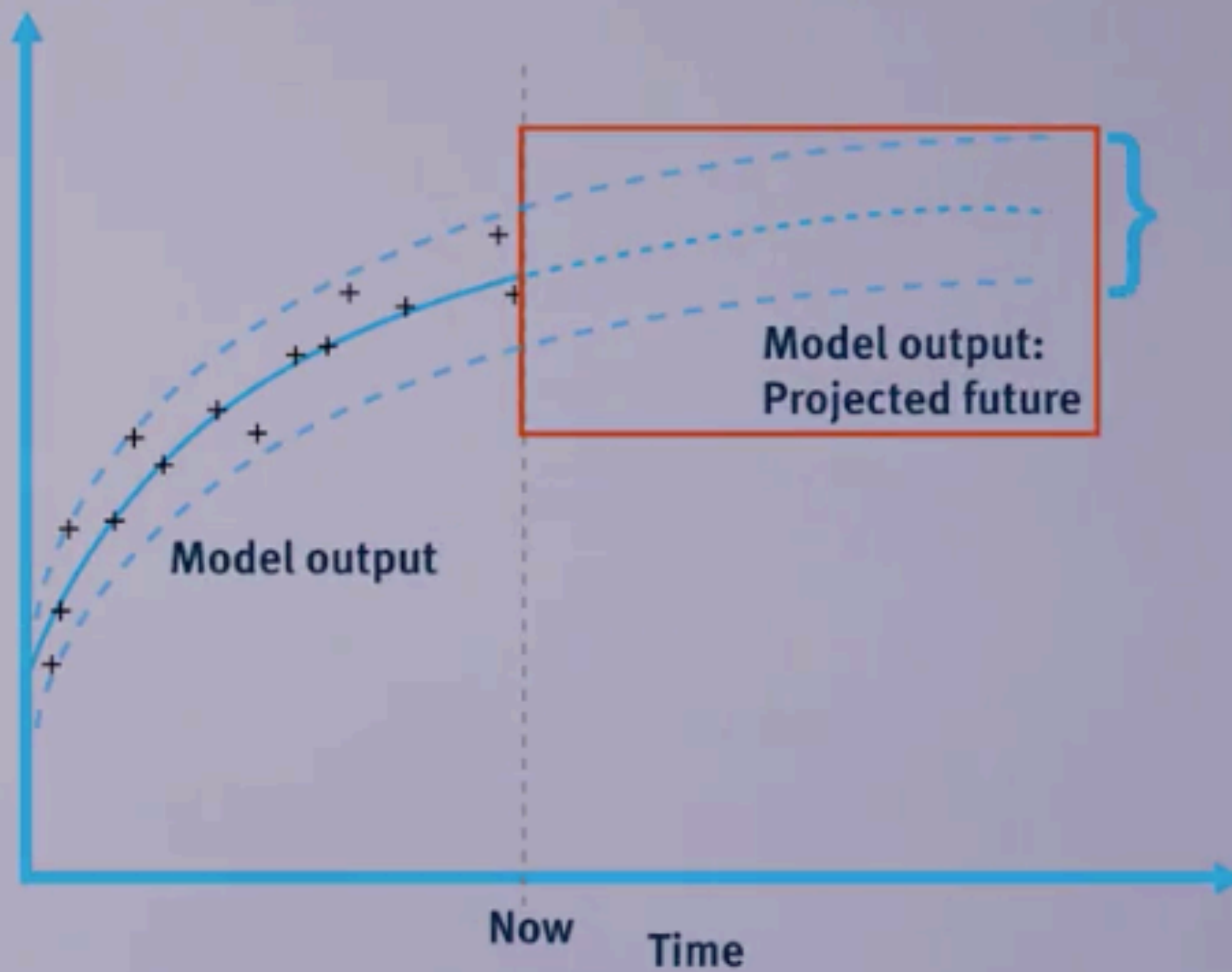


R for Epidemic Models

Dr R Sinuvasan
VIT-AP University

Infectious Disease Modelling

- ✱ Mathematical modelling is increasingly being used to support public health decision-making in the control of infectious diseases
- ✱ All modelling will be conducted in the programming language R, which is widely used in modelling applications today.
- ✱ We focus on compartmental modelling, using the 'SIR model', a cornerstone of infectious disease modelling
- ✱ Behind every infectious disease, there is the basic mechanism of a host passing on the infection to other individuals. We call this transmission dynamics.
- ✱ We are bringing together our best understanding of an infectious disease, and using a model to project what this information means for future scenarios.



