expression, there are thin arrows going from the susceptible and the infectious compartments to the new infections arrow.

3. Double click on the circle of the flow arrow labelled 'new infections'. You should see the following window, which has the equation for the number of new infections per unit time.

These terms can be used in the expressions for the flow arrows.



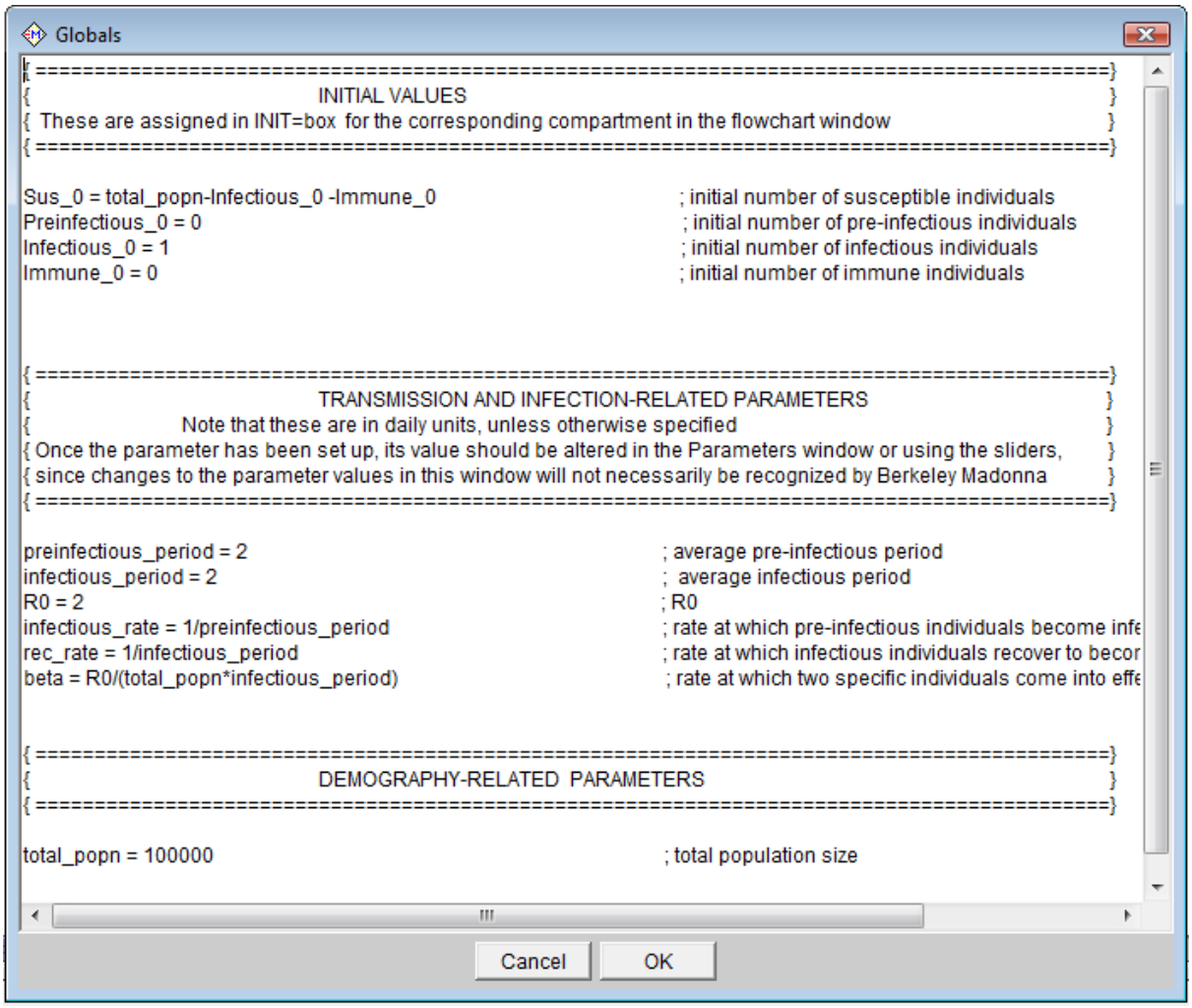
This contains the expression for the number of individuals who are newly infected per unit time.

4. If you have not yet done so, click on the cancel button to exit the window. In a similar way, double click on the new infectious and newly recovered arrows to see the equations for the number of individuals who become infectious or who recover per unit time respectively.

5. Return to the flowchart window and click on the globals button 

on the toolbar.

You should now see the following window, which contains the input parameters.



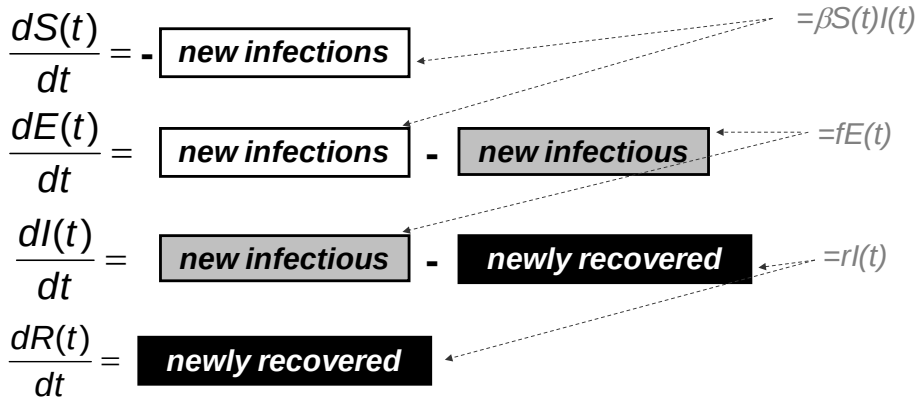
Comments that help you follow the code, but are ignored by Berkeley Madonna, can be included in curly brackets; anything on the same line after a semi-colon is ignored and therefore any comment can be included on the same line as an equation, as long as it is preceded by a semi-colon.

Notice that there is a space on either side of the equals sign in the equations. Berkeley Madonna requires these spaces to be inserted as otherwise, the equations are ignored.

6. Click on cancel to return to the flow diagram. To see the differential equations underlying the flowchart, select the Equations option that is available from the Model option on the main menu. **Note that it is not possible to edit the equations or parameters in the Equations**

window of the flowchart version of the model – any equations have to be edited either in the globals window or in the equations in the flowchart itself (e.g. in the arrows).

The relationship between the equations in the flow arrows and the differential equations in the book is as follows:



where

- *S(t), E(t), I(t)* and *R(t)* are the numbers of susceptible, pre-infectious, infectious and immune (recovered) individuals at time *t*
- *β* is the rate at which two specific individuals come into effective contact per unit time (and equal to R_0/(total population size × duration of infectiousness).
- *f* is the rate at which pre-infectious individuals come infectious (equal to 1/average pre-infectious period)
- *r* is the rate at which individuals recover from being infectious (and equal to 1/infectious period).

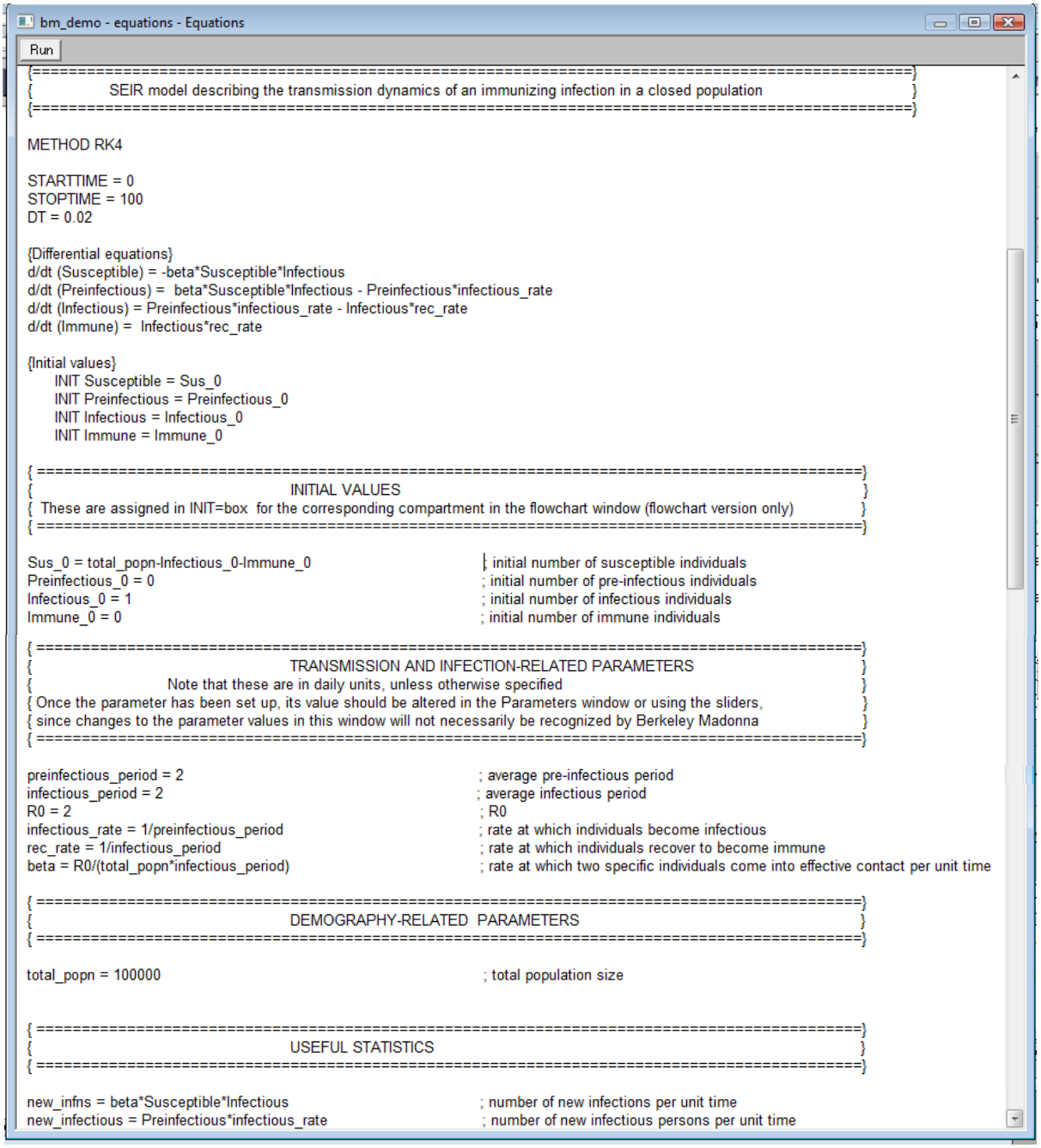
1.2.4 The equation editor window (equation editor version only)

The next page shows what you should see in the equation editor window of Bm_demo – equations.mmd.

The first few lines of the window provide details of the method used to convert the differential equations into difference equations (this is currently set to be “Runge-Kutta 4” or “RK4”), together with the time period over which the model is run and the size of time step used in the difference equations (set by STARTTIME, STOPTIME and DT – see section 1.2.5 on the “Parameters window” for further details).

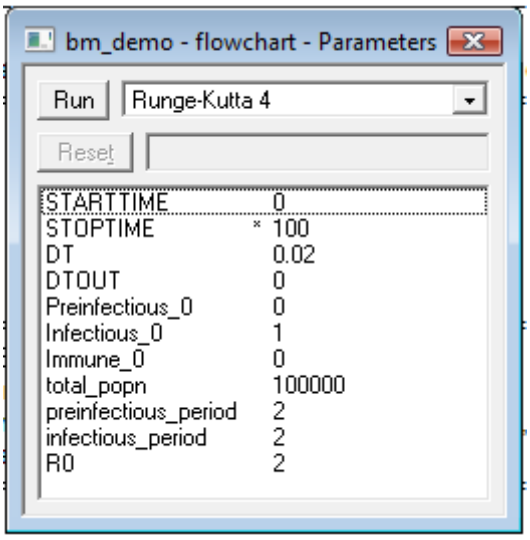
This is usually followed by the differential equations for the model, the initial values for the compartments in the model and the values for the model’s input parameters. The latter are written in the same way that they are written in the globals window of the flowchart editor.

As is the case for the globals window in the flowchart window, comments can be included in curly brackets and anything on the same line after a semi-colon is ignored. Therefore any comment can be included on the same line as an equation, as long as it is preceded by a semi-colon.


bm_demo - equations - Equations
Run
{==========}
{ SEIR model describing the transmission dynamics of an immunizing infection in a closed population }
{==========}
METHOD RK4
STARTTIME = 0
STOPTIME = 100
DT = 0.02
{Differential equations}
d/dt (Susceptible) = -beta*Susceptible*Infectious
d/dt (Preinfectious) = beta*Susceptible*Infectious - Preinfectious*infectious_rate
d/dt (Infectious) = Preinfectious*infectious_rate - Infectious*rec_rate
d/dt (Immune) = Infectious*rec_rate
{Initial values}
INIT Susceptible = Sus_0
INIT Preinfectious = Preinfectious_0
INIT Infectious = Infectious_0
INIT Immune = Immune_0
{==========}
{ INITIAL VALUES }
{ These are assigned in INIT=box for the corresponding compartment in the flowchart window (flowchart version only) }
{==========}
Sus_0 = total_popn-Infectious_0-Immune_0 ; initial number of susceptible individuals
Preinfectious_0 = 0 ; initial number of pre-infectious individuals
Infectious_0 = 1 ; initial number of infectious individuals
Immune_0 = 0 ; initial number of immune individuals
{==========}
{ TRANSMISSION AND INFECTION-RELATED PARAMETERS }
{ Note that these are in daily units, unless otherwise specified }
{ Once the parameter has been set up, its value should be altered in the Parameters window or using the sliders, }
{ since changes to the parameter values in this window will not necessarily be recognized by Berkeley Madonna }
{==========}
preinfectious_period = 2 ; average pre-infectious period
infectious_period = 2 ; average infectious period
R0 = 2 ; R0
infectious_rate = 1/preinfectious_period ; rate at which individuals become infectious
rec_rate = 1/infectious_period ; rate at which individuals recover to become immune
beta = R0/(total_popn*infectious_period) ; rate at which two specific individuals come into effective contact per unit time
{==========}
{ DEMOGRAPHY-RELATED PARAMETERS }
{==========}
total_popn = 100000 ; total population size
{==========}
{ USEFUL STATISTICS }
{==========}
new_infns = beta*Susceptible*Infectious ; number of new infections per unit time
new_infectious = Preinfectious*infectious_rate ; number of new infectious persons per unit time

1.2.5 The Parameters window

The parameters window shows the following:



For most models, it lists the following:

a) The method which Berkeley Madonna uses to convert the differential equations to difference equations and the adjustment made for errors in this conversion. This is normally set to be "Runge-Kutta 4" which we will be using in most of the models. For some exceptional models (see files for chapter 5), we use the "Euler" method using a small step size. This method is equivalent to setting up the same difference equations in Excel.

b) The time period over which the model makes predictions (specified by the values of STARTTIME and STOPTIME, which are set to be 0 and 10 by default). *Note that the units of the start and stop time are determined by those of the parameters used in the flow arrows or, equivalently, the differential equations. If these are in units of per day, then the start and stop times are also in daily units.*

c) The time step which Berkeley Madonna uses in solving the equations (specified by the value of DT which is set to be 0.02 time units by default).

d) DTOUT, which specifies the number of output steps generated in each model run. The default (0) means that the output is generated for each time step i.e. every 0.02 days if DT is set to be 0.02 days.

e) The number of individuals present in various compartments at the start, if they are specified by a number rather than a parameter value (not applicable for the bm_demo model).

f) The current values for all the parameters in the model, e.g. R_0, the pre-infectious and infectious periods etc.

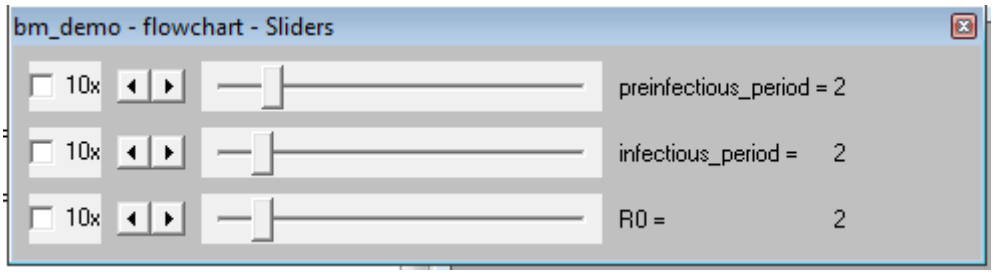
The values for the input parameters can be changed in this window by clicking on the parameter and typing in the new value in the white bar next to the reset button or by using

the sliders (see below). Any parameter whose value has been changed either in the Parameters window or in the sliders, and which therefore differs from the value in the globals or in the Equation editor window has an asterisk to the left of its value in the parameter list. The parameter can be reset to the value in the globals window or in the equation editor by clicking on the parameter and then clicking on the reset button.

Note that the value of the parameter that Berkeley Madonna uses to run the model is the value given in the Parameters window and ***not*** *the value in the globals or equation editor windows.*

1.2.6 The sliders window

The following shows the sliders window that is set up for the file “Bm_demo.mmd”.



Values for parameters can be changed by clicking on the slider for that parameter and dragging it to the left or to the right, as appropriate, using either the mouse or the arrow keys. For example, clicking on the vertical bar  of the slider for the pre-infectious period in the file “Bm_demo”, and then dragging it to the right will result in the assumed value for the pre-infectious period *increasing*; clicking and dragging it to the left will result in the assumed value for the pre-infectious period *decreasing*.

The settings for the sliders can be changed by double clicking on the slider; if the sliders have been accidentally closed, they can be recovered by selecting the Parameters option from the main menu and selecting the “Show sliders” option.

1.2.7 Running the model

The following lists the options that are available for running the model (whereby Berkeley Madonna uses the equations to generate predictions):

1. Click on the Run button in the Parameters window.
2. Click on the Run button in the Figures window.
3. Select the Model option from the main menu and select the run option.
4. Select the Compute option from the main menu and select the run option.
5. Move the slider for a given parameter to the desired position.
6. Click the Run button in the equation window (available from the Model option in the main menu if it is not visible).

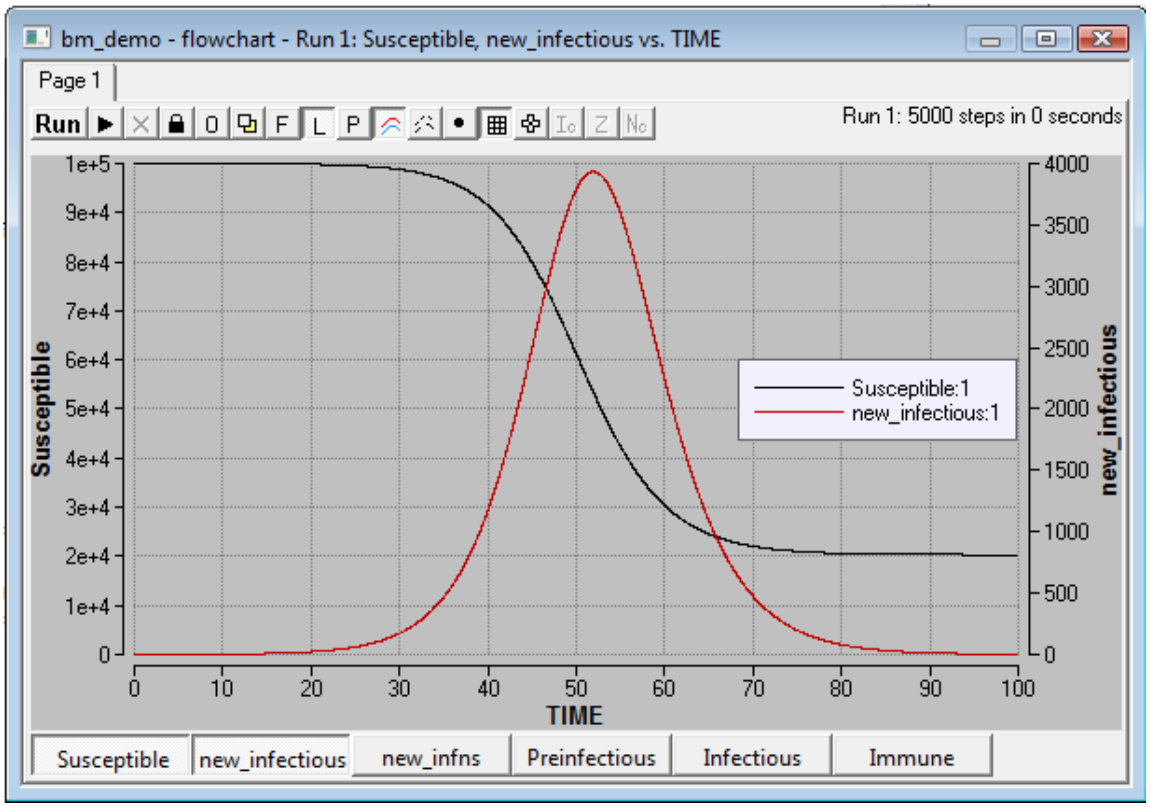
The Figures window is updated each time that the model is run.

When you run the model, Berkeley Madonna may come up with a prompt saying that existing runs will be discarded if the model is recompiled (or a related message about recompiling). You should click on OK (or Yes, if appropriate) to continue.

1.2.8 Viewing the results of model runs – the Figures window

Overview of the layout

The following shows the figure that you will see when you run either the flowchart or equation editor version of "Bm_demo.mmd" (assuming that you have not changed any values for the input parameters).



The buttons along the top of the Figures window allow you to change the display (see Table 1 on page 16 for the definitions of these buttons).

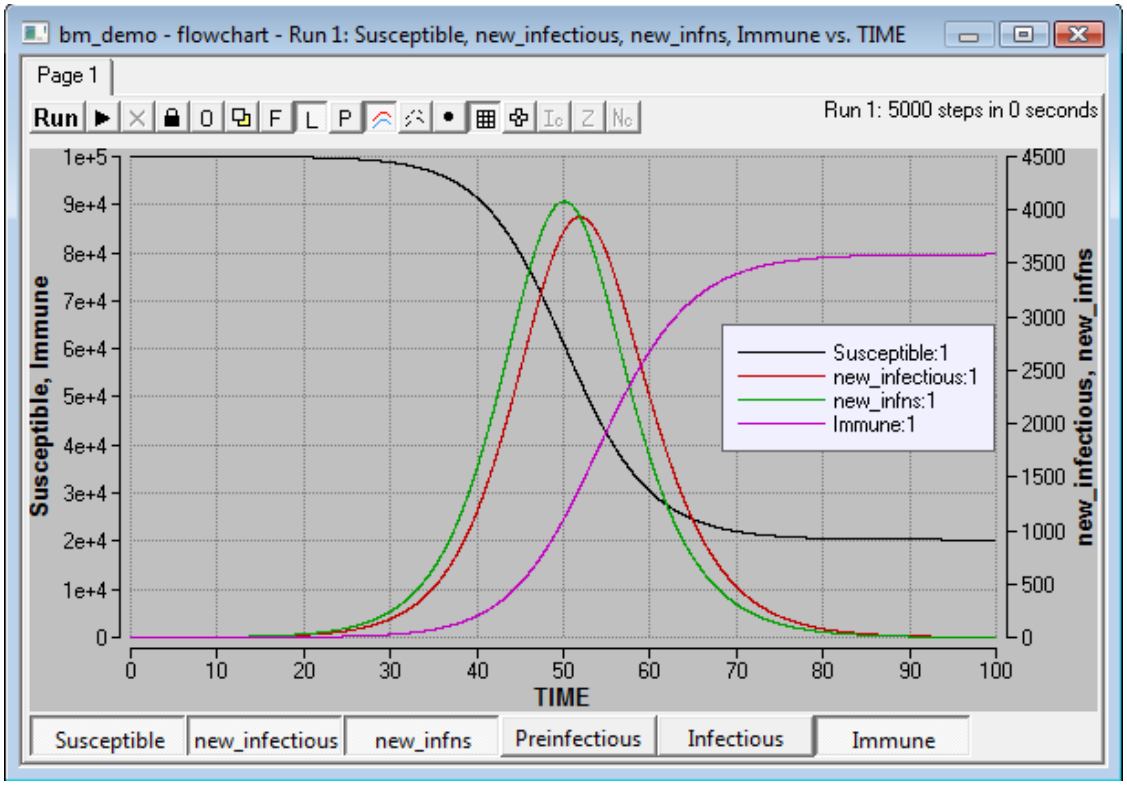
At present, the Figures window in the file Bm_demo.mmd plots the number of susceptible persons and the number of new infectious persons per day.

NB Please avoid clicking on the X button on the top right-hand side of this window! Clicking on this button will result in the window being deleted and once this happens, the window cannot be recovered (unless you have a previous version of the model with the figures set up).

Changing the variable plotted

The Figures window includes buttons at the bottom of window for variables that can be added to the plot. If a variable is already included in the plot, clicking on the button for that

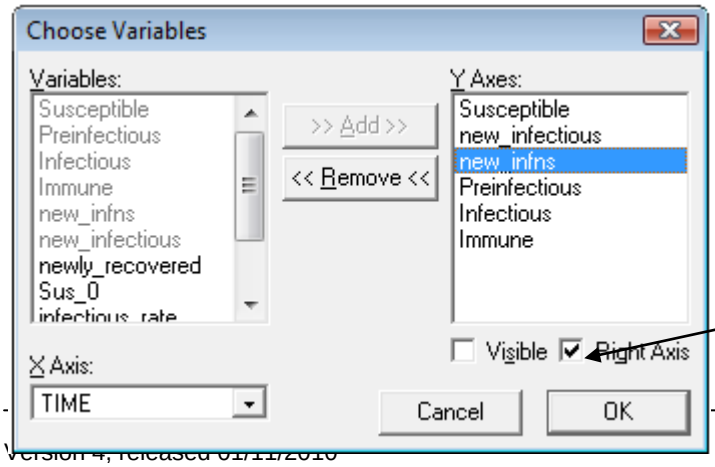
variable will remove the variable from that plot. If a variable is not already plotted, clicking on the button for that variable will result in that variable being added to the plot.
For example, to add a plot of the number of immune individuals and new infections over time, click on the "*Immune*" and "*new infns*" buttons. By default, time (t) is plotted on the x-axis. At this stage, your Figures window should resemble the following:



It is possible to add buttons for variables that are not listed at the bottom of the window. To do this, double click in the middle of the figures window. This will open up a new window called "Choose variables". Double clicking on the variable of interest in the left hand side of this window will add that variable to the list under the "Y Axes" section on the right hand side of window. This section lists the components for which buttons will be set up at the bottom of the figures window. Clicking on OK will result in the button for that variable appearing at the bottom of the window; subsequently re-run the model will add the variable to the plot.

Changing the axis on which a variable is plotted

In the above plot, the number of new infections and the number of new infectious persons per unit time are plotted on the right hand y-axis. To change the plot so that the number of new infections per unit time appears on the left-hand y-axis, double click in the middle of the window, select new infns from the list (called Y-axes) on the right-hand side of the window, uncheck the box next to the Right Axes, and click on OK.



Changing the scales on the x- or y-axis

The scale on the x-or y-axis can be changed by double clicking on one of the axes. In the window which then appears, click on the "Scales" tab, deselect the auto button for the axis that you're interested in and change the minimum and maximum values. The value of #Div specifies how many gridlines are provided for the given axis.

Comparing the output obtained using different sets of assumptions

It is possible to compare the results from different runs of the model by clicking on the overlay button  in the Figures button and then running the model for the assumptions of interest. Multiple lines will then appear in the figure, one for each assumption (see below).

Clicking on the overlay button again and then running the model again will remove all the results from previous runs from the plot.

Parameter plot

It is possible to produce a plot which shows how some outcome varies for different values of a given parameter. For example, suppose we wanted to explore how the size of the peak number of infectious persons during an epidemic varies according to the size of R_0. To do this, we would need to follow the steps below using our bm_demo model:

1. Select the Parameter plot option from the Parameters option in the main menu. This will open a Parameter plot window. :
2. Select R_0 from the drop-down box in the Parameter options in this window.
3. Select number of runs to be 20, with the initial value of 1 and the final value to be 20. This specifies the number of values for R_0 that you will explore and the range in which they lie.
4. Double click on "new infectious" from the list under the "Variables" section. This adds it to the list under the "Y-axis" section.
5. Click on the box to the left hand size of the "Maximum" option to specify that you're interested in the maximum number of new infectious persons seen during an epidemic predicted for an infection with a given value for R_0.
6. Click on Run to run the model.

Table 1: Summary of the buttons which are available on the graphical toolbar:

	Definition	**Description**
Run	Run	Runs the model.
▶	New page	Inserts another page of graphs.
✕	Delete	Deletes the current page of graphs.
🔒	Lock	Locks the current page of figures: the figures will remain unchanged if you re-run your model with other parameter values.
0	Overlay plots	If this button is pressed down, then the results of further runs of your model will be added to the current figure.
	Table	Presents the values for the series being plotted. To export these values e.g. to Excel, click on the "Save Table as" option from the File option on the main menu. You can then import the resulting file into Excel or other programs.
F	Fast Fourier transform	Carries out a Fast Fourier transform of the model output – consult the manual for Berkeley Madonna for further details.
L	Legend	Adds the legend to the figure.
P	Parameters	Adds the parameter values to the figure.
	Colours	Changes the colour scheme from "colour" to "black and white".
	Dashed lines	Changes the lines from being solid to dashed.
•	Data points	Presents the series plotted as individual data points rather than as a continuous line.
	Grid	Adds gridlines to the figure.
	Readout	When this is pressed, a cross in a circle appears in the Figure when you click anywhere on the Figure; the co-ordinates of this cross are shown on the top right hand corner of the window. Clicking anywhere in the window will move this cross to that location. Note that Berkeley Madonna labels the co-ordinates using the titles on the x- and y-axes. For example, if the y-axis title is labelled "new infns", then the y- co-ordinate is labelled as "new infns=", irrespective of the position of that co-ordinate.
Z	Zoom out	Note that you can zoom in on a portion of the graph by clicking with the left mouse button anywhere on the graph when the pointer is a solid black cross and, holding the left mouse button down, dragging across to the point that you're interested in. Clicking on the zoom out button will return the view of the figure to normal.

1.3 Possible problems when running models

Floating point error

Once this error comes up, Berkeley Madonna usually asks whether you would like reduce DT and start again. **You should always decline this offer and then identify the source of the problem**. Otherwise Berkeley Madonna will reduce the size of the time step used to solve the differential equations until it can run it, and this will increase the run time, without necessarily solving the problem. For models which include a β parameter, this error can result from the β parameter being too large (i.e. you may be using an inappropriate value for β). An alternative source of the problem may be that the value of some parameter equals zero and Berkeley Madonna is trying to divide by this parameter.

Matrix reserve error

This kind of error usually comes up when you're trying to run a model and you've generated several figures. The reason behind it is that each time that you run your model, Berkeley Madonna tries to store the values for all the variables (over time) included in the buttons at the bottom of the each page of figures. If you have many figures and many buttons, then the package is unable to store the values for all the variables and says that it has a Matrix reserve problem. To solve the problem, delete any unnecessary figures (by clicking on the "x" button on the toolbar of the figures page), and delete the buttons for any variables that you're not interested in. To do the latter, double click in the middle of the figure, and double click on the corresponding variable listed on the right hand side of the "Choose variables" window. This will remove the button for that variable. You should find that your model will run once sufficient buttons and figures have been deleted.

Changes to parameters do not seem to affect the output

Note that the parameter values which Berkeley Madonna uses when running the model are those given in the Parameter window (accessible via the Parameter option in the main menu) **or in the Sliders**. Changes to the parameters made in the sliders will be reflected in the parameter window. H**owever, changes made to the values of the parameters (and sometimes the equations) in either the globals window or the equation editor are sometimes not recognized unless you first recompile the model. To do this, click on the model option in the main menu, and choose the compile option.**

Documentation and suggested exercises

Chapter 2

How are models set up? I. An introduction to difference equations

2.1 Model 2.1

2.1.1 Overview of the model

This model, set up using difference equations in Excel, describes the course of an influenza epidemic in a closed population, and can be used to reproduce Figure 2.8 in the book.

The key features of the model are:

- One infectious person is introduced into a population where all other 99,999 persons are susceptible.
- There are no births into or deaths out of the population.
- The parameters are currently set to be those for influenza (R_0=2, average pre-infectious period=average infectious period=2 days).

The layout of the spreadsheet is as follows:

1. **Yellow cells (cells A12:I21).** The cells in column F hold the values for the key input parameters in the model. These include the size of the time step used in the difference equations (see below) in cell F15, the total population size (cell F16), average pre-infectious and infectious periods (cells F17 and F18 respectively) and the basic reproduction number (R_0) (cell F19).

   Notice that the cells have been assigned the names provided in column G. Therefore, if any cell has an equation which uses this cell, the name rather than the cell location can be used. For example, cell F16 has been assigned the name "tot_popn", and so the term "tot_popn", rather than the location of the cell, can be used in any equation (see, for example, the equation in cell F22).

2. **Turquoise cells (cells A22:I27).** The cells in column F hold the values for β (cell F22), the average rate of onset of infectiousness (cell F23), and the average recovery rate (F24). These are calculated using the values in the yellow cells and equations 2.11 and 2.16 in the book.

3. **Lilac cells (row 52 onwards).** These hold equations for the number of susceptible, pre-infectious, infectious and immune individuals on a given day, expressed in terms of the number on the preceding day. These equations are the Excel equivalent of equations 2.1, 2.2, 2.3 and 2.4 in the book.

4. **Figure 1.** This shows a plot of the number of susceptible and immune individuals over time, together with the number of new infectious persons per time step (this is identical to Figure 2.8 in the book).

2.1.2 Suggested exercises

1. a) How do you think the population size in this model changes over time? Check your answer by setting up a suitable expression in the lilac cells of column I which sum up the numbers of individuals in each compartment at each time point.

b) How do you think predictions of the number of new infections per day should differ from the number of new infectious persons per day? Check your answer by setting up an appropriate expression in cell F56 and copying it down until the 200th day. A line showing the number of new infections per day should now appear in Figure 1.

c) How should the plot of the number of new infectious persons per day differ from that of the number of infectious persons? Check your hypothesis by adding a line for the number of infectious persons (given by the values of the lilac cells in column D) to Figure 1.

