

Spread of Disease

Dr Julia Collins and Nadia Abdelal

Introduction	4
History	4
The Basic Exponential Model	7
The Reproduction Number R_0	8
The SI model	9
The SIR model	11
Graphing the SIR Model	15
Epidemics and vaccinations	21
Conclusion	26
References	26
Solutions to Exercises	27



Spread of Disease

These notes accompany the video on Spread of Disease.

Click on this link to watch the video: **Spread of Disease Video**

Introduction

Epidemics have been some of the most devastating events in human history. The Black Death in medieval Europe wiped out a third of the total population and the Spanish Flu in 1918 killed more people than the First World War. The 2009 pandemic of Swine Flu caused worldwide panic, while outbreaks of the Ebola and Zika viruses are still serious problems in the world today.

Mathematics and statistics are at the forefront of how scientists, hospitals and governments deal with the issues of epidemics and the spread of disease. The subject of epidemiology models the behaviour of diseases and the steps we can take to control their impact, including issues like when to vaccinate and quarantine people.

One of the most important tools in epidemiology is the differential equation. These equations model how a disease behaves over time, taking into account factors such as the number of susceptible people in the population and the number of people who are immune. These notes focus on exploring how differential equations can accurately model different diseases, and the lessons they teach us about the importance of vaccination and herd immunity.

History

The development of infectious diseases

In the early stages of human history, civilisations were essentially comprised of nomadic tribes. These were not large enough nor settled in any place long enough to sustain the spread of microbial infection. However, with the beginning of the agricultural revolution around 10 000 years ago and the establishment of permanent settlements, people began living in closer proximity to one another as well as to animals. This facilitated the spread

of bacteria and viruses not only between humans, but also between cattle and humans.

An **epidemic** happens when a disease spreads between large numbers of people in a short period of time. When an epidemic goes global, it is called a **pandemic**.

Although bacteria has been linked to a number of epidemics (including the bubonic plague, or 'Black Death', and tuberculosis), the most common cause of epidemics is viruses. Some of the worst pandemics in the world include viruses such as measles, influenza and HIV. The influenza pandemic (Spanish Flu) of 1918 – 1919 is considered one of the greatest natural disasters of all time, infecting half a billion people and killing at least 50 million people worldwide. In Australia alone, the disease killed at least 12 000 people with another 2 million infected.

The improvement of living conditions such as clean water, sewerage systems, better food, and personal hygiene methods (such as hand washing) has significantly reduced the incidence and impact of epidemics. Interventions such as antibiotics, anti-viral drugs and vaccination programs have also done much to prevent the spread of disease, with some diseases, like smallpox, being eradicated altogether.

However, many diseases such as HIV, influenza, hepatitis and measles continue to have a serious impact. There has recently been an increase of reported cases of measles in Australia, calling into question the vaccination rates among the Australian population and highlighting the importance of inoculation and herd immunity among Australian communities.

The beginnings of data collection and analysis

It is estimated that roughly 25 million people died in Europe as a result of the Black Death in about 1350, but it is hard to be sure because there was no systematic way of recording deaths and no censuses to record the initial size of the population. The story was different by the time of the Great Plague of London in 1665.

John Graunt (1620 – 1674) is considered by many to be the world's first statistician. His new approach to the collection and analysis of data became the foundation of modern approaches to demography (the study of changing human populations). In 1662, Graunt published the first-known mortality tables, which combined all the data he could find on the numbers and causes of deaths. Not only does this data help modern historians understand the impact of events such as the plague of 1665 and the Great Fire of London in 1666, it also gave local governments an idea about the size of the population



CAPTAIN JOHN GRAUNT

of London and how quickly it was expanding. This enabled them to build the right infrastructure to support the growing population.

This kind of data collection remains an essential tool for today's epidemiologists. In order to model how a disease will spread and how dangerous it will be, it is necessary to see real statistics of the number of people who become infected during a disease outbreak and how many of those eventually recover or die. We also need to keep statistics of uninfected people, to know what would be 'normal' if the disease were not present in the population.

How smallpox led to epidemiology and vaccines

Smallpox was an infectious disease whose earliest appearance was found on the body of the ancient Egyptian mummy Ramses V. By the 18th century, it had become responsible for about 20 per cent of deaths in Europe. People eventually noticed that individuals who had been sick with the less dangerous disease of cowpox seemed to be immune from getting smallpox. This led to the English doctor Edward Jenner developing the world's first systematic vaccination program in 1798, saving the lives of tens of thousands of people.

Smallpox was also the subject of the first serious mathematical work on epidemiology. This was written by Daniel Bernoulli in 1766 and made use of calculus, which was a relatively new mathematical subject at the time and had been developed by Isaac Newton and Gottfried Wilhelm Leibniz. Bernoulli's approach was very similar to that taken by mathematicians today, using a system of differential equations to model how the number of infectious and immune people in a population varied over time, based on parameters such as the birth rate, death rate and recovery rate. Using statistics such as those compiled by Graunt, Bernoulli crunched the numbers in his equations and concluded that if the entire population could be vaccinated at birth, then life expectancy would increase by more than three years. This kind of analysis paved the way for governments to invest in vaccination programs like the one developed by Jenner.

In the early 1950s there were still an estimated 50 million deaths worldwide per year due to smallpox, but in 1980 the World Health Organisation announced the total eradication of the disease. The hope is that this powerful combination of biological discoveries, mathematical modelling and public health initiatives can help the world eradicate other diseases.

The Basic Exponential Model

The spread of a contagious disease depends on both the amount of contact between individuals and the chance that an infected person will transmit the disease to someone they meet. If the **transmission risk** of the disease is 100 per cent and each infectious person meets two other people before they recover, the disease will soon begin to spread very quickly. Assuming that recovery takes one day, this situation will result in the number of sick people doubling each day. We can model this situation by the following equation:

$$y = 2^{t-1} \quad (1)$$

where

y = the total number of people infected, and

t = the time, in days, which has elapsed since the initial outbreak.

We can also describe the situation graphically:

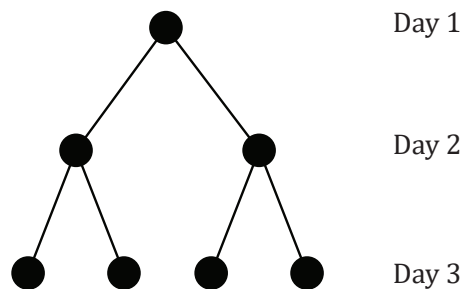


Figure 1: The spread of disease with an exponential growth of $y = 2^{t-1}$

Such a scenario, where the number of infections multiplies by a constant factor each day, is called **exponential growth**. It can be used as a simplistic model of how an infection could potentially spread through a population.

Exercise 1

- Using the situation represented in Figure 1, how many people would be infected at the end of the first week (that is, on Day 7)?
- If the total population of susceptible people were 2 500 000, how long would it take before the entire population became infected?

- c In the above example, the disease had a 100 per cent transmission risk. If the transmission risk were only 50 per cent, that is, each infected person only spreads the disease to half the people they come into contact with, how long would it take before the entire population became infected?

The Reproduction Number R_0

The **basic reproduction number** measures how many people an infected individual will transmit the disease to before they recover, and it is given the symbol R_0 . More precisely, it is the number of secondary infections produced by an infected individual in a population that is totally susceptible. This number is important in determining how quickly an infection will spread through a population.

For example, if the value of R_0 for measles is 14, then each case of measles would produce 14 new secondary cases. This would spread through the population much faster than in our previous example, where the value of R_0 was only equal to 2.

