

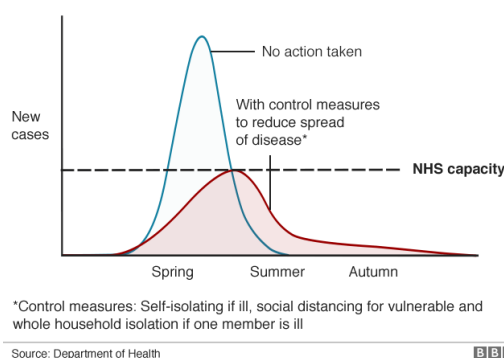
Using A Level Maths to Model the Spread of an Infectious Disease

A graph that shows the number of cases in an infectious disease outbreak has a distinctive shape: the number of cases rises to a peak and then tails off. You can see examples of real cases below.



As a result of the COVID-19 pandemic, we have become very familiar with this characteristic curve. Media and government agencies have shown images of a *flattened curve*, or a *delayed peak*, based on mathematical models.

How control measures may reduce spread of virus in UK



If you have an interest in Maths, you have probably wondered:

- What are these models?
- Is it really possible to find an equation for a virus?
- How do they know where the peak is?
- How is the curve *mathematically* flattened?

The challenge for us:

Could we use our A level Maths skills, to find an equation, or a set of equations, that will produce the characteristic curve of an infectious outbreak?

We are going to develop three epidemic models that increase in sophistication:

- The Geometric model
- The Susceptible – Infective model
- The Susceptible – Infective – Removed model

As each model is improved, the mathematics becomes more challenging. However, these three models can be fully understood by an A level Mathematician.

Background

Mathematicians have developed epidemic models that allow them to predict how the outbreak of an infectious disease may behave. This means that there is already a basic set of equations that will produce the right curve characteristics. By altering particular values in the equations, mathematicians can produce a unique curve to correspond with a particular outbreak. This can give useful information such as an estimate of the transmission rate.

In the early stages of an outbreak, mathematicians plot a graph of real data as it becomes available. At a very basic level, mathematicians can trial values in their equations to find a curve that is a close match to the early stage. As more data becomes available, the model can be refined. The resulting graph is not a guaranteed prediction of the future. But it can give an idea of what may happen during the infectious outbreak if containment measures are not put in place.

The peak of the curve is of particular interest to a community because it gives an idea of the maximum number of people who may be infectious at one point. If this number is greater than the health care system can provide for, then it is vital that intervention is considered.

If the peak can be reduced, hospitals may cope better with the maximum number of cases.

If the peak can be shifted and delayed, the health service has more time to prepare. In the case of the COVID-19 pandemic, the Nightingale hospitals were built in anticipation of the peak.

Mathematicians who study human behaviour and social interactions can use their model to test how interventions such as social distancing may change the shape of a curve and so alter the impact of the outbreak.

Further areas of interest to explore and develop could include: Herd Immunity, Social Interactions, the Impact of a Vaccine.

Modelling assumptions and key terms

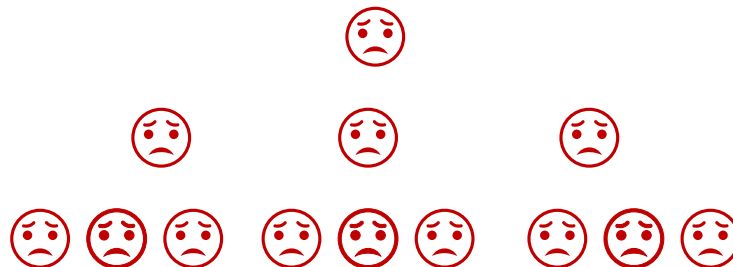
- An infectious disease has a Basic Reproduction number, R_0 . This is the average number of people who will catch the disease from an infectious person. The R_0 value may change for an infectious disease because it depends on the interaction between those who are infected and those who are susceptible.
 - If $R_0 > 1$ an epidemic will occur and if $R_0 < 1$ the infectious disease will die out.
 - Initially, the whole population, P , is susceptible to an infectious disease. The population is considered to be a constant value.
 - An infected individual is immediately infectious.
-

The Geometric Model

Consider a reproduction rate, $R_0 = 3$.

