Worksheet 1: Basic epidemiological models

u1624423

January 25, 2022

All plots and calculations are performed in `Assignment1.m`

Question 1

(a)

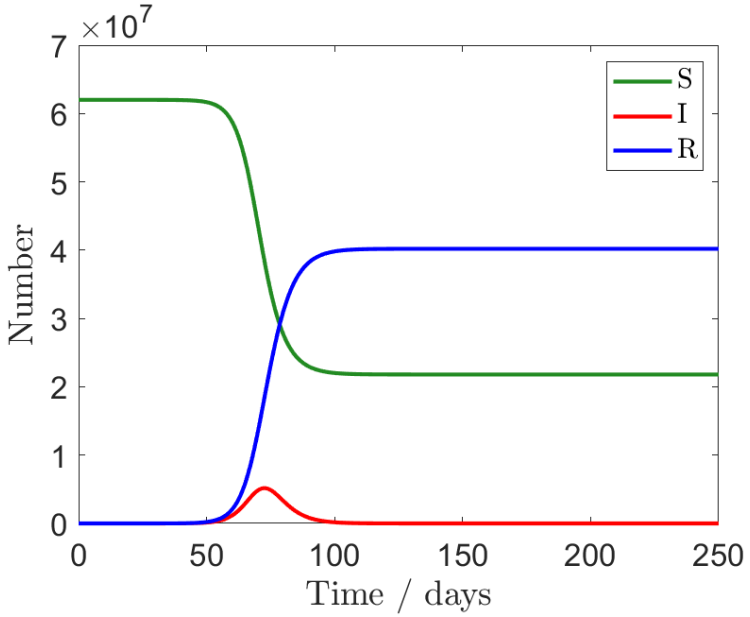

Figure 1: Infected and recovered people in the first 250 days of the epidemic.

(b)

(i) We can use the implicit final size equation, given by:

$$N - R_\infty = (N-1)\exp\left(-\frac{R_0 R_\infty}{N}\right)$$

From here, we can use "fsolve" to numerically solve for R_∞, giving us $R_\infty = 40,198,770$. Alternatively, and more conveniently, we can just use the final number of recovered individuals at the end of the simulation, which assumes that no infected individuals remain.

$$\text{Final size} = R(\text{end}) = 40,198,770$$

These values are, of course, rounded to the nearest whole number.
(ii) By first principles:

$$R_0 = \frac{\beta}{\gamma} = \frac{0.62}{\left(\frac{1}{2.6}\right)} = 1.612$$

(iii) Using "find", we can find the last day with any given condition. Here, we want the last day with at least 1 infected person, i.e. $I(t) \geq 1$. From here, we must ceiling this value because, for example, the 1.5^{th} day would be on the 2^{nd} day. To counteract having to do this, when using "ode45" in the model, we can simply split the date range into integer values corresponding to days. See `ODE_SIR_model.m`.

$$\text{Duration} = 180 \text{ days}$$

(c)

The equation for $R_e(t)$ is:

$$R_e(t) = \frac{\beta S(t)}{\gamma N}$$

This is also equivalent to the force of infection (FOI) divided by $I\gamma$:

$$\text{FOI(t)} = \frac{\beta S(t) I(t)}{N}$$

$$R_e(t) = \frac{\text{FOI(t)}}{\gamma I(t)}$$

Plotting $R_e(t)$ gives the following:

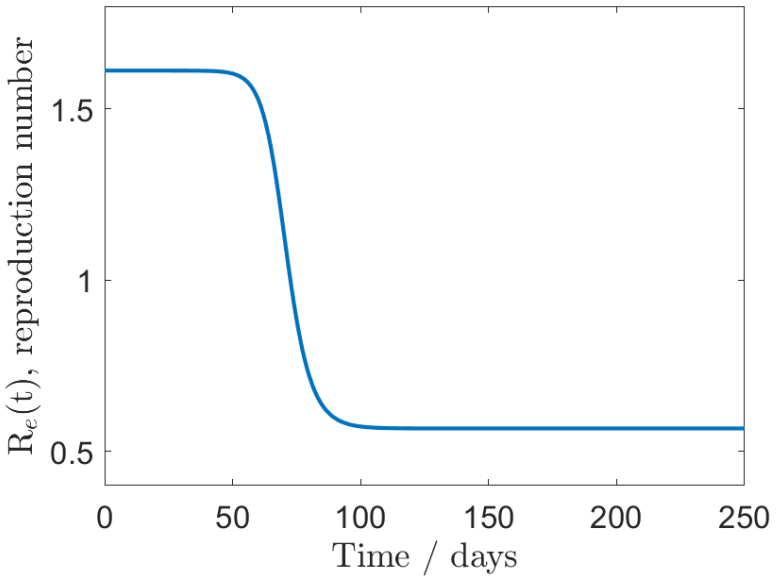

Figure 2: Change in effective reproductive number over time.