2. a) How should model predictions of the number of new infectious persons per day and the number of immune individuals at the end of the epidemic change if you were to use an SIR model, rather than an SEIR model to describe the transmission of influenza?

b) Check your answer by following the steps below:

i) Select columns J and U together, clicking with the right mouse button, selecting the unhide option. You should now see Figure 2, which shows the number of immune and new infectious persons per day which are also shown in Figure 1, and some blank lilac cells.

ii) Set up appropriate expressions for the number of susceptible, infectious, immune individuals and the number of new infectious persons on day 1 in cells L56, M56, N56 and O56 respectively, and copy them down until the 200th day. Figure 2 should now contain a plot comparing predictions from the SEIR and SIR models.

Chapter 3

How are models set up? II. An introduction to differential equations

3.1 Model 3.1

3.1.1 Overview of the model

This model, set up using difference equations in Excel, describes the course of a measles and influenza epidemic in a closed population, and can be used to see the patterns shown in Figure 3.2 in the book. It comprises two worksheets (influenza and measles) and the layout of each sheet is very similar to that in the worksheet in model 2.1.

The key features of the model are:

- One infectious person is introduced into a population where all other 99,999 persons are susceptible.
- There are no births or deaths into or out of the population.
- The parameters in the influenza worksheet are currently set to be the following: R_0=2, average pre-infectious period=average infectious period=2 days. Those in the measles worksheet are as follows: R_0=13, average pre-infectious period=8 days, average infectious period=7 days.

The layout of the spreadsheet is as follows:

1. **Yellow cells (cells A12:I21).** The cells in column F hold the values for the key input parameters in the model. These include the size of the time step used in the difference equations (see below) in cell F15, the total population size (cell F16), average pre-infectious and infectious periods (cells F17 and F18 respectively) and the basic reproduction number (R_0) (cell F19).

   Notice that the cells have been assigned the names provided in column G. Therefore, if any cell has an equation which uses this cell, the name rather than the cell location can be used. For example, cell F16 has been assigned the name "tot_popn", and so the term "tot_popn", rather than the location of the cell, can be used in any equation (see, for example, the equation in cell F22).

2. **Turquoise cells (cells A22:I27).** The cells in column F hold the values for β (cell F22), the average rate of onset of infectiousness (cell F23), and the average recovery rate (F24). These are calculated using the values in the yellow cells and equations 2.11 and 2.16 in the book.

3. **Lilac cells (row 52 onwards).** These hold equations for the number of susceptible, pre-infectious, infectious and immune individuals on a given day, expressed in terms of the number on the preceding day. These equations are the Excel equivalent of equations 2.1, 2.2, 2.3 and 2.4 in the book.

4. **Figure:** This shows a plot of the number of infectious persons over time.

3.1.2 Suggested exercises

1. a) Considering the influenza model, change the size of the time step to be 0.05, 0.1, 0.5, 1 and 2 days to check that you obtain the same patterns in the number of infectious persons over time that are seen in Figure 3.2.

b) How do you think your predictions will change if you reduce the size of the time step further to 0.01 days?

c) How might predictions of the epidemic size (as reflected in the number of individuals that are immune at the end of the epidemic) change as the time step size changes? Check your hypothesis using the model.

d) Repeat questions a) - c) using the measles model.

3.2 Model 3.2

3.2.1 Overview of the model

This model, set up in Excel, describes the transmission dynamics of measles over the course of about 100 years in a population comprising 100,000 persons, incorporating births into and deaths out of the population. The layout of the spreadsheet is very similar to that in model 2.1.

The key features of the model are:

- One infectious person is introduced into a population where all other 99,999 individuals are susceptible.
- All individuals are born susceptible.
- Susceptible, pre-infectious, infectious and immune individuals are assumed to experience the same death rate.
- The infection-related parameters in the spreadsheet are currently set to be those for measles: R_0=13, average pre-infectious period=8 days, average infectious period=7 days. The demographic parameters are: average life expectancy= 70 years, average per capita mortality rate = average per capita birth rate = 3.9×10^{-5} per person per day.
- The equations in the model are the Excel equivalent of those obtained in Exercise 2.2a in the book.

The layout of the spreadsheet is as follows:

1. **Yellow cells (cells A12:I21).** The cells in column F hold the values for the key input parameters in the model. These include the size of the time step used in the difference equations (see below) in cell F15, the total population size (cell F16), average pre-infectious and infectious periods (cells F17 and F18 respectively), the basic reproduction number (R_0) (cell F19), and the life expectancy (cell F20).

   Notice that the cells have been assigned the names provided in column G. Therefore, if any cell has an equation which uses this cell, the name rather than the cell location can be used. For example, cell F16 has been assigned the name "tot_popn", and so the term "tot_popn", rather than the location of the cell, can be used in any equation (see, for example, the equation in cell F22).

2. **Turquoise cells (cells A22:I27).** The cells in column F hold the values for β (cell F22), the average rate of onset of infectiousness (cell F23), the average recovery rate (F24), the average per capita mortality and birth rates (cells F25 and F26 respectively) and the average numbers of births per time step (cell F27). These are calculated using the values in the yellow cells and equations 2.11 and 2.16 in the book.

3. **Lilac cells (row 52 onwards).** These hold equations for the number of susceptible, pre-infectious, infectious and immune individuals on a given day, expressed in terms of the number on the preceding day. These equations are the Excel equivalent of those obtained in Exercise 2.2a in the book

4. **Figure.** This shows a plot of the number of susceptible, immune and new infectious persons per time step over time. This shows that these statistics cycle over time – these patterns are discussed in detail in chapter 4.

3.2.2 Suggested exercises

1. How should the total population size change over time in this model? Check your hypothesis by setting up a suitable expression in cell I55 for the total number of individuals in the population and copying it down until the final time step.

2. Change the infection parameters in the model to be those for rubella in a low transmission setting (R_0=7, average pre-infectious period = 10 days, average infectious period = 11 days).

   a) How do the cycles in rubella incidence compare against those predicted for measles? Suggest possible explanations for the differences in the cycles in the two infections.

   b) Suppose you would like to explore the impact of different interventions that are introduced once the incidence of rubella has reached equilibrium. Suggest possible ways of speeding up how quickly the model reaches equilibrium.

Chapter 4

What do models tell us about the dynamics of infections?

4.1 Models 4.1 and 4.1a

4.1.1 Overview of Model 4.1

This model, set up in Berkeley Madonna, describes the transmission dynamics of measles in a closed population following the introduction of an infectious person. It can be used to see the plots in Figures 4.3, 4.4, 4.5, 4.6 and 4.13 of the book.

The key features of the model are described below, where the sections mentioned can be found in the globals window of the flowchart version, or in the equation window of the equation editor version of the model:

1. One infectious person is introduced into a population in which a proportion of the individuals is immune (see the section called “INITIAL VALUES”). This proportion is specified by the value of prop_immune_0, which is defined in the section called TRANSMISSION AND INFECTION-RELATED PARAMETERS.

2. The parameters are currently set to be those for measles (R_0=13, average pre-infectious period=8 days, average infectious period=7 days) - see the section called “TRANSMISSION AND INFECTION-RELATED PARAMETERS”.

3. The population size remains stable over time, with 100,000 individuals and no births or deaths. For further details, see the section called “DEMOGRAPHY-RELATED PARAMETERS”. You can check this by plotting the value for sum_pop over time (defined in the USEFUL STATISTICS section).

4. The variable “Cum infectious” stores the cumulative number of infectious persons over time. Notice the way that this is set up in the flowchart version of the model. For example, the equation in the arrow which goes into the “Cum infectious” compartment is set up so that it equals whatever is in the “new infectious” arrow, thereby keeping track of the total number of individuals who become infectious over time.

5. Several useful statistics have been set up in the "USEFUL STATISTICS" section, such as the proportion of the population that is susceptible or immune. For simplicity, the latter has been defined simply as 1-proportion susceptible.

The window with the figures includes 5 figures:

Page 1: This plots the proportion of the population that is susceptible, immune, R_n (left hand axis) and the number of new infectious persons per day over time (right-hand axis). This can be used to reproduce Figures 4.3 and 4.4 in the book.

Page 2: This is similar to page 1, except that the scale has been changed on the left axis to go from 0 to 12. This can be used to reproduce the figure for measles in Figures 4.3b in the book.

Page 3: This plots the number of new infectious persons per day over time on a natural log scale. This can be used to reproduce Figure 4.5 in the book.

Page 4: This plots the natural logs of the number of new infectious persons per day, the number of infectious persons and the cumulative numbers of infectious persons over time. This can be used to reproduce Figures 4.6 in the book.

Page 5: This plots the proportion immune (left axis) and number of new infectious persons per day (right axis) over a period of 2000 days. This can be used to reproduce Figure 4.13 in the book.

4.1.2 Overview of Model 4.1a

For reference, Model 4.1 has also been set up using difference equations in Excel - see Model 4.1a.xlsx. The layout is very similar to that for model 3.1 except for the following:

1. It includes an additional parameter prop_immune_0 (see cell F20), which reflects the proportion of the population that is immune at the start.
2. The lilac cells in row 52 onwards of Column H hold equations for the cumulative numbers of infectious persons.
3. The lilac cells in row 52 onwards of Columns K, L and M hold equations for the proportion susceptible, immune and the net reproduction number.
4. Figures 1-3 are similar to Figures 4.3-4.5 of the book.
5. Clicking on the ⊞ button above column Z will reveal a plot resembling Figure 4.6 in the book, together with some lilac cells in row 52 onwards. These lilac cells hold equations for the natural logs of the number of new infectious persons per day, the number of infectious persons and the cumulative numbers of infectious persons.
6. Clicking on the ⊞ button above column AL will reveal a plot which shows the patterns in Figure 4.13 in the book, for a given infection.

4.1.3 Suggested exercises

1. Run Model 4.1 and check that it reproduces the patterns for measles shown in Figures 4.3, 4.4, 4.5, 4.6 and 4.13 of the book.

2. Repeat question 1 using the parameters for influenza (R_0=2, average pre-infectious period =average infectious period = 2 days).

3. The following table summarizes the values of the parameters used to produce the plots for the infections in Figure 4.13. Run the model using these values to check that they lead to plots consistent with those in Figure 4.13.

Infection	R_0	Pre-infectious period (days)	Infectious period (days)
Measles	13	8	7
Influenza	2	2	2
Rubella	7	10	11
Polio	7	7	30
Smallpox	4	14	21

4. Carry out a parameter plot (see page 15) to see how the epidemic size (as implied by the proportion of individuals that are immune at the end) varies according to the size of R_0. You should produce a figure which is similar to that on Figure 4.12 in the book.

4.2 Model 4.2

4.2.1 Overview of the model

This model, set up using difference equations in Excel, describes the course of an influenza epidemic in a closed population. It can be used to reproduce Figure 4.15 in the book and to estimate unknown parameters, such as the basic reproduction number, the initial number of infectious persons and the proportion of infectious persons who had sufficient clinical symptoms for them to be reported as cases. It comprises one worksheet whose layout is very similar to that in the worksheet in model 2.1.

The key features of the model are:

- Infectious persons are introduced into a population where 30% of the 5234 persons are immune.
- There are no births or deaths into or out of the population.
- The average pre-infectious period is assumed to equal the average infectious period=2 days.

The layout of the spreadsheet is as follows:

1. **Yellow cells (cells A12:I23).** The cells in column F hold the values for the key input parameters in the model. These include the following:
    - the size of the time step used in the difference equations (cell F15);
    - the total population size (cell F16);
    - average pre-infectious and infectious periods (cells F17 and F18 respectively);
    - the basic reproduction number (R_0) (cell F19);
    - the proportion of the population that is immune at the start (cell F20);
    - the proportion of those infectious persons who are reported as cases (cell F21), and
    - the number of infectious persons in the population at the start (cell F22).

    Notice that the cells have been assigned the names provided in column G. Therefore, if any cell has an equation which uses this cell, the name rather than the cell location can be used. For example, cell F16 has been assigned the name “tot_popn”, and so the term “tot_popn”, rather than the location of the cell, can be used in any equation (see, for example, the equation in cell F24).

2. **Turquoise cells (cells A24:I26).** The cells in column F hold the values for β (cell F24), the average rate of onset of infectiousness (cell F25), and the average recovery rate (F26). These are calculated using the values in the yellow cells and equations 2.11 and 2.16 in the book.

3. **Lilac cells (row 41 onwards).** Columns B-E hold equations for the number of susceptible, pre-infectious, infectious and immune individuals on a given day, expressed in terms of the number on the preceding day. These equations are the Excel equivalent of equations 2.1, 2.2, 2.3 and 2.4 in the book. Column F holds the number of new infectious persons per day; columns G and H hold the number of new cases reported per day and per week respectively.

4. **Figure:** This shows a plot of the number of new infectious persons, together with the number of susceptible and immune persons over time.

5. **buttons next to row 40 and above column Q.** If you click on these buttons, you will see the following:
   a) Orange cells (cells K41:M57) containing data on the number of cases reported each week in Cumberland, USA between August and December 1918. These are identical to the data plotted in Figures 4.9a and 4.15 (dashed lines) in the book.

   b) Lilac cells (cells N41:N57) containing model predictions of the number of cases during the period spanned by the Cumberland data. These are calculated using Excel's VLOOKUP function by referring to the contents of cells H44:H444.

   c) Grey cells which describe the goodness of fit of the model (based on the current values for the parameters). The log-likelihood deviance of the model prediction from the dataset is provided in cell D28. It is calculated using the contents of cells O41:O57 and P41:P57, which hold contributions to the deviance from each data point, based on the "saturated" and transmission models respectively. See section 4.2.5 of the book and the references cited there for further details.

   d) A figure, which is similar to Figure 4.15 in the book, comparing model predictions of the number of cases occurring each week in Cumberland, 1918.

4.2.2 Suggested exercises

1. At present, the model assumes that R_0=2.1, which is consistent with the value obtained using the size of the epidemic (see Example 4.2.4.1 on page 79 of the book).

   a) How might you need to change the value for R_0 in the model if you wanted to improve the fit of model predictions to the data, without changing the values for any of the other parameters?

   b) Without changing any assumptions in the model, follow the steps in section 10.1.1 of this document to check your hypotheses by obtaining the best-fitting values for the basic reproduction number.

   c) What other parameters might you want to change in the model to further improve the fit of model predictions to the data?

2. a) Follow the steps in section 10.1.2 of this document to estimate the values for the R_0, proportion of infectious persons that are reported and the initial numbers of infectious persons which lead to the best-fit to the data and check that these values are consistent with those presented in Example 4.2.5.1 on page 82 of the book.

   b) Follow the steps in section 10.1.5 of this document to calculate 95% confidence intervals on these estimates and check that the values are consistent with those presented in Example 4.2.5.1 on page 82 of the book

3. a) How do you think the best-fitting values for the parameters obtained in question 2b) will change if we were to assume that the pre-infectious and infectious periods are both shorter (e.g. equal to 1.5 days) than currently assumed in the model,? Test your hypotheses by changing the pre-infectious and infectious periods accordingly and refitting the model using the maxlhood03 macro (see section 10.1.5 of this document).

   b) How might the best-fitting estimates values for the parameters change if we were to assume that no one in the population was immune at the start? Check your hypothesis by changing the value for the proportion immune at the start accordingly and refitting the model (after changing the values for the pre-infectious and infectious periods to both equal 2 days).

4. The cases in the dataset from Cumberland were identified actively through house to house surveys. If the study identified 90% of symptomatic persons in population, what proportion of infectious persons would have been symptomatic in the population?

4.3 Models 4.3 and 4.3a

4.3.1 Overview of Model 4.3

This model, set up in Berkeley Madonna, describes the long-term dynamics of measles, taking account of births into and deaths out of the population, as described in Panel 4.3 of the book. It can be used to see the plots in Figures 4.17 and 4.19 of the book.

The key features of the model are described below, where the sections mentioned can be found in the globals window of the flowchart version, or in the equation window of the equation editor version of the model:

1. One infectious person is introduced into a population where all other 99,999 individuals are susceptible (see the section called "INITIAL VALUES").

2. The parameters are currently set to be those for measles (R_0=13, average pre-infectious period=8 days, average infectious period=7 days) - see the section called "TRANSMISSION AND INFECTION-RELATED PARAMETERS".

3. Individuals die at a constant rate, which is equal to 1/(average life expectancy), where the average life expectancy is 70 years. For further details, see the section called "DEMOGRAPHY-RELATED PARAMETERS" and the equations for the numbers of deaths per unit time from each compartment.

4. All newborns enter the susceptible compartment at a rate which equals the mortality rate. For further details, see the section called "DEMOGRAPHY-RELATED PARAMETERS" and the equation for "new births".

5. The population size remains stable over time, with 100,000 individuals, with the *per capita* birth rate equal to the *per capita* mortality rate (so that the numbers of births equals the numbers of deaths per unit time). You can check this by plotting the value for sum_pop over time (see page 5 of the Figures window).

6. Several useful statistics have been set up in the "USEFUL STATISTICS" section, such as the proportion of the population that is susceptible or immune. For simplicity, the latter has been defined simply as 1-proportion susceptible.

The window with the figures includes 5 figures:

Page 1: This plots the number of new infectious persons per day over 25500 days after the start, producing a plot that is similar to Figure 4.17 in the book.

Page 2: This compares the net reproduction number (left-hand axis) against the number of new infectious persons per 100,000 population (right hand axis) between the 40th and 50th years, reproducing a plot that is similar to Figure 4.19a in the book.

Page 3: This compares the proportion of the population that is susceptible (left-hand axis) against the number of new infectious persons per 100,000 population (right hand axis) between the 40th and 50th years, reproducing plots that are similar to Figures 4.19b and 4.20 in the book.

Page 4: This compares the proportion of the population that is immune (left-hand axis) against the number of new infectious persons per 100,000 population (right hand axis) between the 40th and 50th years, reproducing a plot that is similar to Figure 4.19c in the book.

Page 5: This plots the number of Susceptible, Pre-infectious, Infectious and Immune individuals, together with the value for sum_pop over time.

You may notice that the plot of the number of new infectious persons does not cross the plots of the proportion susceptible, R_n and the proportion immune at exactly the same points that they do in the figures in the book. The reason for this is that the plots in the book were generated in Excel using difference equations and a 1 day time step. This is equivalent to using the Euler method with a value of DT of 1 day. You will find that if you change the corresponding settings in the Parameters window of Model 4.3, you will see the same plot that is provided in the book.

4.3.2 Overview of Model 4.3a

For reference, Model 4.3 has also been set up using difference equations in Excel - see Model 4.3a.xlsx. The layout is very similar to that for model 3.2 except that it includes additional figures, corresponding to Pages 2-3 of the Figures window in the Berkeley Madonna version of this model. These figures are labelled Figures 1-3. The equations which are used to calculate the proportion of the population that is susceptible or immune and the net reproduction number for these figures are in columns J-O.

4.3.3 Suggested exercises

1. Run model 4.3 and answer the following questions. You may also want to check that you get consistent answers using Model 4.3a.
   a) According to page 2 of the figures window, what is the size of the net reproduction number whilst the number of new infectious persons per day is:
      i) increasing?
      ii) decreasing?
      iii) at a peak or a trough?

   b) According to page 3 of the figures window, what proportion of the population is susceptible whilst the number of new infectious persons per day is:
      i) increasing?
      ii) decreasing?
      iii) at a peak or a trough?

   How do these values relate to the epidemic threshold, $1/R_0$?

   c) According to page 4 of the figures window, what proportion of the population is immune whilst the number of new infectious persons per day is:
      i) increasing?
      ii) decreasing?
      iii) at a peak or a trough?

How do your answers relate to the herd immunity threshold, 1-1/R_0?

2. Return to page 2 of the figures window and add the numbers of births per unit time ("new births") to the plot. What do you notice about the proportion of the population that is susceptible whilst the number of new infectious persons per day is
   i) above the numbers of births per day?
   ii) below the numbers of births per day?

   Is this what you would expect and why?

3. a) The following table shows possible values for R_0, the average pre-infectious and infectious periods for several infections. For which infection would you expect the inter-epidemic period to be:
   i) the shortest?
   ii) the longest?

	Pre-infectious period (days)(*D'*)	**Infectious period (days) (*D*)**	**R_0**
Measles	8	7	12-18
Varicella	14	7	3-17
Smallpox	14	21	5-7
Rubella	10	11	6-7

   b) Confirm your hypotheses by running the model, after changing parameter values accordingly. Check that the cycles that are predicted are consistent with those presented in Figure 4.16 of the book and with the values predicted by equation 4.31 in the book.

4.4 Model 4.4

4.4.1 Overview

This model, set up in Excel, describes the transmission of measles using the equations that are implicit in Hamer (1906) – see Panel 4.4. It can be used to see the patterns shown in Figure 4.22 in the book. It comprises one worksheet, whose layout is based on that in the worksheet in model 2.1.

The layout of the spreadsheet is as follows:

1. **Yellow cells (cells A12:I20).** The cells in column F hold the values for the duration of the serial interval (cell F15), the number of weeks in the year (cell F16), the number of births per week (cell F17 respectively), the approximate number of cases at a peak (cell F18) and the number of susceptible individuals at a peak (cell F19).

   Notice that the cells have been assigned the names provided in column G. Therefore, if any cell has an equation which uses this cell, the name rather than the cell location can be used. For example, cell F15 has been assigned the name "durn_serial_interval", and so the term "durn_serial_interval", rather than the location of the cell, can be used in any equation (see, for example, the equation in cell F21).

2. **Turquoise cells (cells A21:I26).** The parameters in column F are calculated using the key input parameters which are in the yellow cells. These cells hold the time step size in years (cell F21), the number of births per serial interval (cell F22), and the value for k (cell F23), which is calculated using the expression $k=1/S_0$ (the derivation is discussed in the suggested exercises below).

3. **Lilac cells (cells A53:E656).** These hold equations for the number of generations and years elapsed since the start (columns A and B respectively), and the number of susceptible individuals and cases at a given time (columns C and D respectively), expressed in terms of the number in the preceding generation. These equations are the Excel equivalent of equations 4.28 and 4.29 in the book.

4. **Figure:** This shows a plot of the number of susceptible individuals and cases over time.

5. Several useful statistics have been set up in the "USEFUL STATISTICS" section.

4.4.2 Suggested exercises

1. a) Check the figure in the spreadsheet to see that once the number of new cases in each generation exceeds the birth rate per serial interval, then the number of susceptible individuals is declining.

   b) What is the inter-epidemic period for measles predicted by this model?

   c) Use the equations $C_{t+1} = kS_tC_t$ to obtain the result that $k=1/S_0$, where S_0 is the number of susceptible individuals at time t=0, which is taken to be at a point when the number of cases is at a peak.

d) Estimate the value for the basic reproduction number for measles in London during the 1900s, assuming that the population comprised about 2 million individuals at this time.

e) According to this model, what would have been the inter-epidemic period for rubella in London during the early 20th century, if the serial interval was 3 weeks in the following circumstances:
   i) R_0 was similar to that calculated for measles?
   ii) R_0 was 10?

4.5 Model 4.5

4.5.1 Overview of the model

This model, set up in Berkeley Madonna, describes the long-term dynamics of measles, taking account of births into and deaths out of the population, as described in Panel 4.3. It has been adapted to include changes in effective contact between individuals in the population (β) during the course of a year, as result of school holidays. It can be used to see patterns similar to those shown in Figure 4.24 of the book.

The key features of the model are described below, where the sections mentioned can be found in the globals window of the flowchart version, or in the equation window of the equation editor version:

1. One infectious person is introduced into a population where all other 99,999 individuals are susceptible (see the section called “INITIAL VALUES”).

2. The average pre-infectious and infectious periods are currently set to be those for measles (8 and 7 days respectively) - see the section called “TRANSMISSION AND INFECTION-RELATED PARAMETERS”.

3. R_0 is given by the value of R0_holiday (currently equal to 3.5) during school holidays and equal the value specified by R0_term during term time. The latter differs by a factor (specified by the parameter R0_term_hol_fact) from that during the holiday. See the section called “TRANSMISSION AND INFECTION-RELATED PARAMETERS” for further details and page 3 of the Figures window.

4. There are 4 school holidays during the year, of different durations - see the section called “SCHOOL-RELATED PARAMETERS” for further details. This section also includes an indicator parameter, called “holiday”, which equals 1 if the day of the year falls during a holiday, and is 0 otherwise.

5. Individuals die at a constant rate, which is equal to 1/(average life expectancy), where the average life expectancy is 70 years. For further details, see the section called “DEMOGRAPHY-RELATED PARAMETERS” and the equations for the numbers of deaths per unit time from each compartment.

6. The population size remains stable over time, with 100,000 individuals, with the *per capita* birth rate equal to the *per capita* mortality rate (so that the numbers of births equals the numbers of deaths per unit time). You can check this by plotting the value for sum_pop over time.

7. Several useful statistics have been set up in the “USEFUL STATISTICS” section.

The figures window includes 3 pages:

Page 1: This plots the number of new infectious persons per day from the start of the model run until 200 years thereafter.

Page 2: This plots the number of new infectious persons per day from 100 years to 200 years after the start of the model run. This plot is similar to that in Figure 4.24b in the book. Note that since the population comprises 100,000 persons, the number of new infectious persons per day is identical to the number of new infectious persons *per 100,000* per day (which is the statistic plotted in Figure 4.24b).

Page 3: This plots the value for R_0 over the course of a year, together with the value for the average R_0 and the value for the indicator "holiday". This plot is similar to Figure 4.24a in the book.

4.5.2 Suggested exercises

1. There are several combinations of values for R_0 during the school term and holiday period which, using Model 4.5, eventually result in regular cycles in incidence occurring roughly every 2-3 years (at least within the first 200 years after the start of the model runs).
   a) Check this by running the model for the following values for R0_holiday and R0_term_hol_fact. *You will probably be able to identify several other combinations of values not listed here – for further reading, see Keeling MJ, Rohani P. (2008) Modeling infectious diseases in humans and animals. Princeton and Oxford: Princeton University Press.*

R0_holiday	R0_term_hol_fact
3.5	4.7, 4.9, 5, 5.5
3.25	3.4, 3.5, 3.6, 3.7, 4.97834462 (!) 5.1, 5.2, 5.7
7	2, 2.3, 2.5, 2.6,
11	1.6-1.9

   To put these values into context, studies (including that by Fine and Clarkson (1982) - see Panel 4.23) have suggested that the contact parameter is 20-30% lower during holidays, as compared with that during term-time, which is equivalent to values of R0_term_hol_fact of 1.25-1.4.

   b) What do you conclude about the effect of school holidays on the cycles in measles incidence?

4.6 Model 4.6

4.6.1 Overview of the model

This model, set up in Berkeley Madonna, describes the long-term dynamics of measles taking account of births into and deaths out of the population, assuming that the population size increases over time, as described in Panel 4.3 of the book. It is similar to model 4.3, which assumed that the population size remained unchanged over time. It can be used to produce the patterns in Figure 4.26a (the correct version (!) – see the erratum page on the website).

The key features of the model are described below, where the sections mentioned can be found in the globals window of the flowchart version, or in the equation window of the equation editor version of the model:

1. One infectious person is introduced into a population where all other 99,999 individuals are susceptible (see the section called “INITIAL VALUES”).

2. The parameters are currently set to be those for measles (R_0=13, average pre-infectious period=8 days, average infectious period=7 days) - see the section called “TRANSMISSION AND INFECTION-RELATED PARAMETERS”.

3. The force of infection is calculated using the effective contact rate (calculated as R_0/(duration of infectiousness)), the number of infectious persons and the population size, using equation 2.13 in Panel 2.5. For further details, see the section called “TRANSMISSION AND INFECTION-RELATED PARAMETERS”.

4. Individuals die at a constant rate, which is equal to 1/(average life expectancy), where the average life expectancy is 60 years. For further details, see the section called “DEMOGRAPHY-RELATED PARAMETERS” and the equations for the numbers of deaths per unit time from each compartment.

5. The population size is controlled by the number of newborns per unit time (new_births_pday), who are all assumed to be susceptible. The latter is calculated using the daily per capita birth rate (“daily_pcapita_brate”), which, in turn, is calculated using the annual birth rate per 1000 population (“annual_brate_p1000”). The population size increases over time, with 100,000 individuals at the start (specified by total_popn0). You can check this by plotting the value for sum_pop over time. For further details, see the section called “DEMOGRAPHY-RELATED PARAMETERS”.

6. Several useful statistics have been set up in the “USEFUL STATISTICS” section.

There are 2 figures in this model:

Page 1: This plots the total population size (sum_pop) over time.

Page 2: This plots the number of new infectious persons per 100,000 population per day on the right y-axis and the daily force of infection, plotted on the left-hand y-axis.

Note on the expression for the number of new infections in the flowchart

You may notice that the flowchart for this model does not include an arc arrow between the Infectious category and the "new infections arrow". This is due to the facts that the number of new infections per unit time is expressed in terms of the force of infection and the number of susceptible persons, and a variable for the force of infection has been set up in the globals window. Note that any variable that is set up in the globals window is available everywhere in the model diagram.

4.6.2 Suggested exercises using model 4.6

1. a) Referring to Page 1 of the figures window, run the model assuming an annual per capita birth rate of 40 per 1000 per year, read off the population size 100 years after the start, and use this value to calculate the annual growth rate in this population.

   b) Is this value consistent with what you would expect, based on the assumed values for the life expectancy and the annual per capita birth rate?

   c) Repeat your calculations assuming that the per capita birth rate is 15 and 25 per 1000 per year.

2. a) How should increasing the birth rate in the population affect the cycles in incidence?

   b) Referring to Page 2 of the figures window, check that predictions of the incidence of infectious persons per 100,000 population are consistent with those presented in Figure 4.26 of the book.

3. a) How do you think your predictions of the cycles in the number of new infectious persons per 100,000 per day would change if you were to assume that transmission was density-dependent (see discussion in Panel 2.5)?

   b) Check your hypothesis by changing the equation for the force of infection to be given by the equivalent of the equation $\beta I(t)$, assuming that β is constant and is given by the equation $c_e/N(0)$, where c_e is the effective contact rate and $N(0)$ is the total population size at the start.
   Reminder: *if you're using the flowchart version of the model, the equation for the force of infection* ***must*** *be changed in the globals window (accessed via the globals button* 

in the flowchart window). If you're using the equation editor version of the model, you can change the equation in the equation editor window. In both instances, you will need to run the model before you will see predictions!

4.7 Model 4.7

4.7.1 Overview of the model

This model describes the long-term dynamics of measles, set up in Berkeley Madonna, taking account of births into and deaths out of the population, as described in Panel 4.3 and adapted to include vaccination of newborns into the population. It can be used to see the patterns shown in Figure 4.29b of the book.

The key features of the model are described below, where the sections mentioned can be found in the globals window of the flowchart version, or in the equation window of the equation editor version of the model:

1. One infectious person is introduced into a population where all other 99,999 individuals are susceptible (see the section called "INITIAL VALUES").

2. The parameters are currently set to be those for measles (R_0=13, average pre-infectious period=8 days, average infectious period=7 days) - see the section called "TRANSMISSION AND INFECTION-RELATED PARAMETERS".

3. Vaccination of newborns is introduced 100 years after the start of simulations at an effective coverage specified by the values of prop_immunized and eff_vacc_cov. Note that the effective coverage is interpretable as the product of the vaccination coverage and the vaccine efficacy. For further details, see the section called "VACCINATION-RELATED PARAMETERS", the expressions for new_susceptible_births and new_immunized_births (Flowchart version only) or the equations window.

4. Individuals die at a constant rate, which is equal to 1/(average life expectancy), where the average life expectancy is 70 years. For further details, see the section called "DEMOGRAPHY-RELATED PARAMETERS" and the equations for the numbers of deaths per unit time from each compartment.

5. The population size remains stable over time, with 100,000 individuals, with the *per capita* birth rate equal to the *per capita* mortality rate (so that the numbers of births equals the numbers of deaths per unit time). You can check this by plotting the value for sum_pop over time.

7. Several useful statistics have been set up in the "USEFUL STATISTICS" section.

The figures window includes 4 pages:

Page 1: This plots the number of new infectious persons per day from the start of the model run until 100 years thereafter.

Page 2: This plots the number of new infectious persons per day from ten years before to 60 years after the introduction of vaccination of newborns. This plot can be used to reproduce Figure 4.29b in the book. Note that since the population comprises 100,000 persons, the number of new infectious persons per day is identical to the number of new infectious persons *per 100,000* per day (which is the statistic plotted in Figure 4.29b).

Page 3: This plots the proportion of the population that is susceptible over time.
Page 4: This plots the daily force of infection over time.

4.7.2 Suggested exercises

1. a) Run the model for different levels of effective vaccination coverage among newborns (by changing the value for eff_vacc_cov) and look at Page 1 of the Figures window to check that you see similar patterns in the number of new infectious persons per 100,000 per day to those seen in Figure 4.29.

   b) Calculate the herd immunity threshold for this model and check that when the effective coverage is above this value in the model, transmission of measles stops.

   c) Repeat these steps for rubella and smallpox, using the values for the preinfectious and infectious periods given in question 3 in section 4.3 in this document, assuming values for R_0 of 7 and 5 for these infections respectively.

2. Change the natural history parameters to be those for measles and run the model for levels of effective vaccination coverage of below 92%, whilst looking at Page 3 of the Figures window. What happens to the proportion of the population that is susceptible in the model as you change the effective vaccination coverage? Is this what you expect and why?

3. a) Use the inter-epidemic period predicted by the model for levels of effective vaccination coverage among newborns of 50%, 80% and 90% to estimate the average age at infection . Why should you be cautious about accepting these values?

   b) As discussed in chapter 5, if it is assumed that individuals mix randomly and that the mortality rate is identical for everyone in the population (as is the case in Model 4.7), the average age at infection, the force of infection and the average mortality rate (m) are related through the equation: $A = \frac{1}{\lambda + m}$. Use this equation to calculate what the average force of infection should be in the population, following the introduction of vaccination at the levels of coverage specified in question 3a), and check that these values are consistent with those predicted by the model.