Solving DE using ode()

* First, run the library() function to load these packages

```
library(ggplot2)
```

```
library(deSolve)
```

Example (Population growth: exponential)

The variable of interest in this case, is the size of the population: call this N. We want to model N as it changes over time dN/dt

$$\frac{dN}{dt} = \alpha N$$

N - state variable

alpha- Parameter

ode() function

```
result <- ode(  y = state_vars # initial state variables
               , times = times   # a vector of timepoints
               , func = exp_pop_fn # the DE itself, in the form of a function
               , parms = parms)  # parameters for the equation in func
```

```
library(ggplot2)
library(deSolve)
```

```
state_vars <- c(N = 1)
times      <- seq(0, 40, by = 0.5)
```

```
exp_pop_fn <- function(time, state, parameters) {
```

```
  N <- state['N']
  dN <- parameters['alpha'] * N
  return(list(c(dN)))
```

```
}
```

```
parms <- c(alpha = log(2))
parms['alpha']
```

```
result <- ode(y = state_vars , times = times , func = exp_pop_fn , parms =
parms)
```

```
head(as.data.frame(result))
```

```
result <- as.data.frame(result)
```

```
require(ggplot2)
```

```
expplot <- ggplot(data = result)+ geom_line(aes(x = time, y = N) , colour = "blue")+ labs(x =  
"time (days)")
```

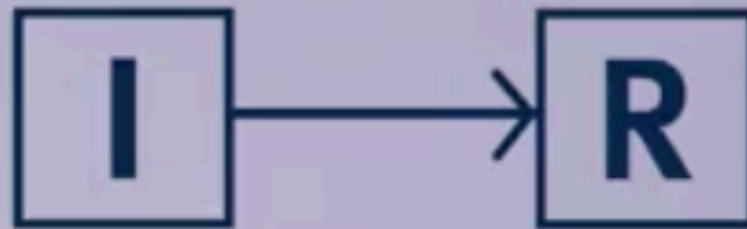
```
expplot
```

More realistic: logistic growth

$$\frac{dN}{dt} = \alpha N \left(1 - \frac{N}{K}\right)$$

```
N      : 1  
alpha  : log(2)  
K      : 1000000
```

Simple Compartmental Models



$$\frac{dI}{dt} = -\gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Question

We are looking at a cohort of ($I=1000000$) currently infected people, and no one has recovered ($R=0$) so far. The average duration of infection is 10 days. How many people will recover from the infection over a 4-week period.

```
library(deSolve) # package to solve the model
```

```
library(reshape2) # package to change the shape of the model output
```

```
library(ggplot2) # package for plotting
```

```
initial_number_infected <- 1000000 # the initial infected population size
```

```
initial_number_recovered <- 0 # the initial number of people in  
# the recovered state
```

```
recovery_rate <- 0.1 # the rate of recovery gamma, in units of days-1
```

```
follow_up_duration <- 28 # the duration to run the model for,  
# in units of days
```

```
initial_state_values <- c(I = initial_number_infected, R = initial_number_recovered)
```

```
parameters <- c(gamma = recovery_rate)
```

```
times <- seq(from = 0, to = follow_up_duration, by = 1)
```

```
cohort_model <- function(time, state, parameters) {
```

```
  with(as.list(c(state, parameters)), {
```

```
    dI <- -gamma*I
```

```
    dR <- gamma*I
```

```
    return(list(c(dI, dR)))
```

```
  })
```

```
}
```

```
ode <- ode(y = c(I = initial_number_infected,  
                R = initial_number_recovered),  
          times = times,  
          func = cohort_model,  
          parms = parameters)
```

```
output <- as.data.frame(ode)
```

Based on the output, how many people have recovered after 4 weeks?

```
output[output$time == 28, c("time","R")]
```

What proportion of the total population does this correspond to?

```
output[output$time == 28,"R"]/(output[output$time == 28,"I"]+  
output[output$time == 28,"R"])
```

Plotting the output

```
output_long <- melt(as.data.frame(output), id = "time")
```

```
ggplot(data = output_long, aes(x = time, y = value, colour = variable, group  
= variable)) +geom_line() + xlab("Time (days)") + ylab("Number of people")  
+ labs(title = paste("Number infected and recovered over time when gamma  
=",parameters["gamma"],"days-1"))
```