When 1 infected person enters a population, they will infect 3 others.

Each of the newly infected individuals will infect 3 others.



We can list the number of new cases: 1, 3, 9, ...

This model produces a Geometric Progression: $a = 1$, $r = 3$

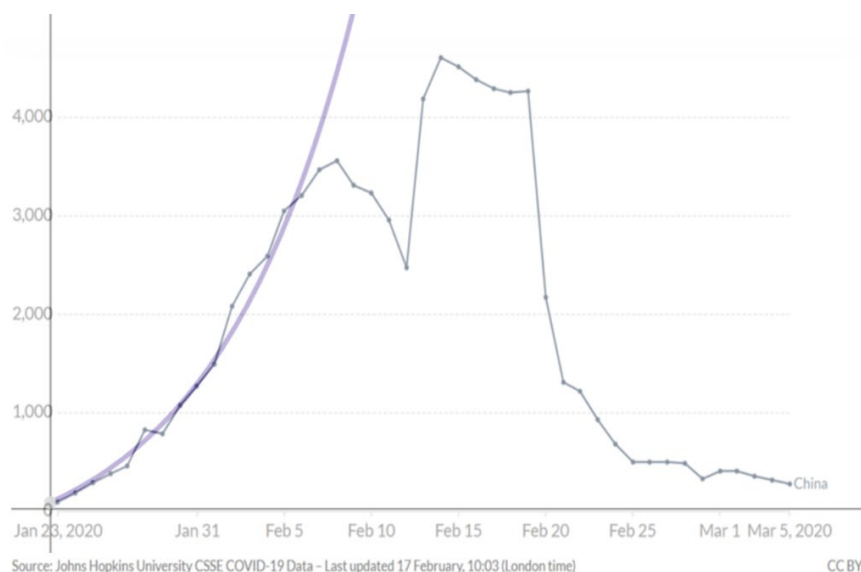
On the n^{th} day the number of cases is given by: 3^{n-1} [using $a r^{n-1}$]

The total number of infected people after n days: $\frac{3^n - 1}{2}$ $\left[S_n = \frac{a(1-r^n)}{1-r} = \frac{1-3^n}{1-3} \right]$

Evaluating the Geometric model

The number of new cases each day and the total number of cases are exponential functions. They are functions that grow very quickly. They show that an infectious disease will spread throughout a population very quickly.

We can evaluate our model by comparing an exponential curve with the characteristic curve of an epidemic model or even the curve produced from the early stages of real data.



If we overlay an exponential curve with the COVID-19 cases reported in China at the beginning of 2020, we can see that initially there is some level of agreement. This model may be useful in the early stages of an outbreak.

However, we know that an exponential function will not flatten out, or reach a peak, so a geometric model cannot represent the entire outbreak.

We can use differentiation to show that our exponential function for the number of new cases is strictly increasing.

If we differentiate the curve equation: $y = 3^{x-1}$ we get $\frac{dy}{dx} = 3^{x-1} \ln 3$

Since $3^{x-1} > 0$ and $\ln 3 > 0$ we know that $\frac{dy}{dx} > 0$.

This shows that the function is strictly increasing and a stationary point does not exist.

Limitations of the Geometric model

- This model does not account for a finite population. In a geometric model the number of new cases continues to rise forever. This is not possible in a limited population because the disease will eventually run out of people to infect.
 - This model ignores the recovery of infected individuals and assumes that once infected, a person will remain infected.
 - There is no peak to our number of new cases. The maximum point is significant to modellers because it indicates the largest number of infected individuals that may require medical treatment at one time. An improved model would give an indication of this value.
-