4.8 Models 4.8 and 4.8a

4.8.1 Overview of model 4.8

Model 4.8, set up in Berkeley Madonna, describes the long-term dynamics of a Susceptible-Infectious-Susceptible infection, such as gonorrhoea, as used by Grassly et al (2005) – see section 4.4 in the book. It can be used to see the plot in Figure 4.31a of the book.

The key features of the model are described below, where the sections mentioned can be found in the globals window of the flowchart version, or in the equation window of the equation editor version of the model:

1. One infectious person is introduced into a population where all other 99,999 individuals are susceptible (see the section called "INITIAL VALUES").

2. The parameters are currently set to be those for gonorrhoea (R_0=1.5, average infectious period=2 months) - see the section called "TRANSMISSION AND INFECTION-RELATED PARAMETERS".

3. Individuals become sexually active at a constant rate, which is assumed to equal the rate at which individuals stop being sexually active, calculated as 1/(average duration that individuals are sexually active). The average number of years which individuals are sexually active is set to be 30 years. For further details, see the section called "SEXUAL-BEHAVIOUR-RELATED PARAMETERS".

4. The population size remains stable over time, with 100,000 individuals. You can check this by plotting the value for sum_pop over time.

5. Several useful statistics have been set up in the "USEFUL STATISTICS" section.

The window with the figures includes 1 figure:

Page 1: This plots the number of new infections persons per month over 50 years after the start, reproducing Figure 4.31 in the book.

4.8.2 Overview of model 4.8a

Model 4.8a, set up in Berkeley Madonna, is identical to Model 4.8, except that it describes the long-term dynamics of a Susceptible-Infectious-Recovered-Susceptible infection, such as syphilis, as used by Grassly et al (2005) – see section 4.4 in the book. It can be used to see the plot in Figure 4.31b of the book.

The difference between this model and Model 4.8 is that infectious individuals recover to enter the recovered compartment, before returning to the susceptible compartment. They stay in the recovered compartment for an average period that is specified by the parameter "durn_immty_years".

4.8.3 Suggested exercises

1. Run Model 4.8a for different values for the duration of immunity and check that it produces patterns in the number of new infections which are similar to those shown in Figure 4.31b of the book.

Chapter 5

Age patterns

5.1 Model 5.1

5.1.1 Overview of the model

This model, set up in Berkeley Madonna, contains a simple catalytic model based on Munch (1959), describing how the proportion of a cohort that is susceptible or who has ever been infected should change with increasing age, assuming that the force of infection is identical in each year of life. It can be used to reproduce the plot in Figure 5.3 in the book.

The key features of the model are described below, where the sections mentioned can be found in the globals window of the flowchart version, or in the equation window of the equation editor version of the model:

1. Everyone is susceptible at birth (see the section called “INITIAL VALUES”).

2. The force of infection is currently set to be 10% per year and constant over time (see the section labelled INFECTION-RELATED PARAMETERS).

The figures window has one page, plotting the proportion of the cohort that is susceptible or has ever been infected, as the cohort ages. Data on the age-specific proportion of individuals that had antibodies to mumps or rubella in England and Wales, presented in Farrington (1990) during the 1980s, that were plotted in Figure 5.1 in the book are also included. These can be viewed by clicking on the buttons #rub_ppos and #mumps_ppos.

5.1.2 Suggested exercises

1. Run the model and check that the predicted proportion of the cohort that is susceptible or (ever) infected as the cohort ages is consistent with that shown in Figure 5.3 of the book, for different values of the force of infection.

2. As discussed on page 108 of the book, the best-fitting values for the average force of infection for the rubella and mumps data that are obtained using this model are about 12% and 20% per year respectively. Check that these values lead to model predictions of the age-specific proportion ever infected that are compatible with the data.

5.2 Model 5.2

5.2.1 Overview of the model

This model shows how a simple catalytic model, based on Münch (1959) and as described in section 5.2.2 of the book, can be set up in Excel. It can be used to see the patterns shown in Figure 5.4 of the book.

The key features of the model are:

- The force of infection is identical for all ages
- All persons are susceptible to infection from birth.
- The mortality rate is identical for susceptibles and those (ever) infected.

The file comprises two sheets. Sheet "mumps" contains the mumps data presented in Figure 5.1 and sheet "rubella" holds the rubella data, extracted from Farrington (1990). The layout of both spreadsheets is identical and is as follows:

1. **Turquoise cells (rows 12-17)** Cell D14 holds the assumed value for the annual force of infection.

   Notice that cell D14 has been assigned the name specified in cell E14. Therefore, if any cell has an equation which uses this cell, the name rather than the cell location can be used. For example, cell D14 in sheet "mumps" has been assigned the name "foi_pyr_mumps", and so the term "foi_pyr_mumps", rather than the location of the cell, can be used in any equation (see, for example, the equation in cell F36).

2. **Yellow cells (cells A29:E79).** These hold data on the observed numbers of individuals that were positive, negative and the total number tested for antibodies in each age group in the given dataset.

3. **Pink cells (cells F29:G79).** Column F in this cell range holds expressions for the proportion of individuals of a given age that are susceptible (i.e. the Excel equivalent of equation 5.2 in the book, $s(a) = e^{-\lambda a}$). Column G in this cell range holds expressions for the proportion who have ever been infected (i.e. the Excel equivalent of equation 5.3 in the book, $z(a) = 1 - s(a)$).

4. **Figure.** This compares data on the observed proportion positive in the spreadsheet against model predictions of the age-specific proportion (ever) infected.

5.2.2 Suggested exercises

1. At present, the models assume that the force of infection is 21% and 11% per year for mumps and rubella respectively. Do you think that the actual average force of infection for these infections was higher or lower than these values?

2. Click on the ⊞ buttons next to row 19 and above column J.

You should see some grey cells which describe the goodness of fit of the model (based on the value for the force of infection in cell D14). The log-likelihood deviance of the model prediction from the dataset is provided in cell D18. It is calculated using the contents of cells H36:H79 and I36:I79, which hold contributions to the deviance from each data point, based on the catalytic model prediction and the "saturated" model respectively. See section 4.2.5 of the book and the references cited there for further details.

Without changing any assumptions in the model, follow the steps in section 10.1.1 of this document to obtain the best-fitting values for the force of mumps and rubella infection, and check whether they are consistent with the values presented on page 108 of the book.

3. Follow the steps in section 10.1.3 of this document to calculate 95% confidence intervals on the estimates for the force of mumps and rubella infection, and check that they are consistent with the values presented on page 108 of the book.

4. a) How do you think your estimates of the force of infection would change if you were to assume that the sensitivity of the antibody test is <100%?

   b) Test your hypotheses by refitting your model after changing your model to assume that i) 95% and ii) 80% of those who have ever experienced infection are serologically positive.

   c) Why might you be cautious about accepting the estimates obtained assuming a test sensitivity of 80%?

5.3 Model 5.3

5.3.1 Overview of the model

This model, set up in Excel, is a variant of the simple catalytic model and assumes that the force of infection is identical for all ages, but that individuals have some immunity during the first few months of life, as described in section 5.2.3.5.1 of the book. It can be used to produce the patterns in Figure 5.9c in the book.

The file comprises two spreadsheets. The spreadsheet "fixed duration maternal immunity" assumes that everyone is immune for the first few months of life and is then susceptible to infection. The worksheet "waning maternal immunity" assumes that maternal immunity wanes at a constant rate, depending on the assumed average duration of maternal immunity.

The layout of each of the spreadsheets is similar to that for model 5.2. The key features are as follows:

1. **Turquoise cells (rows 12-17)**
   Sheet "fixed duration maternal immunity": Cell D14 holds the current value for the annual force of infection, and has been assigned the name "foi_pyr_matfix". Cell D15 holds the assumed fixed duration of maternal immunity, which is currently set to be 0.5 years.
   Sheet "waning maternal immunity": Cells D14, D15 and D16 hold the current value for the annual force of infection, the average duration of maternal immunity, and the average rate at which maternal immunity is lost. These cells have been assigned the names "foi_pyr_matwane", "durn_mat" and "mat_loss_rate" respectively.

2. **Yellow cells (cells A29:E79).** These contain data on the observed numbers of individuals that were positive, negative and the total number tested for rubella antibodies in the UK, extracted from Farrington (1990).

3. **Pink cells (cells F29:G79).** Column F in this cell range contains expressions for the proportion of individuals of a given age that is susceptible. This expression is the Excel equivalent of $s(a) = e^{-\lambda(a-0.5)}$ and $s(a) = \frac{\mu(e^{-\mu a} - e^{-\lambda a})}{\lambda - \mu}$ for assumptions that maternal immunity last for 6 months for all individuals, or declines at a constant rate respectively (see page 118 of the book). The cells in column G hold the Excel equivalent of 1-*s(a)*. In sheet "fixed duration maternal immunity", this equation leads to the value for the proportion (ever infected) at a given age. In sheet "waning maternal immunity", the value obtained using this equation is interpretable as the proportion (ever infected) or with maternal immunity.

4. **Figure.** This compares data on the observed proportion positive in the spreadsheet against model predictions.

5.3.2 Suggested exercises

At present, the force of infection in both sheets is set to be 18% per year.

1. Click on the ⊞ buttons next to row 19 and above column J. You should see grey cells which are similar to those described for exercise 2 in section 5.2.2. Follow the steps in section 10.1.1 of this document to check that the best-fitting values for the force of rubella infection are consistent with the values presented in Table 5.3 of the book.

2. Follow the steps in 10.1.3 of this document to calculate 95% confidence intervals on the data and check that they are consistent with the values in Table 5.3 of the book.

3. How should your estimates change if you were to assume that the duration of maternal immunity is shorter than 6 months? Check your hypothesis by reducing the duration to 4 months and re-fitting the model.

4. Use the formula $a_{\min} = \frac{\ln(\lambda/\mu)}{\lambda - \mu}$ (see exercise 5.6a in the book) to calculate the age at which the smallest proportion of individuals should be immune if the average duration of maternal immunity is fixed at 6 months, for values of the force of infection of 1%, 10%, 15% and 20% per year. Check that the values that you obtain are consistent with predictions from sheet “waning maternal immunity”.

5.4 Model 5.4

5.4.1 Overview of the model

This catalytic model, set up in Excel, assumes that the force of infection differs between the age groups <15 and ≥15 years and that individuals are susceptible from birth, as described in section 5.2.3.5.2 and Panel 5.2 of the book.

The file consists of just one sheet, whose layout is similar to that for models 5.2 and 5.3. The key features are as follows:

1. **Turquoise cells (rows 12-17)** Cells D14 and D15 hold the current value for the annual force of infection for those aged <15 and ≥15 years, which have been assigned the names "foi_u15", and "foi_g15" respectively. Cell D16 holds the proportion of those aged 15 years who are susceptible (assigned the name prop_sus15).

2. **Yellow cells (cells A29:E79).** These contain data on the observed numbers of individuals that were positive, negative and the total number tested for rubella antibodies in different age groups in the UK, extracted from Farrington (1990).

3. **Pink cells (cells F29:G79).** Column F in this cell range contains expression for the proportion of individuals of a given age that is susceptible. The expressions for the age groups 0-14 years are the Excel equivalent of equation 5.27a in the book; those for the age groups ≥15 years hold the Excel equivalent of equation 5.27b. The cells in column G hold the Excel equivalent of 1-*s(a).*

4. **Figure.** This compares data on the observed proportion positive in the spreadsheet against model predictions of the age-specific proportion (ever) infected or the proportion immune.

5.4.2 Suggested exercises

1. Click on the ⊞ buttons next to row 19 and above column J. You should see grey cells which are similar to those described for exercise 2 in section 5.2.2.
   a) First follow the steps in section 10.1.2 of this document to check that the best-fitting values for the force of rubella infection are consistent with the values presented on page 125 of the book (13% and 4% per year for 0-14 and ≥15 years respectively).
   b) Follow the steps in section 10.1.4 of this document to calculate 95% confidence limits on the estimates.

2. a) Rewrite equations 5.27a and 5.27b in the book to assume that individuals have maternal immunity for the first 6 months of life.

2. Change the appropriate expressions in the spreadsheet to assume that individuals are immune during the first 6 months of life and re-fit the model. How do assumptions about maternal immunity affect estimates of the force of infection among children and adults?

5.5 Model 5.5

5.5.1 Overview of the model

This model, set up in Berkeley Madonna, describes the transmission dynamics of rubella in an age-structured population (one with a "Realistic Age Structure", using the approach of Schenzle (1984) - see Panel 5.4 of the book). It can be used to generate the estimates of the proportion of individuals that are susceptible and number of new infections per day at different ages that are shown in Figures 5.16 and 5.21 respectively in the book.

The model has been set up in Berkeley Madonna's equation editor only. It can be set up using Berkeley Madonna's flowchart editor, although the resulting model diagram would be very cumbersome.

The key features of the model are as follows:

- The population is stratified into annual age strata between the ages 0 and 70 years, with each age stratum moving to the subsequent age stratum at the end of each year, as shown in Figure O.5.1.
- Vaccination of newborns is introduced in the year specified by year_start_vacc, with an effective coverage specified by eff_vcov_newborns.
- The infection-related parameters are currently set to be those for rubella in a high incidence setting (average pre-infectious period=10 days, average infectious period= 11 days, R_0=12).

Further technical details of the model are provided below. Those readers who are mainly interested in the application of the model may wish to skip this section and go to the suggested exercises in section 5.5.3.

5.5.2 Further details of the model

The sections mentioned in the description below can be found in the equation window. To improve clarity, the section headings in this window have been pasted into blue-green boxes which cannot be edited.

Equations:

1. The equations are written using difference equations (equivalent to the Euler method) – see the two sections labelled "DIFFERENCE EQUATIONS".

2. Since only newborns are vaccinated in the model (see below), the difference equations differ depending on whether they relate to the population aged ≥1 year or those in their first year of life.

3. The population is stratified into annual cohorts, with the maximum age specified by the parameter "up_age" (currently set to 69 years – see the section labelled "DEMOGRAPHY-RELATED PARAMETERS). This stratification is reflected in the notation used by Berkeley Madonna to write the equations for each age class. For example, Sus[0] refers to those in their first year of life, Sus[1] refers to those aged 1 years, Sus[2] refers to those aged 2 years etc. There are no equations for age strata

greater than 69 years (i.e. >up_age) since individuals leave the model on their 70th (up_age+1) birthday.

Notice that the left hand side of the difference equations for the susceptible population is written Sus[up_age..1]. This notation simply reflects the fact that the equations for Sus[up_age] are evaluated first, followed by those for Sus[up_age-1] etc.

4. The population in each age stratum moves into the subsequent age stratum at the end of each year, and therefore the difference equations used depend on whether or not it is the end of the year. The latter is specified by the value of the variable year_end, which equals 1 if the end of the year has been reached and 0 otherwise (see the section on "TIME-KEEPING VARIABLES").

5. Newborns enter the population only at the end of year, at which time vaccination also occurs (see below).

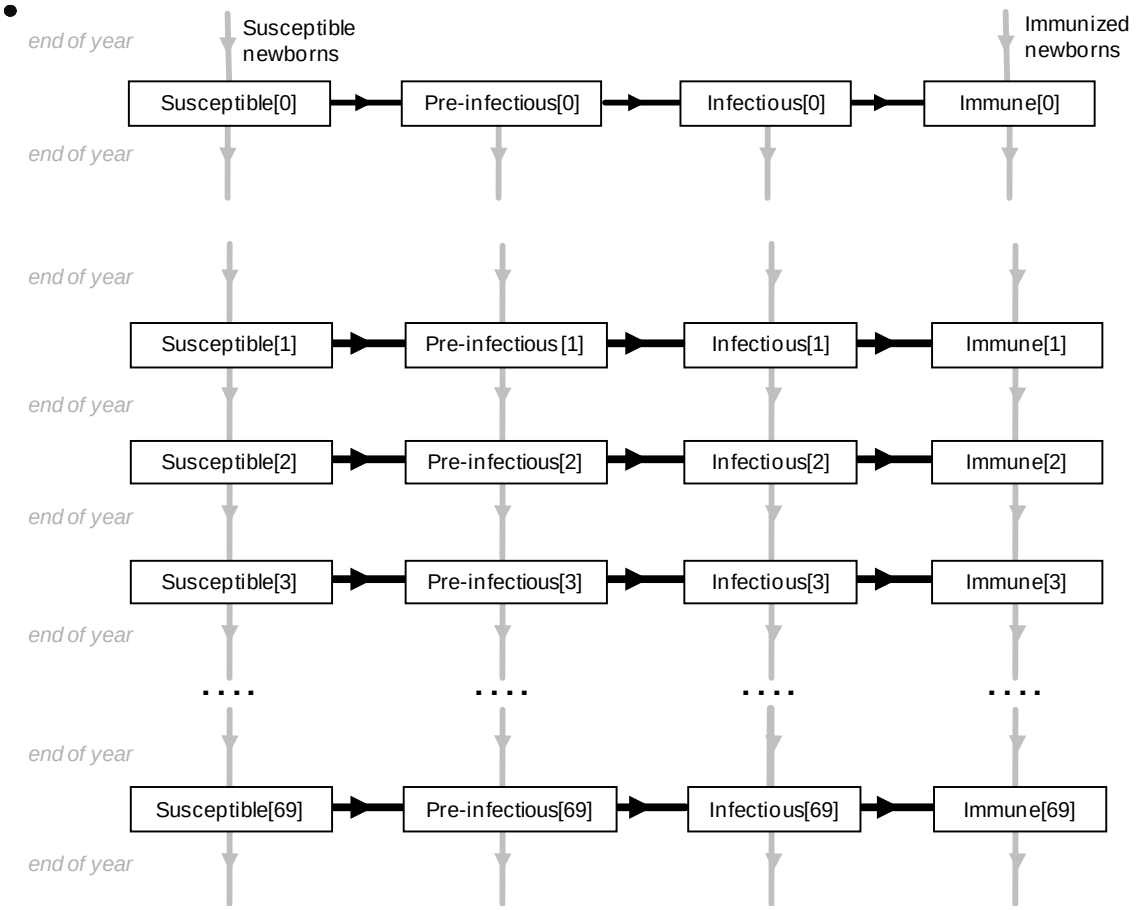


Figure O.5.1: General structure of Model 5.5

Demography:

1. The population has a rectangular age distribution with 1000 persons in each age stratum and hence 70,000 individuals in the whole population (see Pages 1 and 2 of the Figures window and question 1 in the suggested exercises). The variables

tot_pop[0], tot_pop[1], tot_pop[2] equal the total number of individuals who are in their first, second, third etc year of life in the model (i.e. in the age stratum 0, 1, 2 years etc). Tot_pop[0], tot_pop[1] etc are defined in the "USEFUL SUMMARY STATISTICS" section.

Transmission and infection:

1. The infection-related parameters are currently set to be those for rubella in a high incidence setting (pre-infectious period=10 days, infectious period= 11 days, R_0=12) - see the section called "TRANSMISSION AND INFECTION-RELATED PARAMETERS".

2. The force of infection in the population is defined in the section called "TRANSMISSION AND INFECTION-RELATED PARAMETERS" and is given by the expression:

   *beta*ARRAYSUM(Infous[*])*

   Here *ARRAYSUM(Infous[*])* is the total number of infectious individuals in the population. This is the notation that Berkeley Madonna uses to calculate the sum of all the age strata of the Infectious compartment, i.e. Infous[0]+Infous[1]+Infous[2]+ ..+ Infous[upage].

Vaccination:

1. Vaccination of newborns is introduced some years after the start (specified by the value of year_start_vacc). The proportion of newborns that is immunized each year is stored in the parameters prop_newborns_immunized and eff_vcov_newborns. Note that the effective coverage is interpretable as the product of the vaccination coverage and the vaccine efficacy. For further details, see the section called "VACCINATION-RELATED PARAMETERS".

   Unvaccinated newborns are added to the Sus[0] compartment at the end of each year; *vaccinated* newborns are added to the Imm[0] compartment at the end of each year (see the section called "DIFFERENCE EQUATIONS – PERSONS IN THEIR FIRST YEAR OF LIFE".

Miscellaneous:

1. Several useful statistics have been set up in the "USEFUL STATISTICS" section, such as the proportion of 5, 10, 15 etc year olds who are susceptible, the daily numbers of new infections per 100,000 in various age groups.

Figures:

The figures window has 5 pages:

Page 1: This plots the number of individuals in each age stratum over time.

Page 2: This plots the number of individuals in each age stratum at the end (with the age stratum on the x-axis).

Page 3: This plots the daily force of infection over time.

Page 4: This plots the proportion of the population that is susceptible in different age groups and overall over time and can be used to reproduce Figure 5.16 in

the book (after including an appropriate level of vaccination coverage among newborns).

Page 5: This plots the daily number of new infections per 100,000 per day and can be used to reproduce Figure 5.21 in the book (after including an appropriate level of vaccination coverage among newborns).

5.5.3 Suggested exercises

Questions 2, 3a, and 3d are intended to improve understanding of the relationship between some of the equations in the book and the outputs in the model. The equations provide a useful check to see that the model is behaving as it should, which is particularly helpful if you are developing your own models!

Readers who are most interested in looking at the effect of vaccination among newborns on the proportion susceptible and the number of new infections in different age groups may want to focus on questions 1, 3b, 3c and 4.

1. Run the model and click on pages 1 and 2 of the figures window to check that the population size remains unchanged over time and that the age distribution is rectangular.

2. The model is currently set up so that R_0=12 and no newborns are vaccinated.

   a) Use the equation $R_0=\lambda L$ (equation 5.22 in the book) to calculate what you would expect the average force of infection to be in the population (before the introduction of vaccination), and check that it is consistent with that predicted by the model on Page 3 of the figures window.

   b) Repeat part b) assuming that R_0=7.

   c) Use your estimates of the average force of infection (calculated in part a)) to calculate the following, and check that the values you obtain are consistent with those shown on pages 4 and 5 of the figures window:
      i. The average proportion of 20 year olds who should be susceptible (using equation 5.2 in the book).
      ii. The average overall proportion of the population that should be susceptible (using either equation 5.14 or 5.19 in the book).
      iii. The average number of new infections per 100,000 population among those aged 20 years (adapting equation 5.26).

3. Run the model assuming that 50% of newborns are immunized once vaccination has been introduced, assuming that R_0 =12.

   a) Read off the long-term average daily force of infection after the introduction of vaccination from page 3 of the figures window, and check that the values that you obtain lead to value of R_0 which are consistent with the value predicted by equation 5.28 in the book ($R_0 = \frac{\lambda' L}{(1-v)(1-e^{-\lambda' L})}$).

b) Referring to pages 4 and 5 of the Figures window, check that predictions of the proportion susceptible and the daily number of new infections per 100,000 at different ages are consistent with those in Figures 5.16 and 5.21 in the book.

c) Calculate the herd immunity threshold and run the model for different values for the effective vaccination coverage which are below or above the herd immunity threshold. What happens to the long-term proportion susceptible at different ages as the effective vaccination coverage increases and why?

d) Use equations 5.31 and 5.36 in the book to calculate the average proportion susceptible and the number of new infections per 100,000 per day among 20 year olds in the long-term after the introduction of vaccination among newborns, and check that these values are consistent with those shown in pages 4 and 5 of the Figures window.

4. Patterns similar to those in Figures 5.14 and 5.22 in the book can be generated using a parameter plot in Berkeley Madonna.

a) Check this by carrying out a parameter plot (see page 15 of this document) of the (final) long-term average force of infection and the (final) long-term average number of new infections per 100,000 per day among those aged 20, 30 and 40 years, against the effective vaccination coverage among newborns.
Note: you may prefer to set the number of runs to be 11 in the Parameter plot window, with the effective coverage ranging between 0 and 1.

b) In which setting should you be most cautious about introducing rubella vaccination among newborns?

5.6 Model 5.6

5.6.1 Overview of the model

This model, set up in Berkeley Madonna, is very similar to Model 5.5, except that it allows vaccination of any age group. It is currently set up so that 13 year olds are vaccinated and can be used to reproduce Figure 5.23b in the book.

The main differences between the equation window of this model and that for model 5.5 are described below; the sections mentioned can be found in the equation window:

1. Vaccination begins some years after the start (specified by the value of year_start_vacc). The proportion of individuals of age *i* that is immunized each year is stored in the parameter prop_immunized[i]. At present, only 13 year olds are vaccinated, at a level determined by eff_vcov_13 (see the section called "VACCINATION-RELATED PARAMETERS".

2. Vaccination occurs at the end of each year, when immunized individuals in the i-1th age stratum move into the i^{th} age stratum (see the two sections on difference equations in the equations window).

The figures window includes the same 5 pages that were in Model 5.5:

Page 1: This plots the number of individuals in each age stratum over time.
Page 2: This plots the number of individuals in each age stratum at the end (with the age stratum on the x-axis).
Page 3: This plots the daily force of infection over time.
Page 4: This plots the proportion of the population that is susceptible in different age groups and overall over time.
Page 5: This plots the daily number of new infections per 100,000 per day and can be used to reproduce Figure 5.21 in the book (after including an appropriate level of vaccination coverage among 13 year olds).

5.6.2 Suggested exercises

1. a) Run the model assuming that 100% of 13 year olds are immunized and that R_0= 7 and read off the value for the long-term average force of infection. Check that the value for R_0 that you obtain after substituting this value into equation 5.37 in the book is consistent with that assumed in the model.

   b) Check that predictions of the number of new infections per 100,000 population shown on page 5 of the Figures window are consistent with those in Figure 5.23b.

   c) According to pages 4 and 5 of the Figures window, how does immunization of 100% of 13 year olds influence the following and why:
      i. The proportion of children and adults that is susceptible?
      ii. The number of new infections per 100,000 population in different age groups?

   d) Use the force of infection read off in part a) to calculate the reduction in the long-term average force of infection resulting from the introduction of vaccination.

e) Repeat the calculation in part d) using the force of infection obtained after running the model assuming that R_0=12. How does the reduction in the force of infection predicted for this setting after the introduction of vaccination among 13 year olds compare against that predicted for a setting in which R_0=7? Suggest possible reasons for this difference, referring to pages 4 and 5 of the Figures window if necessary.

2. Carry out a parameter plot (see page 15 of this document) of the (final) long-term average force of infection and the number of new infections among 20, 30 and 40 year olds assuming that R_0=7 or R_0=12, and compare your findings against those obtained for question 4 of the exercises for model 5.5. Why might you be happier about introducing vaccination of 13 year olds against rubella than about introducing vaccination of children soon after birth?

Chapter 6

An introduction to stochastic modelling

6.1 Model 6.1

Readers may notice that model 6.1 is not mentioned in the text. We originally intended to refer to it in the caption to Figure 6.1. However, since the model was identical to Model 2.1, we thought it would be better just to refer to Model 2.1 in the caption, rather than duplicate Model 2.1 and rename it to Model 6.1.

6.2 Model 6.2

6.2.1 Overview of the model

This model, set up in Excel, illustrates how a stochastic model describing the course of an outbreak one infectious person enters a population comprising 10 individuals, can be set up using method 1 (see section 6.3 in the book). It can be used to see the patterns shown in Figures 6.3, 6.6 and 6.9 in the book.

The file comprises one sheet. The key features of the model are as follows:

1. **Yellow cells (cells A12:G19).** These hold the key input parameters in the model, specifically, R_0 (cell E15), the probability of two specific individuals coming into effective contact per time step, p (cell E17), and the size of population into which the infection is introduced (cell E18). p is calculated from R_0 using equation 6.2 in the book. p is constrained so that it does not go above 1.

   Notice that the yellow cells in column E have been assigned the name specified in column F. Therefore, if any cell has an equation which uses these cells, the name rather than the cell location can be used. For example, cell E15 been assigned the name “R0”, and so the term “R0”, rather than the location of the cell, can be used in any equation (see, for example, the equation in cell E17).

2. **Coloured cells in rows 20-42.** These cells implement the steps for method 1 for each time step. At present only the calculations for time steps 0 and 1 are visible. The cells are as follows:

   a) The individuals in the population are listed in rows 25-35.

   b) Column F in rows 25-35 indicates the status of each individual at the start of each time step. At the start, only one person is a case and the others are susceptible.

   c) Rows 37-39 sum the number of susceptible, cases and immune persons at the start of each time step.

   d) Row 41 uses the values for the number of cases at the start of the time step and p to calculate the risk of infection during each time step using equation 6.2 in the book (the Reed-Frost equation).

   e) A random number for each susceptible person at the start is drawn in cells P25-P35. Cells G25-G35 use these random numbers to determine whether the person remain susceptible or becomes a case, using the risk of infection calculated for the first time step in cell F41.

      For example, the equation in cell G25, which refers to the initial case, is as follows:

      =IF(P25="--","imm",IF(P25<F$41,"case","sus"))

This equation first checks if a random number has been drawn for that person. If no random number has been drawn for that person then that person becomes immune by the end of the time step. This should have occurred if that person had either been a case or was immune at the start of the time step),

Otherwise, if the random number in cell P25 is less than the risk of infection, then the person becomes a case; otherwise the person remains susceptible.

Pressing the F9 key generates a new set of random numbers in column P.

6.2.2 Suggested exercises

1. Press the F9 key and check that the status of each individual by the start of the first time step changes in the way that you expect.

2. Click on the ⊞ buttons above columns P and column Y. You should now see the calculations for the remaining time steps and the final epidemic size, and a graph showing the numbers of cases, susceptible and immune persons over time. Press the F9 key several times, for different values for R_0.

   How does the size of R_0 affect the shape of the epidemic curve and the size of the outbreak?

3. The spreadsheet has been set up so that it is possible to collect and summarize the numbers of cases in each time step from several runs of the model by following the steps in 6.2.3. Follow these steps; in doing so, you should be able to reproduce figures that are similar to Figures 6.6 and 6.9 in the book.

6.2.3 Collecting the results from multiple runs of model 6.2

The following steps guide you through the method used in Model 6.2 to collect and summarize the results from multiple runs of the model.

Step 1: Click on the ⊞ button next to row 2106.

You should now see the following:

i. Buttons providing options for running different numbers of simulations,
ii. An empty table in rows 101 onwards showing the numbers of cases seen in each time step and the outbreak size for each model run.
iii. A figure (currently empty) plotting the number of cases predicted in each time step for 50 runs of the model.

Step 2: Click on the appropriate button to run the model once. If you get a warning message about the macro not being enabled, enable the macro by following step 2 in section 10.1.1. If you can't see the Security warning mentioned in that step, you may need to close and reopen the spreadsheet again in order to see it.

This button is linked to a macro, which copies the contents of cells F38:O38 to the bottom available row of the grey table. You can view the code for this macro by clicking on the View option on the main menu and selecting the Macro button, followed by the "View macro button. Then select the macro "Run_1_sim" from the list and click on the Edit button.

Step 3: Click on the appropriate button to run the model 50 times (setting R_0=2). This button links to a macro (called "Run_50_simulations") which runs the Run_1_sim macro 50 times.

You should now find that the figure shows the number of cases predicted in each time step for 50 model runs.

3a) According to the plot, when do you think the peak in the outbreak is most likely to occur?

Insight into when the peak (or other outcome of interest) is sometimes provided from the mean and 95% ranges of the outcomes predicted. As discussed in section 6.3.4 of the book, the output from different runs from a stochastic model is usually pooled to calculate these statistics.

Step 4: Click on the ⊞ button next to row 63 to see some cells calculating the average and 95% range of the number of cases predicted for each time step during the first 50 runs of the model, together with a plot showing these statistics.

3b) What do you conclude about the time when the outbreak is most likely to peak?

Step 5: Now run the model 2000 times by clicking on the appropriate button near row 100. You may find that this macro will run for 5-10 minutes, depending on the speed of your computer. If the macro hasn't finished 15 minutes after you have started it, press the Escape button to stop it.

3c) How does this affect your answer to the last question?

The spreadsheet has also been set up to calculate the frequency distributions of outbreak sizes for different numbers of model runs.

Step 6: Click on the ⊞ button next to row 95 to see these calculations, together with a plot which should resemble Figure 6.6 in the book.

3d) What do you conclude about outbreak size that is most likely to be seen if R_0 equals 2?

Step 7: Run the model 500 times using values for R_0 of 0.5 and 1.0 and check to see that the distribution of outbreak sizes are consistent with those in Figure 6.9 in the book.

6.3 Models 6.3, 6.3a and 6.3b

6.3.1 Model 6.3

6.3.1.1 Overview of the model

This model, set up in Excel, illustrates how a stochastic model describing the course of an outbreak after one infectious person is introduced into a population comprising 10 individuals, can be set up using method 2 (see section 6.3 in the book).

The file comprises one sheet. The key features of the model are as follows:

1. **Yellow and turquoise cells (cells A12:G22).** These hold the key input parameters in the model, which are identical to those in Model 6.2. Specifically they hold the size of population into which the infection is introduced, total_popn (cell F16), R_0 (cell F18), the probability of two specific individuals coming into effective contact per time step, p (cell F20). p is calculated from R_0 using equation 6.2 in the book. p is constrained so that it does not go above 1.

   Notice that the yellow and turquoise cells in column F have been assigned the name specified in column G. Therefore, if any cell has an equation which uses these cells, the name rather than the cell location can be used. For example, cell F18 been assigned the name "R0", and so the term "R0", rather than the location of the cell, can be used in any equation (see, for example, the equation in cell F20).

2. **Lilac and purple cells (row 50 onwards)**, which hold equations for the number of susceptible, infectious and immune individuals over time. At present, only the cells for the start are visible. The contents of these cells are as follows:

   a. Columns B, C and D hold the number of susceptible, infectious and immune persons.

   b. Column E holds the risk of infection, calculated using the Reed-Frost equation (equation 6.2 in the book) and constrained to be below 1.

   c. Column F holds a random number, which is used to determine the number of new infections in the current time step.

   d. Column G calculates the number of new infections in the current time step, using Excel's "CRITBINOM" distribution. This works as the inverse of the binomial distribution, i.e. it identifies the number of infections which will be seen for given values for the cumulative probability, the number of susceptibles at time t (S_t) and λ_t . The general notation is as follows:

   $$=\text{CRITBINOM}(S_t, \lambda_t \text{ , cumulative\_prob})$$

   This equation is discussed in further detail in the suggested exercises below.