(d)

There are multiple ways to calculate $r(t)$. The first being:

$$r(t) = \frac{dI(t)}{dt} = \frac{\beta S(t) I(t)}{N} - \gamma I(t)$$

The second is:

$$r(t) = \gamma I(t)(R_e(t) - 1)$$

Both of these give the same plot:

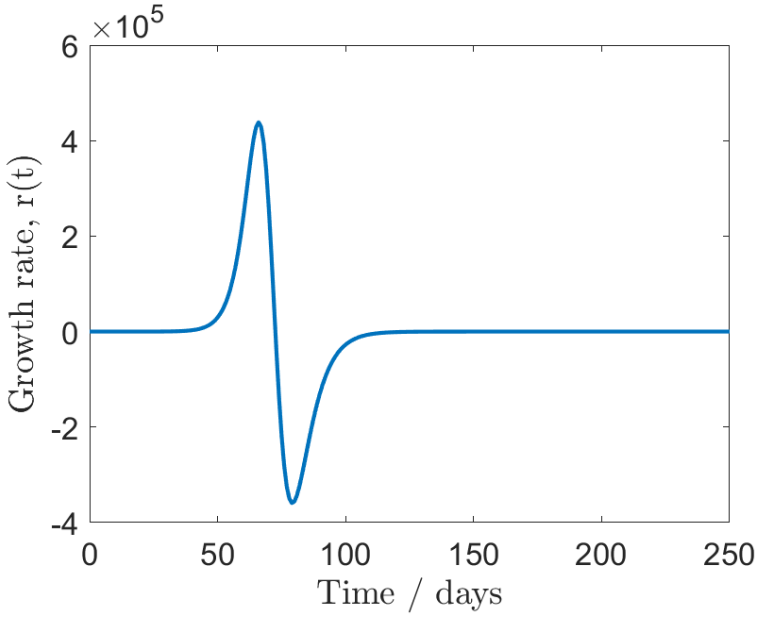

Figure 3: Change in growth rate over time.

Question 2

(a)

The parameter "p" represents the probability that an infected individual has symptoms that are severe enough to be reported. It is the effective proportion of the infected population that has such symptoms.

The new initial conditions given 13% of the population has prior immunity are:

$$(S, I_r, I_n, R) = (0.87N - 1, 0, 1, 0).$$

(b)

The new SIR model:

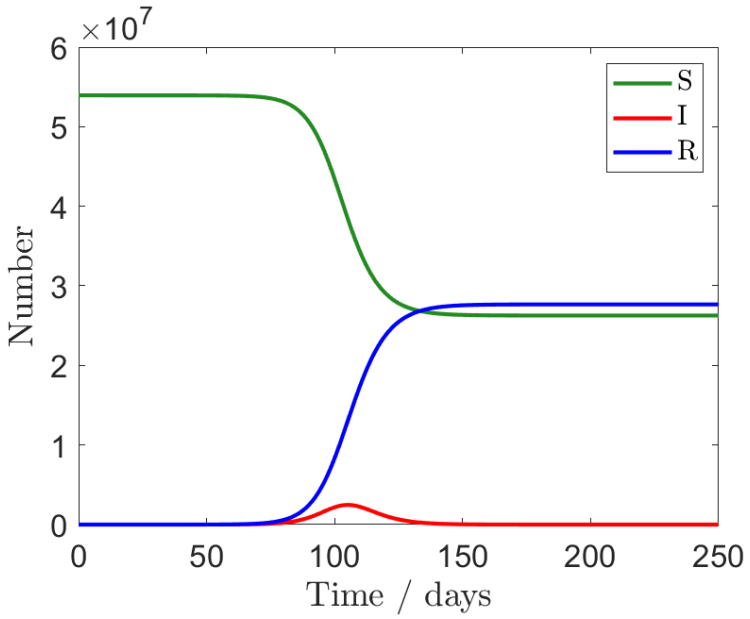

Figure 4: Infected and recovered people in the first 250 days of the epidemic, given an initial 13% immunity.

To compute daily reported incidence, there are a few ways. One way to calculate the incidence in new infections between two time points t_1 and t_2 can be computed by integrating over the force of infection term:

$$\text{New infections between } t_1 \text{ and } t_2 = \int_{t_1}^{t_2} \frac{\beta SI}{N}\, dt$$

So from here, we can deduce that given a day, t, taking p into account:

$$\text{Daily incidence} = \int_{t-1}^{t} \frac{p\beta SI}{N}\, dt$$

Which can be evaluated using "trapz".