Model 1 The Geometric Model: *explore*

Modelling assumptions and key terms

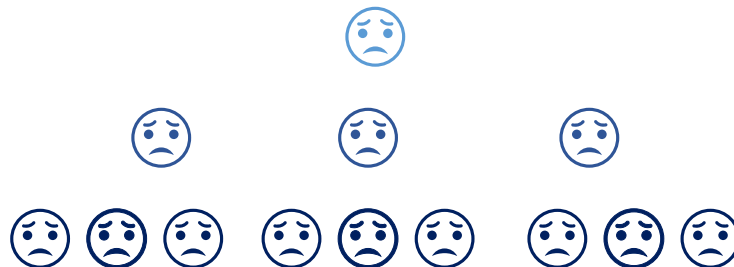
- An infectious disease has a Basic Reproduction number, R_0 . This is the average number of people that will catch the disease from an infectious person. The R_0 value may change for an infectious disease because it depends on the interaction between those who are infected and those who are susceptible.
 - If $R_0 > 1$ an epidemic will occur and if $R_0 < 1$ the infectious disease will die out.
 - Initially, the whole population, P , is susceptible to an infectious disease. The population is considered to be a constant value.
 - An infected individual is immediately infectious.
-

Task 1

Consider a reproduction rate, $R_0 = 3$.

When 1 infected person enters a population, they will infect 3 others.

Each of the newly infected individuals, will infect 3 others.



The number of new cases forms the terms of a geometric progression: 1, 3, 9, ...

- a) State the first term and common ratio for this geometric progression.
- b) How many new cases will there be after 7 progressions?
- c) How many new cases will there be after n progressions?
- d) Sketch a graph for the number of new cases against time.
- e) Use differentiation to show that the number of new cases will continue to rise.
- f) Explain a limitation of this model.

Task 2

A boarding school has 500 students and 60 staff who all live on the school site. The school is in a remote location and is modelled as having no physical interaction with any person outside of the school community. At the beginning of term, one student arrives showing symptoms of an infectious disease. The infectious disease has a Basic Reproduction value of 2.5. Each day represents a new progression. No measures are taken to isolate the individual.

Using a geometric model,

- predict the number of new cases after 3 days.
- predict how long it will take for the entire population to become infected with the disease.

After 3 days, containment measures are put in place to protect the school community. It is assumed that the Basic Reproduction value reduces instantly to 0.8.

- Using a new geometric model, predict the total number of individuals in the school community who will become infected.
- Explain the significance of $R_0 < 1$ in a model that is based on a geometric progression.

Task 3

- The table gives an estimated range of values for the [Basic Reproduction](#) number of different diseases.

Discuss: Would an outbreak of Chickenpox spread more rapidly through a local community than COVID-19?

Disease	R_0
Measles	12-18
Chickenpox	10-12
COVID-19	3.3-5.7
Common cold	2-3
Ebola	1.5-1.9

- The graph shows the number of cases of COVID-19 recorded by China and has been overlaid with an exponential curve. In the early stages of the outbreak, the exponential curve follows the data closely.

Discuss: Why does the data stop following an exponential curve? What factors should be considered in an improved model?



Model 2 The S-I model: *understanding*

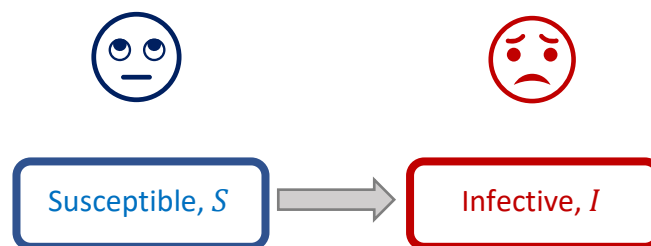
Modelling assumptions and key terms

- An infectious disease has a Basic Reproduction number, R_0 . This is the average number of people who will catch the disease from an infectious person. If $R_0 > 1$ an epidemic will occur and if $R_0 < 1$ the infectious disease will die out.
- Initially, the whole population, P , is susceptible to an infectious disease. The population is considered to be a constant value.
- An infected individual is immediately infectious for a fixed number of days, T . This is the infectious period.
- a is the transmission rate. It is a number that represents how quickly the disease spreads.

$$a = \frac{\text{average number of people likely to be infected}}{\text{number of infectious days}} = \frac{R_0}{T}$$

The S-I Model

This is the most basic compartmental model where the population, P , is divided into two groups: those who are susceptible to the disease and those who are infected with the disease and are infectious to others. Initially, every individual is in the Susceptible group until they pass into the Infective group. Once in the Infective group, the individual will remain there.