3. Clicking on the ⊞ buttons next to rows 49 and 105 will reveal a figure showing the number of susceptible, infectious and immune individuals over time, some summary statistics, together with calculations of the number of individuals in various categories beyond the first time step.

6.3.1.2 Suggested exercises

1. Click on the ⊞ button above column AA.

   You should now see some purple cells calculating the distribution of the number of new infections that could occur during the first time step, calculated using the Excel equivalent of equation 6.7 in the book, using the risk of infection and the number of susceptibles at the start. You should also see two figures, one of which plots this distribution, and the other which plots the cumulative probability distribution.

   a) Use the random number generated for the start and the plot of the cumulative distribution to work out the number of new infections that should occur during time step 0 and check that the value you obtain is consistent with the value in cell G54.

   b) Press the F9 key several times to generate a new set of random numbers and check that the number of new infections predicted for time step 0 is consistent with that calculated using the CRITBINOM function.

2. Change the population to comprise 1000, 10,000, 100,000 or 1 million individuals, whilst keeping R_0 equal to 2 and the number of infectious persons at the start equal to one, and look at the epidemic curve predicted by the model.
   Warning note: the CRITBINOM function may fail to give sensible results for very large population sizes (e.g. >1 million).

   a) How does increasing the population size affect the shape of the epidemic curve and the number of generations of cases in the outbreak and why?

   b) How do you think the distribution of outbreak sizes (or the proportion of the population that has been infected by the end) predicted for a population of 1 million will differ from that shown in Figure 6.9 in the book?

      You can check your hypotheses by first clicking on the ⊞ button next to row 616, which will reveal buttons which, if pressed, will copy the contents of cells B26:D26 down to the table in rows 116 from either 1 or 500 runs of the model. If you get a warning message about the macro not being enabled, enable the macro by following step 2 in section 10.1.1. If you can't see the Security warning mentioned in that step, you may need to close and reopen the spreadsheet again in order to see it.

      The results from these runs are used to calculate distributions plotted in the figures in rows 131-152. The buttons are linked to macros called Run_1_sim_model63 and Run_500_sims_model63, which can be viewed by clicking on the View option on the main menu and selecting the Macro button, followed by the "View macro button. Then select the macro from the list and click on the Edit button.

3. As discussed in section 6.3 in the book, the Reed-Frost equation leads to similar values for the risk of infection as the equation $\lambda_t = \beta I_t$ (where $p \approx \beta$), if the population size is large. Explore the relationship between the difference in the risk of infection calculated using these two equations by following the steps below.

   Step 1: Click on the ⊞ button above column AF.

   This will reveal some cells which calculate the risk of infection using the equation $\lambda_t = pI_t$ (cells AD54:AD104), the percentage difference between the risk of infection calculated using this equation and the Reed-Frost equation (cells AE54:AE104) and the range of this percentage difference for the given simulation (cells AD46:AE46).

   Step 2: Press the F9 key several times using different values for the total population size of between 10 and 1 million, to identify the range in which the difference in the risk of infection obtained using the two equations lies.
   Note: If you wish to do this systematically, click on the ⊞ button next to row 679, where you will see buttons allowing you to collect the output on the minimum and maximum percentage difference in the risk of infection calculated using the two equations. These are linked to macros called "Run_1sim_minmax" and "Run_50sim_minmax". See question 2b for details of how these macros can be viewed.

6.3.2 Model 6.3a

6.3.2.1 Overview of the model

This model, set up in Excel, illustrates how the approach used to set up Model 6.3 ("method 2" in the book) can be extended to deal with time steps which are of less than one serial interval and to include additional transitions, such as infectious persons recovering to become immune, as discussed in section 6.5 in the book.

The layout of the spreadsheet is similar to that of Model 6.3 and is as follows:

1. **Yellow and turquoise cells (cells A12:G22).** These are identical to those in Model 6.3, except that additional cells have been set up for the time step size, (cell F15), the average infectious period (cell F17) and the average rate at which infectious persons recover to become immune per time step (cell F21). The parameters are currently set to be similar to those for influenza, with R_0=2 and an average infectious period of 2 days.

2. **Lilac and purple cells (row 50 onwards),** which hold equations for the number of susceptible, infectious and immune individuals over time. At present, only the cells for days 0 and 0.5 are visible. The contents of these cells are as follows:

    a. Columns B, C and D hold the number of susceptible, infectious and immune persons at a given time, set up using difference equations, and using the numbers of new infections and the individuals who become immune in the current time step in columns H and I.

    b. Column E holds the risk of infection, calculated using the Reed-Frost equation (equation 6.2 in the book). This is constrained to be below 1.

    c. Columns F and G hold random numbers, which are used to determine the number of new infections and the number of infectious persons who become immune in the current time step.

    d. Columns H and I hold the number of new infections and the number of infectious persons who recover during the current time step using the CRITBINOM function.

4. Clicking on the ⊞ buttons next to rows 45 and 91 will reveal a figure showing the number of susceptible, infectious and immune individuals over time, some summary statistics, together with calculations of the number of individuals in various categories beyond the first time step.

6.3.2.2 Suggested exercises

1. Check that the effect of the population size on the outbreak size and duration of the outbreak predicted in model 6.3a is consistent with that seen using model 6.3.
   Note: if you wish to do this systematically, macros which are similar to those mentioned in question 1b) in the exercises for model 6.3 have been set for model 6.3a. They are

accessible via buttons which can be seen if you click on the  *button next to row 774. The buttons are linked to macros called Run_1sim_model63a and Run_500_simulations_model63a.*

6.3.3 Model 6.3b

This model (included for reference) illustrates how a stochastic model, set up using method 2 (with an extension to allow for several transitions and for time steps of less than one serial interval) can be set up in Berkeley Madonna.

The model is identical to Model 4.1, except for the following:

1. The number of new infections and the number of infectious persons who recover per time step are calculated by sampling random numbers from the Binomial distribution (equivalent to using the random number and the CRITBINOM function in Excel). For further details, see the section labelled “STOCHASTIC TRANSITIONS”.

2. The equations are calculated using difference equations using the Euler method (see the Parameters window and the Equations window, and click on the reservoir icons for each compartment)

There is one page in the Figures window, which plots the number of new infections per time step, and the number of susceptibles and immune persons over time.

6.4 Model 6.4

6.4.1 Overview of the model

This model, set up in Excel, illustrates how a stochastic model describing the course of an outbreak following the introduction of an infectious person in a population comprising 10 individuals, can be set up using method 3 (see section 6.6 in the book). It can be used to see the patterns shown in Figures 6.6 in the book.

The key features of the model are as follows:

1. **Yellow and turquoise cells (cells A12:G22).** These hold the key input parameters, which are identical to those in Model 6.3, except that additional cells have been set up for the average infectious period (cell F17) and the average rate at which infectious persons recover to become immune per time step (cell F21). The parameters are currently set to be similar to those for influenza, with R_0=2 and an average infectious period of 2 days.

2. **Coloured cells (row 50 onwards),** which hold equations for the number of susceptible, infectious and immune individuals over time. These are based on the summary of steps for method 3 shown on page 168 of the book. At present, only the cells for day 0 and the first time step are visible. The contents of these cells are as follows:

    a. Column A holds the time at which an event (new infection or an infectious person recovering) occurs.

    b. Columns B, C and D hold the number of susceptible, infectious and immune persons at a given time, set up using difference equations, and updated according to whether the event at the time step considered is a susceptible person being infected, or an infectious person recovering to become immune. The type of event that occurs at this time step is indicated in column M (see below).

    c. The cells in columns E, F and G calculate the hazard rate for a susceptible person to become infected and for an infectious person to recover to become immune, and for the total hazard rate for an event occurring (M_t). The equation for the latter is based on equation 6.9 in the book.

    d. The cells in column H draw a random number to determine the time after which the next transition event occurs. The cells in column I calculate the time after which the event occurs, using equation 6.10 in the book.

    e. Columns J and K calculate the probability that the next event is a susceptible person becoming newly infected, or an infected person becoming immune. The random number drawn in column L is then referred to these probabilities to determine the event which occurs next. The event which will occur next is indicated in column M.

3. Clicking on the ⊞ buttons next to rows 49 and 256 will reveal a figure showing the number of susceptible, infectious and immune individuals over time, some summary statistics, together with calculations of the number of individuals in various categories beyond the first time step.

6.4.2 Suggested exercises

1. Check that the effect of the population size on the outbreak size and duration of the outbreak predicted in model 6.4 is consistent with that seen using model 6.3.
   Note: if you wish to do this systematically, macros which are similar to those mentioned in question 1b) in the exercises for model 6.3 have been set for model 6.4. They are accessible via buttons which can be seen if you click on the ⊞ button next to row 774. The buttons are linked to macros called Run_1sim_model64 and Run_500_simulations_model64. See question 2b of the suggested exercises for Model 6.3 for details of how these macros can be viewed.

2. What are the advantages of models developed using method 3, as compared with those set up using method 2? What are the disadvantages of method 3?

Chapter 7

How do models deal with contact patterns

7.1 Model 7.1

7.1.1 Overview of the model

This model, set up in Berkeley Madonna, describes the transmission of influenza in a population in which persons are stratified into the young or old. The model is referred to in section 7.4.2.2 of the book. It can be used to obtain the WAIFW matrix used in exercise 7.5 of the book by fitting predictions to data from an epidemic curve. The model can also be used to explore the impact of vaccinating children and/or adults before the start of a epidemic.

The data used in this example come from a GP practice in Wales from the 1957 (Asian) influenza pandemic, which are plotted in Figure 7.3 of the book, after aggregating the data into two age groups.

The key features of the model are described below, where the sections mentioned can be found in the globals window of the flowchart version, or in the equation window of the equation editor version of the model:

1. The population is stratified into the “young” and the “old”. These are intended to reflect those aged <15 and >15 years in the dataset plotted in Figure 7.3 of the book and they are also referred to as “children” and “adults” in the description below.

2. Infectious young and old persons are introduced into the population at the start (see the section called “INITIAL VALUES”). The number introduced at the start is specified by the parameters Infous_y0 and Infous_o0. These values are unknown but can be estimated by fitting model predictions to the available data (see the suggested exercises below).

3. No individuals are assumed to be immune because of natural infection at the start (see the section called “INITIAL VALUES”)

4. The parameters are currently set to be those for influenza (average pre-infectious period=average infectious period=2 days) - see the section called “TRANSMISSION AND INFECTION-RELATED PARAMETERS”.

5. A proportion of all infectious persons (specified by frac_rep - see the section called TRANSMISSION AND INFECTION-RELATED PARAMETERS) are assumed to have clinical symptoms and are reported to the GP practice. This proportion is unknown and can be estimated by fitting model predictions to the observed data (see suggested exercises below).

6. The variables “Cum_reported_y” and “Cum_reported_o” store the cumulative number of cases reported over time. Notice the way that this is set up in the flowchart version of the model. For example, the equation in the arrow which goes into the “Cum_reported_y” compartment is set up so that it equals whatever is in the “new infectious” arrow multiplied by the value of frac_rep, thereby keeping track of the total number of infectious persons who are reported over time.

7. The force of infection differs between the young and old, with the WAIFW matrix determined by the values of b1 and b2 - see the section called “TRANSMISSION AND INFECTION-RELATED PARAMETERS”. The values of b1 and b2 are unknown and are estimated by fitting model predictions to the available data (see the suggested exercises below).

8. The population size remains stable over time, with 2639 and 5361 young and old persons respectively and no births or deaths (see the section called “DEMOGRAPHY-RELATED PARAMETERS”).

9. The model allows for a proportion of young and old persons to be vaccinated at the start, with a coverage specified by the values for vacc_cov_y and vacc_cov_o respectively, and a vaccine efficacy of vacc_eff (see the section on VACCINATION-RELATED PARAMETERS. The coverage is currently set to be 0%.

10. The section on “AGGREGATING OUTPUT AND FITTING MODEL PREDICTIONS” does the following:
    a. It reads in data on the weekly numbers of young and old cases reported over time in the GP practice. These data are stored in the file called “flu_data.txt”, which can be viewed in Notepad or another editor.

    b. It aggregates model predictions of the number of cases reported each day into predictions of the numbers of cases reported each week. The variables “wkly_reported_y” and “wkly_reported_o” store model predictions of the number of young and old cases respectively that occurred between the current time in the model run and the start of the week. The values for these variables therefore change with every time step and are initialized at the start of each new week, which accounts for the jagged lines seen when these variables are plotted (see Page 4 of the Figures window) with the default setting for DTOUT of 0. It is possible to remove the jaggedness by changing the value for DTOUT to 7, which results in model output being presented every 7 days.

c. It calculates two statistics describing the goodness of fit of model predictions to the weekly numbers of young and old cases reported over time in the GP practice. The goodness of fit statistics that are calculated are:
    i. SSQ - the sum of squares of the difference between model predictions of the number of cases in the young and old and those observed in the dataset.
    ii. Minus_llhood - the negative (Poisson) loglikelihood. For convenience, the constant term in this expression has been dropped (see the code for further details).

11. Several useful statistics have been set up in the "USEFUL STATISTICS" section, such as the total number of young and old persons in the population.

The window with the figures includes 4 figures:

Page 1: This plots the proportion of young persons who are susceptible or immune, and the number of new infectious young persons per 100,000 per day over time.

Page 2: This is similar to page 1, except that it refers to old persons.

Page 3: This compares model predictions of the cumulative numbers of young and old cases reported against those observed. These are stored in the variables #cum_rep_y and #cum_rep_o, which have been imported from the text files cum_rep_y.txt and cum_rep_o.txt. These files can also be viewed in Notepad or another text editor.

Page 4: This compares model predictions of the weekly number of reported young and old cases against the corresponding numbers observed.

Note that plotting the values in flu_data.txt is not straightforward in Berkeley Madonna, since this file has more than two columns and Berkeley Madonna is not designed to do plot data from such files easily. The files weekly_cases_y.txt and weekly_cases_o.txt, which both consist of 2 columns, have therefore been imported into the model so that model predictions and the actual numbers of reported young and old cases each week can be plotted in the same figure. To see these data, click on the buttons labelled #weekly_cases_y and #weekly_cases_o in the Figure. Set the value for DTOUT to be 7 (days) if to remove the jagged effect of the plot of model predictions of the weekly numbers of reported cases.

7.1.2 Suggested exercises

The first question illustrates how model predictions can be fitted to observed data to estimate unknown parameters. If you wish just to explore how assumptions about contact between individuals affect the impact of control, you may like to skip to question 2.

1. Several parameters in the model, namely the contact parameters for the WAIFW matrix (b1 and b2), the proportion of infectious persons that have clinical symptoms and are reported to the GP practice (fract_rep) and the initial numbers of infectious young and old persons (infous_y0 and infous_o0) are unknown in the model. These parameters

can be estimated using Berkeley Madonna's "curve fit" or "optimize" functions, which are described in section 10.2.

a) Follow the steps in section 10.2.1 of this document to use the Curve fit option to estimate all of these parameters at the same time. You may like to use the values for guess 1 and guess 2 (and try any others that you prefer) on the next page.

b) Follow the steps in section 10.2.2 of this document to use Berkeley Madonna's optimize function to estimate the above parameters by minimizing:
   i) ssq (the sum of squares of the difference between model predictions and the observed data) and
   ii) the minus_llhood.

   As for part a), you may wish to use the following values for Guess 1 and 2:

Parameter	**Guess number**	**Starting set of parameter values**		
		A	**B**	**C**
b1 (per day)	1	3.60×10^{-4}	3.32×10^{-4}	2.00×10^{-4}
	2	4.00×10^{-4}	3.40×10^{-4}	6.00×10^{-4}
b2 (per day)	1	5.40×10^{-4}	4.00×10^{-4}	3.00×10^{-4}
	2	6.70×10^{-5}	7.00×10^{-5}	5.00×10^{-5}
frac_rep	1	0.60	0.45	0.30
	2	0.45	0.49	0.80
Infous_y0	1	1	3	7
	2	3	10	20
Infous_y0	1	2	2	10
	2	6	9	15

c) What can you conclude about the parameters that you have estimated?

2. Incorporate the following values for b1 and b2 into the model: b1= 3.38×10^{-4} per day and b2=7.14×10^{-5} per day. Assuming that there are sufficient doses of vaccine for 2500 individuals in the population before the outbreak, use estimates of the total numbers of cases reported to determine whether it would be best to provide the vaccines to:
   i) children only;
   ii) adults only;
   iii) the same proportion of children and adults;
   iv) equal numbers of children and adults.

   Compare your answer against that obtained for question 7.5 in the book or using Model 7.6.

   Note: you may find that incorporating certain levels of vaccination coverage and extending the stoptime in the model will result in error messages. These may be due to Berkeley Madonna struggling to calculate several statistics in the section on "AGGREGATING OUTPUT AND FITTING MODEL PREDICTIONS". If this occurs, you can simply comment out this entire section by inserting a curly bracket at the start and end of the section.

7.2 Model 7.2

7.2.1 Overview of the model

This model, set up in Berkeley Madonna, describes the transmission dynamics of rubella in an age-structured population (one with a "Realistic Age Structure", using the approach of Schenzle (1984) - see Panel 5.4 of the book).

There are two versions of this model, both of which are similar to model 5.5; these incorporate the age-dependent contact patterns described by matrices R1 and R2, which are calculated in section 7.4.2.1.1 of the book. The file called "model 7.2 - matrix R1.mmd" incorporates the age-dependent contact parameters described in Matrix R1; the file called "model 7.2 - matrix R2.mmd" incorporates the age-dependent contact parameters described in Matrices R2.

The files can be used to generate the predictions of the number of new infectious persons with rubella per 100,000 population that are shown in Figure 7.11 in the book.

The key features of the model are as follows:

- The population is stratified into annual age strata between the ages 0 and 75 years, with each age stratum moving to the subsequent age stratum at the end of each year, as shown in Figure O.5.1.
- Vaccination of newborns is introduced in the year specified by year_start_vacc, with an effective coverage specified by eff_vcov_newborns.
- The infection-related parameters are currently set to be those for rubella (average pre-infectious period=10 days, average infectious period= 11 days).
- The force of infection differs between those aged 0-14 years (the young) and those aged ≥15 years.

Further technical details of the model are provided below.

7.2.2 Further details of the model

Equations:

These are identical to those described for Model 5.5 (see page 50), except for the following:

- The value for "up_age" (the maximum age of persons in the population) is set to be 74 years.
- The expression for the number of new infections in the difference equations for the Susceptible and Pre-infectious persons for each age stratum is given by force_of_infn[i]*Sus[i]*dt.

Demography:

1. The population comprises 55 million persons, as specified by the value of total_popn (see the section on "Demography-related variables".

2. The age distribution is rectangular. The number of persons in each single year age stratum (and therefore the number of births each year) is given by total_popn/(up_age+1).

3. The variables tot_pop[0], tot_pop[1], tot_pop[2] equal the total number of individuals who are in their first, second, third etc year of life in the model (i.e. in the age stratum 0, 1, 2 years etc). Tot_pop[0], tot_pop[1] etc are defined in the "USEFUL SUMMARY STATISTICS" section.

Transmission and infection :

1. The infection-related parameters are currently set to be those for rubella (pre-infectious period=10 days, infectious period= 11 days) - see the section called "TRANSMISSION AND INFECTION-RELATED PARAMETERS".

2. The rate at which two specific individuals come into effective contact per unit time differs between those aged 0-14 years (the "young") and those aged ≥15 years (the "old". It is determined by the values beta_yy, beta_yo, beta_oy and beta_oo, which in turn, are determined by the values of beta_1 and beta_2.

3. The force of infection differs between the young and the old. For those aged 0-14 years, it is calculated using the equation:

   force_of_infn[0..14] = beta_yy*Tot_infous_y + beta_yo*Tot_infous_o

   For those aged ≥15 years, it is calculated using the equation:

   force_of_infn[15..up_age] = beta_yo*Tot_infous_y + beta_oo*Tot_infous_o

   Here, tot_infous_y and tot_infous_o are the total numbers of young and old persons in the population, and are calculated in the "USEFUL SUMMARY STATISTICS" section

Vaccination:

The assumptions are identical to those used in Model 5.5, as follows:

1. Vaccination of newborns is introduced some years after the start (specified by the value of year_start_vacc). The proportion of newborns that is immunized each year is stored in the parameters prop_newborns_immunized and eff_vcov_newborns. Note that the effective coverage is interpretable as the product of the vaccination coverage and the vaccine efficacy. For further details, see the section called "VACCINATION-RELATED PARAMETERS".

2. *Unvaccinated* newborns are added to the Sus[0] compartment at the end of each year; *vaccinated* newborns are added to the Imm[0] compartment at the end of each year (see the section called "DIFFERENCE EQUATIONS – PERSONS IN THEIR FIRST YEAR OF LIFE".

Miscellaneous:

1. Several useful statistics have been set up in the "USEFUL STATISTICS" section, such as the proportion of 5, 10, 15 etc year olds who are susceptible, the daily numbers of new infections per 100,000 in various age groups.

Figures:

The figures window has 6 pages:

Page 1: This plots the number of individuals in each age stratum over time.

Page 2: This plots the number of individuals in each age stratum at the end (with the age stratum on the x-axis).

Page 3: This plots the daily force of infection among the young and old over time.

Page 4: This plots the number of infectious young and old persons over time.

Page 5: This plots the proportion of the population that is susceptible in different age groups and overall over time.

Page 5: This plots the number of infectious young and old persons in the population overall over time.

Page 6: This plots the daily number of new infections per 100,000 per day over time. By changing the value for eff_cov_newborns to 0.72, you should be able to reproduce Figure 7.11 in the book.

7.2.3 Suggested exercises

1. Run the model using matrix R1, assuming that no persons are vaccinated and check the following:

   a) Click on pages 1 and 2 of the figures window to check that the number of persons in each age stratum equals the value that you expect.

   b) Click on page 3 of the figures window to check that the values for the average daily force of infection among the young and the old are consistent with the values estimated in the serological data for England and Wales (13% and 4% per year respectively), which were also used to calculate the β parameters.

   c) Click on page 4 of the figures window to check that the values for the number of infectious young and old persons are consistent with the values calculated in Panel 7.4 of the book (18,965 and 2,859 respectively).

   d) Repeat steps a)-c) using matrix R2.

2. You may recall from chapter 5 that if vaccination among newborns is introduced into the population, the force of infection decreases and the proportion of adults in a given age group who are susceptible increases.

   Run model 7.2 using both matrices R1 and R2, assuming that 50% of newborns are vaccinated and answer the following questions.

   a) Referring to page 3 of the figures window, read off the long-term average force of infection obtained for the young and old.

   b) Why does the force of infection predicted for old persons after vaccination is introduced among newborns for the population mixing according to matrix R1 differ

from the corresponding value obtained for the population in which individuals mix according to matrix R2?

c) For which assumption about contact between individuals would you expect the proportion of 20 year olds who are susceptible in the long-term to be highest following the introduction of vaccination for a given level of coverage among newborns? Check your hypothesis by referring to page 5 of the figures window.

d) You may also recall from chapter 5 that if vaccination among newborns is introduced with a coverage which is below the herd immunity threshold, then the overall proportion of the population that is susceptible is identical to that before the introduction of vaccination. These predictions were based on the assumption that individuals mix randomly.

Look at page 5 of the figures window of both models to check if this result still holds if individuals in the population are assumed to mix non-randomly.

3. a) Run both models assuming that 72% of newborns are immunized. Check that predictions of the daily number of new infections per 100,000 population are consistent with those in Figure 7.11.

b) As mentioned on page 210 of the book, the basic reproduction number associated with matrices R1 and R2 are 3.5 and 4.75 respectively ***(NB As stated in the erratum page, the R_0 for matrix R1 is about 3.5 and not 3.6, as stated on page 210 of the book.).***

Calculate the herd immunity thresholds that are associated with these matrices and check that if you incorporate an effective vaccination coverage among newborns which is:

i) slightly greater than the herd immunity threshold for the given mixing pattern, then transmission ceases;

ii) slightly less than the herd immunity threshold for the given mixing pattern, then transmission continues.

4. In chapter 5, we also explored how the number of new infections in different adult age groups might change as the vaccination coverage among newborns increased,

a) Carry out a parameter plot (see page 15 of this document) of the long-term daily number of new infectious persons per 100,000 population for those aged 20, 30 and 40 years against the effective vaccination coverage among newborns using matrices R1 and R2 and compare them against that obtained using Model 5.5 (see question 4 of the suggested exercises for Model 5.5). ***Note: you may prefer to set the number of runs to be 21 in the Parameter plot, with the effective coverage ranging between 0 and 1; you may also want to set the STOPTIME to equal 400,000 days.***

b) How does the assumption that individuals contact each other according to mixing pattern R1 influence conclusions about the number of new infections per 100,000 population in different adult age groups, and the effect of vaccination of newborns?

7.3 Model 7.3

7.3.1 Overview of the model

This model, set up in Excel, can be used to calculate the basic reproduction number that is associated with a given WAIFW matrix or Next Generation Matrix, in which the population is stratified into two different groups. It can be used to see the patterns shown in Figures 7.13, 7.14 and 7.15 in the book.

The file comprises two spreadsheets. The spreadsheet "matrix A or B" can be used to calculate R_0 given input just on the Next Generation Matrix. The spreadsheet "matrix R1 or R2" is similar to "matrix A or B" except that the Next Generation Matrix is calculated from input on the WAIFW matrix, the duration of infectiousness and the size of the two groups considered in the population.

Differences between the layouts of sheets "Matrix R1 or R2" and "Matrix A or B" are described in suggested exercise 4 (see below); the layout of the worksheet "Matrix A or B" is as follows:

1. **Yellow cells (cells A11:K22):** These hold the number of infectious young and old persons introduced into the population (cells G20 and G21 respectively).

   Notice that these have been assigned names. Therefore, if any cell has an equation which uses a named cell, the cell name rather than the cell location can be used in the equation. For example, cell G20 has been assigned the name "infous_y0", and so the term "infous_y0", rather than the location of the cell, can be used in any equation.

2. **Turquoise cells (cells A29:K24):** these hold the Next Generation Matrix in cells D32:E33.

   Notice that the cells have been assigned the names R_yy, R_yo, R_oy and R_oo. To see the name, click on the cell and look at the white bar to the left of the formula bar below the ribbon.

3. **Lilac cells (row 38 onwards):** These hold the number of infectious persons in each generation. Column J holds the number of infectious young and old persons introduced at the start in the simulated population that is used to calculate R_0; column K holds the numbers of infectious young and old persons in the first generation, which are calculated using the Next Generation Matrix and the numbers of infectious young and old persons at the start.

7.3.2 Suggested exercises

1. Use the values for R_yy, R_yo, R_oy and R_oo, together with the number of infectious young and old persons in the first generation to calculate the number of infectious young and old persons, and the total number of infectious persons in the second generation.

You can check your answer by clicking on the ⊞ button above column U, which will reveal the number of infectious persons until the 10th generation, together with Figure A, which holds a plot of these numbers.

2. Calculate the following:
   a) The ratio between the number of infectious persons in generation 8 and that in the preceding generation. Repeat your calculations considering generations 9 and 10 to see whether this ratio converges to the value specified on page 209 of the book (about 0.7).

   b) The proportion of the infectious persons in generations 8, 9 and 10 who are young.

   You can check your answers by clicking on the ⊞ buttons above column AD and next to row 50. This will reveal the ratio between the numbers of infectious persons in successive generations, the proportion of the persons in each generation who are young and old, together with plots of these values. These plots should be similar to those in Figures 7.14 and 7.15 of the book.

3. a) Change the initial numbers of infectious young and old persons to be any values you choose (e.g. the values presented in the legend to Figure 7.13 in the book), and check to see that the ratio between the numbers of infectious persons in each generation, together with the proportion of the infectious persons in each generation who are young always converge to the same value.

   b) Repeat your calculations using Matrix B presented on page 206 of the book, i.e. $\begin{pmatrix} 1.6 & 0.1 \\ 0.1 & 0.1 \end{pmatrix}$ or any other matrix that you choose. *Note: you may find that matrices which have the following kinds of structures* $\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$ *or* $\begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix}$ *(i.e. ones which have non-zero entries along one of the diagonals of the matrix, but zeros elsewhere) do not behave as you might expect. These matrices are discussed on page150 in the solutions to question 2c) for section 7.3.2.*

4. Click on the spreadsheet labelled "Matrix R1 or R2". Its layout is identical to that for the spreadsheet "Matrix A or B", except that it has additional yellow cells, which contain the following:
   - The number of young and old persons which are used to calculate the Next Generation Matrix (cells G14 and G15);
   - The duration of infectiousness (G16);
   - A WAIFW matrix in cells D26:E27, for which the β parameters are currently set to be those for matrix R1 (see section 7.4.2.1.1).

   a) Check that the value obtained for R_0 is consistent with the value presented for matrix R1 on page 210 of the book (see also errata page), i.e. a value of about 3.5.

b) Change the values for the β parameters in the WAIFW matrix to be those for matrix R2 (see page 198 of the book) and check that the value that you estimate for R_0 is consistent with the value presented for matrix R2 on page 210 of the book, i.e. a value of about 4.75.

7.4 Model 7.4

7.4.1 Overview of the model

This file, set up in Excel, illustrates how R_0 for matrices R1 and R2 (see section 7.4.2.1.1 of the book) can be calculated using the simultaneous equation and matrix determinant approaches described on pages 212 and 213 of the book.

The file comprises a single sheet, called “matrix R1 or R2”.

The layout of the worksheet is as follows**:**

1. **Yellow cells (cells A11:K21):** These hold the following:
    - The numbers of young and old persons in the population (cells G14 and G15 respectively);
    - The average infectious period (cell G16);
    - The proportion of the typical infectious person which is young (cell G18)
    - A cell which will eventually hold the value for R_0 (cell G19);
    - A WAIFW matrix in cells D26:E27, for which the β parameters are currently set to be those for matrix R1 (see section 7.4.2.1.1).

    Notice that these have been assigned names. Therefore, if any cell has an equation which uses a named cell, the name rather than the cell location can be used. For example, cell G14 has been assigned the name “N_y”, and so the term “N_y”, rather than the location of the cell, can be used in any equation.

2. **Turquoise cells (cells A29:K24):** these hold the Next Generation Matrix in cells D32:E33.

    Notice that the cells have been assigned the names R_yy, R_yo, R_oy and R_oo. To see the name, click on the cell and look at the white bar to the left of the formula bar below the ribbon.

3. **Cells which are set up for calculating R_0 using the simultaneous equations approach (rows 43-54) and the matrix determinant approach (rows 58-64).** The calculations are described in the exercises below.

7.4.2 Suggested exercises

1. As described on pages 212-213 in the book, if the population is stratified into the young and the old, R_0 and the proportion of the typical infectious person that is young (x) satisfy the following equations:

$$R_{yy}x + R_{yo}(1-x) = R_0 x \quad (7.37)$$

$$R_{oy}x + R_{oo}(1-x) = R_0(1-x) \quad (7.38)$$

    Cells D51 and D52 hold the left hand sides of equations 7.37 and 7.38 respectively; cells E51 and E52 hold the right hand sides of equations 7.37 and 7.38 respectively. The

values for x and R_0 used in these equations are given by the values of x and R0_est in cells G18 and G19 respectively.

We can use Excel's Solver function (see section 10.1.1 of this document) to find values for R_0 and x which satisfy these equations by identifying the values for R0_est and x (cells G19 and G18 respectively) which result in both the value of cell D51 being equal to F51, and in the value of cell D52 being equal to F52.

For this purpose, the grey cells I51, I52 and I55 have been set up as follows:

- Cell I51 measures the square of the difference between the values of cells D51 and F51;
- Cell I52 measures the square of the difference between the values of cells D52 and F52;
- Cell I55 takes the sum of cells I51 and I52. Consequently, this cell should be equal to zero once R0_est equals the value for R_0 that is associated with the Next Generation Matrix, and once x equals the proportion of the typical infectious person that is young.

As discussed in section 10.1.1, when using Solver to identify unknown parameters, we need to specify some single "Target cell" which has to be minimized (maximized or set to zero) by varying the parameters that we are interested in. In our case, we need to vary R0_est and x in order to find the minimum value for cell I55.

a) Set the value for R0_est to be about 3, and x to be about 0.5 and use the Solver function (accessed from the Data option in the main menu) to find R_0 and x. The values that you obtain should be identical to those obtained for Matrix R1 using the simulation approach (see model 7.3) i.e. 3.53 and about 0.31 respectively.

b) Reset R0_est and x to equal 1.5 and 0.5 respectively, and re-run Solver. What do you notice about the value for x and R_0? How might you fix the problem?

To fix the problem encountered in part b), we can set up constraints in Solver so that the parameters that it finds are in a plausible range.

c) Run Solver after setting up the constraints that $x \geq 0$ and $x \leq 1$ to check that you obtain plausible values for R_0 and x.

2. As described on page 212 of the book if the population is stratified into the young and the old, the value for the R_0 associated with a given Next Generation Matrix is the value which results in the determinant of the following matrix being equal to zero:

$$\begin{pmatrix} R_{yy} - R_0 & R_{yo} \\ R_{oy} & R_{oo} - R_0 \end{pmatrix}$$

This matrix is set up in cells D63:E64, with R_0 taking the value R0_est. The determinant of this matrix is calculated in cell H63.

a) Set the value for R0_est to equal 3 and run Solver to find values for R0_est which result in the determinant of the above matrix being equal to zero.

b) Reset R0_est to equal 2 and re-run Solver. What do you notice about the value for *R0_est*?

c) What are the merits of using the matrix determinant, simultaneous equations and simulation approaches for calculating R_0?

7.5 Model 7.5

7.5.1 Overview of the model

This model, set up in Excel, is designed to calculate R_0 and the net reproduction number for measles in England in 1994/5 following the method used by Gay et al, as described in section 7.5.4.1 of the book. It can also be used to answer some parts of question 7.4 in the book.

The spreadsheet comprises 2 sheets:

1. Sheet "R0_Rn_calcs_fin" holds all the calculations for the Next Generation Matrix, R_0 and the net reproduction number.
2. Sheet "R0_Rn_calcs_empty" is identical to "R0_Rn_calcs_fin" except that the cells holding key calculations are left blank. You will be able to fill in the cells by following the steps in the suggested exercises below.