The basic reproduction number is affected by several factors:

- The rate of contact between individuals in the host population;
- The probability of the infection being transmitted during contact;
- The duration of infectiousness.

The magnitude of R_0 not only indicates the speed of how a disease will spread, but whether it not it will spread at all. If it is the case that

$$R_0 > 1$$

then the infection will spread throughout the population. But if

$$R_0 < 1$$

then the infection will not be able to take hold and will eventually die out. Generally, the greater the value of R_0 , the harder it is to control an epidemic.

Exercise 2

- a Research the basic reproduction numbers of the following diseases and fill in the table below:

Disease	R_0
Measles	
Ebola	
Influenza	
Smallpox	
Whooping cough	

- b** Why is it that even though some diseases have a higher value for R_0 , they are not as prevalent in certain communities as others are?

The SI model

A more realistic way of modelling the spread of disease is using the **SI model**. This is the simplest one among epidemic models. At each time t (measured in days), we divide the population into the number who are **susceptible**, $S(t)$, and the number who are **infectious**, $I(t)$. We let N be the total population size, which we take to be a constant.

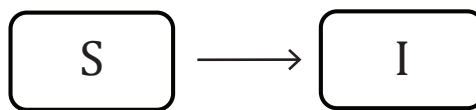


Figure 2: The SI model of the spread of disease.

In our model, we assume that every member of the population is either susceptible or infectious, giving us the equation

$$S(t) + I(t) = N.$$

Differentiating both sides with respect to t , and remembering that N is a constant, we get a relationship between how the numbers of susceptible and infected people are changing over time:

$$\frac{dI}{dt} = -\frac{dS}{dt}.$$

It often makes sense to study the proportions of infected and susceptible people rather than the raw numbers. So we define new variables as follows:

$$i(t) = \frac{I}{N}, \text{ the proportion of the population who are infected, and}$$

$$s(t) = \frac{S}{N}, \text{ the proportion of the population who are susceptible.}$$

Then we know that $i + s = 1$ or, rearranging, that $s = 1 - i$.

In the SI model we assume that any infected people will continue to spread the disease until the end of the epidemic; that is, they never recover. The rate at which the disease spreads will be proportional to s (the more susceptible people there are, the faster the disease can spread) and also to i (the more infected people there are, the faster the disease will spread). This observation leads us to write down our first differential equation:

$$\begin{aligned}\text{Rate of infection} &= \beta \times (\text{proportion of susceptibles}) \times (\text{proportion of infecteds}) \\ &= \beta s(t)i(t), \text{ or } \beta si\end{aligned}$$

Here β is a constant called the **transmission rate** and is the average number of people each infectious person spreads the disease to each day. It can be calculated by multiplying the transmission risk with the average number of contacts per day.

We can rewrite this equation solely in terms of i :

$$\frac{di}{dt} = \beta(1 - i)i. \quad (2)$$

Every person who becomes infected is one less person who is susceptible. In other words, the fraction of susceptible people will go down at the same rate as the fraction of infected people goes up. This gives us a second equation:

$$\frac{ds}{dt} = -\frac{di}{dt} = -\beta si. \quad (3)$$

A way of representing this model graphically is shown below.

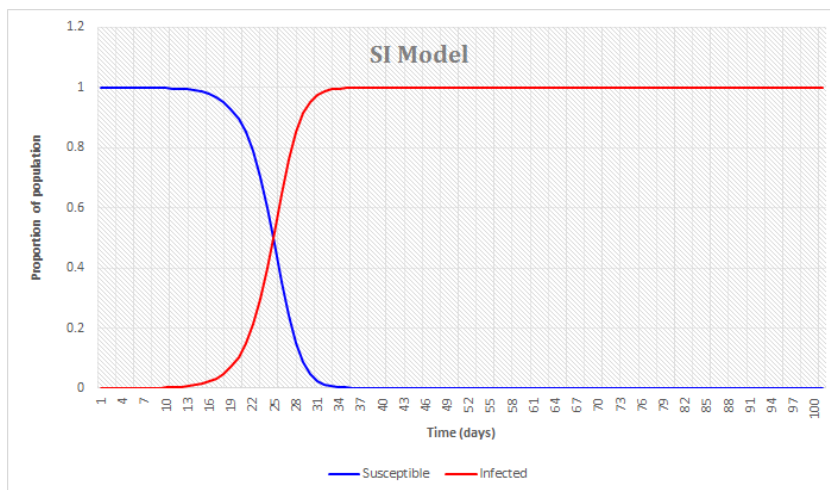


Figure 3: The SI model of the spread of disease, showing the proportion of susceptible (blue) and infected (red) people over 100 days.

Exercise 3

- a Why is the number of susceptible people in Figure 3 dropping?
- b In your own words, explain why $i + s = 1$.
- c Why is the SI model not effective in accurately modelling the spread of disease? What factors would you want to include to make it more accurate?

Exercise 4

Solving the differential equation: If we know from Equation 2 and Equation 3 that

$$\frac{di}{dt} = \beta si = \beta(1 - i)i$$

and that $i = i_0$ at $t = 0$, show that:

$$i(t) = \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}}.$$

The SIR model

The **SIR model** more accurately represents how an infection would spread through a population because it takes into consideration that some people will recover from the disease and no longer be susceptible. This model assumes that people who recover from the infection become immune and cannot become infected a second time.

First we can investigate how the graphs of susceptible and infectious people might be different if we assume that infectious people recover after a fixed amount of time. In Figure 4 below, we look at the evolution of infection *A*, which has a reproduction number R_0 of 3, an infectious duration of 5 days, and an initially infected proportion of 1 per cent of the population.

The **blue line** is the proportion of the population who are susceptible, $s(t)$.

The **red line** is the proportion of the population who are actively infected, $i(t)$.

In Figure 5, we have zoomed in to get a closer look at what is happening over the first five days.

Exercise 5

- a Do the values for i and s still add to 1 in this model of infection *A*? Explain.
- b What happens to infection *A* after Day 14?