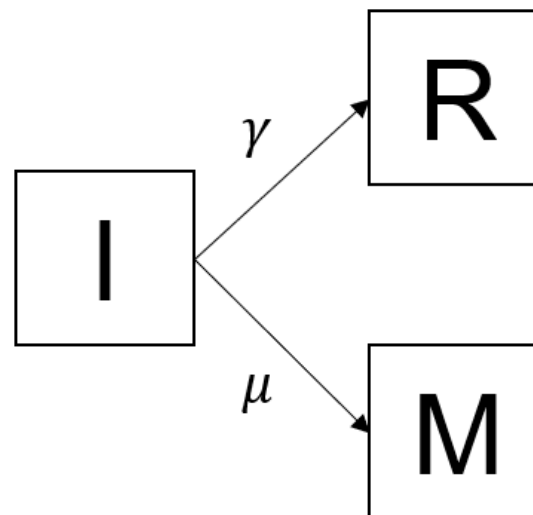
```
ggplot(data = output_long,  
       aes(x = time, y = value,  
           colour = variable, group = variable))  
+geom_line()  
+ xlab("Time (days)")  
+ ylab("Number of people")  
+ labs(title = paste("Number infected and recovered over time when gamma  
=",parameters["gamma"],"days-1"))
```

Simulating competing hazards

The differential equations for this model look like this:

$$\begin{aligned}\frac{dI}{dt} &= -\gamma I - \mu I \\ \frac{dR}{dt} &= \gamma I \\ \frac{dM}{dt} &= \mu I\end{aligned}$$

This corresponds to the following model diagram:



The **model function** describing this set of differential equations looks like this:

```
cohort_model <- function(time, state, parameters) {  
  
  with(as.list(c(state, parameters)), {  
    dI <- -gamma * I - mu * I  
    dR <- gamma * I  
    dM <- mu * I  
  
    return(list(c(dI, dR, dM)))  
  })  
}
```

Defining model input and timesteps

```
initial_state_values <- c(I = 1000000,
```

```
      R = 0,  
      M = 0)
```

```
parameters <- c(gamma = 0.1,      # the recovery rate gamma is 0.1 days-1  
                mu = 0.2)         # the mortality rate mu is 0.2 days-1
```

```
times <- seq(from = 0, to = 4*7, by = 1) # model the course of the  
                                           # infection over 4 weeks = 4*7 days
```

```
# remember that the rates are in units of days-1
```

After 4 weeks, do you expect more people to have recovered or more people to have died, and why?

```
library(deSolve)
library(reshape2)
library(ggplot2)
output <- as.data.frame(ode(y = initial_state_values,
                           times = times,
                           func = cohort_model,
                           parms = parameters))

output_long <- melt(as.data.frame(output), id = "time")

ggplot(data = output_long, # specify object containing data to plot
       aes(x = time,
           y = value,
           colour = variable,
           group = variable)) + # assign columns to axes and groups
  geom_line() +                # represent data as lines
  xlab("Time (days)") +        # add label for x axis
  ylab("Number of people") +   # add label for y axis
  labs(colour = "Compartment") # add legend title
```


Based on the model output, what proportion of the initially infected cohort died before recovering over the 4 week period?

output

output[output\$time == 28,]

output[29,"M"]/1000000

Now use the competing hazards formula to calculate the case fatality rate. Does this agree with your answer to the previous question?

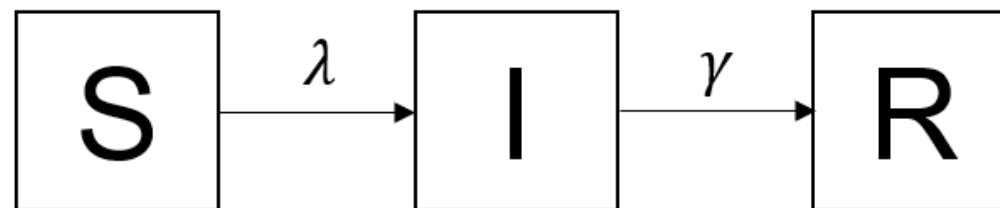
The formula is: $CFD = \frac{\mu}{\mu + \gamma}$

parameters["mu"]/(parameters["mu"]+parameters["gamma"])