The layout of the spreadsheet "R0_Rn_calcs_fin" is as follows:

1. **Yellow cells (cells A11:I35):** These hold the following:
    - The duration of infectiousness (in years) in cell F13;
    - The values for the β parameters (units of per year) and alpha, which is the scalar used by Gay et al to define the amount of contact between individuals aged 10-14 years. These are in row 17 and the cell for each β parameter has been assigned a different colour;
    - The WAIFW matrix (in cells C22:G26) which was used by Gay et al in the analyses. This is set up using the contents of row 17 and the cells have the same colour coding that i used for the cells in this row. Figure 1 plots this WAIFW matrix.
    - The number of persons in each age group in England in 1994/5 (row 31);
    - The estimated numbers of susceptible persons in each age group in England in 1994/5 (row 32) – see the footnote to Table 7.3 in the book for details of how these were calculated;
    - The number of persons introduced at the start in the simulated population which will be used to calculate R_0 and R_n (see below)

    Notice that the cells have been assigned names. Therefore, if any cell has an equation which uses a named cell, the name rather than the cell location can be used. For example, cell B17 has been assigned the name "beta_1", and so the term "beta_1", rather than the location of the cell, can be used in any equation.

2. **Turquoise cells (cells A36:I43):** Clicking on the ⊞ button next to row 44 will reveal the Next Generation Matrix in cells D39:H43, which can be used to calculate R_0. Notice that the cells have been assigned the names R_11, R_12, R_13 etc. To see the name, click on the cell and look at the white bar to the left of the formula bar below the ribbon.

3. **Lilac cells (row 67 onwards):** These hold the calculations for R_0 using the simulation approach. The cells are as follows:

- Row 71 holds the numbers of infectious persons that are introduced at the start in each age group in the simulated population which is used to calculate R_0.
- Columns B-G in row 72 holds the numbers of infectious persons in each age group and overall in the simulated population in the first generation. These are calculated using the Next Generation Matrix and the numbers of infectious persons in each age group at the start.
- Column I holds the ratio between the number of infectious persons in the first generation and that in the preceding generation.
- Rows 73-81 (seen by clicking on the ⊞ button next to row 82) will reveal the number of infectious persons in each generation until the 10th generation, together with ratio between the numbers of infectious persons in successive generations (i.e. calculations of R_0). This ratio is plotted in Figure 2, which can be seen by clicking on the ⊞ button next to row 66.

4. **Calculations of the net reproduction number:** clicking on the ⊞ button above column T will reveal the following:
    - The Next Generation Matrix in cells N39:R43, which can be used to calculate R_n (see section 7.5.4 in the book for the expression used to calculate each element of the Next Generation Matrix). Notice that the cells have been assigned the names Rn_11, Rn_12, Rn_13 etc.
    - Lilac cells in columns K-S in row 67 onwards. These hold calculations of R_n, using the Next Generation Matrix in cells N39:R43. The calculations are analogous to those used to calculate R_0 in columns B-I.
    - Figure 3, which plots the ratio between the number of infectious persons in a given generation and that in the preceding generation.

7.5.2 Suggested exercises

Questions 1-2 and 4a-c involve setting up equations yourself in the worksheet "R0_Rn_calcs_empty"*. If you prefer to avoid doing this, you may wish to skip to questions 3 and 4d-e, where you can think how assumptions about the amount of contact between 10-14 year olds (determined by the value for α) influence R_0, R_n and conclusions about the potential for a measles epidemic to occur*

1. In the worksheet "R0_Rn_calcs_empty", use the WAIFW matrix provided together with the number of individuals in each age group (row 31), and the duration of infectiousness (cell F13) to set up the Next Generation Matrix in cells D39:H43.

   You can check your answer against the values in the corresponding cells in the worksheet "R0_Rn_calcs_fin". The values that you obtain should be consistent with those for the solutions to question 7.4b ii) in the book.

2. a) In the worksheet "R0_Rn_calcs_empty", use the Next Generation Matrix that you have just calculated, together with the number of infectious persons in each age group at the start to calculate the following for the simulated population:
    i. The number of infectious persons in each age group and the total number of individuals in the first generation (in cells B72:G72);

ii. The ratio between the total number of infectious persons in the first generation and that at the start (in cell I72).

b) Copy the equations that you have just set up down to the 10th generation, and check that, assuming that α=1.5, the ratio between the number of infectious persons in the current and preceding generation converges to the same value as that in sheet "R0_Rn_calcs_fin".

If α=1.5, this ratio (and therefore the value for R_0) should be 9.9 (see also the solution to question 7.4b ii) in the book).

3. How does changing the size of α change the value for R_0, and why?

4. Click on the ⊞ button above column T in the worksheet "R0_Rn_calcs_empty". You should now see some blank turquoise and lilac cells (rows 36-43 and 67 respectively onwards).

a) Use the values for the number of susceptible persons in each age group in 1994/5 in cells C32:G32, the duration of infectiousness and the WAIFW matrix provided to calculate the Next Generation Matrix that you need in order to calculate the net reproduction number.

The values that you obtain should be identical to those presented on page 217 of the book (see also the spreadsheet "R0_Rn_calcs_fin".

b) Use your answer to part a) to set up equations in the appropriate cells for the number of infectious persons in each age group in the first generation in columns and an equation for the ratio between the numbers of infectious persons in the first generation and that in the preceding generation.

c) Copy the equations for the first generation down until the 10th generation and check that the value for R_n for 1994/5 that is obtained assuming that α=1.5 are consistent with the value of about 1 (see Figure 7.17 in the book).

d) Explore how changing the value of α affects the estimates that you obtain for the net reproduction number.

e) Do you think that there was potential for a measles epidemic to occur in 1994/5 in England?

7.6 Model 7.6

7.6.1 Overview of the model

This model, set up in Excel, is designed to calculate R_0 and the net reproduction number for influenza for the example in question 7.5 in the book, for different assumptions about the distribution of the vaccine doses between the young and old.

The spreadsheet comprises 2 sheets:

1. Sheet "R0_Rn_calcs_fin" holds all the calculations for the Next Generation Matrix, R_0 and the net reproduction number.
2. Sheet "R0_Rn_calcs_empty" is identical to "R0_Rn_calcs_fin" except that the cells holding key calculations are left blank. You will be able to fill in the cells by following the steps in the suggested exercises below.

The layout of the spreadsheet "R0_Rn_calcs_fin" is as follows:

1. **Yellow cells (cells A11:I29):** These hold the following:
    - The duration of infectiousness (in days) in cell F13.
    - The WAIFW matrix (in cells C19:D20) which was estimated using Model 7.1 and is provided in the exercise. Figure 1 plots this WAIFW matrix.
    - The number of persons in each age group in the population (row 24).
    - The numbers of young and old persons vaccinated (row 25).
    - The minimum number of susceptible young and old persons.
    - The number of persons introduced at the start in the simulated population which will be used to calculate R_0 and R_n (see below).

    Notice that the cells have been assigned names. Therefore, if any cell has an equation which uses a named cell, the cell name rather than the cell location can be used in that equation. For example, cell C19 has been assigned the name "b_yy", and so the term "b_yy", rather than the location of the cell, can be used in any equation. To see the name, click on the cell and look at the white bar to the left of the formula bar below the ribbon.

2. **Turquoise cells (cells A30:I34):** Clicking on the ⊞ button next to row 35 will reveal the Next Generation Matrix in cells C33:D44, which can be used to calculate R_0. Notice that the cells have been assigned the names R_yy, R_yo, R_oy and R_oo.

3. **Lilac cells (row 61 onwards):** Clicking on ⊞ button next to row 76 will reveal the calculations for R_0 using the simulation approach. The cells are as follows:

    - Row 65 holds the numbers of young and old infectious persons that are introduced at the start in each age group in the simulated population which is used to calculate R_0.
    - Columns B-D in row 66 holds the numbers of young, old and total infectious persons in the simulated population in the first generation. These are calculated using the Next Generation Matrix and the numbers of infectious persons at the start.

- Column F holds the ratio between the number of infectious persons in the first generation and that at the start.
- Rows 67-75 hold the number of infectious persons in each generation until the 10th generation, together with ratio between the numbers of infectious persons in successive generations (i.e. calculations of R_0). This ratio is plotted in Figure 2.

4. **Calculations of the net reproduction number:** clicking on the ⊞ button above column U will reveal the following:
   - The Next Generation Matrix in cells N33:O34, which can be used to calculate R_n. Notice that the cells have been assigned the names Rn_yy, Rn_yo, Rn_oy and Rn_oo etc.
   - Lilac cells in columns K-T in row 65 onwards. These hold calculations of R_n, using the Next Generation Matrix in cells N33:O34. The calculations are analogous to those used to calculate R_0 in columns B-F.
   - Figure 3, which plots the ratio between the number of infectious persons in a given generation and that in the preceding generation.

7.6.2 Suggested exercises

You can practice setting up calculations for the basic and net reproduction number yourself in the spreadsheet "R0_Rn_calcs_empty", by following the steps provided below for questions 1-3. Alternatively, you may like to skip to question 4, where you can think about the impact of different options for the distribution of the vaccine doses between the young and old.

1. In the worksheet "R0_Rn_calcs_empty", use the WAIFW matrix provided, the number of young and old persons in the population (row 24), and the duration of infectiousness (cell F13) to set up the Next Generation Matrix in cells C43:D44.

   You can check your answer against the values in the corresponding cells in the worksheet "R0_Rn_calcs_fin". These values should be consistent with those for the solutions to question 7.5a in the book.

2. a) In the worksheet "R0_Rn_calcs_empty", use the Next Generation Matrix that you have just calculated, together with the number of infectious persons at the start to calculate the following for the simulated population:
      i. The number of young and old infectious persons and the total number of infectious persons in the first generation (in cells B66:D66);
      ii. The ratio between the total number of infectious persons in the first generation and that at the start (in cell F66).

   b) Copy the equations that you have set up for the first generation down to the 10th generation, and check that the ratio calculated in column F (and therefore the value for R_0) converges to about 1.85. This value should be consistent with that in the solutions to question 7.5a in the book.

3. Click on the ⊞ button above column U in the worksheet "R0_Rn_calcs_empty".

a) Use the values for the number of young and old susceptible persons in cells C26:D26, the duration of infectiousness and the WAIFW matrix provided to calculate the Next Generation Matrix that you need in order to calculate the net reproduction number.

b) Use your answer to part a) to set up equations in the appropriate cells for the number of infectious young and old persons in the first generation and an equation for the ratio between the numbers of infectious persons in the first generation and that in the preceding generation.

c) Copy the equations for the first generation down until the 10th generation, and check that the ratio calculated in column F (and therefore the value for R_n) converges to a value which is equal to R_0 (i.e. 1.85) if the Next Generation Matrix is calculated assuming that everyone is susceptible at the start. In the following question, you will be able to explore the effect of different assumptions about the number susceptible when calculating R_n.

4. Assuming that there are sufficient vaccination doses for only 2500 individuals for use before the start of a pandemic, use estimates of the net reproduction number to decide whether it would be best to provide the vaccines to:
    i) children only;
    ii) adults only;
    iii) the same proportion of children and adults;
    iv) equal numbers of children and adults.

   What do you conclude about the potential for an epidemic to occur for each of these vaccination scenarios?

   You can test your hypotheses by running model 7.1 after incorporating the levels of coverage that are associated with these vaccination scenarios (see suggested exercise 2 for Model 7.1).

Chapter 8

Sexually transmitted infections

8.1 Model 8.1

8.1.1 Overview of the model

This model describes the short-term dynamics of a non-immunising curable sexually transmitted infection, such as gonorrhoea, in a population with homogeneous risk behaviour, as described in **Section 8.3** of the book.

The key features of the model are:

i) Closed population (no birth, death, migration). Stable population size
ii) Single gender
iii) Assumes homogeneous (same) risk behaviour
iv) Population size = 20 million
v) One infectious person is introduced into a population where all the other individuals are susceptible
vi) Mean partner change rate in population (per year) = 2
vii) Transmission probability per partnership = 0.75
viii) Duration of infection (years) = 0.167 (2 months)

There are two windows with figures in the model file.

The window with two figures includes:

Page 1: This plots the prevalence of infection over time and R_0 with the x and y-axes set so it will reproduce **Figure 8.5** in the book.

Page 2: This plots the prevalence of infection over time and R_0 with the left y-axis scale set to between 0 and 1 so shows the infection prevalence clearly as the duration of infection or partner change rate is increased.

The other window contains one figure showing a parameter plot of how the endemic prevalence and R_0 vary as the mean partner change rate is increased from 2 to 10.

8.2 Model 8.2

8.2.1 Overview of the model

This model describes the short-term dynamics of a non-immunising curable sexually transmitted infection, such as gonorrhoea, in a population with heterogeneity in risk behaviour and proportionate mixing, as described in **Section 8.4** of the book.

The key features of the model are:

1. Closed population (no birth, death, migration). Stable population size
2. Single gender
3. Assumes **heterogeneous** risk behaviour
4. **Assumes proportionate mixing**
5. Population size = 20 million
6. One infectious person is introduced into a population where all the other individuals are susceptible
7. **2% of population in high activity group**
8. **Overall mean partner change rate (per year) = 2, but rate in high activity group = 31.4 and rate in low activity group = 1.4**
9. Transmission probability per partnership = 0.75
10. Duration of infection (years) = 0.167 (2 months)

There are two windows with figures in the model file.

The window with two figures includes:

Page 1: This plots the prevalence of infection over time in the low and high activity group, and the overall population, reproducing **Figure 8.8 (left)** in the book.

Page 2: This plots the force of infection, or incidence rate over time in the overall population, reproducing **Figure 8.8 (right)** in the book.

The other window contains one figure showing a parameter plot of how the endemic prevalence in the low and high activity group, and in the total population varies as the difference (heterogeneity) in partner change rate between the two activity groups increases, reproducing **Figure 8.9** in the book. The figure is actually a mirror image of that shown in book, because I could not (easily) work out how to plot it the correct way round (!). Suggestions welcome.

8.3 Model 8.3

8.3.1 Overview of the model

This model describes the short-term dynamics of a non-immunising curable sexually transmitted infection, such as gonorrhoea, in a population with heterogeneity in risk behaviour and variable mixing, as described in **Section 8.5** of the book.

The key features of the model are:

1. Closed population (no birth, death, migration). Stable population size
2. Single gender
3. Assumes heterogeneous risk behaviour
4. **Allows 'with-unlike' (disassortative), 'proportionate', or 'with-like' (assortative) mixing between activity groups**
5. Population size = 20 million
6. One infectious person is introduced into a population where all the other individuals are susceptible
7. 2% of population in high activity group.
8. Overall mean partner change rate (per year) = 2, but rate in high activity group = 31.4 and rate in low activity group = 1.4
9. Transmission probability per partnership = 0.75
10. In default scenario duration of infection (years) = 0.167 (2 months). This can be varied to keep R_0 constant as the mixing pattern is changed, see below.

Mixing can be varied using Q (after *Gupta, Anderson et al., AIDS, 1989*), where

$Q < 0$ models with-unlike mixing
$Q = 0$ models proportionate mixing
$0 < Q \leq 1$ models with-like mixing

Note, in the examples in the book, to avoid having to alter partner change rates, the minimum 'with-unlike' mixing level is constrained to be Q = -0.46, see Section 8.5.4

There is one window with one figure and one window with a figure and a table in the model file. The window with one figure shows a plot of R_0 as the population mixing is varied between 'More with-unlike' (Q= – 0.46), through proportionate (Q = 0), to purely 'with-like' mixing (Q = +1). This reproduces the figure shown in **Figure 8.13** and shows that as mixing becomes more with-like, R_0 increases.

The other window with a figure and a table includes:

Page 1: shows a plot of overall prevalence of infectious individuals over time. It can be used to reproduce **Figure 8.14** and show prevalence assuming 'more with-unlike' (Q = – 0.4), proportionate (Q = 0), or 'more with-like' (Q = +0.4) mixing between activity groups while keeping R_0 constant. To do this change Q to equal these values, while ensuring R_0 remains = 1.36 by re-setting the duration of infection to be D = 0.340, 0.167 and 0.097 years, respectively.

Page 2: The table window shows R_0, and the number of secondary infections in high-activity group members generated by an infected high-activity group member (R_{HH}), the number of secondary infections in low-activity group members generated by an infected high activity group member (R_{LH}), the number of secondary infections in high-activity group members generated

by an infected low-activity group member (R_{HL}), and the number of secondary infections in low-activity group members generated by an infected low-activity group member (R_{LL}).

8.4 Model 8.4

8.4.1 Overview of the model

This model describes the short-term dynamics of a non-immunising curable sexually transmitted infection, such as gonorrhoea, in a population with heterogeneity in risk behaviour and variable mixing, as described in **Section 8.5.7** of the book.

The model is identical to Model 8.3 but it is set up to show the effects of changing the mixing pattern between activity groups on equilibrium STI prevalence for a given STI natural history and partner change rates. Thus, this might illustrate the possible effects of an intervention that changes population mixing patterns, but does not affect rates of partner change or STI natural history.

There is one window with a figure and one window with a table in the model file.

The window with a figure shows a parameter plot, on the $\log_{10}$ scale, of the equilibrium prevalence in the low and high activity groups, and overall, by mixing pattern for the default STI natural history and partner change rates. By default the model will reproduce **Figure 8.15c,** the plot for a STI with a 'Moderate' R_0 (= 1.36 when proportionate mixing is assumed).

The model can be used to reproduce **Figure 8.15a and 8.15b** by setting R_0 when proportional mixing to lower (0.50) and higher (4.00), by setting the duration of infection to be 0.062 and 0.493 years, respectively.

As for Model 8.3, the window with the table window shows R_0, and the number of secondary infections in high-activity group members generated by an infected high-activity group member (R_{HH}), the number of secondary infections in low-activity group members generated by an infected high activity group member (R_{LH}), the number of secondary infections in high-activity group members generated by an infected low-activity group member (R_{HL}), and the number of secondary infections in low-activity group members generated by an infected low-activity group member (R_{LL}).

8.5 Model 8.5

8.5.1 Overview of the model

This model describes the short-term dynamics of a non-immunising curable sexually transmitted infection, such as gonorrhoea, in a population with heterogeneity in risk behaviour, variable mixing, and STI screening, as described in **Section 8.5.8** of the book.

The model is very similar to Model 8.3 but screening has been implemented very simply by adding a term to the equations determining the rate of change of the infectious and susceptible individuals in the high and low activity groups, to simulate a higher rate of recovery. These two rates are then altered so that the number of screenings is kept constant but the screenings are targeted at the high or low activity group, or distributed randomly.

The rates in the two groups are calculated from the number of screens per year, '*ns',* and whether screening will be targeted randomly or at the low or high activity group.

There is one window with a figure and one window with a table in the model file.

The window with a figure shows a parameter plot of the equilibrium STI prevalence in the overall population as the number of screenings per year is increased from 0 to 20 million.

By default the model assumes proportionate mixing (ie *Q*=0) and random targeting. If you run this model you will get the 'Target randomly' line of figure **Figure 8.16(middle)**, showing the impact on STI prevalence if proportionate mixing is assumed.

To show how much more effective targeting the high activity group would be, replace

```
{== To target randomly uncomment this code: }
y_H = ns/N
y_L = ns/N

{== To target the high-activity group uncomment this code: }
;y_H = ns/N_H
;y_L = 0comment out (insert a ';') the two lines below
```

With

```
{== To target randomly uncomment this code: }
;y_H = ns/N
;y_L = ns/N

{== To target the high-activity group uncomment this code: }
y_H = ns/N_H
y_L = 0
```

Targeting the low-activity group can be achieved in a similar way.

To reproduce **Figure 8.16(left)** showing the relative impact if more with-unlike mixing is assumed; repeat the three targeting strategies above after setting Q= -0.4 and D = 0.340.

To reproduce **Figure 8.16(right)** showing the relative impact if more with-like mixing is assumed; repeat the three targeting strategies above after setting Q= +0.4 and D = 0.097.

As for Model 8.3, the table window shows R_0, and the number of secondary infections in high-activity group members generated by an infected high-activity group member (R_{HH}), the number of secondary infections in low-activity group members generated by an infected high activity group member (R_{LH}), the number of secondary infections in high-activity group members generated by an infected low-activity group member (R_{HL}), and the number of secondary infections in low-activity group members generated by an infected low-activity group member (R_{LL}).

8.6 Model 8.6

8.6.1 Overview of the model

This model describes the long-term dynamics of HIV in a population with heterogeneous risk behaviour and proportionate mixing, as described in **Section 8.8** of the book.

The key features of the model are:

1. Onset of sexual activity and non-HIV death
2. Single gender
3. Population size = 10,000
4. Assumes heterogeneous risk behaviour
5. 15% of population in high activity group
6. Rate of partner change per year in high activity group = 8 and rate in low activity group = 0.2
7. One infectious person is introduced into a population where all the other individuals are susceptible
8. Transmission probability per partnership = 0.05
9. Duration of sexual activity in absence of HIV (years) = 35
10. Duration of HIV infectiousness (years) = 9
11. Duration of AIDS stage in which no sexual activity is assumed (years) = 1

There is one window with four figures and a table in the model file.

Page 1: This plots the trends in the incidence and prevalence of HIV (left y-axis) and cumulative AIDS deaths (right y-axis), to reproduce **Figure 8.20a** in the book

Page 2: This plots the trends in the mean partner change rate in the population to reproduce **Figure 8.20b** in the book

Page 3: This plots the trends in the numbers of new HIV infections and deaths of HIV infecteds to reproduce **Figure 8.20c** in the book

Page 4: This table is of the trend in the number of HIV infected individuals and can be used to estimate the doubling time of the epidemic, see **Section 8.8.1.2**

Page 5: This plots the trends in the numbers of individuals in the low and high activity groups and the total population size, showing the impact of the HIV epidemic

Chapter 9

Special topics in infectious disease modelling

9.1 Model 9.1

This will be available in future updates...apologies for any inconvenience!

9.2 Model 9.2

This will be available in future updates...apologies for any inconvenience!

9.3 Model 9.3

This will be available in future updates...apologies for any inconvenience!

9.4 Model 9.4

9.4.1 Overview of the model

This model, set up in Berkeley Madonna, describes the long-term dynamics of HIV and a cofactor STI in a population with heterogeneous risk behaviour and proportionate mixing, as described in **Section 9.4** of the book.

The key features of the model are:

- Onset of sexual activity and non-HIV death
- Single gender
- Population size = 10,000
- Assumes heterogeneous risk behaviour
- 15% of population in high activity group
- Rate of partner change per year in high activity group = 8 and rate in low activity group = 0.2
- One HIV infectious person is introduced into a population in year 0 where all the other individuals are susceptible to HIV
- HIV transmission probability per partnership = 0.05
- Duration of sexual activity in absence of HIV (years) = 35
- Duration of HIV infectiousness (years) = 9
- Duration of AIDS stage in which no sexual activity is assumed (years) = 1
- **One cofactor STI infectious person is introduced into a population in year -50 where all the other individuals are susceptible to the cofactor STI. The cofactor STI is introduced 50 years before HIV to allow the cofactor STI to become endemic**
- **Cofactor STI transmission probability per partnership = 0.80**
- **Duration of cofactor STI infectiousness (years) = 0.22**
- **Per partnership STI cofactor for HIV acquisition and transmission = 3.1. Derived from *Hayes et al, J Trop Med Hyg, 1995*, see Panel 9.3.**

There is one window with four figures and one table in the model file.

1. Page 1: This plots the trends in the incidence and prevalence of HIV (left y-axis) and cumulative AIDS deaths (right y-axis), to reproduce Figure 9.9a and Figure 9.9b in the book. By default is shows the more rapid spread of HIV in the presence of the STI cofactor to reproduce Figure 9.9b. To remove the effect of the STI cofactor, set the per partnership STI cofactor for HIV acquisition and transmission = 1. This will now reproduce Figure 9.9a.
2. Page 2: This plots the trends in the prevalence of the cofactor STI, to reproduce Figure 9.9c in the book.
3. Page 3: This plots the trends in the numbers of individuals in the low and high activity groups and the total population size, showing the impact of the HIV epidemic
4. Page 4: This plots the trends in the numbers of new HIV infections and deaths of HIV infecteds
5. Page 5: This table is of the trend in the number of HIV infected individuals and can be used to estimate the doubling time of the epidemic, see Section 9.4.1 of the book.

Appendix

10.1 Estimating parameters by fitting models to data in Excel

10.1.1 Estimating a single parameter

Excel has an inbuilt function ("Solver") which allows users to estimate model parameters by fitting model predictions to data. Solver is used with model 4.2 to estimate the unknown parameters, such as R_0. It is also used with the catalytic models in chapter 5 to estimate the force of infection for the catalytic model which leads to best fits to the data in the yellow cells. The steps for using Solver are as follows:

Step1: Click on the Data option on the main menu and select the Solver option. If the Solver option is not available, click on the Microsoft Office Button  , and then click on "Excel Options". Then click on "Add-Ins", and, in the "Manage" box, select "Excel Add-ins", before clicking on "Go". In the "Add-Ins available" box, check the "Solver Add-in" box, and then click on the OK button.

Step 2: If there is a security warning below the ribbon stating "Macros have been disabled", click on the "Options" button next to this warning and then select the "Enable this content" option, before clicking on OK.

Step 3: If you have not yet done so, select the Solver option.

You should then see a dialogue box in which you specify which (Target) cell should be minimized, maximized or set equal to some value, and what cell should be changed in order to attain this minimum etc.

Step 4:

a) Set the Target cell to point to the deviance. For model 4.2, the deviance is located in cell D28; for the catalytic models in chapter 5, it is in cell D18.

b) Specify that the deviance should be minimized by selecting the "Min" option under the "Equal to" option.

c) Specify that the parameter that you're estimating should be changed. For model 4.2, you might type in R_0 or "F19" (without the inverted commas) in the "By Changing Cells" box. For models 5.2 or 5.3 you would specify that the force of infection should be changed by typing "D14" or the cell name (without the inverted commas) in the "By Changing Cells" box.

d) Click on the "Solve" button.

Solver will then try different values for the parameter that you're estimating until it finds the one which leads to the smallest deviance. If it succeeds, a box will come up informing you that Solver has found a solution and asking you to click on OK to accept the value.

If Solver fails, this may be due to a mistake in the expression for the model prediction, or because the search for the best-fitting value started too far from the best value, which made it too difficult for Solver to find it. The solution to this problem is to start Solver after having changed the value for the parameter being fitted (e.g. foi_pyr_mumps if you're fitting the force of infection for mumps using file model 5.2) to a value which, by eye, looks as though it gives a good fit to the data.

As is the case for most fitting routines (see also section 10.2.3), the best-fitting parameter values usually depend on the starting values used when fitting model predictions to data. Several different starting values should be used and the overall best-fitting values should be taken as those which result in the smallest overall value for the goodness of fit statistic.

10.1.2 Estimating two or more parameters (models 4.2, 5.3 and 5.4)

The methods are identical to those described above for estimating a single parameter, except that step 4c has to be changed as follows:

Specify that the parameters that you're estimating should be changed. For model 4.2, you might type in "R0,prop_reported", or "F19,F21" (without the inverted commas) in the "By Changing Cells" box, if you are estimating R_0 and the proportion of infectious persons who were reported. If you're estimating the force of infection among those aged <15 and ≥15 years in model 5.4, you would type "D14:D15" or the cell names, separated by a comma (i.e. "foi_u15,foi_g15" without the inverted commas).

10.1.3 Calculating the 95% confidence interval for a single parameter (models 5.2 and 5.3)

A macro (called maxlhood01) has been set up to calculate the 95% confidence interval on the force of infection estimate. This macro refers to specific locations in the spreadsheet where it expects to find the force of infection and the deviance, and it will not work if the cell locations are changed from their original settings; also, the macro only works if only one parameter is being estimated.

10.1.3.1 Calling the macro

The macro can be called by one of the following two ways:

a) (Models 5.2 and 5.3 only) Click on the large grey button in cell G23. If it is not visible, click on the ⊞ button next to row 29.
b) Select the View option on the main menu. Click on the macro button, selecting the "View macros" option and choose the maxlhood01 option.

Note that for this macro to run, you need to have enabled the macros in the spreadsheet. You will have done this already if you followed the steps in section 10.1.1 (see step 2). If you get a warning message about the macro not being enabled, enable the macro by following step 2 in section 10.1.1. If you can't see the Security warning mentioned in that step, you may need to close and reopen the spreadsheet again in order to see it.

The macro uses Solver and, if the macro is able to find Solver (see section 10.1.6.1 if it cannot do so), a dialogue box will appear when Solver finds the best-fitting force of infection and the upper and lower confidence limit. You will need to click on OK whenever this occurs to allow the macro to continue to work.

10.1.3.2 Method used by the macro

The mechanics of how the macro maxlhood works is as follows:

1. Solver is first used to fit the expression for the proportion (ever) infected to the proportion seropositive by minimizing the log-likelihood deviance. The best-fitting value for the force of infection is then pasted into cell C25 and the deviance that results from this force of infection (referred to as the "optimal deviance") is pasted into cell C27.

2. The upper 95% confidence limit on the force of infection is then obtained as follows:
    a. The macro sets the value for the force of infection in cell D14 to be slightly above the best-fitting value.
    b. Solver then looks for a value for the force of infection which leads to a deviance which differs from optimal deviance by 3.84. Note that the difference between the deviance resulting from the force of infection in cell D14 and the optimal deviance is held in cell D21.
    c. Once it finds this value (which equals the upper limit of the 95% confidence interval), it is pasted into cell E25.
3. The lower 95% confidence limit is calculated in a similar way, except the value for the force of infection (in step a) is set to be slightly lower than the best-fitting value.

4. The best-fitting value for the force of infection is copied back from cell C25 into cell D14.

You can also view the code for the macro by clicking on the macro button from the View tab, clicking on the macro name (maxlhood01) and then selecting the Edit button.

10.1.4 Calculating 95% confidence intervals for two parameters (model 5.4)

A macro (called maxlhood02) has been set up in model 5.4 which calculates 95% confidence intervals on two parameters. As for the macro maxlhood01, it points to specific locations in the spreadsheet where it expects to find the parameters (in this case, foi_u15 and foi_g15) and the deviance. Consequently the macro will not work if the cell locations are changed from their original settings; also it only works if two parameters are being estimated.

The method for calling the macro is identical to that described for macro maxlhood01 (see section 10.1.3.1). The steps for calculating the 95% confidence intervals are similar to those described for maxlhood01:

1. Solver is first used to fit the expression for the age-specific proportion (ever) infected to the age-specific proportion seropositive by minimizing the log-likelihood deviance. The resulting best-fitting values for the force of infection among those aged <15 and ≥15 years are then pasted into cells C25 and C26. The deviance that results from these values (referred to as the optimal deviance) is pasted into cell C27.

2. The upper 95% confidence limit on the force of infection for those aged <15 years is then obtained as follows:
    a. The macro sets the value for foi_u15 in cell D14 to be slightly above the best-fitting value, with foi_g15 fixed at the best-fitting value.
    b. Solver then looks for a value for foi_u15 which leads to a deviance which differs from optimal deviance by 3.84. Note that cell D21 holds the difference between the deviance resulting from the force of infection in cell D14 and the optimal deviance.
    c. Once it finds this value (which equals the upper limit of the 95% confidence interval), it is pasted into cell E25.

3. The lower 95% confidence limit is calculated similarly, except the value for foi_u15 (in step a) is set to be slightly above the best-fitting value. The lowest 95% confidence limit is pasted into cell D25.

4. The value for the force of infection among those aged <15 years in cell D14 is then reset to equal the best-fitting value (currently in cell C25).

5. The upper and lower 95% confidence limits for foi_g15 are then calculated using the method described in step 2, but with foi_u15 fixed at its best-fitting value. The upper and lower confidence limits are pasted into cells E26 and D26 respectively.

6. The best-fitting values for the force of infection are copied back from cell C25 and C26 into cells D14 and D15 respectively.

10.1.5 Calculating 95% confidence intervals for three parameters (model 4.2)

A macro (called maxlhood03) has been set up in model 4.2 which calculates 95% confidence intervals on three parameters. As for the macros maxlhood01 and maxlhood02, it points to specific locations in the spreadsheet where it expects to find the parameters (in this case, R_0, the proportion of infectious persons who are reported as cases, and the initial number of infectious persons) and the deviance. Consequently the macro will not work if the cell locations are changed from their original settings; also it only works if two parameters are being estimated.

The method for using the macro is similar to that described for macro maxlhood01 (see section 10.1.3.1). The macro can be called by one of the following two ways:

a) Click on the large grey button in cell G33. If it is not visible, click on the ⊞ button next to row 39.
b) Select the View option on the main menu. Click on the macro button, selecting the "View macros" option and choose the maxlhood03 option.

Note that for this macro to run, you need to have enabled the macros in the spreadsheet. You will have done this already if you followed the steps in section 10.1.1 (see step 2). If you get a warning message about the macro not being enabled, enable the macro by following step 2 in section 10.1.1. If you can't see the Security warning mentioned in that step, you may need to close and reopen the spreadsheet again in order to see it.

The macro uses Solver and, if the macro is able to find Solver (see section 10.1.6.1 if it cannot do so), a dialogue box will appear when Solver finds a best-fitting value and the upper and lower confidence limit. You will need to click on OK whenever this occurs to allow the macro to continue to work.

The steps for calculating the 95% confidence intervals are similar to those described for maxlhood01 and maxlhood02:

1. Solver is first used to fit model predictions of the number of cases reported each week to the observed data. The resulting best-fitting values for R_0, the proportion of infectious persons who are reported and the initial numbers of infectious persons are then pasted into cells C34:C36. The deviance that results from these values (referred to as the optimal deviance) is pasted into cell C38.

2. The upper 95% confidence limit on R_0 is then obtained as follows:
    a. The macro sets the value for R_0 in cell F19 to be slightly above the best-fitting value, with the other parameters fixed at the best-fitting value.
    b. Solver then looks for a value for R_0 which leads to a deviance which differs from optimal deviance by 3.84. Note that cell D31 holds the difference between the deviance resulting from the R_0 in cell F19 and the optimal deviance.
    c. Once it finds this value (which equals the upper limit of the 95% confidence interval), it is pasted into cell E35.