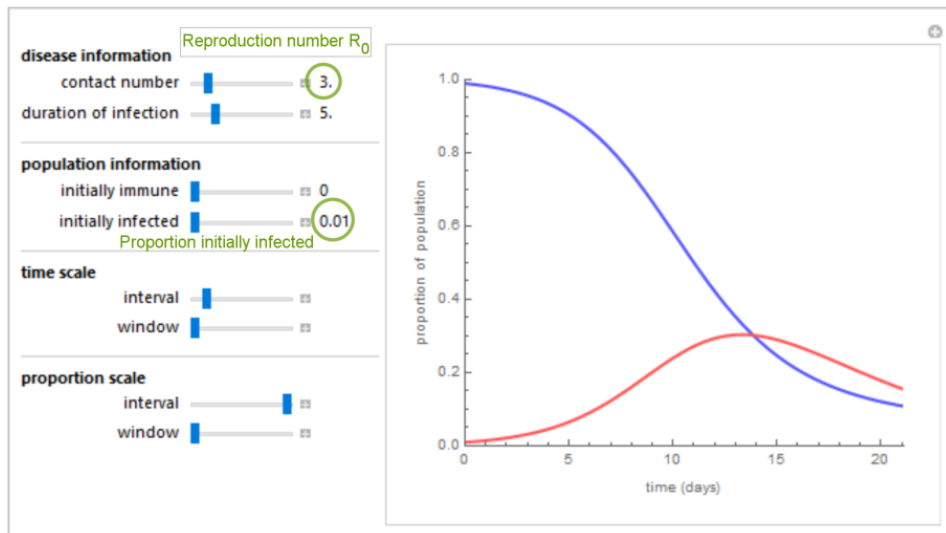


Figure 4: SIR model for infection A

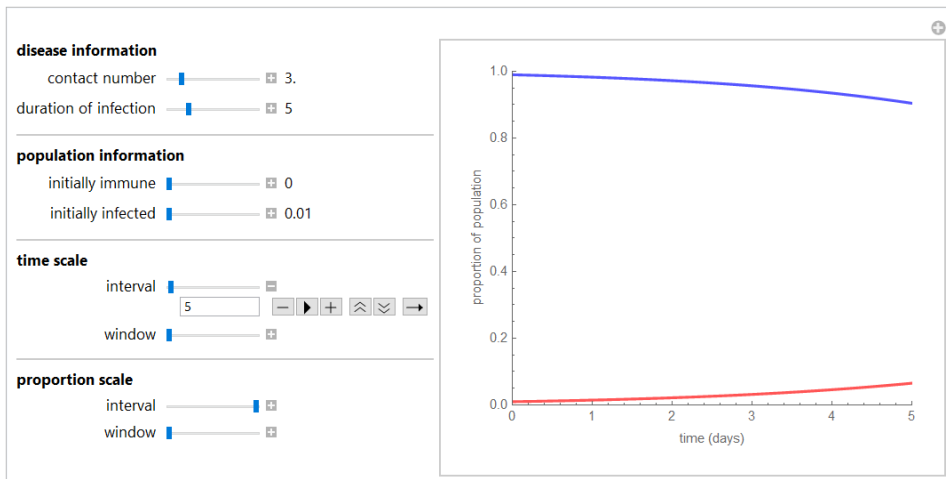


Figure 5: SIR model showing infection A in the first five days

Exercise 6

For each of the graphs given in Figure 6, answer the following questions assuming a population of $N = 50\,000$ people:

- How many people were infected and how many were susceptible at the initial outbreak of the infection?
- How long after the initial outbreak did the number of infected people exceed the number of susceptible people?
- How long did it take for all of the susceptible people to become infected?

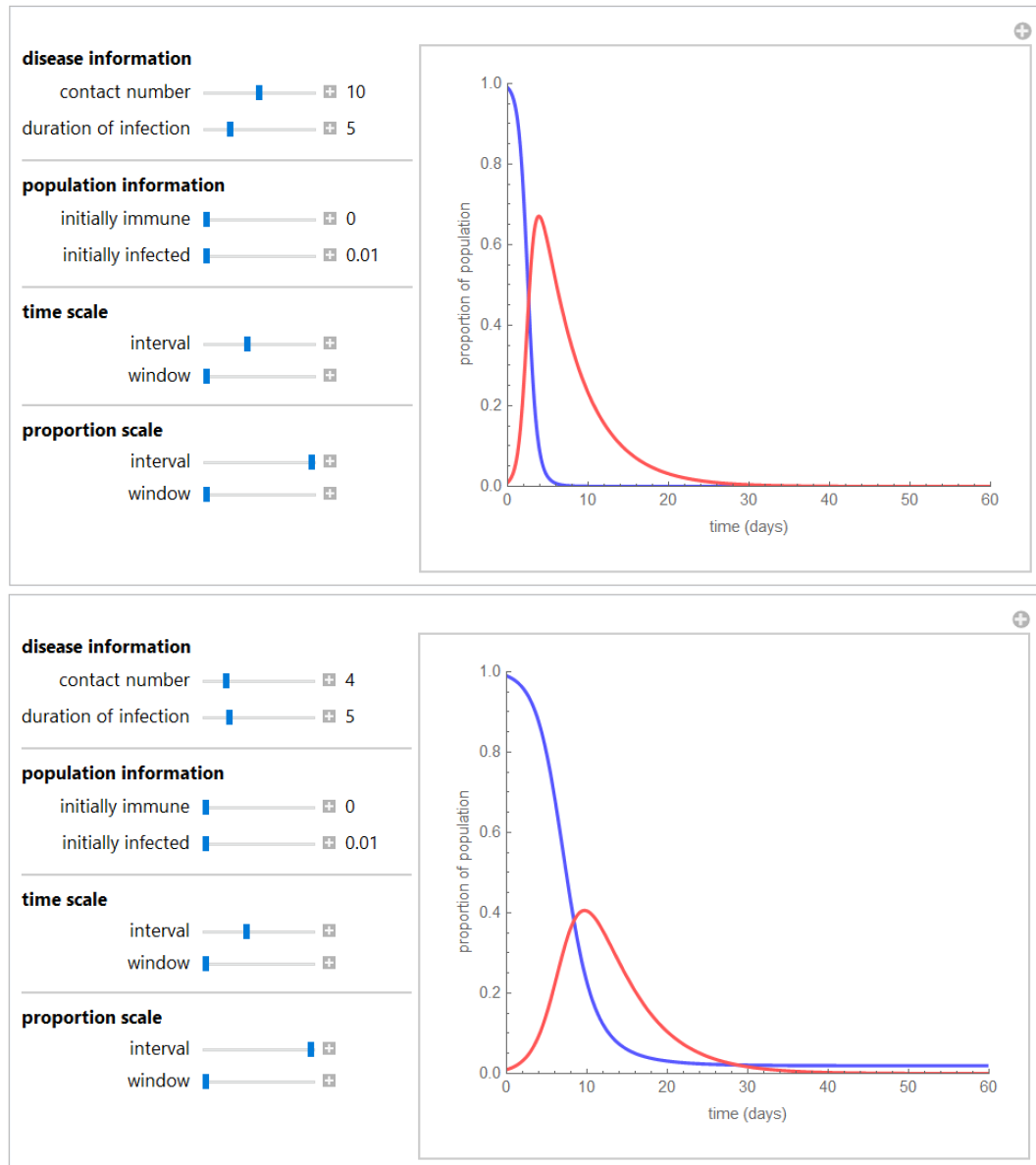


Figure 6: SIR model showing two different infections over the first 60 days

In an SIR model, the population is divided into three types:

- Susceptible (S) (not infected),
- Infectious (I), and
- Recovered (R) (that is, vaccinated or recovered with immunity).



Figure 7: The SIR model of the spread of disease.

We will make three assumptions of how these categories of the population relate to each other:

- 1 The number of infected people increases at a rate proportional to both the number of infected and the number of susceptible people. The number of susceptible people decreases at this same rate. The ratio involved is the **transmission rate** β (beta), the same as in the SI model.
- 2 The number of recovered people increases at a rate proportional to the number of infected people. The ratio involved is called the **recovery rate** γ (gamma).
- 3 A susceptible person who catches the disease becomes infectious immediately.

We model each of S , I and R as functions of a time variable t , which is measured in days. Our variables then become

$S = S(t)$ = number of susceptible individuals at time t

$I = I(t)$ = number of infected individuals at time t

$R = R(t)$ = number of recovered individuals at time t .

Just as in the SI model, it makes more sense to consider the proportions of each type of individual rather than their actual number. If the total population again has size N , the new variables become:

$s(t) = \frac{S(t)}{N}$ = the proportion of susceptible individuals at time t

$i(t) = \frac{I(t)}{N}$ = the proportion of infected individuals at time t

$r(t) = \frac{R(t)}{N}$ = the proportion of susceptible individuals at time t .

Then we can write the equation $s(t) + i(t) + r(t) = 1$.

Exercise 7

- a Based on the assumptions given, how do you think $s(t)$, $i(t)$ and $r(t)$ will vary over time?
- b Sketch what you think each of these functions will look like, doing the three drawings on the same graph.
- c Explain why the functions $s(t)$, $i(t)$ and $r(t)$ sum to one.

