

# Modeling an Infected Cohort

## What is the Model

In this activity, I will create a simple model to find out how long it takes for a cohort of infected people to recover. To do this, I will note 2 distinct populations: those that are infected ( $I$ ) and those that have recovered ( $R$ ). Infected people recover at a rate  $\gamma$  (*gamma*). We also need to make the following assumptions:

- It is equally likely that any person in the cohort recover from the disease
- Infected persons can only recover

The differential equations describing this are:

$$\frac{dI}{dt} = -\gamma I \quad (1)$$

$$\frac{dR}{dt} = \gamma I \quad (2)$$

## Building the Model

We will begin by running the packages necessary.

```
library(deSolve) # to solve the model
library(reshape2) # change the shape of the model output
library(ggplot2) # to provide detailed plotting
```

We will look at a cohort of  $10^6$  infected people, with no one having recovered so far. Let us say that the average duration of infection is 10 days. We want to know how many people will recover from the infection over a 4 week period. Using this information, we will create variables to hold this information and for later analysis.

```
initial_number_infected <- 1000000

initial_number_recovered <- 0

recovery_rate <- 0.1 # the rate of recovery gamma, in units of days-1, so in this case, 10-1

follow_up_duration <- 28 # the duration to run the model for in days, so 7 days times 4 weeks
```

Next, we'll combine the data into objects that are recognized by the deSolve package as **model input**.

```
# the initial state values are stored as a vector and each value is assigned a name
initial_state_values <- c(I = initial_number_infected,
                          R = initial_number_recovered)

# parameters are stored as a vector with names and values
parameters <- c(gamma = recovery_rate) # we only have one parameter, gamma
```

We also need to specify the times, or days in this case, we want the model to run for.

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

Next, we will specify the actual model. We will create a model function with the differential equations specified above with the intention of feeding this function as a parameter into the `ode()` function in `deSolve`.

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

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

    dI <- (-1) * gamma * I
    dR <- gamma * I

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

Now, we can solve this set of equations using the `deSolve` package. We will create a result variable called `output`, which will calculate and store the number of infected and recovered people in each timestep.