SIR model with a constant force of infection

The differential equations for the simple SIR model with a constant force of infection look like this

$$\begin{aligned}\frac{dS}{dt} &= -\lambda S \\ \frac{dI}{dt} &= \lambda S - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$



susceptible (S), infected (I), recovered (R)

λ -Force of infection

γ -recovery rate

Initial number of people in each compartment:

$$S = 10^6 - 1, I = 1 \text{ and } R = 0$$

Parameters:

$\lambda = 0.2 \text{ days}^{-1}$ (this represents a force of infection that's constant at 0.2)

$\gamma = 0.1 \text{ days}^{-1}$ (corresponding to an average duration of infection of 10 days)

We want to run the model for 60 days.


```
output_long <- melt(as.data.frame(output), id = "time")
```

```
ggplot(data = output_long,  
      aes(x = time, y = value, colour = variable, group = variable)) +
```

```
geom_line() +  
xlab("Time (days)") +  
ylab("Number of people") +
```

```
labs(colour = "Compartment")
```

Based on the plot, describe the pattern of the epidemic over the 2 month period. How does the number of people in the susceptible, infected and recovered compartment change over time? After how many days does the epidemic reach its peak? After how many days does it end?

The number of infected people quickly increases, reaching a peak of 500000 infected people after around 7 days, before steadily decreasing again.


The number of recovered people starts to rise shortly after the first people become infected. It increases steadily (but less sharply than the curve of infected people) until the whole population has become immune - by day 53, 99% are in the R compartment, and nearly no susceptible people remain after 60 days.

SIR dynamics with varying parameters

C: Average number of contacts a susceptible makes per day

I/N: Proportion contacts that are made with an infectious person

p: Per contact with an infectious person, probability of a susceptible person becoming infected

$$\lambda = \beta \frac{I}{N}$$


Not usually measured directly, but estimated indirectly from data

We are modelling a disease where every infectious person infects 1 person on average, every 2 days, and is infectious for 4 days.

$$\beta = 1 \text{ person}/2 \text{ days} = 0.5 \text{ days}^{-1}$$

$$\gamma = 1/4 = 0.25 \text{ days}^{-1}$$

Modelling this epidemic for a duration of 100 days

```

library(deSolve)
library(reshape2)
library(ggplot2)

initial_state_values <- c(S=999999,I=1,R=0)
parameters <- c(beta =0.5,gamma =0.25 )
times <- seq(from = 0, to = 100, by = 1)

sir_model <- function(time, state, parameters) {

  with(as.list(c(state, parameters)), {

    N <- S+I+R
    lambda <- beta* I/N

    dS <- -lambda * S
    dI <- lambda * S - gamma * I
    dR <- gamma * I

    return(list(c(dS, dI, dR)))

  })

}

```

```
output <- as.data.frame(ode(y = initial_state_values,  
                           times = times,  
                           func = sir_model,  
                           parms = parameters))
```

```
output_long <- melt(as.data.frame(output), id = "time")
```

```
ggplot(data = output_long,  
       aes(x = time, y = value, colour = variable, group = variable)) +  
  geom_line() +  
  xlab("Time (days)") +  
  ylab("Number of people") +  
  labs(colour = "Compartment")
```

Now imagine an intervention is introduced to control infection, for example infected people are isolated so that they cannot spread infection. As a result, β drops to 0.1. Model the epidemic under this new scenario.

What do you observe?

What do you observe when β is reduced to 0.1 instead, with γ remaining at 0.25?

Under this set of conditions, no epidemic occurs - the number of infected people does not increase following the introduction of a first infectious case.