The two equations that we created under the SI model can now be rewritten for the SIR model to incorporate the number of recovered individuals, and can be expressed as follows:

$$\frac{ds}{dt} = -\beta si. \quad (4)$$

$$\frac{di}{dt} = \beta si - \gamma i \quad (5)$$

$$\frac{dr}{dt} = \gamma i. \quad (6)$$

Exercise 8

- a What does the negative sign in Equation 4 represent with respect to the number of susceptible individuals over time?
- b Why are there now two terms in Equation 5?
- c Calculate the sum of Equations (4), (5) and (6) and write a brief comment on what it suggests about the population of the epidemic model.

Graphing the SIR Model

The evolution of the equations in the SIR model can be represented graphically.

The following graph shows the development of an infection where the average infectious period is ten days and each infected person spreads the infection to one person every two days. At the beginning of the infection, when $t = 0$, the number of susceptible individuals is $S = 5\,000\,000$, while the number of infected individuals is $I = 10$. We can assume that

no one has recovered from the infection in its initial stage of outbreak, so to start with we have $R = 0$.

The transmission rate β is 0.5 because this is the average number of people an infectious person passes the disease to each day.

The recovery rate γ is 0.1 because if each person is infectious for 10 days then we would expect a tenth of the people to recover each day. The recovery rate is always the inverse of the infectious period.

The **blue line** represents the number of susceptible individuals.

The **red line** represents the number of infected individuals.

The **yellow line** represents the number of recovered individuals.

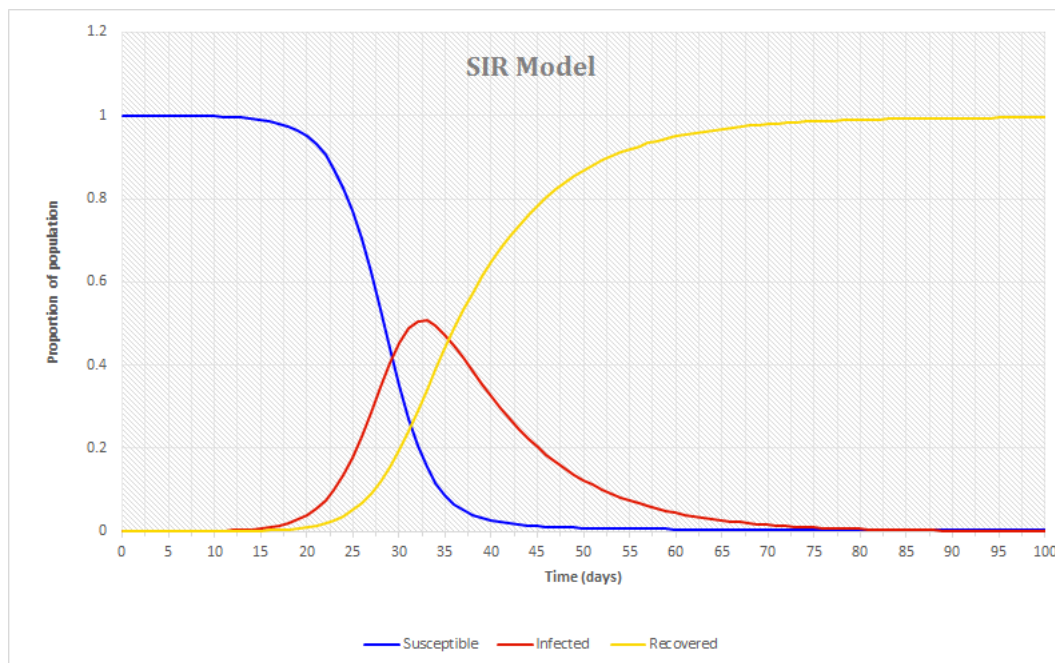


Figure 8: SIR model graph over 100 days with an initial susceptible population of 5 000 000, an initial infected population of 10, an initial recovered population of 0, a transmission rate β of 0.5, and a recovery rate γ of 0.1.

As the proportion of recovered people increases, the proportion of susceptible people decreases, as shown respectively by the yellow and blue lines. The red line represents the trend of infected individuals, and how their numbers change over time. In this example, both the blue and red lines decrease until they hit zero. This means that the entire population has become infected with the disease and moved into the recovered phase. When the blue line does NOT reach zero, this means the disease has died out before everyone in the population has contracted it.

In the SIR model, the basic reproduction number R_0 can be calculated using the ratio of the transmission rate to the recovery rate:

$$R_0 = \frac{\beta}{\gamma}.$$

Exercise 9

Figure 9 below shows the graph of an outbreak of the Severe Acute Respiratory Syndrome (SARS) virus in Hong Kong in 2003. Use the graph and your knowledge of the relationships between $S(t)$, $I(t)$ and $R(t)$ to help you answer the following questions:

- a Complete the following table of values

Percentage of the population who are susceptible at $t = 0$	
Average infection period	
How often an infected person encounters a susceptible person	
Time taken for number of infected individuals to reach zero	

- b Which day has the highest rate of infection? Why doesn't the number of infected individuals continue to increase after this point?
- c Describe what happens to the gradient of the yellow line over the first 45 days.
- d What happens to the blue line after the majority of infection has passed?
- e Why does the blue line never reach zero? What does this say about the number of infected, susceptible and recovered individuals within the population?
- f Complete the following table of values.

	0 days	40 days	60 days
Susceptible Individuals, S			
Infected Individuals, I			
Recovered Individuals, R			
$S + R + I$			

- g What do you notice about the total $S + R + I$? Explain this result.
- h At the height of the epidemic, how many people were infected?
- i Using the graph, can you determine whether the entire population was infected with the virus before it ran its course? Explain why or why not.
- j How many people, if any, escaped infection?

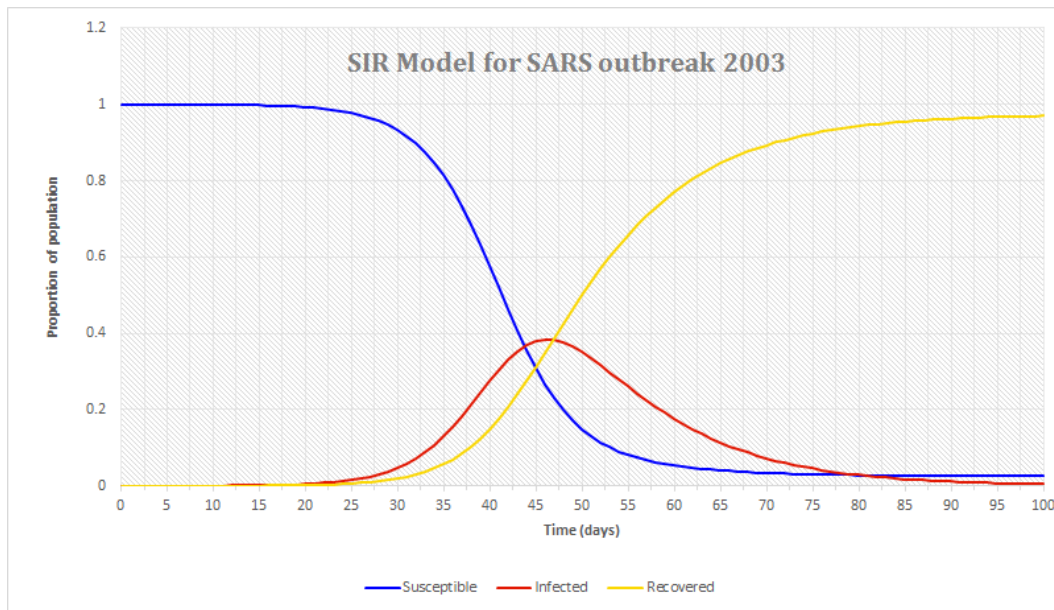


Figure 9: Outbreak of the SARS virus in Hong Kong over 100 days. The initial susceptible population is 7 400 000, the initial infected population is 10, the initial recovered population is 0, the transmission rate β is 0.36, and the recovery rate γ is 0.1.