We can form the equation:
$$S + I = P$$

Whilst the values of S and I will change, the population, P , is assumed to remain at a constant number.

Our maths will become easier later on if we consider **proportions** of the population rather than the size of a group.

For example, 70% are susceptibles and 30% are infectives

0.7 are susceptibles and 0.3 are infectives.

Dividing by P produces the proportion who are susceptible and the proportion who are infective:

$$\frac{S}{P} + \frac{I}{P} = \frac{P}{P}$$

$$\frac{S}{P} + \frac{I}{P} = 1$$

We can introduce new variables, s and i , to represent the proportion of each group which leads to:

$$s + i = 1 \quad \text{equation [1]}$$

In other words, the proportion of susceptibles (s) and the proportion of infectives (i) adds up to the whole population.

In the $S - I$ model we study the **growth** of the disease by considering how quickly the Infective group grows. To study 'how quickly' we need to set up a differential equation.

First, we consider the rate that people **leave** the Susceptible group. This is proportional to $s \times i$ which represents how interactive the two groups are with each other.

The rate of people leaving the Susceptible group will also depend on the transmission rate of the disease, a . We say:

$$\frac{ds}{dt} = -asi \quad \text{equation [2]}$$

From equation [1] we also know that:

$$s + i = 1$$

Differentiate each term with respect to time, t :

$$\frac{ds}{dt} + \frac{di}{dt} = 0$$

This model tells us that:

$$\begin{aligned} \frac{di}{dt} &= -\frac{ds}{dt} \\ \frac{di}{dt} &= asi \quad \text{equation [3]} \end{aligned}$$

This result may appear intuitive:

people leave the Susceptible group at the same rate as they join the Infective group.

But it does depend on the assumption that the total Population is constant.

Our equations so far:

$$s + i = 1 \quad \text{equation [1]}$$

$$\frac{ds}{dt} = -asi \quad \text{equation [2]}$$

$$\frac{di}{dt} = asi \quad \text{equation [3]}$$

If we can solve equation [3], we will have a relationship between the proportion of the population who are infected and time. We can plot a graph that shows how the proportion of infectives changes over time and compare it with the characteristic curve for an epidemic.

When we solve the differential equation, we will have a general solution for any infectious disease outbreak.

If we say that initially the proportion of infectives is i_0 then we can solve our differential equation and form a general solution for any infectious disease outbreak.

Equation [3] is a differential equation with three variables: s, i, t . We need to eliminate the variable, s , so that we can find i in terms of t .

Use equation [1] to eliminate s from equation [3]

$$\begin{aligned} \frac{di}{dt} &= asi \\ \frac{di}{dt} &= a(1-i)i \end{aligned}$$

We must *separate the variables* to solve this equation. Remember that a is a number and not a variable.

$$\int \frac{1}{(1-i)i} di = \int a dt \quad \text{equation [4]}$$

To integrate the left-hand side, we need to separate $\frac{1}{(1-i)i}$ into partial fractions.

$$\frac{1}{(1-i)i} = \frac{A}{1-i} + \frac{B}{i}$$

$$1 = Ai + B(1-i)$$

$$\text{Let } i = 0, \quad 1 = B$$

$$\text{Let } i = 1, \quad 1 = A$$

Now replace in equation [4]

$$\int \frac{1}{1-i} + \frac{1}{i} di = \int a dt$$

$$-\ln|1-i| + \ln|i| = at + c$$

$$\ln \left| \frac{i}{1-i} \right| = at + c$$

When $t = 0$, $i = i_0$ this leads to $\ln \left| \frac{i_0}{1-i_0} \right| = c$

We have a particular solution but still need to rearrange to get i in terms of t .

$$\ln \left| \frac{i}{1-i} \right| = at + \ln \left| \frac{i_0}{1-i_0} \right|$$

$$\ln \left| \frac{i}{1-i} \right| - \ln \left| \frac{i_0}{1-i_0} \right| = at$$

$$\ln \left| \frac{i(1-i_0)}{i_0(1-i)} \right| = at$$

$$\frac{i(1-i_0)}{i_0(1-i)} = e^{at}$$

$$i(1-i_0) = e^{at} i_0 (1-i)$$

$$i - i i_0 = e^{at} i_0 - e^{at} i i_0$$

$$i - i i_0 + e^{at} i i_0 = e^{at} i_0$$

$$i(1 - i_0 + i_0 e^{at}) = e^{at} i_0$$

Hence, we have an equation that links the proportion of infectives with time.