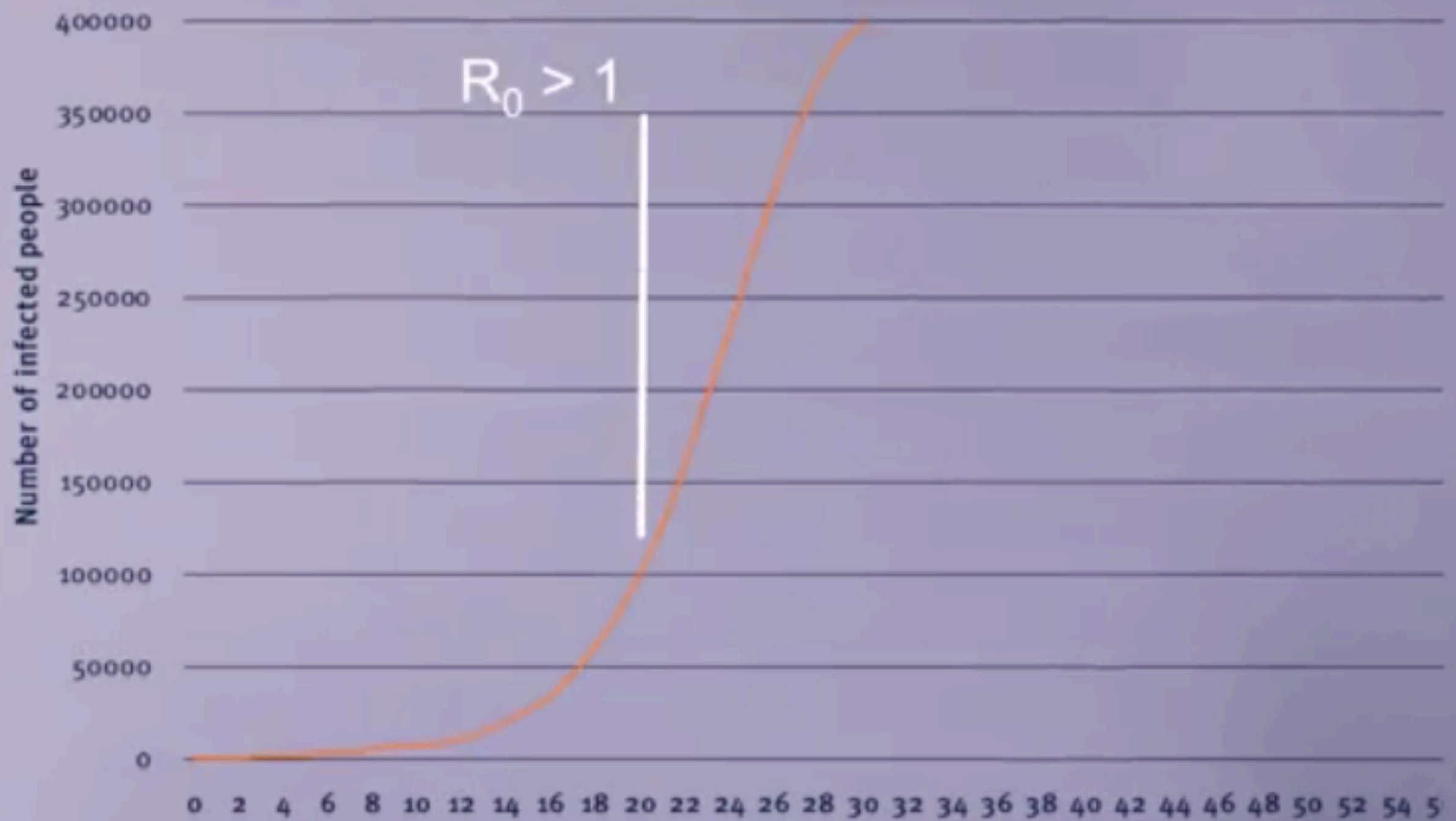
Assuming β equals 0.1, what value of γ do you need in order to get an epidemic? In real life, what could give rise to this change in γ ?