Euler's Method: A numerical method for solving differential equations

The SIR model is referred to as a **compartment model** since it is useful to refer to people moving from one compartment to another. These compartment models can often be modelled using recursive relationships of the form:

$$\text{New Value} = \text{Old Value} + \text{Gain} - \text{Loss}$$

We can consider a very simplistic situation where we have an outbreak of measles within a population of 50 000 people. When the infection is first recognised, and treatment and analysis is initiated, there are already 2000 people infected and 250 recovered. Assuming that $\beta = 0.00005$ and $\gamma = 0.06$, we can use numerical calculations to see the predictive power of the SIR model as follows.

To begin with, we have

$$S + I + R = 50\,000$$

$$S = 50000 - 2000 - 250 = 47\,750$$

Using equations 4, 5 and 6, but substituting S , I and R for s , i and r , we can approximate the rates of change at the initial stage of identification; that is, at time $t = 0$:

$$\frac{dS}{dt} = -\beta SI = -0.00005(47750)(2000) = -4\,775 \text{ people per day}$$

$$\frac{dI}{dt} = \beta SI - \gamma I = 0.00005(47750)(2000) - (0.06)(2000) = 4\,655 \text{ people per day}$$

$$\frac{dR}{dt} = \gamma I = (0.06)(2000) = 120 \text{ people per day}$$

After a small change in time subsequent to the initial outbreak, we can assume the following:

$$S = 47750 - 4775 = 42\,975 \text{ people}$$

$$I = 2000 + 4655 = 6\,655 \text{ people}$$

$$R = 250 + 120 = 370 \text{ people}$$

We can see here that as the number of infected and recovered people increases, the number of susceptible individuals decreases at the same rate. We now have 42 975 susceptible people, 6 655 infected people and 370 recovered people. These will now represent our new values as we continue with modelling the spread of this particular disease over a period of time.

The process of calculating the values of S , I and R in this way, by putting the results from the last calculation into the next one, is called a **recursion relation**. The calculation we just did can be written using the following equations.

$$S(1) = S(0) + [-\beta S(0)I(0)]$$

$$I(1) = I(0) + [\beta S(0)I(0) - \gamma I(0)]$$

$$R(1) = R(0) + [\gamma I(0)]$$

Here we have used values of S , I and R at time $t = 0$ to calculate the values at time $t = 1$, advancing in a time increment of 1 day. The method would continue by using the values $S(1)$, $I(1)$ and $R(1)$ to find the values for S , I and R at time $t = 2$ and so on. But we can get more accurate results by using a smaller time increment or time step.

Let us choose a time step of h . Then $t_n = t_{n-1} + h$. That is, time t_n is one time step h after time t_{n-1} . We can also think of t_n as $t_0 + nh$. That is, n time steps after the initial time t_0 . Then we can define $S_n = S(t_n)$, $I_n = I(t_n)$ and $R_n = R(t_n)$.

This notation allows us to write our recursion equations as:

$$S_n = S_{n-1} + h[-\beta S_{n-1}I_{n-1}]$$

$$I_n = I_{n-1} + h[\beta S_{n-1}I_{n-1} - \gamma I_{n-1}]$$

$$R_n = R_{n-1} + h[\gamma I_{n-1}]$$

The smaller the time step h , the more accurate the equations will be in modelling the spread of the disease. However, smaller time steps also mean a longer calculation time, so mathematicians must balance speed with accuracy.

Yet another way of writing the three formulas is:

$$S_n = S_{n-1} + h \frac{dS}{dt}(S_{n-1}, I_{n-1}) \quad (7)$$

$$I_n = I_{n-1} + h \frac{dI}{dt}(S_{n-1}, I_{n-1}) \quad (8)$$

$$R_n = R_{n-1} + h \frac{dR}{dt}(S_{n-1}, I_{n-1}). \quad (9)$$

These three equations are called **Euler's formulas** and can be used to generate a series of values that will reflect what is happening with the epidemic over a period of time. To calculate something from these formulas, we must have explicit values for β , γ , S , I , R and h .

A spreadsheet such as Excel can be used to compute the outputs of these equations and show the SIR graphs for any given epidemic. For the purposes of this module, one has already been created for you and will be available on the website. Figures 8 and 9 are both examples of graphs created by this document.

Figure 10 shows a screenshot of the Excel document attached to this module. It shows the variables that are used in the graphing of the SIR equations outlined above. Users can change the interaction rate (cell H6) and the recovery rate (cell H9) to see how these variables affect the progress of the disease. The spreadsheet automatically calculates the new values of β and the basic reproduction number R_0 , and updates the graph instantly.

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3											
4	Time	Susceptible	Infected	Recovered	total		Total	1			
5	0	0.999998	0.000002	0	1		Infection Rate	1	beta=		0.5
6	1	0.999997	2.8E-06	2E-07	1		Interaction Rate	0.5			
7	2	0.9999956	3.91999E-06	4.8E-07	1		Initial Infected	0.000002	R0=		5
8	3	0.99999364	5.48798E-06	8.71999E-07	1		Delta Time	1			
9	4	0.999990896	7.68316E-06	1.4208E-06	1		Recovery Rate	0.1 v			
10	5	0.999987055	1.07564E-05	2.18911E-06	1						

Figure 10: Screenshot of the Excel document for modelling SIR graphs using Euler's method.

The Hong Kong Flu

In early 1968, a new influenza virus was detected in Hong Kong that caused a global outbreak. By the end of that same year it had reached pandemic proportions. The virus

spread rapidly throughout Southeast Asia, the Panama Canal Zone, the United States and the United Kingdom as well as Europe, Australia, Japan, South America and various African countries. It is reported to have caused an estimated four million deaths throughout the world, making it the most significant outbreak of influenza since the 1918 Spanish Flu Pandemic.

If we take New York City in 1968 with a population at the time of 7 900 000, we can get an idea of the effects that the virus would have had on its susceptible population. The new nature of the virus meant that almost everyone in the population was susceptible, which is what caused its devastating initial effects. However, we can assume that there may have been a trace level of the infection already present. We will estimate this to be 10 people.

Exercise 10

- a Using the information above, state the initial values when $t = 0$ for the population variables $S(t)$, $I(t)$ and $R(t)$.
- b If each individual is infectious for an average of three days and we assume that they make contact with a susceptible person every two days, state the values of the parameters β and γ .
- c Using the Excel document provided, create your own graph for the Hong Kong Flu pandemic and write a paragraph explaining its effect on the population of New York City. You may wish to include the following pieces of information:
 - i What was the peak of the epidemic?
 - ii How many days after the initial infection did the epidemic die out?
 - iii Did all susceptible individuals get infected? If not, approximately how many people escaped infection?
 - iv Explain how the evolution of the Hong Kong flu differed from that of the 2003 SARS outbreak.