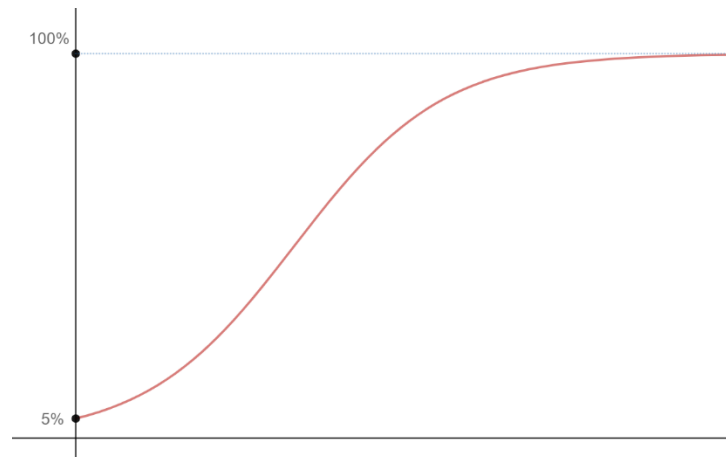
$$i = \frac{e^{at} i_0}{1 - i_0 + i_0 e^{at}}$$

Evaluating the S-I model

If we know a value for the transmission rate of a disease, a , and we know what proportion of the population are infected at the outbreak, i_0 , then we can sketch a graph and compare it to an epidemic curve.

For example, consider an infectious disease with a transmission rate of 0.6, where 5% of the population are thought to have the disease initially.

Using values of $a = 0.6$, $i_0 = 0.05$ would lead to $i = \frac{0.05e^{0.6t}}{1 - 0.05 + 0.05e^{0.6t}}$



Our model has produced a curve that flattens which is an improvement on the geometric model.

In this model, we can consider what will happen over a long period of time:

As $t \rightarrow \infty$, $e^{at} \rightarrow \infty$ and so $i \rightarrow 1$.

Our model shows that over a long enough period of time, the whole population moves into the Infective compartment. This is an improvement on the Geometric model because the number of infectives does not increase beyond the total population.

Limitations of the S-I model

The $S - I$ model predicts that the whole population will become infected and remain infected. This is not the case in real life. An improved model would account for recovery from the disease; the proportion of individuals infected would reduce and the curve would slope down.

Modelling assumptions and key terms

- An infectious disease has a Basic Reproduction number, R_0 . This is the average number of people who will catch the disease from an infectious person. If $R_0 > 1$ an epidemic will occur and if $R_0 < 1$ the infectious disease will die out.
- Initially, the whole population, P , is susceptible to an infectious disease. The population is considered to be a constant value.
- An infected individual is immediately infectious for a fixed number of days, T . This is the infectious period.
- a is the transmission rate. It is a number that represents how quickly the disease spreads.

$$a = \frac{\text{average number of people likely to be infected}}{\text{number of infectious days}} = \frac{R_0}{T}$$

Task 1

The table shows values of R_0 for different diseases along and the infectious period.

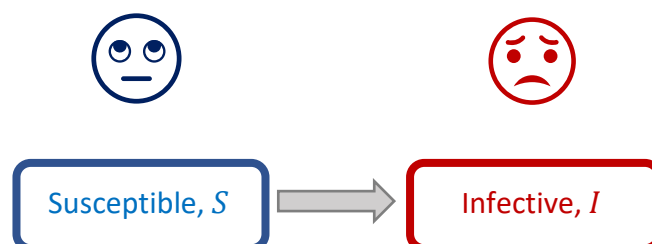
- a) Work out the corresponding transmission rates.