Alternatively, we can use the fact that the total amount of infections is cumulative and we can instead calculate the cumulative infections using "cumtrapz" to evaluate the total reported cases, "TRC":

$$\text{TRC} = \int_{t_1}^{t_{fin}} \frac{p\beta SI}{N}\, dt$$

Where t_1, t_{fin} are the first and last days of the simulation, respectively.
From here, given a day, t:

$$\text{Daily incidence}(t) = \text{TRC}(t) - \text{TRC}(t-1)$$

Both of these produce the following:

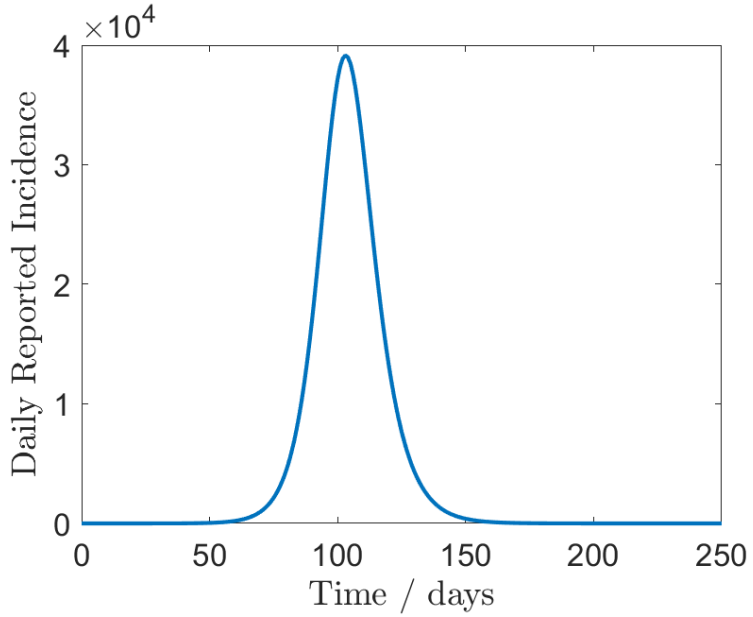

Figure 5: Daily reported incidence over 250 days of the epidemic.

(c)

The cumulative number of reported cases is given by TRC, as described above.

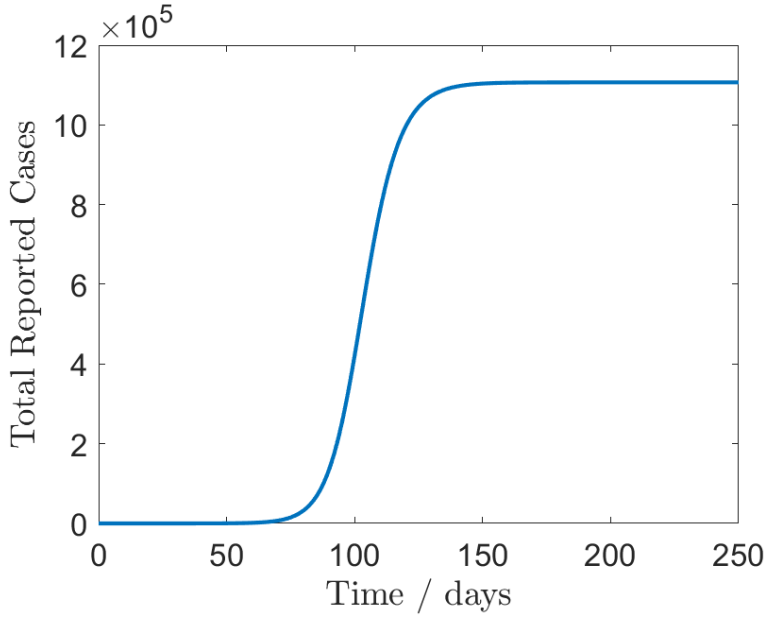

Figure 6: Total reported cases (cumulative) over 250 days of the epidemic.

Here, `ODE_SIR2_model.m` is used to include p.

Question 3

(a)

The data in "SwineFluCaseData.csv" spans from the 24th week of 2009 to the 1st week of 2010, so for convenience, I have changed the week number of week 1 of 2010 to be effectively the 54th week of 2009. Then 24 is subtracted from these when plotting so we can start at week 0 of the epidemic.

The model in question 2 only has one peak in infections, whereas in the data given , there are 2 peaks, with the first of the two being higher and having a faster increase in rate of infection and the second being lower than the first and with a lower rate of infection. This implies that at the first peak, some factor likely reduced the contact rate, β, such as a lockdown.