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

head(output)
```

```
##   time      I      R
## 1    0 1000000.0    0.00
## 2    1  904837.4 95162.58
## 3    2  818730.8 181269.25
## 4    3  740818.2 259181.78
## 5    4  670320.0 329679.96
## 6    5  606530.7 393469.34
```

Now, we can see a dataframe with columns `time` (containing the time vector), `I` (containing the number of infected people at each timestep), and `R` (containing the number of recovered people at each timestep), from `time = 0` to 28 days.

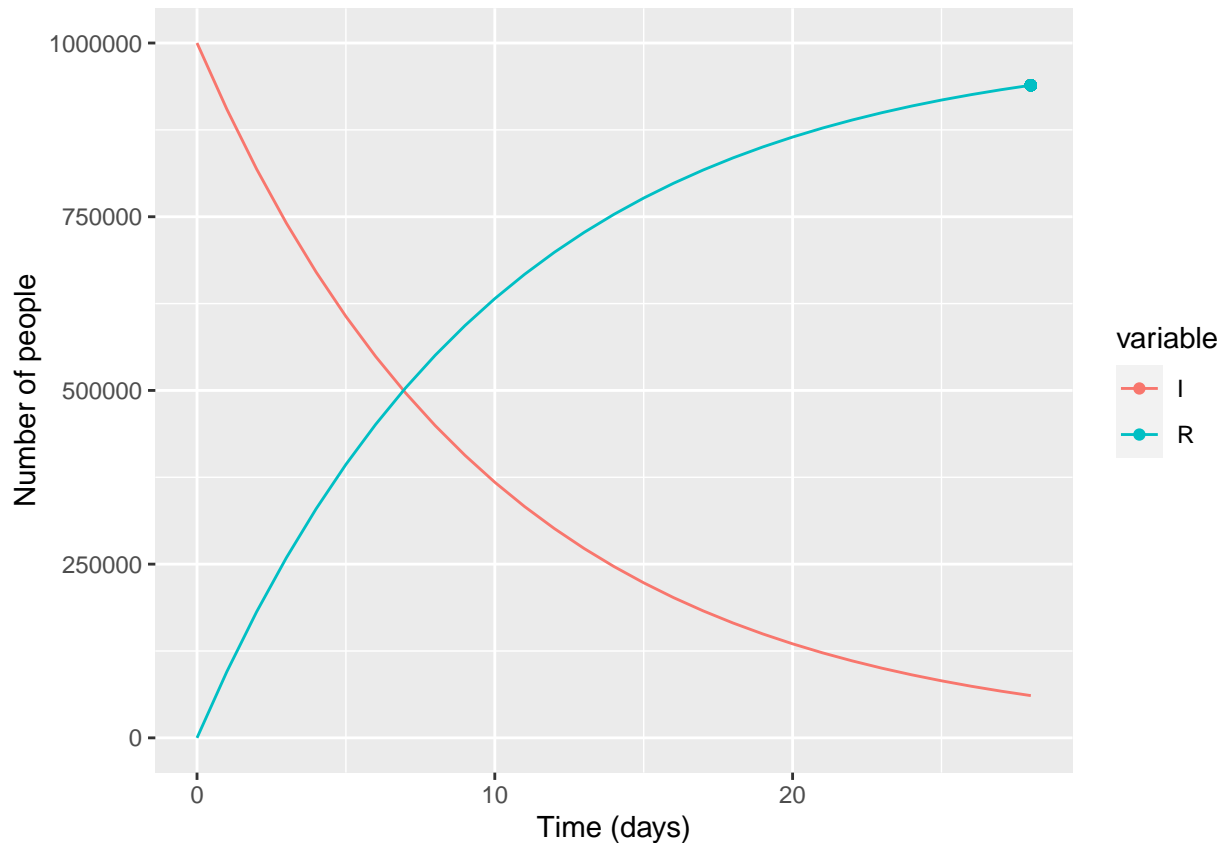
## Questions

**Based on the output, how many people have recovered after 4 weeks? What proportion of the total population does this correspond to?**

We can answer this question by plotting the model output, with time on the x axis and the number of infected and recovered people on the y axis.

```
output_long <- melt(as.data.frame(output), id = "time") # we turn this dataset into a long format,
                                                         # so that the number in each compartment at each time.

ggplot(data = output_long,
       aes(x = time, y = value, group=variable, color=variable)) +
  geom_line() +
  xlab("Time (days)") +
  ylab("Number of people") +
  geom_point(aes(x = 28, y = 939189.94))
```



We can see from the plot above that nearly all members of the cohort have recovered. We can use `output_long` to see exactly how many have:

```
total_recovered <- as.numeric(output_long[output_long['variable'] == 'R' & output_long['time'] == '28'])
total_recovered
```