Disease	R_0	T	a
Measles	12	9	
Chickenpox	10	7	
COVID-19	3	7	
Common cold	2	10	

- b) Discuss the statement: *A disease with a higher transmission rate will spread more quickly within a community.*

Task 2

In the most basic compartmental model, the population, P , is divided into two groups: those who are susceptible to the disease and those who are an infected with the disease and are infectious to others. Initially every individual is in the susceptible group until they pass into the infective group. Once in the infective group, the individual will remain there.



The variable, s , represents the proportion of the population who are susceptible to a new infectious disease. The variable, i , represents the proportion of the population who are infected and infectious to others.

a) Explain why $s + i = 1$

The rate that people leave the Susceptible group and join the Infective group is proportional to:

- $s \times i$ which represents the interaction between the two groups
- the transmission rate, a

In a particular outbreak, the transmission rate, a , is taken to be 0.6.

Hence the rate of change of the two groups is given by:

$$\frac{ds}{dt} = -0.6si \quad \text{and} \quad \frac{di}{dt} = 0.6si$$

b) Explain what the signs represent in the differential equations.

c) Explain why the total rate of change is zero.

d) Show that: $\frac{di}{dt} = 0.6i(1 - i)$

e) Initially 5% of the population are thought to be infected with the disease. Solve the differential equation to find i in the terms of t .

f) Use a graph-plotter to sketch the shape of i against t , $t \geq 0$. Use the graph to describe how the infected population changes over time.

Modelling assumptions and key terms

- R_0 is the average number of people who are infected by an individual who has an infectious disease. $R_0 > 1$ an epidemic will occur and if $R_0 < 1$ the infectious disease will die out.
- Initially, the whole population, P , is susceptible to an infectious disease. The population is considered to be a constant value.
- An infected individual is immediately infectious for a fixed number of days, T . This is the infectious period.
- a is the transmission rate. It is a number that represents how quickly the disease spreads.

$$a = \frac{\text{average number of people likely to be infected}}{\text{number of infectious days}} = \frac{R_0}{T}$$

- b is the recovery rate and represents how quickly a person recovers from the disease.

$$b = \frac{1}{\text{number of infectious days}} = \frac{1}{T}$$

The S-I-R Model

In this compartmental model the population, P , is divided into three groups: those who are susceptible to the disease; those who are infected with the disease and are infectious to others and those who are removed from the Infective group. The Removed category is for those who are no longer infected. We assume that they are now immune to the disease and do not need to return to the Susceptible category.



To model the outbreak of a new infectious disease we assume that initially everyone is susceptible. An individual can only move from the Susceptible category directly to the Infective category. The individual will then move into the Removed category where they will remain.

We know that together these categories form the whole population:

$$S + I + R = P$$

However, it is more useful to consider the proportion that each group makes of the total population so we will define the variables, s , i , r as the proportion of each group and state that :

$$s + i + r = 1$$

Differentiating with respect to time shows the link between the rates of change:

$$\frac{ds}{dt} + \frac{di}{dt} + \frac{dr}{dt} = 0$$

We are able to improve on the $S - I$ model by including the recovery rate of the disease, b , and the removed group, r , in our differential equations.

- As before, individuals leave the Susceptible group at a rate proportional to the transmission rate and the interactions between susceptibles and infectives.

$$\frac{ds}{dt} = -asi$$

- The rate of change for the Infective group is made up of those entering the group, asi , and those leaving the group. The rate that individuals are leaving the group is proportional to the recovery rate and the size of the Infective group.

$$\frac{di}{dt} = asi - bi$$

- The Removed group is increasing at a rate proportional to the number of infectives and the recovery rate of the disease.

$$\frac{dr}{dt} = bi$$

Together we can see that the sum of the rates of change is zero: there is no overall change in the population.

We cannot solve these three differential equations using A level Maths techniques. A process of numerical analysis is required which is beyond the scope of our course. Note that we have used numerical methods, such as Newton-Raphson, to find a solution to an equation that we are unable to solve any other way.