3. The lower 95% confidence limit is calculated in a similar way to the upper 95% confidence limit, except the value for R_0 (in step 2a) is set to be slightly below the best-fitting value. The lowest 95% confidence limit is pasted into cell D35.

4. The values for R_0 and the other parameters are then reset to equal the best-fitting values (currently in cell C35:C37).

5. The upper and lower 95% confidence limits for the proportion of infectious persons who are reported are then calculated using the method described in step 2, but with the other parameters (R_0 and initial numbers of infectious persons fixed at their best-fitting values. The upper and lower confidence limits are pasted into cells E36 and D36 respectively.

6. The values for all the parameters are then reset to equal the best-fitting values (currently in cell C35:C37).

7. The upper and lower 95% confidence limits for the initial numbers of infectious persons are then calculated using the method described in step 2, but with the other parameters (R_0 and proportion of infectious persons who are reported fixed at their best-fitting values. The upper and lower confidence limits are pasted into cells E37 and D37 respectively.

10.1.6 Known issues with the confidence interval macros...

10.1.6.1 Locating Solver

The macros sometime have problems with locating Solver. If, when running the macro, the macro opens up a window with the macro code and with the Word "Solver" highlighted, and a dialogue box appears saying that there is a compile error and that the project or library cannot be found, try the following steps:

a) Close the dialogue box and click on the reset button  in the window in which the macro has paused to stop the macro. This button resembles a blue box on a grey background.
b) Select the tools option from the main menu and select the References option, followed by the Available references option. Deselect the box next to Missing:Solver.XLA, and click on the box next to "SOLVER".

After exiting the dialogue box, return to the main spreadsheet and run the macro as before.

10.2 Estimating parameters by fitting models to data in Berkeley Madonna

There are two main ways of fitting models to data in Berkeley Madonna, namely using the Curve Fit and Optimize functions.

10.2.1 The curve fit function

10.2.1.1 Using the curve fit function

The easiest method of fitting models to data in Berkeley Madonna is to use the "Curve fit" function. This function can be used with Model 7.1 to fit model predictions of the cumulative numbers of reported young and old cases against the corresponding observed data to estimate the unknown parameters, namely the initial numbers of infectious persons (Infous_y0 and Infous_o0), the contact parameters (b1 and b2) and the proportion of infectious persons that were reported to the GP practice (frac_rep). The steps are as follows:

1. Choose the Curve fit option from the Parameter option on the main menu.

2. Select the parameters that you would like to estimate by double clicking on them in the list in the "Available" box. They should then appear in the "Parameters" box.

3. Click on each parameter box and in the boxes to right of the box, specify the minimum and maximum possible values for the parameters, together with the value for the first and second guesses.

4. In the lower half of the box, select the variables and the data to which you wish to fit the model. In the case of Model 7.1, we wish to fit the variable Cum_reported_y to the data which are stored in #cum_rep_y and the variable Cum_reported_o to the data which are stored in #cum_rep_o. This can be done as follows:

    a. Click on the "Multiple fits box".

    b. Select Cum_reported_y from the drop-down list labelled "fit variable" and select "#cum_rep_y" from the drop down list labelled "To dataset" and then click on the "Add button". Cum_reported_y and #cum_rep_y should now appear in the box under "Multiple fits"

    c. Do the same for "Cum_reported_o" and "#cum_rep_o".

5. To fit the model, click on the OK button.

Berkeley Madonna will then fit model predictions of the cumulative numbers of young and old cases to the observed data. The value for Berkeley Madonna's inbuilt goodness of fit statistic will appear once the fitting has finished if you had previously selected the option "Pause after Curve fit". This option can be found through the "Preferences" option from the "Edit" option in the main menu. See section 10.2.3 for further important general considerations when fitting models.

The best-fitting parameter values can be seen either in the parameters window or on the sliders. In general, the best-fitting values will depend on the values taken for guess 1 and guess 2 for the parameters, although they should all lead to a similar goodness of fit, and similar parameter values. The overall best-fitting values should be taken as those which result in the smallest overall value for the goodness of fit statistic.

10.2.1.2 Disadvantages of the curve fit function

Berkeley Madonna's curve fitting option is relatively straightforward to use. However, its main disadvantage is that it is not designed to work easily with predictions of the number of new infections or cases per day or per week and we have to use it to fit to the *cumulative* number of cases and by default, all of the datapoints are given equal weight in the fitting. This means that the curve-fit function will preferentially try to find parameter values which lead to a good fit to the data on the cumulative numbers of cases towards the end of the outbreak, since these are the largest, and therefore potentially contribute most to the goodness of fit statistic.

10.2.2 Fitting models using the Optimize function

To overcome the problems with Berkeley Madonna's curve fit function (see section 10.2.1.2), we can use Berkeley Madonna's optimize function to fit model predictions of the numbers of new infections or cases per week to observed data. Considering model 7.1, we can do this by setting up variables which aggregate output on the daily numbers of cases into the weekly numbers of cases (see the code in Model 7.2 for calculating wkly_rep_y and wkly_rep_o in section on "AGGREGATING OUTPUT AND FITTING MODEL PREDICTIONS"). We then need to set up an appropriate goodness of fit statistic in Berkeley Madonna and tell the optimize function to minimize it by identifying parameter values that we are interested in.

The steps for using the Optimize function are as follows:

1. Choose the Optimize option from the Parameter option on the main menu.

2. Select the parameters that you would like to estimate by double clicking on them in the list in the "Available" box. They should then appear in the "Parameters" box.

3. Click on each parameter box and in the boxes to right of the box, specify the minimum and maximum possible values for the parameters, together with the value for the first and second guesses.

4. In the box labelled "Minimize Expression", type in the name of the expression for the goodness of fit statistic that you wish to minimize. In model 7.1, this can be either "ssq" or "minus_llhood".

5. To fit the model, click on the OK button.

Berkeley Madonna then fits model predictions to the observed data to estimate the unknown parameters. As for the "Curve fit" function, Berkeley Madonna will display the value for the goodness of fit statistic if the option "Pause after Curve fit" has been selected before the start of the fitting. This option can be found through the "Preferences" option from the "Edit" option in the main menu. Similarly, the best-fitting parameter values can be seen in the parameters window or in the sliders.

10.2.3 General issues when fitting models using Berkeley Madonna

In general, the best-fitting values obtained using the optimize or curve fit options depend on the values for guesses 1 and 2 for the parameters and on the goodness of fit statistic that is used. However, the best-fitting values are usually all fairly similar and they should all lead to a similar goodness of fit. In any case, when trying to estimate unknown parameter values, several different values for guesses 1 and 2 should be used and the overall best-fitting values for the parameters should be taken as those which result in the smallest overall value for the goodness of fit statistic.

As a general rule, it is also advisable to set up the same model in another package (e.g. Excel, Matlab etc) and/or set up the model using a programming language which calls up tested fitting routines (such as those provided in the Numerical Recipes book) to help identify any errors in the code.

Solutions to suggested exercises

Exercises for Model 2.1

1. a) The total population size is given by the sum of the number of susceptible, pre-infectious, infectious and immune individuals. This remains unchanged over time, with the value equal to the value for tot_popn (100,000 individuals).

b) Figure O2.1 shows the plot that you should have obtained, comparing the numbers of new infections per day and the number of new infectious persons per day. The Excel expression for day 1 is given by =beta*B55*D55, i.e. the Excel equivalent of $\beta S_t I_t$, The peak in the number of new infections per day occurs approximately two days before that in the number of new infectious persons per day, i.e. on day 56, as compared with day 58. You would expect the number of new infections per day to peak roughly two days before the peak in the number of new infectious persons per day, given that newly infected individuals are assumed to become infectious after an average period of two days.

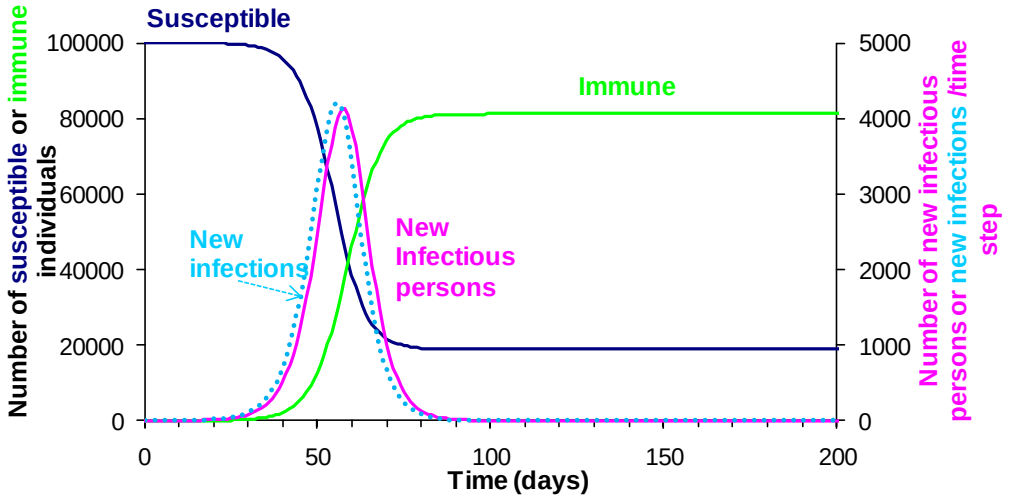


Figure O2.1: Comparison between predictions of the number of new infections per day and the number of new infectious persons per day, predicted using Model 2.1. One infectious person is introduced into the population comprising 99,999 susceptible individuals at the start, R_0=2, pre-infectious period=infectious period =2 days, no individuals are assumed to be born into or to die from the population. The time step size is taken to be 1 day.

c) Cells D55:D255 hold the number of infectious persons at any given time until the 200th day (using a 1 day time step). They are approximately two-fold greater than the number of new infectious persons per day, as illustrated in Figure O2.2. This is to be expected, given that prevalence ≈ incidence × duration of the condition. Therefore, the number of infectious

persons at any given time equals the number of new infectious persons per day ×duration of infectiousness (=2 days).

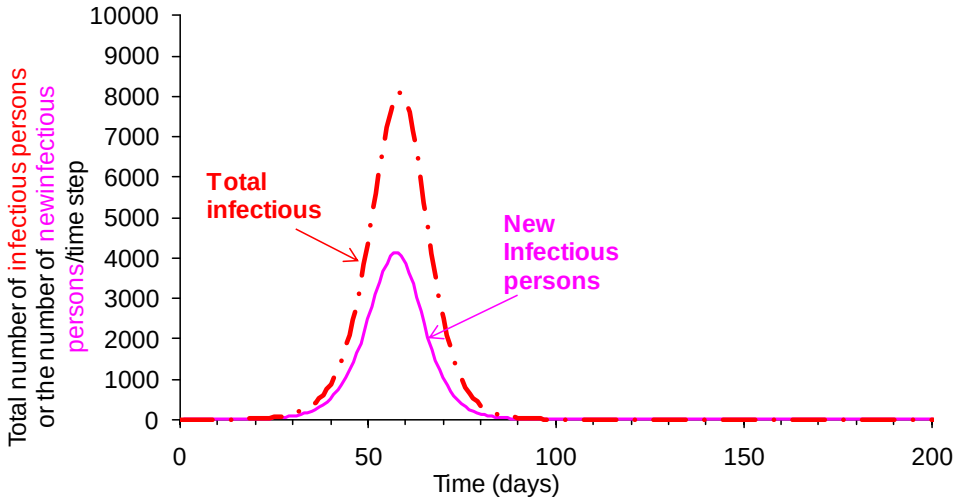


Figure O2.2: Comparison between predictions of the number of new infectious persons per day, and the total number of infectious persons, predicted using Model 2.1. See the caption to Figure O2.1 for further details.

2 a) and b) The following are the expressions that you should have set up for day 1 in part b) - see also "Model 2.1a.xlsx":

Number susceptible:	=L55-beta*M55*L55
Number infectious:	=M55+L55*M55*beta-M55*rec_rate
Number immune:	=N55+M55*rec_rate
Number of new infectious persons per day:	=beta*L55*M55

Figure O2.3 compares predictions of the number of immune individuals and the number of new infectious persons per day obtained using the SEIR model (from question 1) and the SIR model.

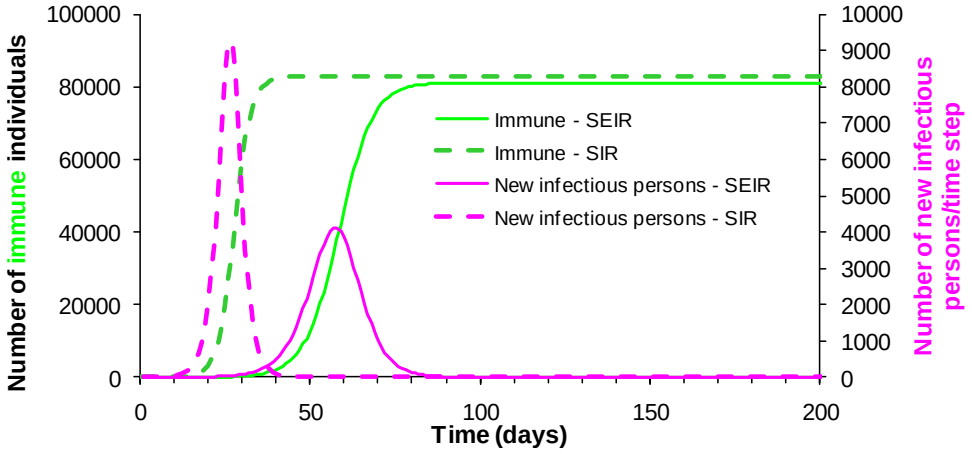


Figure O2.3: Comparison between predictions of the number of immune individuals and the new infectious persons with influenza per day obtained using an SEIR model and an SIR model. See the caption to Figure O2.1 for further details.

The SIR model predicts that the number of new infectious persons per day increases more rapidly than does the SEIR model. This is to be expected since the SIR model assumes that individuals are infectious immediately after infection, whereas the SEIR model assumes that, on average, individuals are infectious 2 days after infection.

Notice that the number of infectious persons peaks once a sufficient proportion of susceptible individuals has been depleted. For both models, this occurs once about 50% of the population is immune. Consequently, the peak predicted by the SIR model occurs sooner than that predicted by the SEIR model. For example, the susceptible population is depleted more rapidly in the SIR model than in the SEIR model, and therefore it reaches the threshold required for the epidemic to peak more quickly than does the SEIR model. These issues are discussed in further detail in chapter 4 of the book.

The number of immune individuals predicted by the end of the epidemic by the SIR model is slightly higher than that predicted by the SEIR model. In fact, they should be very similar. Differences between the number of immune individuals predicted by the two models are largely due to rounding error – you should find that the numbers are almost identical if the time step size is very small (e.g. 0.001 days). Issues relating to the time step are discussed in chapter 3 of the book.

Exercises for Model 3.1

1. b) Predictions obtained using a time step of 0.01 days are very similar to those obtained using a time step of 0.05 days for both the measles and influenza models. This is to be expected, given that predictions obtained using a time step of 0.1 and 0.05 days are already similar. Therefore reducing the time step size further should not influence predictions any further.

c) Table O3.1 shows the number of immune individuals predicted at the end of the epidemic for the measles and influenza models. As might be expected (given your observations of the number of infectious individuals during the course of the epidemic for different time step sizes), this is relatively insensitive to the time step size if the time step is relatively small (<0.1 days).

For the measles model, the number of immune individuals predicted at the end of the epidemic when the time step is 3 or more days is unrealistic, as it exceeds the total population size of 100,000. This highlights the importance of using a suitably small time step when setting up models using difference equations. However, note that even in these situations, the population size as given by the sum of the number of susceptible, pre-infectious, infectious and immune individuals (the values given in the lilac cells in column I) still sum to 100,000.

Table O3.1: Summary of the epidemic size for measles and influenza (as given by the number of immune individuals at the end of the epidemic) obtained using Model 3.1 for different values of the time step size. The last two columns provide the percentage difference between the estimate obtained using a given step size and that obtained using a step size of 0.01 days.

Time step (days)	Number of immune individuals at the end of the epidemic		% difference from the estimate obtained using a time step of 0.01 days	
	Influenza	Measles	Influenza	Measles
0.01	79676.7	99898.3	--	--
0.05	79751	99999.8	0.09	0.10
0.1	79821	99999.8	0.18	0.10
0.5	80396.1	99999.9	0.90	0.10
1	81157.6	100000	1.86	0.10
2	86871.4	100000	9.03	0.10
3	--	100000.3	--	0.10
4	--	100454.7	--	0.56
5	--	--	--	--

Exercises for Model 3.2

1. Since the *per capita* birth rate is assumed to equal the *per capita* death rate, and all individuals are assumed to experience the same death rate, you would expect the population size to remain constant over time, with 100,000 persons. You should find that this is indeed the case when you sum up the number of susceptible, pre-infectious, infectious and immune individuals at each time step.

2.a) The peaks and troughs in rubella incidence occur less frequently than those in measles incidence, and it takes longer for the incidence to reach equilibrium following the introduction of one infectious person with rubella into a totally susceptible population than it does following the introduction of one infectious person with measles. This might be expected given that the average pre-infectious and infectious periods for rubella (and therefore the serial interval) are longer than those for measles. These cycles are described in further detail in chapter 4.

b) To speed up how quickly the model reaches equilibrium, one approach would be to increase the time step size. However, this may affect the accuracy of the model, depending on the amount by which the step size increases.

Another possible approach would be to set the initial numbers of susceptible, infectious, and immune individuals to values which differ slightly from the equilibrium values. For example, we could set the number of immune individuals at the start to equal 80000, and set the number of infectious or susceptible individuals to equal 1 and tot_popn-1-80000 respectively. In this situation, you should find that the cycles in rubella incidence establish themselves more quickly than in the situation when the population is assumed to be totally susceptible at the start.

As discussed in chapter 4, the equilibrium values for the number of susceptible and immune individuals in the absence of vaccination are given by tot_popn/R_0 and tot_popn×(1-1/R_0) respectively. Substituting for tot_popn=100000 and R_0=7 into this equation leads to equilibrium numbers of susceptible and immune individuals of about 14283 and 85714 respectively.

Exercises for Model 4.2

1 a) If R_0 is assumed to equal 2.1, then model predictions of the number of new reported cases underestimate the observed data particularly around the peak of the epidemic. To improve the fit of model predictions to the data you would need to increase the value for R_0, if the other parameters are left unchanged.

b) If you estimate R_0 by fitting model predictions to the data using Solver, you would get a values for R_0 of about 2.4, which should be consistent with your answer to question 1a).

c) To improve the model fit, you might want to change the assumed value for the number of infectious persons at the start, the proportion of persons who are reported, the proportion immune at the start and the values for the pre-infectious and infectious periods, as none of these statistics are known reliably.

3 a) If you reduce the average pre-infectious and infectious periods to both equal 1.5 days without changing anything else, you should notice that the predicted numbers of reported cases begin to increase earlier than when the average pre-infectious and infectious periods are assumed to equal 2 days. This follows from the fact that, if the pre-infectious period is 1.5 days, individuals become infectious sooner than when the pre-infectious period is assumed to be 2 days.

For model predictions to match the data assuming that the pre-infectious and infectious periods are 1.5 days, you would therefore need to assume that infectious persons are less infectious (i.e. R_0 should be reduced) than when the pre-infectious and infectious periods are 2 days, since this will slow the rate at which new infectious persons appear in the population.

This hypothesis is confirmed in Table O.4.1, which summarizes the best-fitting parameter values obtained by fitting the model to the data assuming that the pre-infectious and infectious periods were 1.5 days. Specifically, the best-fitting value for R_0 is about 2.2, which is lower than that obtained assuming a pre-infectious period of about 2 days (2.6).

Table O.4.1: Summary of the best-fitting parameter values obtained by fitting model 4.2 to the Cumberland data, using the macro maxlhood03, assuming that the pre-infectious and infectious periods are both 1.5 days.

Parameter	**Estimate (95% confidence interval)**
R_0	2.24 (2.23,2.25)
Proportion of infectious persons who are reported as cases	0.9 (0.86,0.94)
Initial number of infectious persons	0.59 (0.55,0.63)
Deviance	153 (on 10 degrees of freedom)

3b) If no individuals are assumed to be immune at the start, then the model predicts that the number of cases should increase more rapidly at the start than if the proportion of the population that is assumed to be immune at the start equals 0.3.

To think about how the parameters considered would need to change in order for model predictions to match the data, you may first want to consider the following issues:

i) What is the net reproduction number at the start of the epidemic? You might expect that the net reproduction number at the start should be similar for different best-fitting models obtained assuming different values for the proportion immune at the start.

   For the existing assumptions in the model, 70% of the population is assumed to be susceptible at the start. If the best-fitting value for R_0 is 2.57, this means that the net reproduction number at the start must equal 2.57×0.7=1.8. If we now change the model to assume that no individuals are immune initially, then for the net reproduction number to equal 1.8 at the start, the basic reproduction number would have to equal 1.8. You might therefore expect that if you refit the model assuming that no one is immune at the start, then the best-fitting value for R_0 will be about 1.8.

ii) How do the three parameters being estimated affect the epidemic curve? You should notice that, of the three parameters considered, R_0 has a bigger effect on the shape of the epidemic curve than the other two parameters. For example, the initial number of infectious persons mainly influences the timing of the peak epidemic, whereas the proportion of infectious persons who are reported influences the number of cases each week, without greatly affecting the general shape. The extent to which the proportion of infectious persons who are reported and the initial numbers of infectious persons would need to change in order for model predictions to match the data is difficult to predict.

You should have obtained the best-fitting values shown in Table O.4.2 after refitting the model assuming that no individuals are immune at the start. As hypothesised, the value of R_0 is indeed about 1.8.

Table O.4.2: Summary of the best-fitting parameter values obtained by fitting model 4.2 to the Cumberland data, using the macro maxlhood03, assuming that the pre-infectious and infectious periods are both 2 days and no individuals are immune at the start

Parameter	**Estimate (95% confidence interval)**
R_0	1.8 (1.79,1.81)
Proportion of infectious persons who are reported as cases	0.54 (0.51,0.56)
Initial number of infectious persons	1.14 (1.06,1.23)
Deviance	146 (10 degrees of freedom)

4. In total, 2085 cases were reported during the epidemic in Cumberland. If only 90% of the symptomatic persons were reported, then the number of symptomatic persons in the population would have been 2085/0.9 = 2317.

If we use the best-fitting value for prop_reported that was obtained assuming that the pre-infectious and infectious periods were 2 days (i.e. 77%), then we can estimate that the total number of persons who had been infectious in the population during the epidemic would have been 2085/0.77 = 2606.

The proportion of infectious persons who were symptomatic can then be estimated as the number of symptomatic cases in the population divided by the total number of infectious persons in the population, i.e. 2317/2606=0.89. This calculation assumes that all of those who were symptomatic were infectious, which is not necessarily the case!

If we use the best-fitting value for prop_reported obtained assuming that the pre-infectious and infectious periods were both 1.5 days (i.e. 90%), then we would have estimated that 100% of those infectious would have been symptomatic.

Exercises for Model 4.3

1. The values for R_n when the daily number of new infectious persons is increasing, decreasing, the proportion of the population that is susceptible or immune should be as follows:

Number of new infectious persons per day	R_n	Proportion susceptible	Proportion immune
Increasing	>1	>0.07	<0.93
Decreasing	<1	<0.07	>0.93
At a peak or a trough ("constant")	1	~0.07	~0.93

0.07 is approximately equal to the epidemic threshold ($1/R_0$), if R_0 equals 13; 0.93 is approximately equal to the herd immunity threshold ($1-1/R_0$).

2. Whilst the number of new infectious persons is *above* the number of births per day, the proportion of the population that is susceptible is *decreasing*; whilst the number of new infectious persons per day is below the number of births per day, the proportion of the population that is susceptible is increasing. The explanation for this is provided on page 87 of the book.

3. i) and ii) Following the logic on page 93 of the book, you would expect the inter-epidemic period to be shortest for measles since its basic reproduction number is higher than that of the infections listed in the table. Conversely, the inter-epidemic period for smallpox should be the longest, since its basic reproduction number is the smallest. However, there is much variation in the basic reproduction number for varicella, and, plausibly, its inter-epidemic period could be longer than that for smallpox.

iii) The inter-epidemic periods are predicted to be as follows:

	Pre-infectious period (days)(*D'*)	Infectious period (days) (*D*)	R_0	Inter-epidemic (years) period predicted by using	
				Model*	Equation 4.31
Measles	8	7	12	3.3	3.2
			18	2.5	2.6
Varicella	14	7	3	~10	8.9
			10	4	4.2
			17	3.3	3.2
Smallpox	14	21	5	~9	8.1
			7	6.7	6.6
Rubella	10	11	6	~5	5.6
			7	5	5.1

* Calculated by counting the number of cycles occurring during a 10 year period once the cycles appear to occur at regular have stabilized and then dividing 10 by this number. Note that for low values of R_0, the value for STOPTIME in the model needs to be extended in order to see regular cycles in incidence.

Estimates of the inter-epidemic period obtained using the two methods are reasonably consistent. Those for measles, smallpox and rubella are also consistent with the data plotted in Figure 4.16 of the book. The data for varicella are compatible with estimates obtained assuming a basic reproduction number of 10. However, note that estimates of the inter-epidemic period obtained using both the model and equation 4.31 are based on simple assumptions (random mixing, average life expectancy of 70 years).

Exercises for Model 4.4

1. b) An epidemic occurs approximately every two years, i.e. the inter-epidemic period is 2 years.

c) If the number of cases in the population is at a peak at time t, then $C_{t+1}=C_t$. If this occurs at time t=0, then $C_1=C_0$. Substituting for $C_1=C_0$ into the equation $C_1 = kS_0C_0$, we obtain the result:

$$C_0 = kS_0C_0$$

Dividing both sides of this equation by C_0, we obtain the result:

$$1 = kS_0$$

After rearranging this equation, we obtain our intended result that:

$$k = 1/S_0$$

d) The basic reproduction number can be calculated using the expression kN, where N is the population size. This equation can be derived from the equation $R_0=\beta ND$, that is discussed in chapter 2 of the book.

For example, as discussed in Panel 4.4, k is interpretable as the fraction of all contacts (i.e., occurring over the entire infectious period) between a susceptible person and a case which result in the susceptible person becoming a case. k/D should therefore be approximately equal to β, which is defined as the rate at which two specific individuals come into effective contact per unit time. Substituting for $\beta = k/D$ into the equation $R_0=\beta ND$ leads to the equation $R_0=kN$. Substituting for N=2 million and $k=6.67\times10^{-6}$ per serial interval into this equation implies that $R_0\approx13$.

e) To adapt the model to describe the transmission of rubella, it is necessary to change the duration of the serial interval to 3 weeks, as this will then automatically update the number of births per time step. The value for k also needs to be changed, depending on whether R_0 is assumed to be identical to that for measles. In the absence of further information, the number of cases at a peak can be kept to be identical to that for measles.

Considering part i), if R_0 for rubella is assumed to be identical to that for measles, then the values for k for the measles and rubella models are identical. For example, rearranging the equation $R_0=kN$ (see part d)), leads to the result $k=R_0/N$. Since R_0 for rubella is assumed to equal that for measles, this equation implies that k for measles and rubella must also be identical.

Considering part ii), if R_0 is assumed to be 10, then using equation $k=R_0/N$ implies that $k=10/2000000 = 5\times10^{-6}$ per serial interval. Since the value for k changes, the value for S_0 must also change. It can be calculated using the equation $S_0=1/k$, which is obtained after rearranging the equation $k=1/S_0$ (see part c)). Substituting for $k=5\times10^{-6}$ per serial interval into this equation, we obtain the result that S_0 must equal 200,000.

Figure O.4.1 summarizes predictions of the number of cases of rubella per serial interval obtained assuming R_0=13 and 10, showing that the inter-epidemic period is about 2.5 and 3

years for these assumptions respectively. As might be expected, the number of susceptible individuals is lower when R_0 is assumed to equal 13, than in the situation when it is assumed to equal 10.

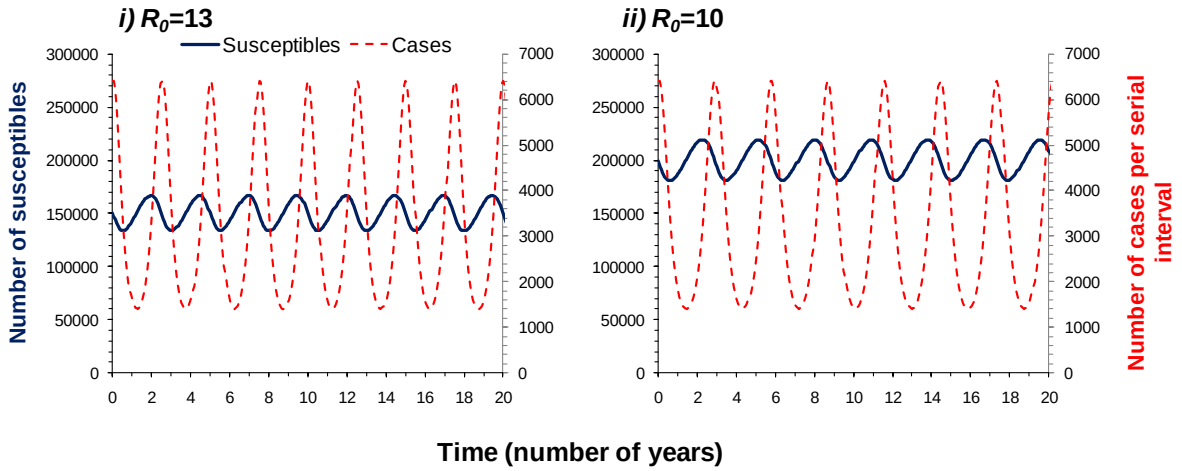


Figure O.4.1: Predictions of the number of susceptibles and the numbers of rubella cases per serial interval obtained using Model 4.4, assuming that R_0 is either 13 or 10.

Exercises for Model 4.5

1b) The model suggests that school holidays have the potential to affect the cycles in incidence, provided that the reduction in contact during school holidays is sufficiently lower than that during school term times.

However, the model is somewhat simple, e.g. it assumes that people mix randomly, it does not account for age-dependent transmission and contact. Also, it assumes that there are only 4 school holidays per year, whereas in reality (at least in the England) children have 1 week holidays during the spring and summer. A further complication is that school holidays (and therefore reductions in contact between schoolchildren) do not always occur simultaneously across the country.

Exercises for Model 4.6

1. The growth rate *m* can be calculated using the expression $\frac{1}{100}\ln\left(\frac{N(100)}{N(0)}\right)$ where $N(0)$ and $N(100)$ are the population sizes at the start and after 100 years respectively. This expression can be obtained after rearranging equation 3.15 (see page 55 of the book).

Annual birth rate per 1000 population	Population size after 100 years	Annual growth rate (%/year)
15	84648	-0.2
25	230098	0.8
40	1031230	2.3

   The growth rate can also be calculated as the difference between the per capita birth rate and the per capita mortality rate (=1/(average life expectancy) = 0.1667 per year). For an annual birth rate of 40 per 1000 per year, this equation leads to values of the growth rate of $100\times(40\times0.001-0.1667) \approx 2.3\%$ per year.

2. As suggested on page 94 of the book, increasing the birth rate should lead to a reduction in the inter-epidemic period. This follows from the fact that an increased birth rate means that an increased number of newborns come in per unit time, which reduces time that is required for the proportion of the population that is susceptible to reach the threshold for an epidemic to occur ($1/R_0$).

3. a) If transmission is assumed to be density-dependent, then, as discussed in Panel 2.5, the average force of infection is assumed to increase as the population size grows. This should have a similar effect to that of increasing the basic reproduction number. Following the logic discussed in section 4.3.2.2 in the book, you would therefore expect the inter-epidemic period to be shorter if transmission is assumed to be density-dependent than if it is frequency-dependent (as assumed in the current model).

   b) The file “model 4.6a.mmd” is identical to model 4.6, except that the force of infection has been changed so that it is given by the equation $\lambda(t)=\beta I(t)$. As shown in Figure O.4.2, for this assumption, the predicted average force of infection increases over time, and the inter-epidemic period is shorter than for the (frequency dependency) assumption.

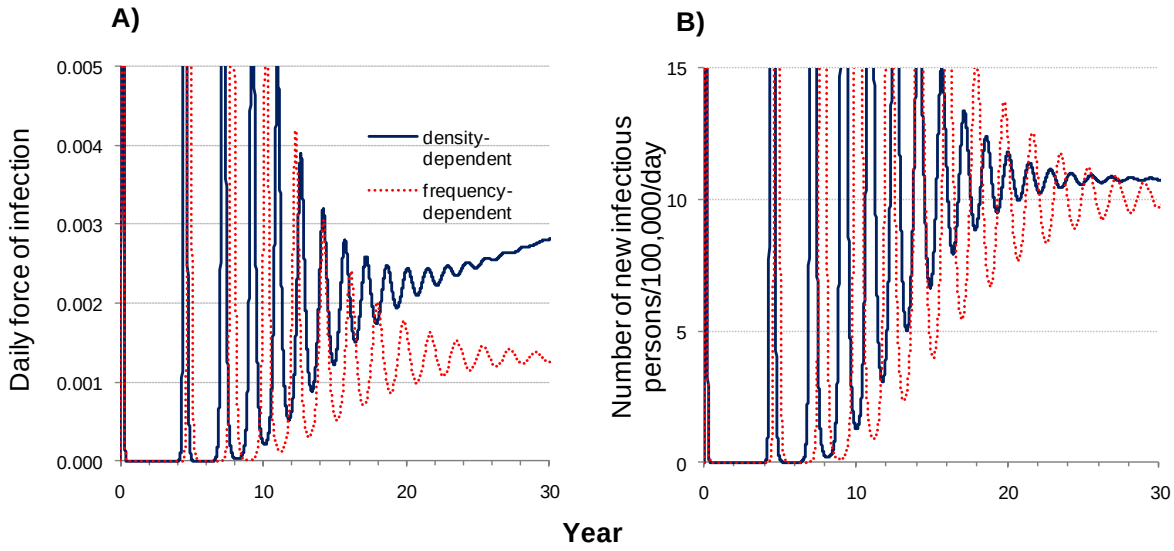


Figure O.4.2: Comparison between predictions of the daily force of infection and the daily number of new infectious persons, obtained using Models 4.6 and 4.6a. The annual birth rate is assumed to be 40 per 1000 population.