```
## [1] 939189.9
```

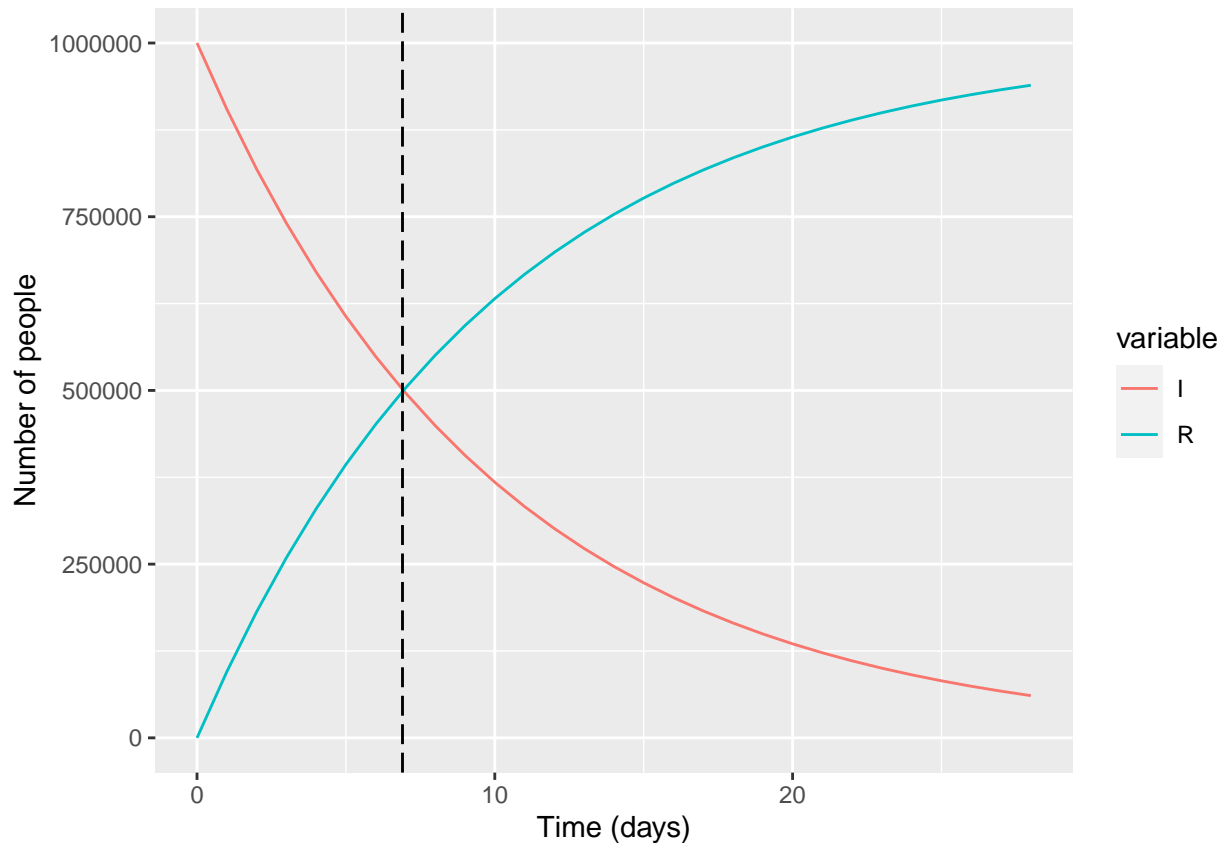
```
proportion <- total_recovered/1000000
proportion
```

```
## [1] 0.9391899
```

We can see 939189.9 members of the cohort have recovered. In terms of a proportion, this can be expressed as  $\frac{939189.9}{1000000} \approx 0.9391899$ .

Based on the plot, at what timepoint were infected and recovered individuals equal in number?