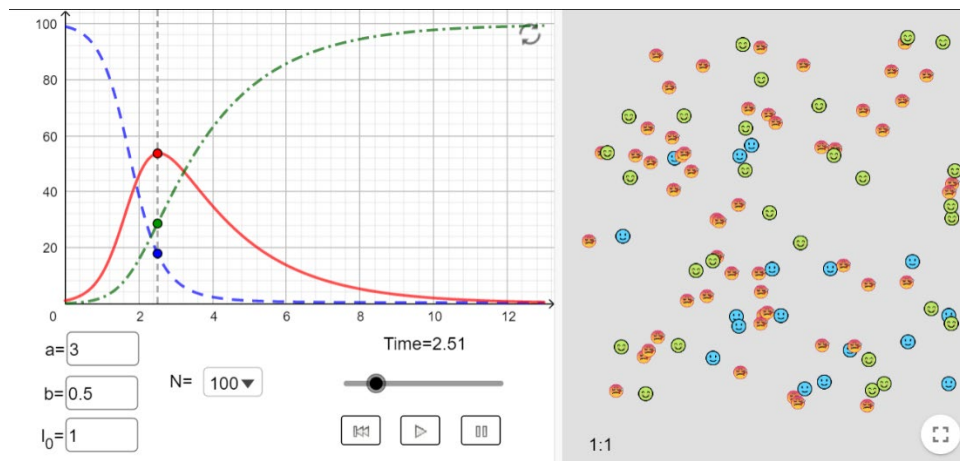
Although we cannot solve the differential equations, we are still able to analyse them using a graph-plotter. Each differential equation gives the gradient at a point in time. We can input these gradient functions to discover the shape of each curve. We can compare our curve for infectives against time, with the characteristic curve for an infectious outbreak.

This [geogebra](#) file is already set up with the differential equations for the S-I-R model. If you are interested in how to input the differential equations then watch this [numberphile](#) video.

With a transmission rate, $a = 3$ and a recovery rate of $b = \frac{1}{2}$ our differential equations would be:

$$\frac{ds}{dt} = -3si \qquad \frac{di}{dt} = 3si - \frac{1}{2}i \qquad \frac{dr}{dt} = \frac{1}{2}i$$

Using these values in the geogebra file produces the corresponding curves below:



The vertical axis, N , is the size of the population. By selecting $N = 100$ we can retain the idea of proportion. This data has been set with 1% of the population as initially infected.



Evaluating the S-I-R model

The shape of the curve for infectives, in red, does resemble the shape given by data for an infectious outbreak. This means that the model may be useful to understand a particular outbreak. The recovery rate, b , for an outbreak is unlikely to change. However, the rate of transmission, a , can vary with interventions. In the case above we can see a peak where just over 50% of the population are infected at the same time. By adjusting values for a , the peak will change. The peak value has an impact on local health services. Various levels of social interventions can be considered that would lead to a more desirable transmission rate.

We can see from the blue and green curves, that the whole population moves from the Susceptible group to the Removed group. This means that every individual in the population will be infected at some point.

Limitations of the S-I-R model

- We may be satisfied that the graph shape for infectives is realistic, but the three groups are still very broad. It does not account for individuals who are exposed to an infection and are not yet infectious.

- In this model we have assumed a constant transmission rate throughout the population. However, the transmission rates within an age group may vary. Individuals in an adult working environment will interact with each other quite differently to children in a primary school.
 - The $S - I - R$ model does not account for natural immunity, herd immunity or the introduction of a vaccine.
-

Modelling assumptions and key terms

- An infectious disease has a Basic Reproduction number, R_0 . This is the average number of people who will catch the disease from an infectious person. If $R_0 > 1$ an epidemic will occur and if $R_0 < 1$ the infectious disease will die out.
- Initially, the whole population, P , is susceptible to an infectious disease. The population is considered to be a constant value.
- An infected individual is immediately infectious for a fixed number of days, T . This is the infectious period.
- a is the transmission rate. It is a number that represents how quickly the disease spreads.

$$a = \frac{\text{average number of people likely to be infected}}{\text{number of infectious days}} = \frac{R_0}{T}$$

- b is the recovery rate and represents how quickly a person recovers from the disease.

$$b = \frac{1}{\text{number of infectious days}} = \frac{1}{T}$$

Task 1

The table shows values of R_0 for different diseases along with the infectious period. Calculate the corresponding transmission rates and recovery rates.