Exercises for Model 4.7

1. b) and c) The herd immunity thresholds for measles, rubella and smallpox can be calculated using the equation $1-1/R_0$, and are as follows:

	Assumed R_0	**Herd immunity threshold**
Measles	13	100×(1-1/13) ≈ 92%
Rubella	7	100×(1-1/7) ≈ 86%
Smallpox	5	100×(1-1/5) ≈ 80%

You should find that introducing vaccination at above the levels of coverage implied by the herd immunity threshold do result in control of transmission.

Note that for practical purposes, the impact of introducing vaccination is best seen when vaccination is introduced once the incidence reaches equilibrium in the model. Model 4.7 has been set up so that one infectious person is introduced into the totally susceptible population and for the values of R_0 assumed for rubella and smallpox, equilibrium is reached more than 100 years after the start of the simulations. For these assumptions, it would therefore be appropriate to change the value for year_start_vaccination so that it is above 100 years.

Alternatively, as discussed in the solution to question 2b) for the exercises for Model 3.2, you speed up the time at which the incidence reaches equilibrium by changing the values for the initial number of susceptible and immune persons to values that are close to equilibrium. For example, changing the value for Immune_0 so that it equals total_popn*(1-R0) and setting Susceptible_0 to equal total_popn-Infectious_0-Immune_0 will result in the equilibrium being more apparent less than 100 years after the start of model runs, than in the situation whereby one infectious person is introduced into the population at the start.

2. Figure O.4.3 shows the plot of the proportion susceptible over time, for different levels of effective vaccination coverage which are below the herd immunity threshold. On average, it remains unchanged over time.

An unchanging average value for the proportion susceptible is to be expected since, if the coverage is below the herd immunity threshold, the infection is still endemic and therefore the average net reproduction number still equals one. As discussed in section 4.6 of the book, the net reproduction number is related to the proportion susceptible through the equation:

$$R_n = R_0 s$$

Therefore, substituting for R_n=1 into this equation and rearranging it leads to the result:

$$s = 1/R_0$$

i.e. the proportion susceptible remains $1/R_0$ if the effective vaccination coverage is below the herd immunity threshold.

If the coverage is above the herd immunity threshold, transmission should cease and the proportion susceptible is just given by 1-proportion effectively vaccinated.

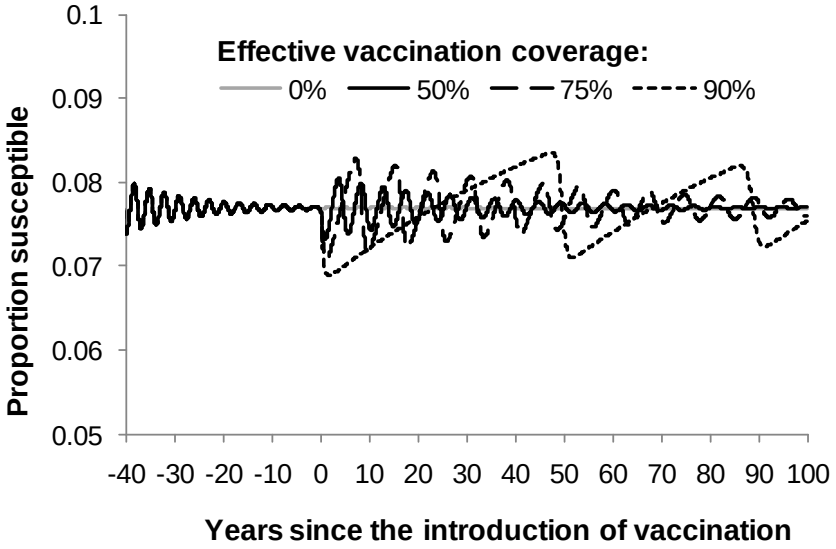


Figure O.4.3: Predictions of the proportion of the population that is susceptible to measles for different levels of effective vaccination coverage among newborns (introduced in year 0), obtained using Model 4.7.

3 a) As discussed in the answer to question 4.2b in the book, the average age at infection can be calculated from the inter-epidemic period using the equation: $A = \frac{T^2}{4\pi^2(D+D')}$.

Values for the inter-epidemic period can be obtained from the plot in the model using the formula: 10/(number of complete cycles over a 10 year period. The values obtained using this method, together with the values for the average age at infection obtained using the above equation, are summarized in Table O.4.3.

Note that in Table O.4.3, *the number of complete cycles over a 10 year period was read off from the figure at a point after the cycles in incidence had stabilized, i.e. when they occurred at regular intervals (albeit at a low level). In some instances, regular cycles in incidence could be seen if the value for STOPTIME in the model was extended to up to 146000 days.*

You should be cautious about accepting the values for the average age at infection since the model makes several simplifying assumptions, e.g. that individuals mix randomly, the mortality rate is identical for all persons in the population, and it does not account for maternal immunity.

b) Rearranging the equation $A = \frac{1}{\lambda + m}$ leads to the following equation for the average force of infection:

$$\lambda = \frac{1}{A} - m$$

As shown in Table O.4.3, the values for the average force of infection obtained using the average age at infection calculated in part a) are generally consistent with the values predicted by the model.

Table O.4.3: Summary of predictions of the inter-epidemic period, the average age at infection and the average force of infection predicted in the long-term after the introduction of vaccination of newborns in Model 4.7.

Effective vaccination coverage among newborns (%)	**No. of cycles over a 10 year period**	**Inter-epidemic period after the introduction of vaccination (years)**	**Average age at infection (years)***	**Average daily force of infection☺**	**Average daily force of infection predicted by the model**
50%	~2.6	~3.8	~9	$\sim 2.6\times10^{-4}$	$\sim 2\times10^{-4}$
80%	~1.5	~6.7	~28	$\sim 6\times10^{-6}$	$\sim 6\times10^{-5}$

* calculated using the equation $A = \dfrac{T^2}{4\pi^2(D+D')}$

☺ calculated using the equation $\lambda = \dfrac{1}{A} - m$. (Note that the units used for A and m must be consistent when doing the calculation!).

Exercises for Model 5.2

1. For both mumps and rubella, model predictions underestimate the observed data for teenagers and overestimate those for adults. It is therefore difficult to predict whether the true average force of infection was higher or lower than the value currently assumed!

4. You would expect the best-fitting force of infection to greater if the test sensitivity is assumed to be <100% than when the test sensitivity is assumed to be 100%. By definition, when the test has a sensitivity of <100%, not all of those who have been infected are identified as such. Therefore, in order for model predictions of the proportion seropositive to match the observed seropositive, the model needs to assume that a higher proportion of individuals of a given age have been infected than in the situation when the sensitivity is assumed to be 100%.

b and c) To incorporate 80% test sensitivity in the model, the expression in the pink cells of column G would need to be multiplied by 0.8. The values in this column are then interpretable as the predicted proportion of the population in the given age group that is seropositive. You should find that when the new expressions are fitted to the data, the estimated average force of infection increases (as you should have predicted in question 4a) to about 22% per year for rubella and 40% for mumps.

You should be cautious about accepting the estimates obtained assuming that the sensitivity is 80% since they lead to a poor fit of model predictions to the data. This poor fit should be clear from the graph and from the fact that the deviance is much greater than that obtained when the test sensitivity is assumed to be 100%. On the other hand, if the sensitivity is assumed to be 95%, the fit of the model to the data is considerably better than that obtained assuming 100% sensitivity.

In fact, using techniques described for model 5.4 for fitting models to data and estimating confidence intervals for two unknown parameters to the data, it is possible to estimate the sensitivity of the test for rubella and mumps antibodies to be about 94% (95%CI: 93-95%) and 97% (95% CI: 97-98%) respectively.

The following compares the best-fitting force of infection estimates obtained assuming a sensitivity of 100%, 95% and 80%, together with the deviance:

	Sensitivity	Force of infection (95% CI)	Deviance
Mumps	100%	19.8 (19.1, 20.5)	337
	95%	27.6 (26.3, 28.9)	133
	80%	39.6 (37.2, 42.3)	1294
	97% (best-fitting)	25.3 (24.3, 26.5)	81
Rubella	100%	11.6 (11.1, 12.1)	133
	95%	14.5 (13.7, 15.3)	67
	80%	21.8 (20.3, 23.5)	274
	94% (best-fitting)	14.9 (14.1, 15.8)	66

Exercises for Model 5.3

3. If the duration of maternal immunity is reduced to 4 months, then you would expect the estimated value for the force of infection to be lower than that estimated when you assume that it is 6 months.

This difference follows from the fact that if the duration of maternal immunity is assumed to be short, the model assumes that individuals of a given age have been susceptible for a longer period of time than for the situation when the duration of maternal immunity is assumed to be long. The model is therefore able to reproduce the observed proportion seropositive by that age using a lower force of infection, than in the situation when the duration of maternal immunity is assumed to be long.

Assumption about maternal immunity	**Duration**	**Force of infection (%) (95% CI)**
Waning immunity	6 months (average)	12.0 (11.5 ,12.5)
	4 months (average)	11.8 (11.3 ,12.4)
Fixed duration	6 months	12.1 (11.6 ,12.6)
	4 months	11.9 (11.4 ,12.4)

4. The following table summarizes the values for the age at which the smallest proportion of individuals should be immune if the average duration of maternal immunity is fixed at 6 months, for different values for the force of infection, based on the expression $a_{\min} = \frac{\ln(\lambda/\mu)}{\lambda - \mu}$. These should be consistent with the values predicted by the catalytic model.

	Assumed force of infection			
	1%/yr	5%/yr	10%/yr	20%/yr
a_{min}	2.7 years	1.9	1.6	1.3

Exercises for Model 5.4

2. To incorporate maternal immunity, equations 5.27a and 5.27b for the proportion susceptible at age *a* would need to be amended as follows:

$$s(a) = \begin{cases} e^{-\lambda_1(a-0.5)} & a < 15 \text{ years} \\ e^{-14.5\lambda_1} e^{-\lambda_2(a-15)} & a \geq 15 \text{ years} \end{cases}$$

These expressions can be incorporated into the spreadsheet by changing the expression for prop_sus15 (cell D16) to =exp(-14.5*foi_u15), and changing the expression for the proportion of 0.5 years that are susceptible to =exp(-foi_u15*(A36-0.5)), before copying it down until age 14.5 years.

The best-fitting estimates should then be as follows:

	Force of infection (% per year) (95% CI)	
Assumed duration of maternal immunity	**<15 year olds**	**≥15 year olds**
No immunity	13.3 (12.7, 13.9)	4.2 (2.9, 5.6)
6 months	14.1 (13.5, 14.7)	3.8 (2.5, 5.2)

As discussed on pages 119-120 of the book, incorporating maternal immunity into the model leads to an increased estimate for the force of infection (at least for those aged <15 years).

Exercises for Model 5.5

2. a) Rearranging the equation $R_0=\lambda L$ leads to the equation $\lambda=R_0/L$. Substituting for $R_0=12$ and $L=70$ years into this equation implies that the average force of infection equals 12/70 $\approx$0.17 per year, which is equivalent to $0.17/365=4.7\times10^{-4}$ per day. This value is reasonably consistent with that predicted by the model (taken as the midpoint between the highest and lowest values) of about 4.6×10^{-4} per day.

b) Similarly, the equation $\lambda=R_0/L$ implies that the average force of infection is about *0.1* per year or $0.1/365 = 2.7\times10^{-4}$ per day, which is reasonably consistent with the value predicted by the model (2.6×10^{-4} per day).

c) The following table compares model predictions against estimates obtained using given equations. You should find that the values predicted by the model are fairly consistent with those obtained using the equations (as you would expect!). In fact, comparing model predictions against values predicted by the equations provides a good way of checking that the model has been set up correctly.

R_0	Method used for estimation	Proportion susceptible		Daily number of new infections per 100,000 (20 yr olds) (calculated as $100{,}000\times\lambda e^{-\lambda 20}/365$)
		20 year olds (calculated as $e^{-\lambda 20}$)	Overall (calculated as $1/R_0$)	
7	Predicted by an equation	0.135	0.142	3.7
	Model prediction	0.13	0.145	3.5
12	Predicted by an equation	0.032	0.083	1.5
	Model prediction	0.030	0.084	1.33

3. a) The average force of infection in the long-term after the introduction of immunization of 50% of newborns is about 2.3×10^{-4} per day. Substituting this value into the equation $R_0 = \frac{\lambda' L}{(1-v)(1-e^{-\lambda' L})}$, and assuming that L=70 years, v=0.5 implies that R_0 is given by the equation $R_0 = \frac{2.3\times10^{-4}\times365\times70}{(1-0.5)(1-e^{-2.3\times10^{-4}\times365\times70})} \approx 12$, which is consistent with the value assumed in the model.

c) If R_0=12, the herd immunity threshold equals $1-1/R_0 = 1-1/12 \approx 0.917$.

For children in a given age group, as the effective vaccination coverage among newborns increases (and is still below the herd immunity threshold) the proportion susceptible in the long-term decreases, because of the removal of the cohort at birth through vaccination (as discussed on page 130 of the book).

For adults, as the vaccination coverage increases and it is below the herd immunity threshold, the proportion susceptible increases. This follows from the fact that the force of infection decreases as the vaccination coverage increases. Therefore (following the arguments on page 133 of the book) the proportion of unvaccinated individuals who reach adulthood still susceptible increases as the vaccination coverage among newborns

increases. This outweighs the removal of the cohort of individuals at birth through vaccination; the net effect is that the proportion of all adults in a given age group who are susceptible increases.

If the effective vaccination coverage is above the herd immunity threshold, then eventually, the proportion of the population in a given age group that is susceptible equals 1-effective vaccination coverage (e.g. 3% if the effective vaccination coverage is 97%). This follows from the fact that once the effective vaccination coverage exceeds the herd immunity threshold, transmission stops in the model (see page 3 of the Figures window) and therefore the proportion susceptible will just depend on the vaccination coverage.

Note: to identify whether long-term predictions from the model following the introduction of vaccination are correct, you can check that model predictions for the proportion of a given age group that is susceptible are consistent with those estimated using equation 5.31 in the book. Similarly, predictions of the number of new infections per day can be compared against those predicted using equation 5.36 in the book.

4. The parameter plot should produce predictions of a general decline in the force of infection and the daily number of new infections in each age group in the long-term with increasing effective vaccination coverage among newborns which are consistent with those shown in Figures 5.14 and 5.22 in the book (see Figure O.5.2 and Figure O.5.3).

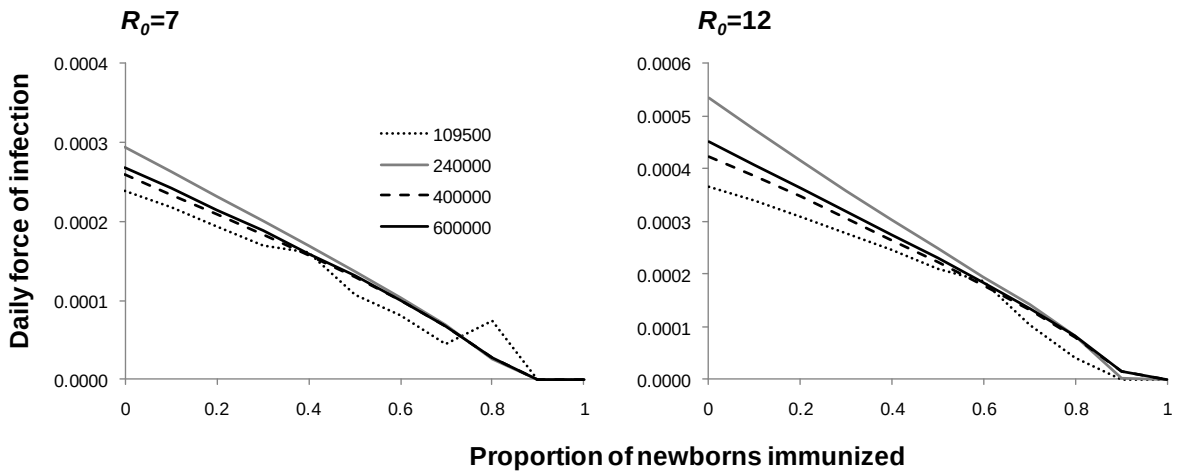


Figure O.5.2 Predictions of the final daily force of infection obtained using a parameter plot for Model 5.5 for different values for the effective vaccination coverage among newborns introduced in year 200 in the model. The different lines are obtained using different values for the stoptime, assuming that R_0=7 or 12. The number of runs used in the parameter plot was 11.

However, you should find that the plot depends on the value for the stoptime used when running the model, since the stoptime determines whether or not the model has reached equilibrium by the end of the run time (i.e. the point taken for the final force of infection and the number of new infections per day in the age groups of interest).

This is illustrated in Figure O.5.2 and Figure O.5.3. For example, considering the predictions obtained assuming that R_0=7, the average daily force of infection and the number of new infections per 100,000 in a given age group are predicted to be higher when the effective vaccination coverage among newborns is 80% than when the latter is 70% and when the

stoptime in the model was 109,500 days. This follows from the fact that an epidemic is predicted to peak shortly before day 109,500, and therefore the model has not yet reached equilibrium by this time. In contrast, these patterns are not predicted when the stop time is 240,000, 400,000 or 600,000 days.

The disadvantage of using such long stoptimes is that the time required to run the model can be prohibitively long. To reduce this problem, it would also be reasonable to bring forward the year in which vaccination is introduced (e.g. introduce it from the start). For example, the outcome of interest is the final force of infection, and if the time period during vaccination has been in place is sufficiently long in the model, the final force of infection obtained should be is independent of the year in which the vaccination is introduced.

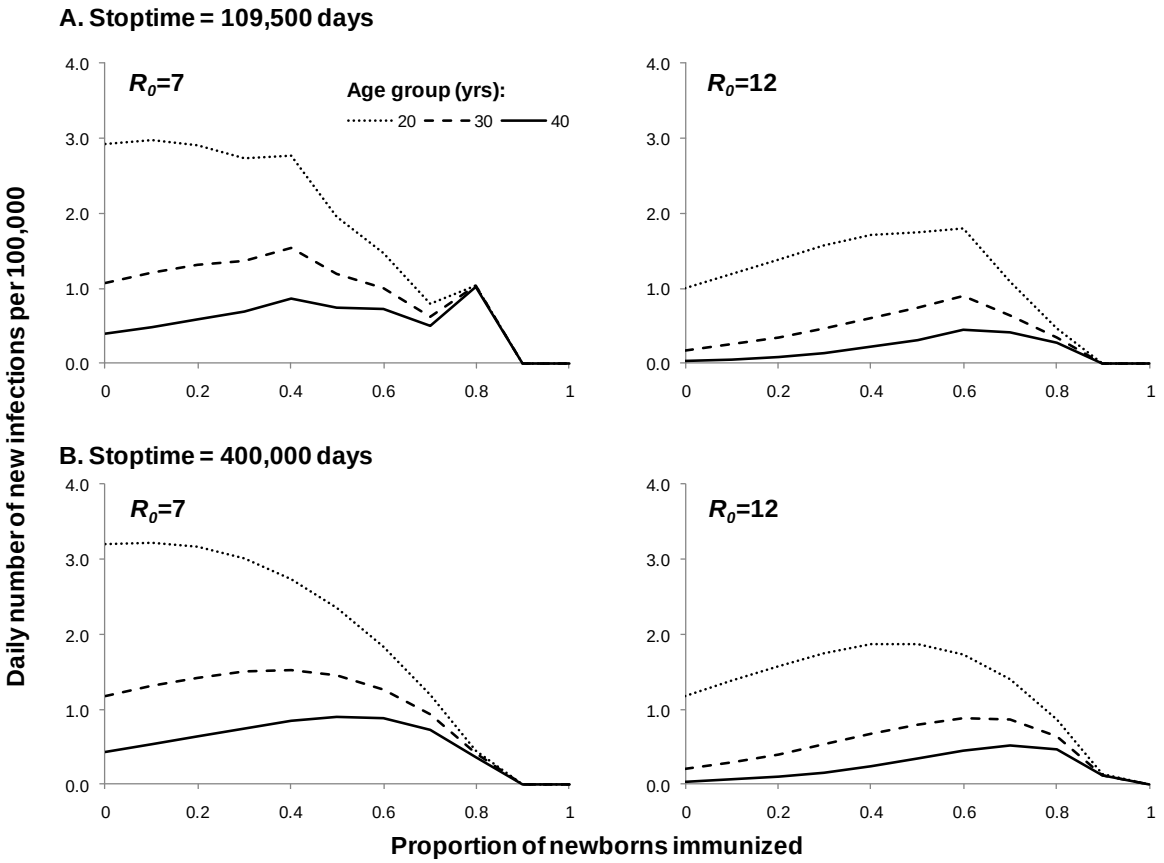


Figure O.5.3: Predictions of the final daily number of new infections per 100,000 at different ages obtained using a parameter plot by Model 5.5 for different values for the effective vaccination coverage among newborns introduced in year 200 in the model, and using different values for the stoptime, assuming that R_0=7 or 12. The number of runs used in the parameter plot was 11.

b) Based on these parameter plots, you should be most cautious about introducing rubella vaccination among newborns in the high transmission setting (R_0=12), since it leads to a greater relative increase in the number of new infections per 100,000 population in adult age groups than in a low transmission setting. Therefore, unless women are vaccinated before they reach children-bearing age, the introduction of rubella vaccination among newborns could, potentially, lead to an increase in the burden of Congenital Rubella Syndrome (see pages 140-141 of the book).

Exercises for Model 5.6

1a) The long-term average force of infection following the introduction of 100% immunization of 13 year olds is about 1.14×10^{-4} per day. Substituting this value for λ', and for L=70 years and v=1 into equation 5.37 leads to the following value for R_0: $R_0 = \frac{1.14\times10^{-4}\times70\times365}{1-1\times e^{-1.14\times10^{-4}\times13\times365}-0} \approx 7$, which is consistent with the value assumed in the model

c) i) The proportion of children that are susceptible increases, e.g. from 35% before the introduction of 100% immunization of 13 year olds to about 64% for 10 year olds. This increase follows from the fact that the introduction of vaccination leads to a reduction in the force of infection, which increases the proportion of individuals who reach a given age still susceptible. Notice that 10 year olds are below the age at vaccination and do not benefit from the direct effect of vaccination.

The proportion of individuals who are susceptible drops to zero because in the model, all of them would have been immunized.

ii) The number of new infections among children decreases following the introduction of 100% immunization of 13 year olds, largely resulting from the reduction in the force of infection, which, in this situation, outweighs the increase in the proportion susceptible seen in part c) i).

The number of new infections among adults goes down to zero in the model once the first cohort of vaccinees reaches a given adult age group. This follows from the fact that all of them would have immunized.

d) The force of infection before the introduction of vaccination was about 2.64×10^{-4} per day. The reduction in the average force of infection as a result of immunization of 100% of 13 year olds in the model is therefore $(2.64\times10^{-4} - 1.14\times10^{-4}) \times100/ 2.64\times10^{-4} = 57\%$.

e) If R_0=12, the average force of infection before and after vaccination is about 4.85×10^{-4} and 3.74×10^{-4} per day respectively. This corresponds to a reduction of 23%, i.e. the reduction in the force of infection is smaller in the high transmission setting than in a low transmission setting.

This result follows from the fact that, in the absence of vaccination, a considerably smaller proportion of individuals in the high transmission setting reach age 13 years still susceptible to infection, than in the low transmission setting. As a result, since only those susceptible to infection at age 13 years are vaccinated in the model, the proportion of the *overall* population that is vaccinated in the high transmission setting is much lower than in the low transmission setting. Therefore the effect on the overall amount of transmission in the high transmission setting is much smaller than in the low transmission setting.

2. Figure O.5.4 shows the parameter plot obtained for the daily force of infection using a stoptime of 109,500 days. The plot for R_0=7 is generally consistent with the plot shown in Figure 5.23 in the book.

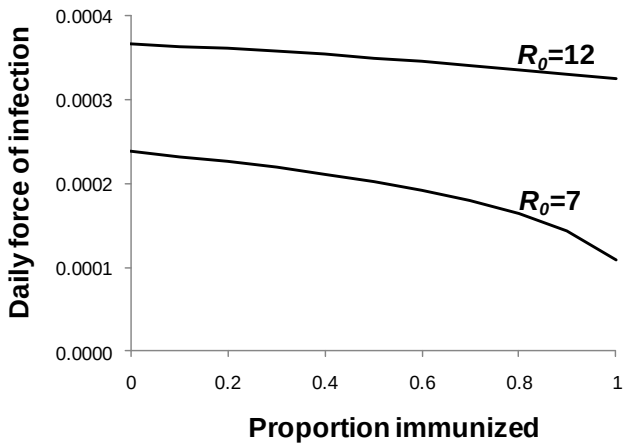


Figure O.5.4: Parameter plot of the (final) daily force of infection obtained using Model 5.6 predicted for different values of the proportion of 13 year olds that are immunized.

This plot again highlights the fact that vaccinating 13 year olds has a smaller impact on the force of infection in high transmission settings than in low transmission settings, although, in both settings, the impact on the overall force of infection is smaller than that of vaccinating newborns (see Figure O.5.2).

As shown in Figure O.5.5, the number of new infections per 100,000 among those aged 20, 30 and 40 years (and therefore the burden of Congenital Rubella Syndrome) is predicted to be lower in the long-term for all levels of vaccination coverage among 13 year olds, than for the situation when only newborns are vaccinated. You might therefore be happier about introducing vaccination among 13 year olds than about introducing vaccination of newborns.

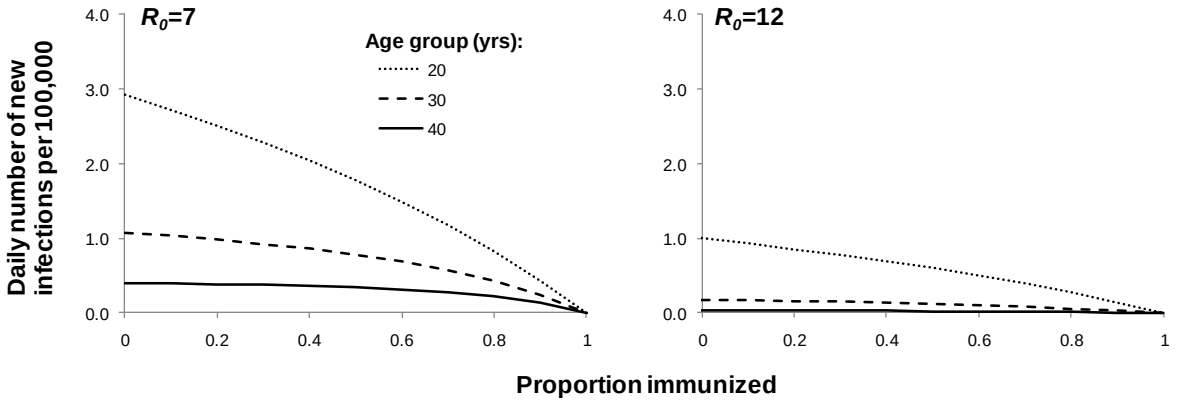


Figure O.5.5: Parameter plot of the (final) daily number of new infections per 100,000 among those aged 20, 30 and 40 years obtained using Model 5.6 for different values of the proportion of 13 year olds that are immunized.

Exercises for Model 6.2

1. You should notice that high values for R_0 are associated with short durations of outbreaks and large outbreak sizes.

The following are the solutions to the questions in section 6.2.3 of this document, which relate to model 6.2.

3. a) Figure O.6.1 summarizes predictions of the number of cases in each time step obtained using 50 runs, assuming that R_0=2. Based on this plot, it is difficult to predict when the peak number of cases is most likely to occur – it seems plausible that it occurs sometime during time steps (or generations) 1-4.

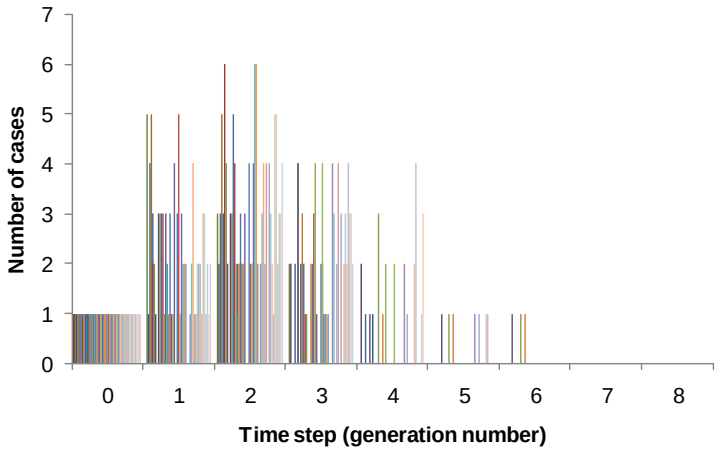


Figure O.6.1: Summary of the number of cases predicted in each time step, obtained using 50 runs of model 6.2, assuming that R_0=2. Each bar for a given time step reflects the output for a single model run.

3b and c) Figure O.6.2 summarizes the average and 95% range in number of cases predicted in each time step for 50 and 2000 model runs, assuming that R_0=2. There is some suggestion that the peak is most likely to occur on the second time step, since the average number of cases for this time step is slightly higher than for the other time steps. However, the 95% range of the predictions overlap with those for most of the other time steps. The predictions are similar when either 50 or 2000 runs were used, suggesting that the variation in the predictions obtained using 50 runs is not just due to the relatively small number of runs.

3d) In about 64% of the model runs, 8-10 cases occurred in the outbreak (at least when 500 or more runs were considered), suggesting that you are most likely to see 8-10 cases in the outbreak. However, for about 10% of the runs, no cases occurred at all, showing that it is possible that no outbreak will occur at all.

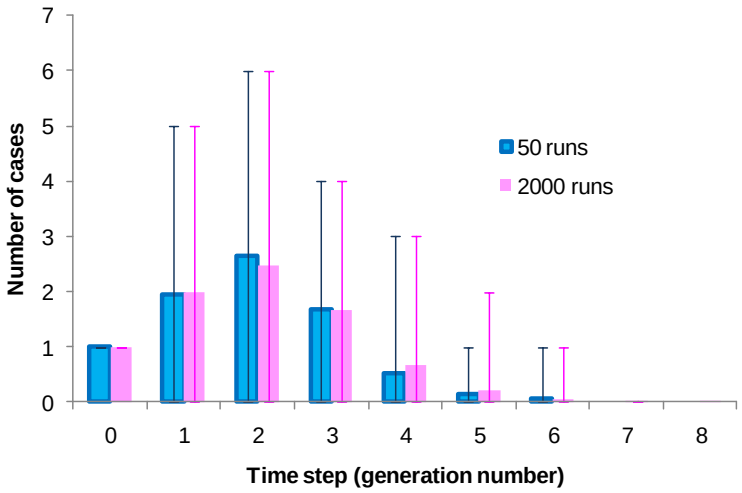


Figure O.6.2: Summary of the average and 95% range in number of cases predicted in each time step in 50 and 2000 runs of model 6.2, assuming that R_0=2.

Exercises for Model 6.3

2 a) You should notice that as you increase the population size, the peak of the epidemic becomes increasingly delayed. This follows from the fact that (for a given value for R_0) if the population is large, it takes longer to deplete sufficient numbers of susceptible individuals to the threshold required for the incidence to decrease than when the population is small. These effects are also predicted by the deterministic model (see e.g. Model 2.1).

You should also notice that if the population is large, then any outbreak predicted by the model is identical to that predicted by a deterministic model and the proportion of the population that has been infected by the end is identical to that predicted by the final size equation (see equation 4.21 or Figure 4.12 in the book).

b) From your observations in part a), you would expect that for large values of the population size, most of the model runs will lead to similar values for the proportion of individuals that are infected by the end. You would also expect that, by chance, no outbreaks will occur at all for a few of the runs considering a large population. This is illustrated in Figure O.6.3 and Figure O.6.4.

3. As shown in Figure O.6.5, the greatest discrepancy between the risk of infection calculated using the Reed-Frost equation and that calculated using the equation $\lambda_t = \beta I_t$ (where $p \approx \beta$) for Model 6.3 occurs when the population size is small (10 individuals), when the values obtained using the two equations differ by up to about 50%. The discrepancy decreases as the population size increases up to a certain level. For the assumptions used in the model (R_0=2), the maximum discrepancy is similar for populations comprising 10,000 or more individuals. For these populations, the risk of infection calculated using these two equations differs by up to about 16%.