Epidemics and vaccinations

Previously, we discussed the basic reproduction number R_0 of a disease as a value that is used to measure the transmission potential of a disease. For example, if the value of R_0 for measles in a population were 14, then we would expect it to produce a total of 14 secondary cases for each primary case. However, the calculation of R_0 assumes that the entire population is susceptible to the disease, and this may not always be the case. Some people will be immune due to a prior infection creating life-long immunity, or as a result of vaccination. To consider this, we use the **Effective Reproduction Number** R . This can

be calculated by multiplying R_0 by the fraction of the population who are susceptible, leading to the equation:

$$R = R_0 s \quad (10)$$

where s is the fraction of the host population who are susceptible.

Example

If $R_0 = 14$ for measles and half the population has been immunised, the effective reproduction rate would be:

$$R = 14 \times 0.5 = 7$$

So, each single case of the measles will produce seven secondary cases.

It is extremely important to maintain knowledge of the value of R in order to control and potentially eradicate a disease from a given population.

Herd Immunity

Herd immunity occurs when a critical proportion of a susceptible population is immunised against a contagious disease, giving overall protection to the remainder of the ‘unprotected’ community (herd). This minimises the chance of an outbreak occurring and allows those who are not eligible for vaccines, such as infants and pregnant women, to also receive some protection from the disease. Herd immunity works because it is more difficult for diseases to spread between individuals if large numbers are already immune, as this breaks the chain of infection. A graphical representation of this is illustrated in Figure 11 below.

In Figure 11 we see that when $R_0 = 3$, one infectious person generates 27 new cases of infection after just three generations of the disease. But if two-thirds of the population are immunised then only 1 of those 27 people gets infected after three generations. This shows that far more people in the population are protected from the disease than just those receiving the vaccination.

The **herd immunity threshold** (v) is the proportion of the population that needs to be immune in order for an infection to be contained. This is reached when each new case leads to just a single new case, leading to the infection becoming stable within the population; that is, $R = 1$. (Recall that when $R > 1$ we have an epidemic.)

To find the value of v for a given disease, we start by using Equation 10 to give the equation:

$$R_0 s = 1. \quad (11)$$

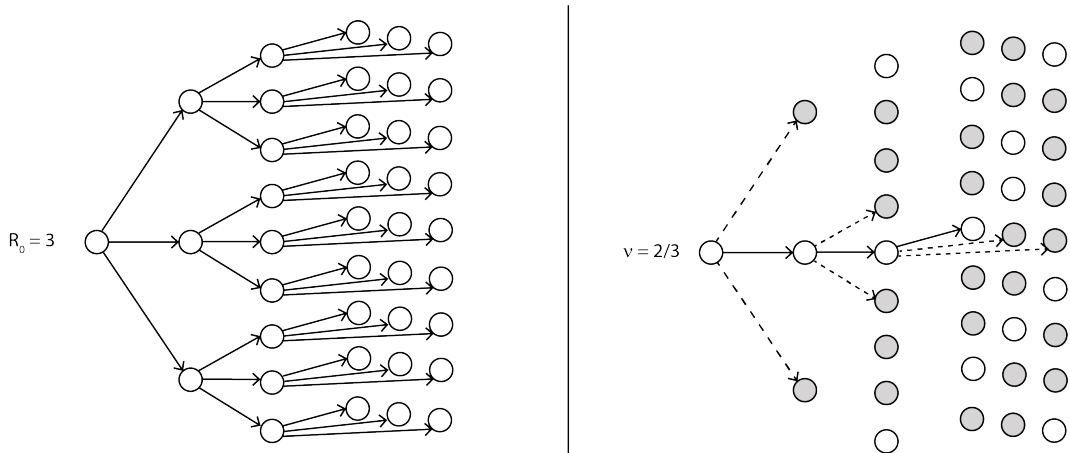


Figure 11: The effects of immunity on a susceptible population. Here a solid line means the disease is transmitted, and a dotted line means the disease is not transmitted. The grey individuals are those with immunity.

Initially everyone in the population is either vaccinated or they are susceptible, so $s + v = 1$. Rearranging this to get $s = 1 - v$ and substituting into Equation 11 gives us

$$R_0(1 - v) = 1.$$

Dividing both sides by R_0 gives

$$1 - v = \frac{1}{R_0}$$

and a final rearrangement gives us a formula for v in terms of R_0 :

$$v = 1 - \frac{1}{R_0}. \quad (12)$$

Case Study: Measles

Measles is currently one of the most highly communicable diseases within the human population. This means that it is highly contagious and difficult to control once an outbreak occurs. Transmission is generally airborne but it can also be spread via direct contact with bodily fluids. The basic reproduction number R_0 for measles is between 12 and 18, which means that one case of measles in a completely susceptible community generates between 12 and 18 new cases.

With measles being such a highly contagious disease, it brings to light the importance of vaccination rates within our communities. In recent times, the safety and efficacy of vaccines has come under fire as a result of numerous unsubstantiated studies and ‘anti-vax’ groups. It is important to note that before a vaccine was developed, there were an estimated 130 million cases of measles outbreaks across the world each year, with a

significant number of these cases resulting in death or in life-long disabilities, including blindness, brain damage and deafness.

In order to control future outbreaks, a certain level of immunity needs to be maintained within our communities. This level can be calculated using the herd immunity threshold represented by Equation 12. All we need to calculate this threshold is the basic reproduction number R_0 , which for measles is between 12 and 18.

For $R_0 = 12$, we have $v = 1 - \frac{1}{12} = 0.917 \approx 92\%$.

For $R_0 = 18$, we have $v = 1 - \frac{1}{18} = 0.944 \approx 94\%$.

This means that governments have to immunise between 92 per cent and 94 per cent of the population to reach the herd immunity threshold. Australia's vaccination program currently achieves a rate of between 93 per cent and 95 per cent, and in 2014 was confirmed by the World Health Organisation (WHO) to have eliminated measles.

Remember that this is the threshold level of containment. If the proportion of immune individuals exceeds this level due to a mass vaccination programme, the disease will die out. This is why the herd immunity threshold is sometimes also called the **critical immunisation threshold**. Epidemiologists use Equation 12 to work out the minimum number of people that must be immunised in a population in order to eradicate a disease.

Eliminating a disease in a particular part of the world, like Australia, does not mean that that region will never experience the disease again. Even in Australia there are occasional outbreaks. These usually happen when an overseas visitor arrives carrying the disease. This happened in Sydney in both 2012 and 2017 when visitors carrying measles arrived from Thailand and Indonesia respectively. Such outbreaks can help to identify which members of the population are still at risk: this may be young children who have not yet been immunised, older people whose vaccine immunity is not as strong as was thought or particular communities who may have missed out on vaccination for different reasons.

Worldwide, deaths from measles have reduced by over 75 per cent since vaccination programs began. The WHO has the aim of eliminating measles in at least five major world regions by 2020, and as of 2018 this has been achieved by only one region: The Americas.

Exercise 11

- a Use the completed table from Exercise 2 to calculate the percentage of the population that would need to be immunised in order for the disease to die out within the susceptible population.

Disease	R_0	ν
Measles		
Ebola		
Influenza		
Smallpox		
Whooping cough		

- b For the case study of the Hong Kong flu on page 20 and the information given in Exercise 10, calculate the herd immunity threshold ν and therefore the proportion of the population that will need to be vaccinated in order for the disease to become stable within the community.

The story of HPV

Human papillomavirus, or HPV, is the most common sexually transmitted infection in the world. Most of us will become infected at some point in our lives without even knowing it. But some instances of the virus produce serious illnesses, including cervical cancer, throat cancer and genital warts.