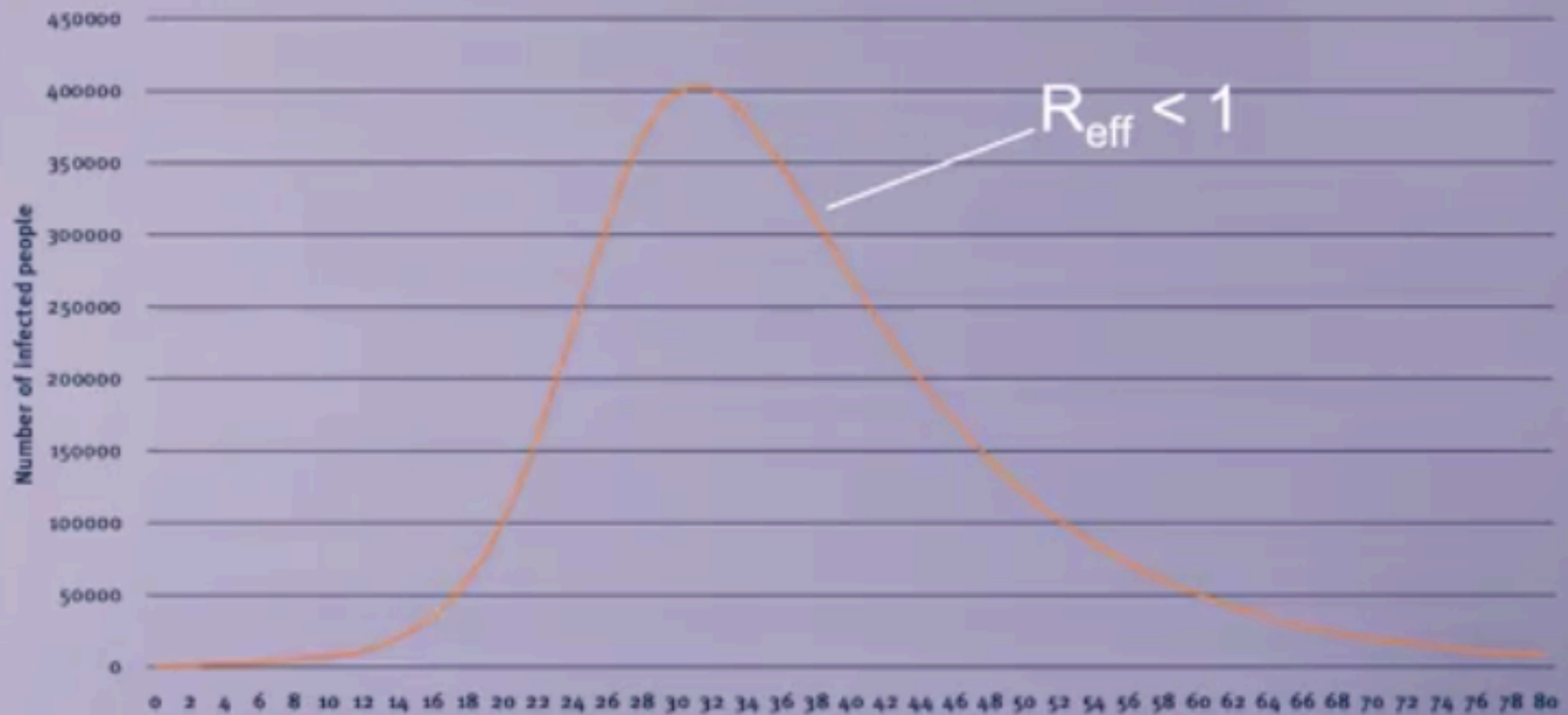
With γ around 0.09 or lower, we start to see a small epidemic if we run the model for long enough (ca. 1000 days). Different mechanisms can lead to such a decrease in the recovery rate, corresponding to an increase of the average infectious period, for example strain evolution of the infectious agent or changes in social behaviour.

Based on your answers to the previous question, can you think of a condition involving β and γ that is necessary for an epidemic? Test this condition using your code above.