Disease	R_0	T	a	b
Measles	12	9		
Chickenpox	10	7		
COVID-19	3	7		
Common cold	2	10		

Task 2

In an S-I-R model, the population is separated into three groups:



- s is the proportion of the population who are susceptible to the outbreak of an infectious disease.
- i is the proportion of the population who are infected and infectious to others.
- r is the proportion of the population who are removed from the infectious group and are no longer susceptible to the disease.

- a) Justify why: $s + i + r = 1$
 b) Explain what the following differential equations represent:

$$\frac{ds}{dt} = -asi$$

$$\frac{di}{dt} = asi - bi$$

$$\frac{dr}{dt} = bi$$

Task 3

- a) Select an infectious disease from the table in Task 1 and write down the transmission rate and recovery rate.

The following differential equations model the rate of change of for each group within the population.

$$\frac{ds}{dt} = -asi$$

$$\frac{di}{dt} = asi - bi$$

$$\frac{dr}{dt} = bi$$

- b) Use your values to form differential equations for the disease.

The differential equations are gradient functions and can be used to plot curves to show how the proportion of Susceptibles, Infectives, and Removed individuals changes over time.

- c) Use this [geogebra](#) file to produce curves for your equations. Note that in this file:
- The size of the population is given by the variable, N . Use $N = 100$ so that the vertical axis can be interpreted using percentages of the population.
 - The value of i_0 is the value for those who are initially infected. Use $i_0 = 1$ to see the disease spread when 1% of the population are initially infected.
- d) Use your curves to make observations such as:
- the maximum proportion of the population who are infected at one time.
 - how long it takes for the whole population to be infected.

An area of mathematics is the *maths of human interaction and social behaviour*. Although the recovery rate for a disease is difficult to change, the transmission rate can be reduced by social interventions.

- e) Compare the peak of the infective curve for different transmission rates.

Transmission rate, a	Recovery rate, b	Use the peak for infectives	
		% infected	day this occurs

- f) List intervention measures that could take place and give reasons why a community may want to *flatten the curve* by reducing and delaying the peak.

Model 1: The Geometric model

Students should:

- know how to find the n^{th} term of a **geometric progression** and the sum of the first n terms.
- understand the condition for a **convergent series** and the formula for the **sum to infinity**.
- understand what is meant by an **increasing function** and know how to **differentiate** a function of the form a^x where a is a constant.
- recognise the shape of an **exponential** curve

Model 2: The S-I model

Students should:

- know how to find **partial fractions** where linear terms form the denominator.
- be able to rearrange equations that involve **exponential** and **log functions**.
- know how to **separate the variables** to find the general solution of a differential equation and also use conditions to find the particular solution.
- understand how to find the **limit** of a function over long periods of time.

Model 3: The S-I-R model

Students should:

- understand how to form equations for **rates of change**.
- be familiar with the idea of numerical methods such as **Newton-Raphson** to find solutions to equations

Further Thought



This is a fantastic place to go for inspiration and relevant mathematics beyond the classroom. There are a host of [COVID-19](#) articles relating to Herd Immunity, the Vaccine and so much more.



Take a look at an interactive [timeline](#) of WHO's response to the COVID-19 outbreak.



The South Pacific nation of Vanuatu makes a fascinating case study.

When the World Health Organisation declared COVID-19 a pandemic, on 11th March 2020, Vanuatu had approximately 300 000 inhabitants spread across 80 islands and not one hospital with a ventilator.

The government took decisive action to close borders and strictly quarantine arrivals. Their regular [government updates](#) announced that on 10 November 2020, the first confirmed case of COVID-19 was detected during routine day 5 testing of an asymptomatic person in quarantine.

Incredibly, to date, this is the only confirmed case for the island nation.

What a fabulous success story!