```
ggplot(data = output_long,
       aes(x = time, y = value, group=variable, color=variable)) +
  geom_line() +
  xlab("Time (days)") +
  ylab("Number of people") +
  geom_vline(xintercept = 6.9, linetype= "longdash")
```



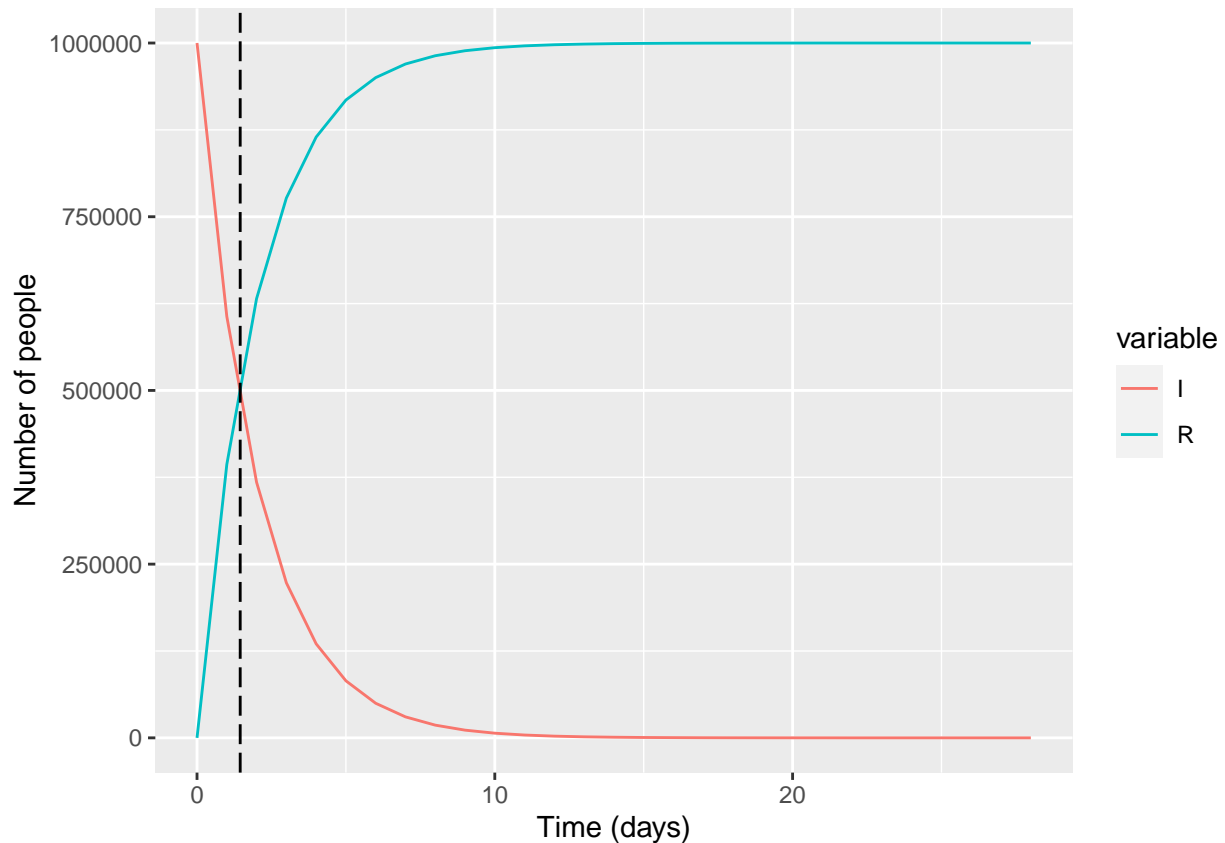
By using the plot, we can see this was at about 7 days, or one week, where there were 500,000 individuals who had recovered and 500,000 who were infected.

However, this is with a  $\gamma$  (gamma) of 0.1. Let's try changing gamma so it corresponds to an average infectious period of 2 days and 20 days. Currently, we were working with a gamma of 0.1 which describes an infectious period of 10 days. We'll start with 2 days.

```
parameters2 <- c(gamma = 0.5)
output2 <- as.data.frame(ode(y = initial_state_values,
                             times = times,
                             func = cohort_model,
                             parms = parameters2))

output_long2 <- melt(as.data.frame(output2), id = "time") # we turn this dataset into a long format, so