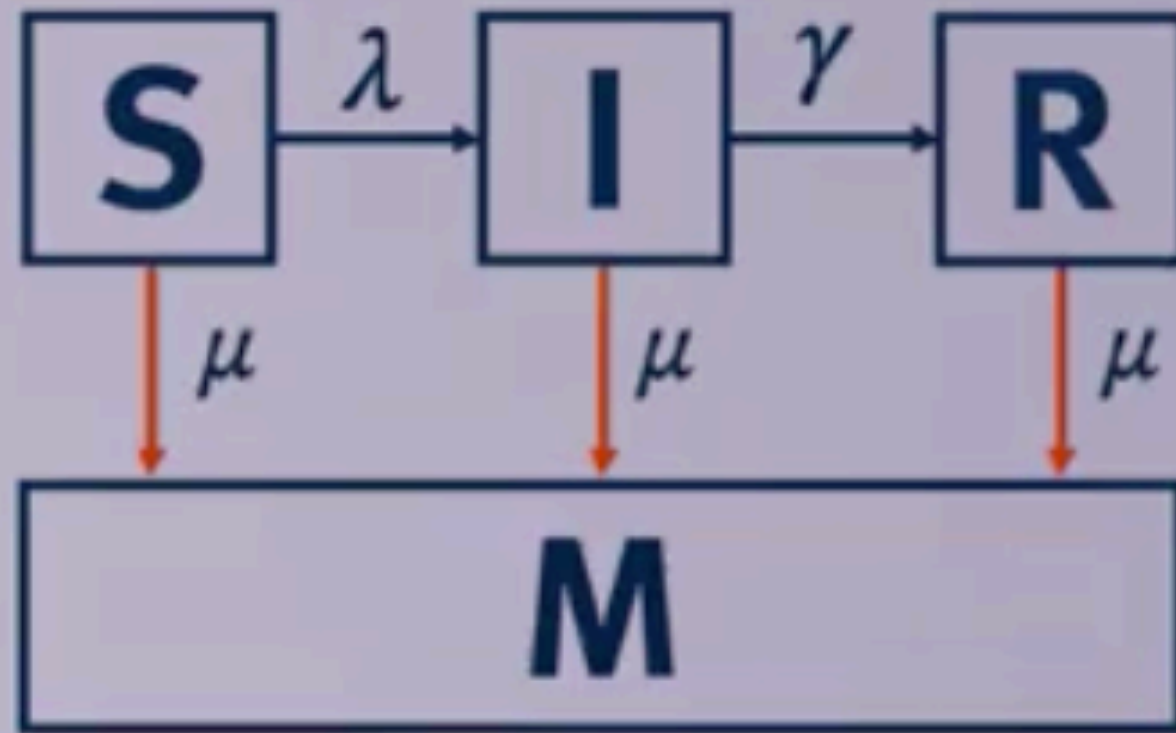
For an epidemic to happen, the ratio β/γ has to be greater than 1.

In other words, to give rise to an epidemic, infectious people have to be infectious enough (β has to be high enough) for long enough (γ has to be low enough) to pass on the pathogen - β has to be higher than γ . Because of the relationship between these two parameters, a low infection rate can still lead an epidemic if infected people are infectious for long enough, as you modelled in the previous question.