In 1991, University of Queensland researchers Ian Frazer, Jian Zhou and Xiao-Yi Sun designed and created the first ever HPV vaccine. Their idea was to assemble proteins into a structure that resembled the HPV shell but which did not contain any of the deadly virus. When injected into a human, the body is tricked into making antibodies against the virus, so that when the real virus comes along the body has immunity.

By 2006 the vaccine had undergone strict clinical trials and was approved for use. In order to be effective, it must be given to the population before they have had the chance to come into contact with the real virus, which means administering it before the population becomes sexually active. In Australia the vaccine is given to students in Years 7 and 8, targeting both boys and girls. So far the results have been very encouraging, with a 90 per cent reduction in genital warts among under 21s, and a 77 per cent decrease in the types of HPV that cause cervical cancer. However, as cervical cancer can take many years to manifest itself, it will be a number of years before there will be definitive results on the efficacy of the vaccine.

There are 170 different types of HPV, making it hard to pin down an exact value for the basic reproductive number. From studies in different countries around the world, R_0 for HPV is thought to be between 1.1 and 1.7, making it far less contagious than a disease like measles but still above the threshold for being an epidemic. The hope is that with the current vaccination program, the effective reproductive number can be brought below 1

to help eliminate all the different types of cancer HPV is associated with.

Conclusion

Infections and diseases remain a persistent and recognised health risk to most communities worldwide. However, our increased understanding about how contagions spread has allowed us to not only contain many outbreaks but to work toward the complete eradication of many previously devastating infections. There will always be new viruses and outbreaks but with the aid of mathematical modelling, epidemiologists will be able to develop a deeper understanding of the potentially harmful consequences of future diseases and to limit the effects they will have on the larger public.

References

History

<http://www.emelbourne.net.au/biogs/EM00473b.htm>

Image of John Graunt can be found at

<https://www.york.ac.uk/depts/maths/histstat/people/#g>.

It is from p.147 of T O'Donnell, *History of Life Insurance in its Formative Years*, Chicago: American Conservation Company 1936. No known copyright restrictions.

Epidemiological Models

<https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/epidemic-theory>

https://deepblue.lib.umich.edu/bitstream/handle/2027.42/111523/brouweaf_1.pdf

<http://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model>

<https://www.britannica.com/event/Hong-Kong-flu-of-1968>

<http://www.who.int/csr/resources/publications/measles/whodscsr991.pdf>

Graphs

Figures 4, 5, and 6 are screenshots from the Wolfram Demonstrations Project, “SIR Epidemic Dynamics”.

<http://demonstrations.wolfram.com/SIREpidemicDynamics/>

Solutions to Exercises

Exercise 1

- a** At the end of the first week, we have $t = 7$. Therefore

$$\begin{aligned} y &= 2^{7-1} \\ &= 2^6 \\ &= 64 \text{ people infected at the end of the first week.} \end{aligned}$$

- b** Let $y = 2,500,000$. Then we have

$$2,500,000 = 2^{t-1}.$$

Take the logarithm of both sides of the equation:

$$\log(2,500,000) = (t - 1) \log(2).$$

Rearranging,

$$t - 1 = \frac{\log(2,500,000)}{\log(2)} = 21.2535$$

This gives $t = 22.2535$, so the entire population would become infected by the 23rd day after the initial outbreak.

- c** Keeping the assumptions that infected individuals recover after a day and that during the day when they are infected they come into contact with only two other people, then only one person becomes infected each day. This means that it will take 2 500 000 days for the disease to spread through the entire population. This is 6844 years (assuming that a year length of 365.2422 days).

Exercise 2

- a** For example, see

https://en.wikipedia.org/wiki/Basic_reproduction_number.

Disease	R_0
Measles	12 – 18
Ebola	1.5 – 2.5
Influenza	2 – 3
Smallpox	5 – 7
Whooping cough	5.5

- b** Some communities may have individuals who are vaccinated or who have natural immunity to a disease. These immune individuals will help to stop the disease spreading as it normally would through the population.

Exercise 3

- a** Susceptible members of the population are gradually becoming infected.
- b** The entire population is made up from only infected and susceptible individuals, so the sum of the two proportions must be equal to 1. We can also work this out from knowing that $S + I = N$, since

$$s + i = \frac{S}{N} + \frac{I}{N} = \frac{1}{N}(S + I) = \frac{1}{N}(N) = 1.$$

- c** In reality, people recover from a disease and stop being infectious after a period of time. We may also want to build into the model the fact that not everyone in the population is susceptible to a disease, since some people may have immunity from either genetics or vaccination.

Exercise 4

Begin with

$$\frac{di}{dt} = \beta i(1 - i)$$

Separate variables and integrate both sides:

$$\int \frac{1}{i(1-i)} di = \int \beta dt$$

Use partial fractions to simplify the left-hand side, then integrate and simplify:

$$\begin{aligned}\int \frac{1}{i} + \frac{1}{1-i} di &= \int \beta dt \\ \Rightarrow \ln i - \ln(1-i) &= \beta t + C \\ \Rightarrow \ln\left(\frac{i}{1-i}\right) &= \beta t + C \\ \Rightarrow \frac{i}{1-i} &= e^{\beta t + C} = Ae^{\beta t}.\end{aligned}$$

We know that $i = i_0$ when $t = 0$. This gives the constant of integration as

$$A = \frac{i_0}{1-i_0}.$$

Finally, we must rearrange the equation to isolate the variable i :

$$\begin{aligned}i &= Ae^{\beta t}(1-i) \\ &= \frac{Ae^{\beta t}}{1 + Ae^{\beta t}} \\ &= \frac{\frac{i_0}{1-i_0}e^{\beta t}}{1 + \frac{i_0}{1-i_0}e^{\beta t}} \\ &= \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}}.\end{aligned}$$

Exercise 5

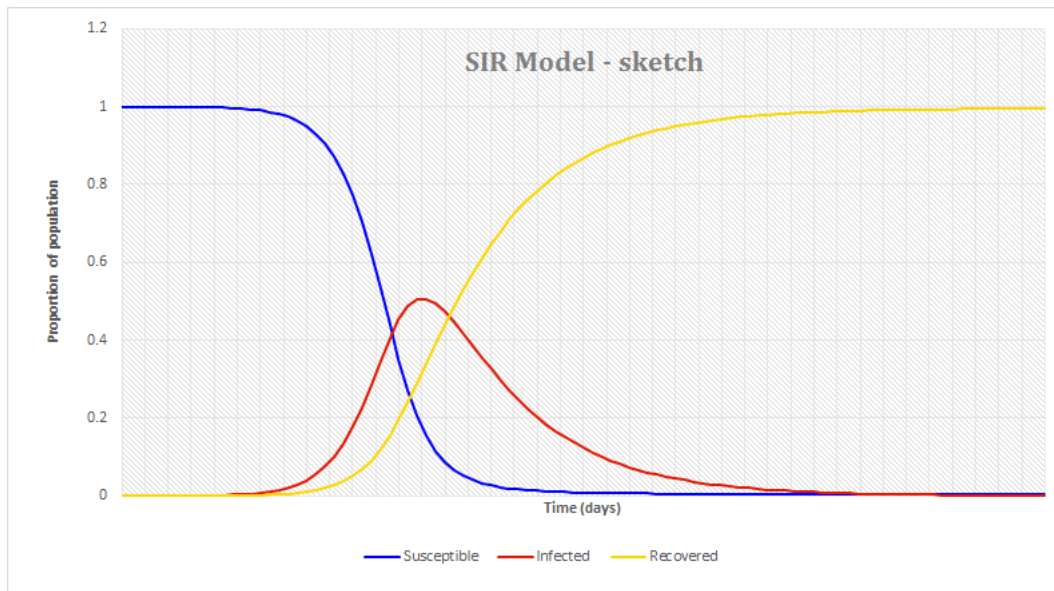
- a After five days, infected individuals begin to recover and will be neither susceptible nor infectious. This means that $s + i$ no longer equals 1.
- b After 14 days the total number of infected individuals begins to fall.