ggplot(data = output_long2,
       aes(x = time, y = value, group=variable, color=variable)) +
  geom_line() +
  xlab("Time (days)") +
  ylab("Number of people") +
  geom_vline(xintercept=1.45, linetype= "longdash")
```



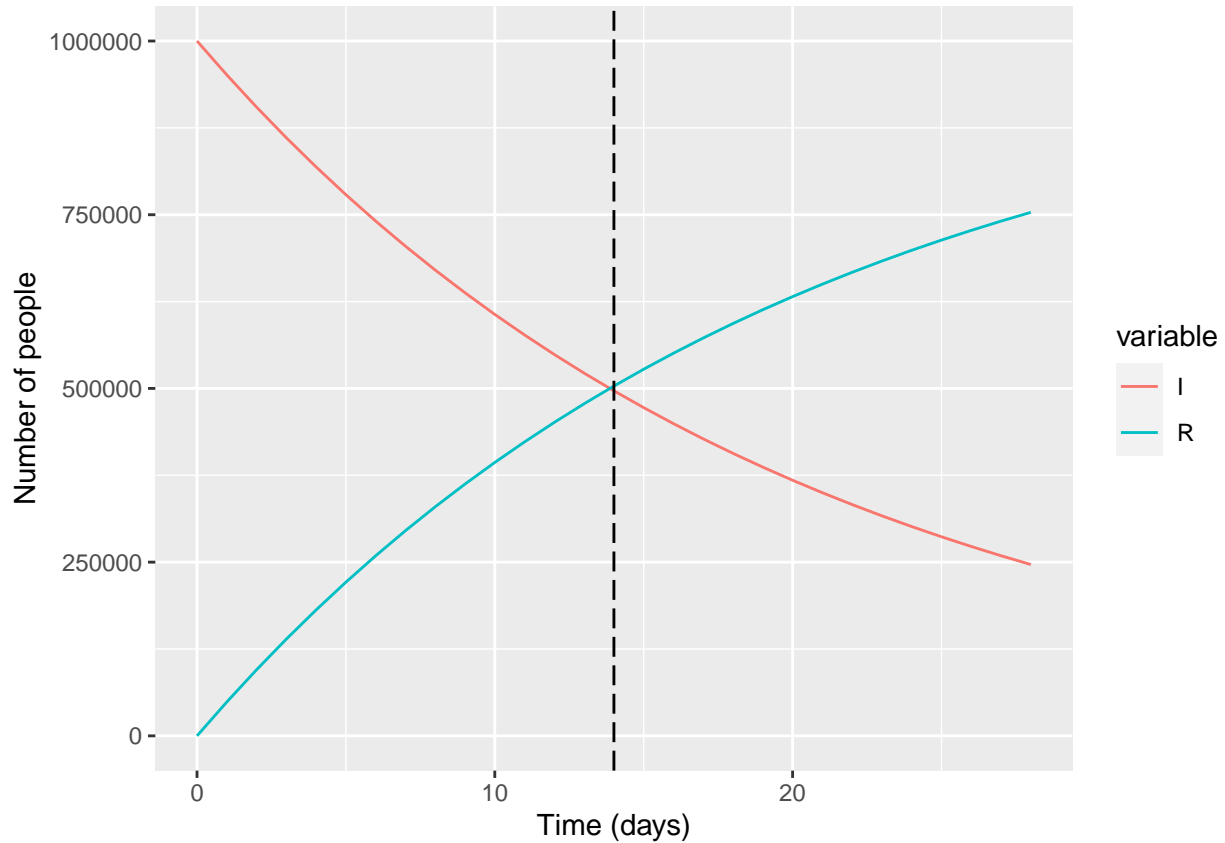
When we use a gamma of 0.5, which corresponds to an infectious period of 2 days, we see that most people recover much more quickly. We also see that there were 500,000 individuals who had recovered and 500,000 who were infected at only 1.5 days.

Now, let's try using a gamma of 0.05, which corresponds to an infectious period of 20 days.

```
parameters20 <- c(gamma = 0.05)
output20 <- as.data.frame(ode(y = initial_state_values,
                             times = times,
                             func = cohort_model,
                             parms = parameters20))

output_long20 <- melt(as.data.frame(output20), id = "time") # we turn this dataset into a long format,

ggplot(data = output_long20,
       aes(x = time, y = value, group=variable, color=variable)) +
  geom_line() +
  xlab("Time (days)") +
  ylab("Number of people") +
  geom_vline(xintercept=14, linetype= "longdash")
```



We see here that unlike the previous two examples, the full cohort is not recovered after 28 days. We also see that there were 500,000 individuals who had recovered and 500,000 who were infected at 2 whole weeks, which makes sense given that our gamma value is very small in comparison.

**What changes do you observe in the transition to the recovered compartment if  $\gamma$  is higher or lower? For example, how long does it take for everyone to recover in both cases?**

If  $\gamma$  is higher, we see an infectious period that is shorter and this a transition to the recovered compartment faster. We see this when we use a  $\gamma = 0.1$  (infectious period of 10 days) and  $\gamma = 0.5$  (infectious period of 2 days). In contrast, if  $\gamma$  is lower, we see an infectious period that is longer and a transition to the recovered compartment slower. We see this in the last example where  $\gamma = 0.05$  (infectious period of 20 days), where the 28 day timeframe is not long enough for the full cohort to recover.