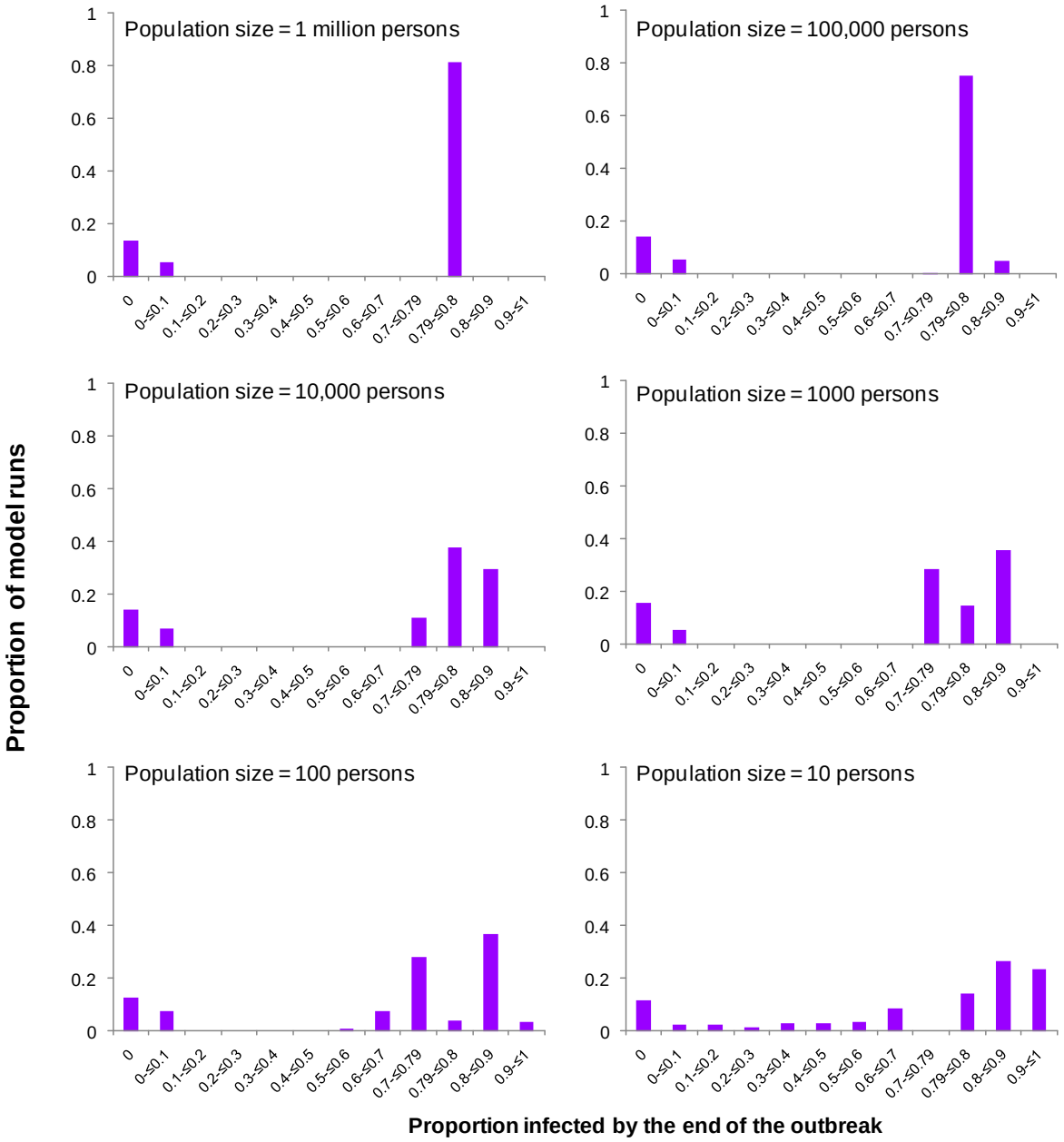


Figure O.6.3: Summary of the proportion of the simulations of Model 6.3, for which different proportions of individuals were infected by the end of the outbreak in populations comprising between 10 and 1 million persons. R_0 as taken to equal 2 and 1 infectious person was introduced into the population at the start. The results are based on 500 runs of the model.

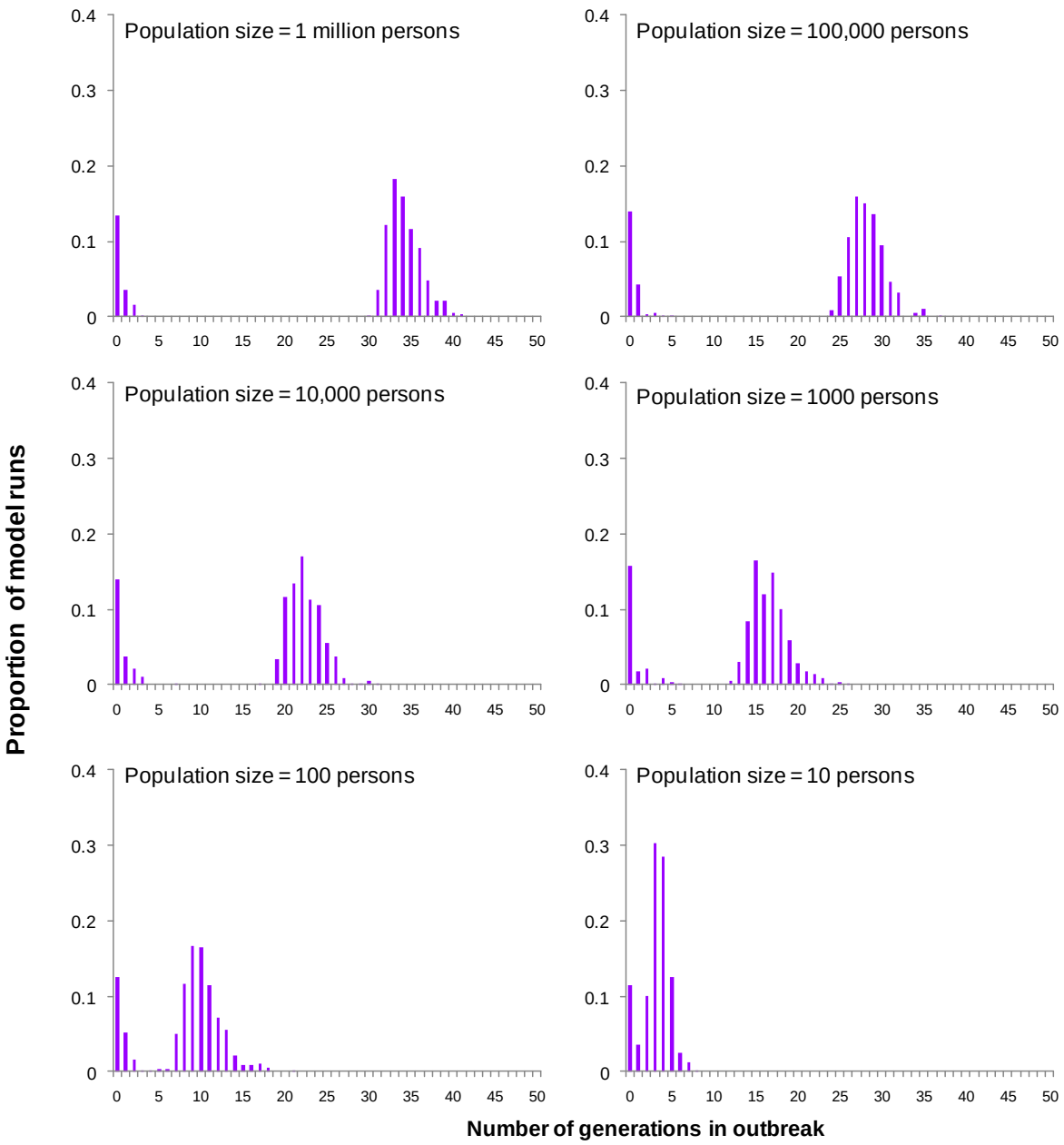


Figure O.6.4: Summary of the proportion of the simulations of Model 6.3, for which different numbers of generations of cases occurred in a population comprising between 10 and 1 million persons. R_0 as taken to equal 2 and 1 infectious person was introduced into the population at the start. The results are based on 500 runs of the model.

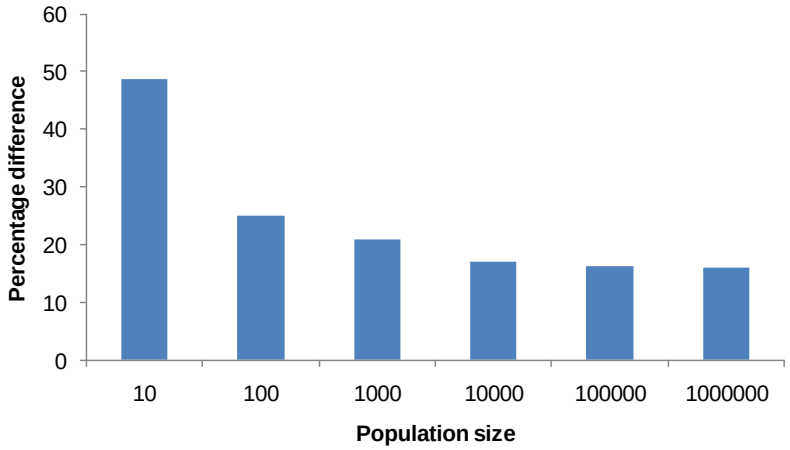


Figure O.6.5: Summary of the range in the percentage difference between the risk of infection obtained using the Reed-Frost equation and the equation $\lambda_t = \beta I_t$ (where $p \approx \beta$), obtained using 50 model runs for values of the population size ranging between 10 and 1 million. R_0 was taken to equal 2 and 1 infectious person was introduced into the population at the start.

Exercises for Model 6.4

2. Models set up using method 3 are slightly more realistic than those set up using method 2, since they describe transitions using variable time steps and individuals do not all become infected or develop disease or become immune at the same time (which is consistent with reality). The main disadvantage of models set up using method 3 is that predicting the course of each outbreak (for example) requires more computations than do models set up using method 2.

Exercises for Model 7.1

1. As shown in Table O.7.1, the best-fitting values obtained by fitting predictions from Model 7.1 to the observed data from the 1957 (Asian) influenza pandemic depended both on the starting set of parameter values and the goodness of fit statistic used. However, in general, the best-fitting values for the contact parameters and the proportion of infectious persons who were reported were similar for all scenarios. There was some variation in the best-fitting values for the numbers of infectious young and old persons, suggesting that these parameters cannot be estimated reliably. As shown in Figure O.7.1, the best-fitting predictions obtained using the different approaches are similar.

Table O.7.1: Summary of the best-fitting values for the contact parameters (b1 and b2), the proportion of infectious persons that were reported (frac_rep) and the initial numbers of infectious young and old persons obtained by fitting predictions from Model 7.1 to data from the 1957 (Asian) influenza pandemic shown in Figure 7.3 using Berkeley Madonna and different goodness of fit statistics.

Goodness of fit statistic used in the fitting	**Starting set of parameter values**	**Parameter**					**Best-fitting value for goodness of fit statistic**
		b1 (per day)	**b2 (per day)**	**frac_rep**	**Infous_y0**	**Infous_o0**	
SSQ	A	3.38×10^{-4}	7.14×10^{-5}	0.49	0.74	34.78	16997
	B	3.39×10^{-4}	6.94×10^{-5}	0.50	6.94	0.02	17620
	C	3.38×10^{-4}	7.14×10^{-5}	0.49	0.74	34.78	16997
Minus_llhood	A	3.23×10^{-4}	7.46×10^{-5}	0.49	6.85	5.90	-7275
	B	3.23×10^{-4}	7.46×10^{-5}	0.49	6.85	5.90	-7275
	C	3.23×10^{-4}	7.46×10^{-5}	0.49	6.85	5.90	-7275
Berkeley Madonna's curve-fit statistic	A	3.34E-04	7.27×10^{-5}	0.48	0.00	29.68	53
	B	3.55E-04	7.39×10^{-5}	0.46	0.00	19.78	56
	C	3.45E-04	7.37×10^{-5}	0.47	1.48	15.35	57

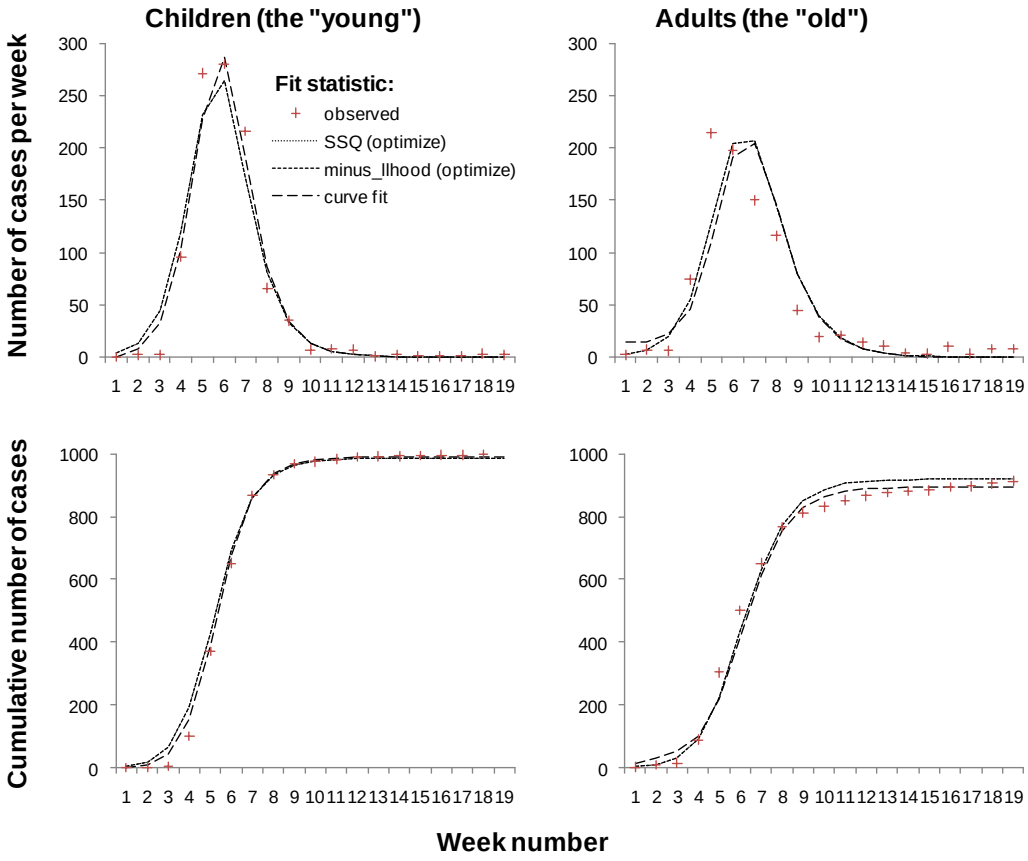


Figure O.7.1: Comparison between best-fitting model predictions of the numbers of cases per week and the cumulative numbers of reported cases against the data observed for the GP practice in Wales from the 1957 (Asian) influenza pandemic, which are plotted in Figure 7.3 of the book, after aggregating the data into two age groups. The different lines reflect the best-fitting predictions obtained either by using the optimize function in Berkeley Madonna (applied to ssq and minus_llhood) or by using Berkeley Madonna's curve fit statistics.

2. As shown in Figure O.7.2 and Table O.7.2, the smallest numbers of reported cases overall in the population are predicted for the assumption that all the vaccine doses are distributed to young, whereas vaccination is predicted to have the smallest impact if the doses are distributed just to the old. The small impact in the population of vaccinating only adults follows from the assumption that the model assumes that there is relatively little contact both between adults and other adults, and between adults and children. Therefore vaccinating adults has little impact on reducing the amount of transmission in either adults or children.

In contrast, the comparatively large reduction in the overall number of reported cases predicted assuming that the vaccine doses are distributed among children follows from the fact that children are assumed to have a large amount of contact with other children. Therefore vaccinating children leads to substantial reductions in the risk of infection for other children and therefore in the total numbers of cases reported in the population.

Notice that the relative sizes of the total numbers of cases reported for each vaccination strategy are consistent with the relative sizes of the reproduction number estimated for these scenarios (see solutions to the exercise 7.5 in the book and exercise 3 for model 7.6). For

example, the largest reproduction number is associated with the strategy of vaccinating only adults, the second largest reproduction number is associated with that of vaccinating an equal proportion of children and adults etc.

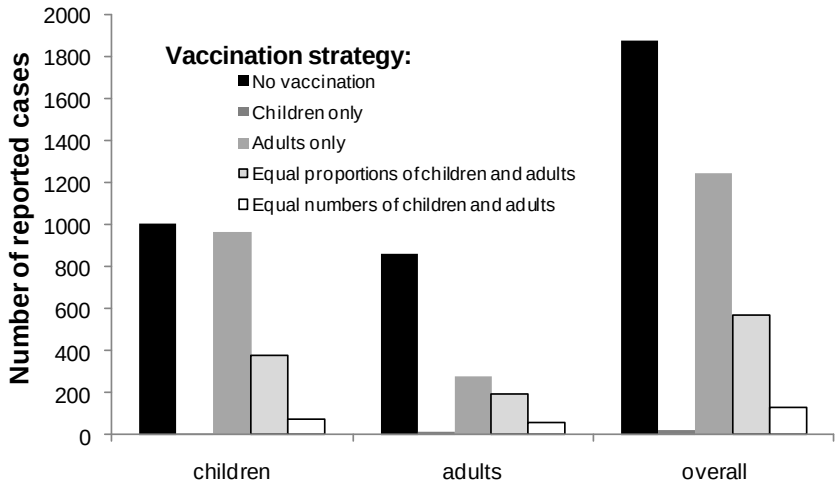


Figure O.7.2: Summary of predictions of the total numbers of cases reported among children (the "young"), adults (the "old") and overall in the population obtained using Model 7.1 for different assumptions about how 2500 vaccine doses are distributed. The predictions are obtained under extreme assumptions about the vaccine efficacy (100%), and therefore reflect the minimum numbers of cases reported for these scenarios.

Table O.7.2: Summary of the vaccination coverage required for 2500 vaccine doses to be distributed according to the various options presented in question 3 for model 7.1, together with the total numbers of cases reported among children (the "young), adults (the "old") and overall in the population. The predictions are obtained under extreme assumptions about the vaccine efficacy (100%), and therefore reflect the minimum numbers of cases reported for these scenarios.

Individuals targeted	**% vaccinated**[+]		**Number of reported cases among:**		
	Children	**Adults**	**Children**	**Adults**	**Total**
No vaccination	0%	0%	1008	866	1874
i) Children only	94.7% (=100×2500/2639)	0%	4	18	22
ii) Adults only	0%	46.6% (=100×2500/5361)	964	282	1246
iii) Same proportion of children and adults*	31.25% (=100× 2500/(2639+5361)	31.25% (=100× 2500/(2639+5361)	376	192	568
iv) Equal numbers of children and adults	47.4% (=100×1250/2639)	23.3% (=100×1250/5361)	75	58	133

[+] These are calculated using the numbers presented in the solutions to question 7.5 of the book

*The proportion of children and adults that need to be targeted with this strategy equals the number of vaccine doses available ÷ population size = 2500/(2639+5361)=31.25%

Exercises for Model 7.2

1. a) You would expect the number of persons in each age stratum to equal the population size divided by the number of age strata. This should equal 55,000,000/75 = 733,333.

b-d) Table O.7.3 shows the average force of infection together with the number of infectious young and old persons that you should have obtained. The values were can be obtained by taking the arithmetic mean of the values for the last 10 years of the model runs (for those interested, the output was exported to Excel, where the average was calculated). Due to the variability in the numbers during the course of a single year, you would have obtained slightly different values if you had simply read off the values approximately halfway between the minimum and the maximum in a given year.

You should notice that the values obtained for the numbers of infectious persons are generally consistent with the values presented in the book.

Table O.7.3: Predictions of the average daily and annual force of infection and the numbers of infectious young and old persons obtained using Model 7.2 in the absence of vaccination.

		Matrix R1	**Matrix R2**
Average daily force of infection	Young	3.54×10^{-4}	3.56×10^{-4}
	Old	1.10×10^{-4}	1.06×10^{-4}
Annual force of infection*	Young	0.13	0.13
	Old	0.04	0.04
Number of infectious persons	Young	18924	18924
	Old	2880	2881

* calculated as the average daily force of infection × 365

2 b) With 50% effective vaccination coverage among newborns, you should find that the average force of infection in the long-term for matrices R1 and R2 is as follows:

	Daily force of infection	
	Matrix R1	**Matrix R2**
Young	1.10×10^{-4}	1.10×10^{-4}
Old	4.53×10^{-5}	6.04×10^{-5}

i.e. the force of infection in the long-term among children is predicted to be similar if persons are assumed to mix according to matrix R1. However, the force of infection among adults is predicted to be higher if persons are assumed to mix according to matrix R2, than if they are assumed to mix according to matrix R1.

b) The slightly higher force of infection among adults that is predicted if individuals are assumed to mix according to matrix R2 than if they mix according to matrix R1 follows from the fact that the β parameter describing contact between the young and the old in matrix R2 (β_{yo} and β_{oy}) is lower than that for matrix R1. Therefore the amount of transmission (and therefore vaccination) among children has a smaller effect on the rate at which adults are infected if individuals in the population are assumed to mix according to matrix R2 than if they are assumed to mix according to matrix R1.

c) As discussed in chapter 5, the higher the value for the force of the infection, the smaller the value for the proportion of unvaccinated individuals who reach a given age still susceptible. In question 2a) we found that the long-term average force of infection among the old following the introduction of vaccination among newborns was higher for the population mixing according to matrix R2 than it was for the population mixing according to matrix R1. You would therefore expect the proportion of 20 year olds who are susceptible in the population mixing according to matrix R1 to be higher than the corresponding proportion for the population mixing according to matrix R2.

d) As shown in Figure O.7.3, if it is assumed that individuals mix non-randomly, then if vaccination is introduced among newborns at a coverage of below the herd immunity threshold, the overall proportion of the population that is susceptible is predicted to be higher after the introduction of vaccination than before vaccination is introduced.

This is intuitively reasonable. For example, the reduction in the force of infection as a result of vaccination (and therefore the proportion of persons in specific age groups who are susceptible), depends on the assumed mixing patterns (see answers to questions 2b and c). This means that the overall proportion of the population that is susceptible must also depend on the assumed mixing patterns.

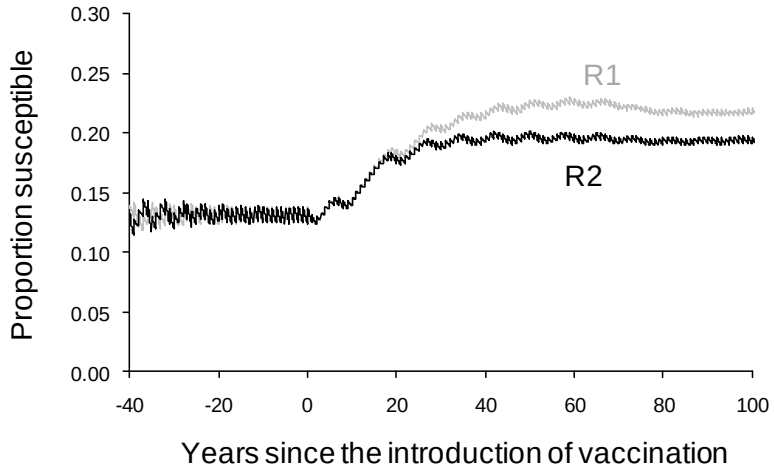


Figure O.7.3: Predictions of the overall proportion of the population that is susceptible following the introduction of vaccination of newborns in year, with an effective coverage of 50%, obtained using Model 7.2, and assuming that individuals mix according to either WAIFW matrix R1 or WAIFW matrix R2 (see section 7.4.2.1.1 of the book).

3. b) The herd immunity threshold associated with matrix R1 is about 1-1/3.5≈0.714. The herd immunity threshold for matrix R2 is about 1-1/4.75≈0.789.

To identify whether or not transmission persists following the introduction of vaccination at some level of coverage, you can click on the Table button of the figure showing predictions of the numbers of new infections per 100,000 per day. Once this number drops to zero, then transmission in the model has stopped. As long as this number is above zero (albeit some very small number), then transmission is still continuing in the model. You may need to extend the stoptime in the model to identify when transmission does actually cease.

Considering the population in which individuals mix according to matrix R1, you should find that if you include an effective vaccination coverage of 71% among newborns, then transmission continues and an epidemic occurs about 247 years after the introduction of vaccination. However, if you include an effective vaccination coverage of 72%, then transmission stops eventually by day 77090 (or 111 years after the introduction of vaccination). This shows that the herd immunity threshold for mixing pattern R1 is between 71% and 72%.

Considering matrix R2, you should find that if you include an effective vaccination coverage of 78% among newborns, then transmission continues and an epidemic occurs about 72 years after the introduction of vaccination. However, if you include an effective vaccination coverage of 79%, then transmission stops eventually by day 644960 (or year 1667 after the introduction of vaccination). This shows that the herd immunity threshold is between 78% and 79% for mixing pattern R2.

4 a & b) Figure O.7.4 compares the predictions that you should have obtained using the parameter plot for matrices R1 and R2 against the predictions obtained assuming that individuals mix randomly. In general, the predicted daily number of new infections per 100,000 among adults is much lower if it is assumed that individuals contact each other according to matrix R1 or R2 than if it assumed that they mix randomly. This is intuitively reasonable, given that the force of infection among adults which is associated with contact patterns R1 and R2 is much lower than that associated with the assumption that individuals mix randomly.

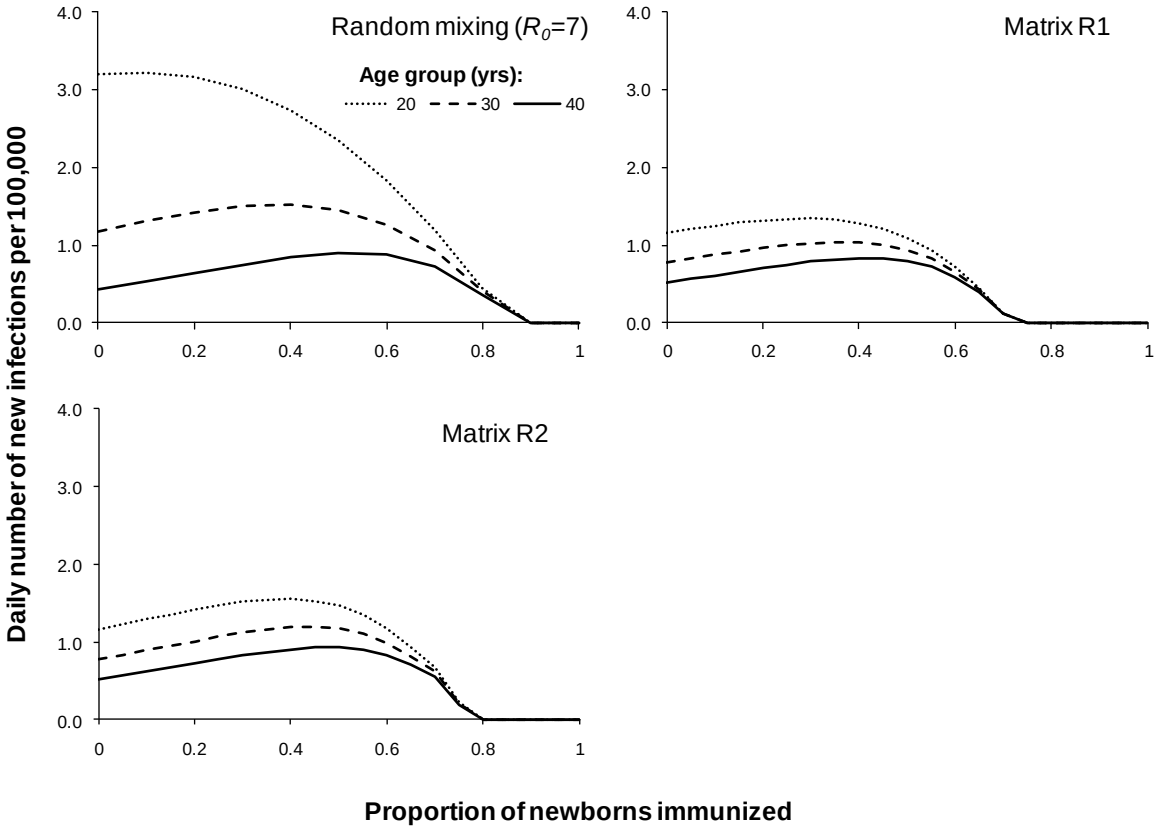


Figure O.7.4: Predictions of the final daily number of new infections per 100,000 at different ages obtained using a parameter plot by Model 7.2 for different values for the effective vaccination coverage among newborns introduced in year 100 in the model. The stoptime was taken to be 400,000 days; 21 runs were used in the parameter plot.

In general, vaccination with a given level of coverage appears to lead to similar relative increases in the daily number of new infections per 100,000 for the assumptions that individuals mix randomly or that they contact each other according to matrix R1 or R2.

Exercises for Model 7.4

1. b) If R0_est and and *x* are set to equal 1.5 and 0.5 respectively then you should find that you get a value for R_0 of 1.08 and a value for *x* of -1.24. These values are implausible since it is unacceptable for the proportion of the infectious person that is infectious to be negative.

2b) If R0_est is set to equal 2, then Solver finds a value for R_0 of 1.08. This highlights the fact there is more than one value of "R_0" for which the determinant of the matrix $\begin{pmatrix} R_{yy}-R_0 & R_{yo} \\ R_{oy} & R_{oo}-R_0 \end{pmatrix}$ equals zero. The value which should be taken to be the "real" R_0 is the largest value for which the determinant equals zero.

c) The advantages of the simultaneous equations and matrix determinant approaches for calculating R_0 is that they require relatively few calculations. However, their disadvantage is that since the basic reproduction number is the largest value for R_0 which satisfies equations 7.37 and 7.38, or for which the determinant of the matrix $\begin{pmatrix} R_{yy}-R_0 & R_{yo} \\ R_{oy} & R_{oo}-R_0 \end{pmatrix}$ is zero, we need to identify all values for "R_0" which satisfy these equations to identify which is the largest. Excel is not yet designed to do this, although some packages, such as Stata, can provide all the values that we are interested in, if we type in an appropriate matrix.

The disadvantage of the simulation approach for calculating R_0 is that it requires more calculations than do the simultaneous equation or matrix determinant approaches. However, as discussed in section B.7.4 of the book, the value obtained almost always converges to the correct value for R_0. Some exceptions to this include matrices which have structures which are similar to the following:

a) $\begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$. For this matrix, if we use the simulation approach to estimate R_0, and start the simulation with persons in only one of the groups, then the ratio of the number of secondary persons in successive generations converges either just to 4 or to 1 (i.e. one of the eigenvalues of the matrix).

b) $\begin{pmatrix} 0 & 1 \\ 4 & 0 \end{pmatrix}$. For this matrix, using the simulation approach will result in the ratio of number of secondary infectious persons in successive generations oscillating between 1.6 and 2.5. For this matrix, R_0 is then the square root of the product of these numbers i.e. $\sqrt{1.6 \times 2.5} = \sqrt{4} = 2$, which would be correctly identified using the matrix determinant or simultaneous equations approach. Further discussion of these matrices can be found in work by Dietz[1] and Heesterbeek and Diekmann[2-4].

It is possible to obtain an intuitive explanation for why we need to take the square root of the product of 1.6 and 2.5 to obtain R_0 by considering the kinds of infections which might be described using matrices with the above structure. The following is an argument adapted from Dietz[1] and Heesterbeek and Diekmann[2-4].

A matrix with the above structure might be used to describe transmission of a parasite between a vector (e.g. mosquito) and a human for which transmission only occurs between the mosquito and the human and there is no direct human-human or vector-vector transmission. Consequently to calculate the number of secondary infectious humans resulting from one infectious human, we need to know how many infectious vectors result from each human, and how many infectious humans result from each vector. The product of the two numbers gives us the number that we are interested in, i.e. the number of secondary infectious humans resulting from each infectious human. The square root of this number gives us the average number of secondary infectious persons or vectors per generation. The latter number is R_0.

Reference List

1. Dietz K. The estimation of the basic reproduction number for infectious diseases. Statistical Methods in Medical Research 1993; 2:23-41.

2. Diekmann O, Heesterbeek JA, Metz JA. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. J Math Biol 1990; 28(4):365-382.

3. Heesterbeek JA. R_0. Amsterdam: Centrum voor Wiskunde en Informatica; 1992. PhD thesis.

4. Diekmann O, Heesterbeek JA. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. Chichester: John Wiley; 2000.

Exercises for Model 7.5

3. You should find that as you increase the value for α, the value for R_0 increases slightly. For example, if α=2, R_0=11.35, whereas if α=1.5, R_0=9.9.

Notice that simply increasing the value for α only affects the amount of contact between those aged 10-14 years and doesn't affect the parameters describing contact between other age groups. Since 10-14 year olds make up a small proportion of the population, increasing the value leads to only small increases in R_0.

4d and e) As discussed in question 7.4 in the book, you should find that as the value for α (and therefore the amount of contact between 10-14 year olds) increases, the value for the net reproduction number increases as follows.

α	1	1.25	1.5	1.75	2
R_n	0.96	0.99	1.04	1.11	1.2

These values are generally consistent with those in Figure 7.17.

The fact that the net reproduction number was very close to 1 in 1994/5 for values for α>1, and for different WAIFW matrices, suggests that, as concluded by Gay et al, there was a potential for an epidemic to occur in 1994/5 in England.

Exercises for Model 7.6

4. The solution to the first part of the question is identical to that provided for exercise 7.5 in the book (reproduced below, for convenience).

The following table summarizes the number of susceptible individuals for each vaccination scenario, the Next Generation Matrix and the values for the reproduction number:

Individuals targeted	**Number of susceptible**		**Next Generation Matrix**	**Reproduction number**
	Children (S_y)	Adults (S_o)		
No vaccination	2639	5361	$\begin{pmatrix} 1.784 & 0.188 \\ 0.383 & 0.766 \end{pmatrix}$	1.85 ($=R_0$)
Children only	139 (=2639-2500)	5361 (=5361-0)	$\begin{pmatrix} 0.094 & 0.010 \\ 0.383 & 0.766 \end{pmatrix}$	0.77
Adults only	2639 (=2639-0)	2861 (=5361-2500)	$\begin{pmatrix} 1.784 & 0.188 \\ 0.204 & 0.409 \end{pmatrix}$	1.81
Same proportion of children and adults*	1814 (=2639×(1-0.3125))	3686 (=5361×(1-0.3125))	$\begin{pmatrix} 1.226 & 0.130 \\ 0.263 & 0.526 \end{pmatrix}$	1.27
Equal numbers of children and adults	1389 (=2639-1250)	4111 (=5361-1250)	$\begin{pmatrix} 0.939 & 0.099 \\ 0.294 & 0.587 \end{pmatrix}$	1.01

*The proportion of children and adults that need to be targeted with this strategy equals the number of vaccine doses available ÷ population size = 2500/(2639+5361)=31.25%

The smallest value for the reproduction number is associated with the strategy of vaccinating only children, which suggests that, of the four strategies, this approach may be the best way of distributing the vaccine stocks. However, we would also need to account for the severity of influenza and the mortality rates in other age groups before making the final decision about which vaccination strategy should be adopted.

There is still potential for an epidemic to occur for each of these vaccination scenarios, since the reproduction number is very close to 1. Note that these calculations assume that the vaccine efficacy is 100%, which is unrealistic. Therefore, if this were to be a real situation, the true reproduction number would have been greater than the values calculated above.