Exercise 6

- a In both graphs, a proportion of 0.01 is initially infected. If $N = 50\,000$, this gives $I = 0.01 \times 50\,000 = 500$ people initially infected, and $S = 50\,000 - 500 = 49\,500$ susceptible people.
- b To answer this question, we must find where the blue and red lines first cross each other. In the first graph, this happens at $t = 3$ days, and in the second graph, this is at $t = 9$ days.
- c The susceptible people are all infected when the blue line reaches zero. In the first graph, this happens at $t = 6$ days, and in the second graph this never happens.

Exercise 7

- a As people start to catch the disease in the beginning, $i(t)$ will increase while $s(t)$ decreases. Once people start to recover, $r(t)$ will increase, causing $i(t)$ to slow its increase and eventually decrease, with $s(t)$ still decreasing but more slowly.
- b This is a sketch of a possible SIR graph.



- c Each of i , s and r are proportions of the total population, and are the only three states that an individual can be in. At any given time, each member of the population is in one of the three groups. Therefore the sum of the three proportions must be 1.

Exercise 8

- a The rate of change of $s(t)$ is in inverse proportion to $i(t)$, so as the proportion of infected individuals increases, the proportion of susceptible people decreases.
- b The proportion of infected individuals is now a function of those who are susceptible and those who have recovered. So $i(t)$ will increase as more susceptible people catch the disease, but will decrease as more people recover.
- c Summing the three equations gives

$$\frac{ds}{dt} + \frac{di}{dt} + \frac{dr}{dt} = -\beta si + (\beta si - \gamma si) + \gamma si = 0.$$

This tells us that the population remains constant.

Exercise 9

- a Remember that the average infectious period is the inverse of the recovery rate, and the contact rate is the inverse of the transmission rate since we assume that the transmission risk is 100%.

Percentage of the population who are susceptible at $t = 0$	100
Average infection period	10 days
How often an infected person encounters a susceptible person	Every 2 days 18 hours
Time taken for number of infected individuals to reach zero	95 days

- b Day 46 has the highest infection rate. After this day, there are more recovered people than susceptible people in the population, making it more difficult for the disease to spread.
- c The blue line stays almost horizontal until about day 25 and then begins to drop dramatically.
- d It continues to decrease, but at a slower rate, eventually becoming asymptotic.
- e The blue line does not reach zero because after day 95 the infection has died out ($i = 0$), so the remaining susceptible people never catch the disease. After this point we must have $s + r = 1$ because everyone is either susceptible (never caught the disease) or recovered.
- f To find the exact values in this table, we modelled the SARS virus on our Excel spreadsheet. Approximate values can be obtained from estimating the values given by the graph.

	0 days	40 days	60 days
Susceptible Individuals, S	7400000	4269806	392200
Infected Individuals, I	10	2027603	1302402
Recovered Individuals, R	0	1102601	5705408
$S + R + I$	7400010	7400010	7400010

- g It is always the same. it represents the total population. This means the disease is non-fatal.
- h On day 46, the number of infected people is roughly $0.39 \times 7400010 = 2886004$, so about 2.9 million people.
- i The entire population was not infected. This is indicated by the asymptotic behaviour of the blue curve, which never reaches zero. This could be due to some people having natural immunity or not coming into contact with infected individuals.

- j On day 95, when the infection dies out, there are roughly $0.02 \times 7400010 = 148000$ people who escaped infection.

Exercise 10

- a Given a population of 7 900 000, we have $I(0) = 10$, $S(0) = 7900000 - 10 = 7\,899\,990$ and $R(0) = 0$.
- b $\beta = \frac{1}{2}$, $\gamma = \frac{1}{3}$.

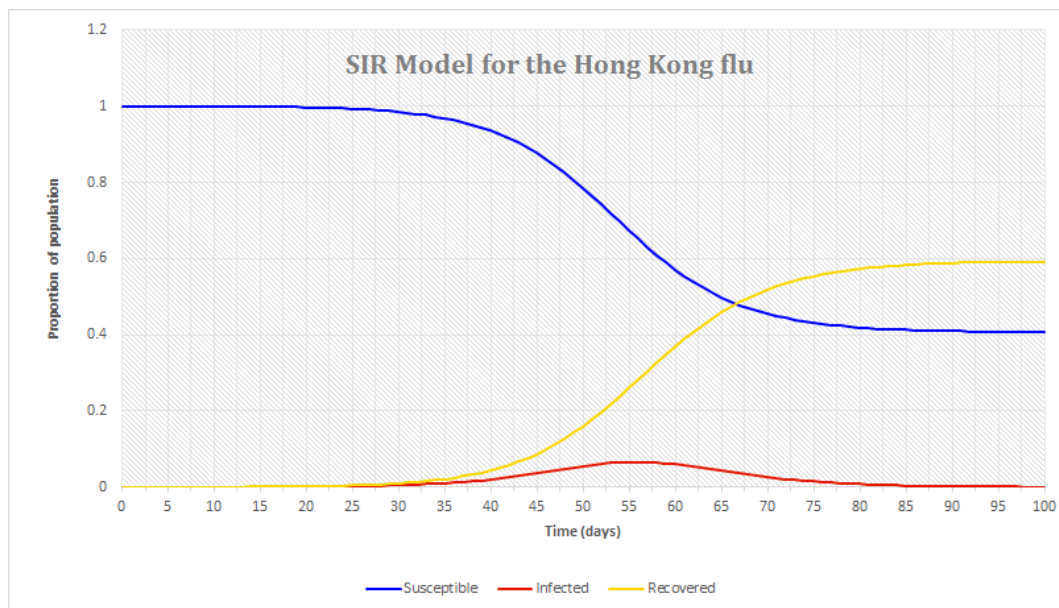


Figure 12: SIR graph modelling the Hong Kong flu

- c
- Day 55 was the worst day of the epidemic, with 6.5% per cent of the population infected.
 - The disease dies out on about day 95.
 - No – after the disease dies out, there is still about 40% of the population (which is 3.1 million individuals) still being susceptible, having not been infected.
 - The HK flu took a similar time to the SARS outbreak to die out – around 95 days. Although the HK flu had a higher transmission rate than the SARS virus (0.5 compared with 0.36), it affected a much lower proportion of the population – 60 per cent compared with 98 per cent. This is because of the much faster recovery rate – people recovered from the HK flu in three days compared with ten days for the SARS virus.

Exercise 11

- a The following values for ν are calculated from the formula $\nu = 1 - \frac{1}{R_0}$.

Disease	R_0	ν
Measles	12 – 18	92 – 94%
Ebola	1.5 – 2.5	33 – 60%
Influenza	2 – 3	50 – 67%
Smallpox	5 – 7	80 – 86%
Whooping cough	5.5	82%

b We first need to calculate the value of R_0 . We have

$$R_0 = \frac{\beta}{\gamma} = \frac{\frac{1}{2}}{\frac{1}{3}} = \frac{3}{2}.$$

This gives us

$$\nu = 1 - \frac{1}{R_0} = 1 - \frac{2}{3} = \frac{1}{3}.$$

Hence for the population of New York, 33 % (or 2.63 million people) need to be vaccinated in order for the disease to become stable.