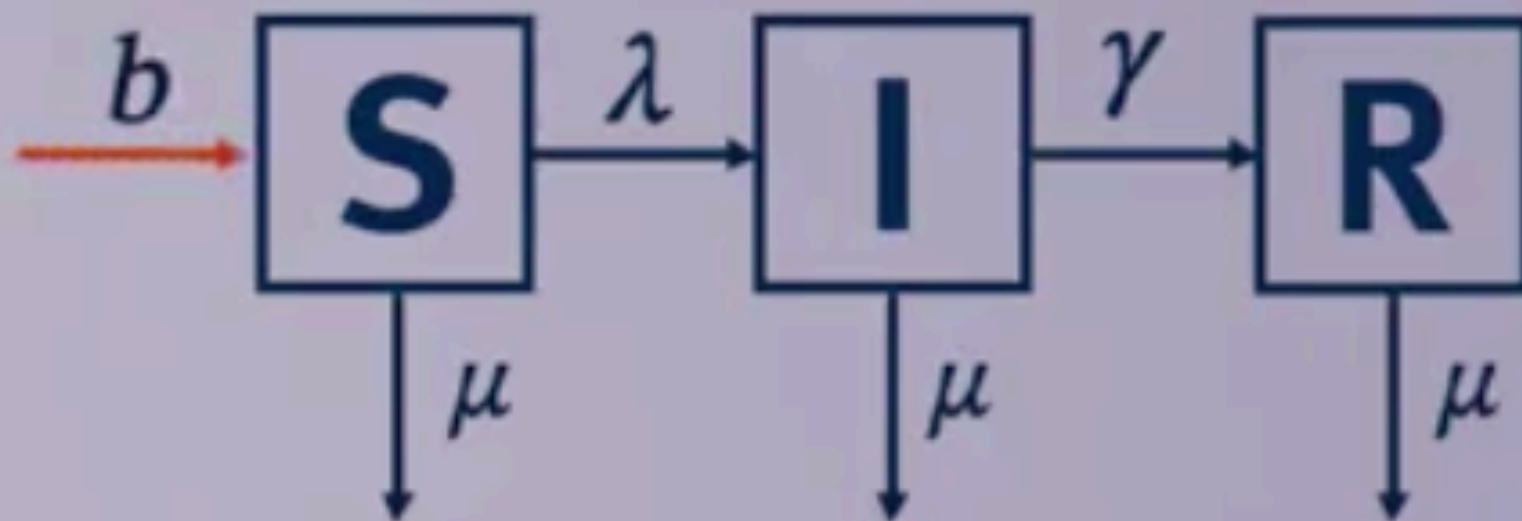
Modelling Population Turnover



$$\frac{dS}{dt} = -\lambda S - \mu S$$

$$\frac{dI}{dt} = \lambda S - \gamma I - \mu I$$

$$\frac{dR}{dt} = \gamma I - \mu R$$



$$\frac{dS}{dt} = -\lambda S - \mu S + bN$$

$$\frac{dI}{dt} = \lambda S - \gamma I - \mu I$$

$$\frac{dR}{dt} = \gamma I - \mu R$$

Vaccination

$$\frac{dS}{dt} = -\lambda S$$

$$\frac{dI}{dt} = \lambda S - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Initial conditions: at time $t = 0$

$$S = (1-p)(N-1), \quad I = 1, \quad R = p(N-1)$$

Does everyone in the population need to be vaccinated in order to prevent an epidemic?

What do you observe if you model the infection dynamics with different values for p ?

Can you explain why?

Modelling a disease where $\beta = 0.4 \text{ days}^{-1}$, $\gamma = 0.1 \text{ days}^{-1}$ and the vaccine coverage $p = 0.5$

```
library(deSolve)
library(reshape2)
library(ggplot2)
```

```
p <- 0.5
```

```
N <- 10^6
```

```
initial_state_values <- c(S = (1-p)*(N-1),  
                          I = 1,  
                          R = p*(N-1))
```

```
parameters <- c(beta = 0.4,  
                gamma = 0.1)
```

```
times <- seq(from = 0, to = 730, by = 1)
```

```
sir_model <- function(time, state, parameters) {  
  with(as.list(c(state, parameters)), {  
  
    N <- S+I+R  
  
    lambda <- beta * I/N  
  
    dS <- -lambda * S  
  
    dI <- lambda * S - gamma * I  
  
    dR <- gamma * I  
  
    return(list(c(dS, dI, dR)))  
  })  
}
```

```
output <- as.data.frame(ode(y = initial_state_values,  
                           times = times,  
                           func = sir_model,  
                           parms = parameters))
```

```
output_long <- melt(as.data.frame(output), id = "time")
```

```
output_long$prevalence <- output_long$value/sum(initial_state_values)
```

```
ggplot(data = output_long,aes(x = time, y = prevalence, colour = variable, group  
= variable)) +geom_line() + xlab("Time (days)") + ylab("Prevalence (proportion)")  
+labs(colour = "Compartment", title = "Prevalence of infection, susceptibility and  
recovery over time")
```

Does everyone in the population need to be vaccinated in order to prevent an epidemic?

What do you observe if you model the infection dynamics with different values for p ?

Can you explain why?

No, not everyone in the population needs to be vaccinated in order to prevent an epidemic.

In this scenario, if p equals 0.75 or higher, no epidemic occurs - 75% is the critical vaccination/herd immunity threshold.

herd immunity describes the phenomenon in which there is sufficient immunity in a population to interrupt transmission. Because of this, not everyone needs to be vaccinated to prevent an outbreak.

What proportion of the population needs to be vaccinated in order to prevent an epidemic if $\beta = 0.4$ and $\gamma = 0.2 \text{ days}^{-1}$? What if $\beta = 0.6$ and $\gamma = 0.1 \text{ days}^{-1}$?

If $\beta = 0.4$ and $\gamma = 0.2 \text{ days}^{-1}$, the herd immunity threshold is 50%.

If $\beta = 0.6$ and $\gamma = 0.1 \text{ days}^{-1}$, the required vaccination coverage is around 83%.

Remember that vaccination changes the effective reproduction number, by reducing the number of people who are susceptible. Based on your answers to the previous questions, can you use the formula for the effective reproduction number R_{eff} to derive a formula for calculating the critical vaccination threshold?

In mathematical modelling terms, herd immunity is just the same as saying that $R_{eff} < 1$. We can derive the herd immunity threshold by solving the formula for R_{eff} for p when $R_{eff} = 1$:

$$R_{eff} = R_0 \frac{S}{N}$$

$$R_{eff} = R_0(1 - p)$$

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

Remember, we can calculate R_{eff} by dividing β by γ .

Thank you