This also affects the total number of infected people, the total at the end of the data is 906,497, whereas in the model, it is 1,106,718. (N.B. these are reported cases, which are only 4% of the total)

This suggests whatever factor affected β significantly reduced the amount of people to become infected. The peak in the model also happens between the two peaks in the data. This is likely because the model has a constant β and so is distributed more centrally within the epidemic's time period.

(b)

We first modify our SIR model to accept a specific date range rather than just an endpoint. Then the daily incidence can be calculated for each of the 3 periods of time.

Although the infection starts on week 17, i.e. day 119, it is best to shift our plot to have time start at the point the infection starts. It is important to note that when simulating each individual period of time, we use the final conditions of the previous period of time as the initial conditions. Here, `ODE_SIR3_model.m` is used to have a start and end time rather than starting at 0.

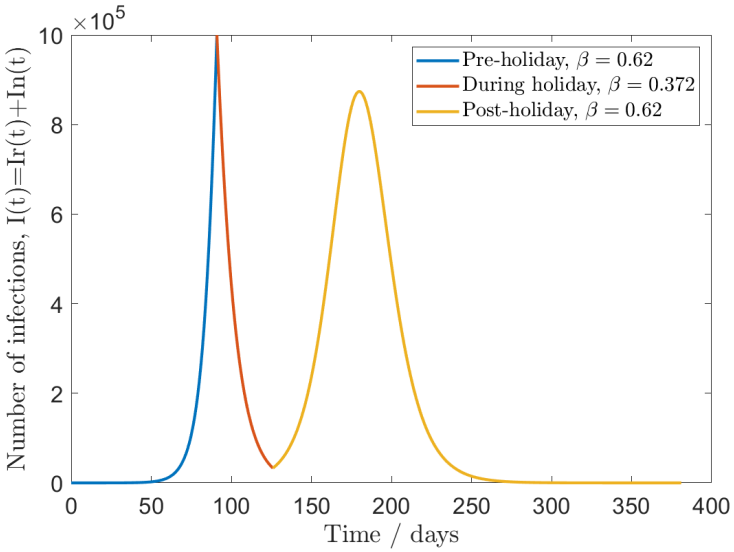

Figure 7: Daily infection incidence (total).

(c)

$R_e(t)$ is affected as such:

$$R_e(t) = \begin{cases} \frac{\beta S(t)}{\gamma N} & \text{pre-holiday} \\ \frac{0.6\beta S(t)}{\gamma N} & \text{during holiday} \\ \frac{\beta S(t)}{\gamma N} & \text{post-holiday} \end{cases}$$

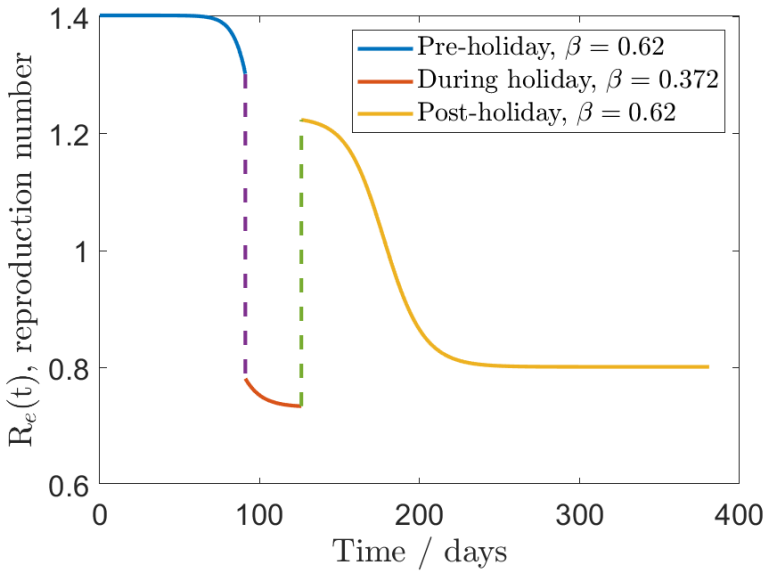

Figure 8: Change in effective reproduction number with holiday.

(d)

We can work out the duration and final size of the outbreak the same way as in question 1.

$$\text{Final size} = 23,106,160,$$

$$\text{Duration} = 376 \text{ days (54 weeks)}$$

It is important to note, the epidemic ends on day epidemiological day 495, but started on day 119, so $495 - 119 = 376$ days is the duration. The number of infected people has been reduced from 57.6% to 50.3% of the total population due to this holiday.

(e)

To plot weekly reported infection incidence, we can either add up each 7 days of our model or we can perform one of the two methods outlined in Question 2 (b), but with a time interval of 7 days instead of just 1.

Either gives the following plot for weekly incidence (the data in "SwineFluCaseData.csv" has also been plotted):

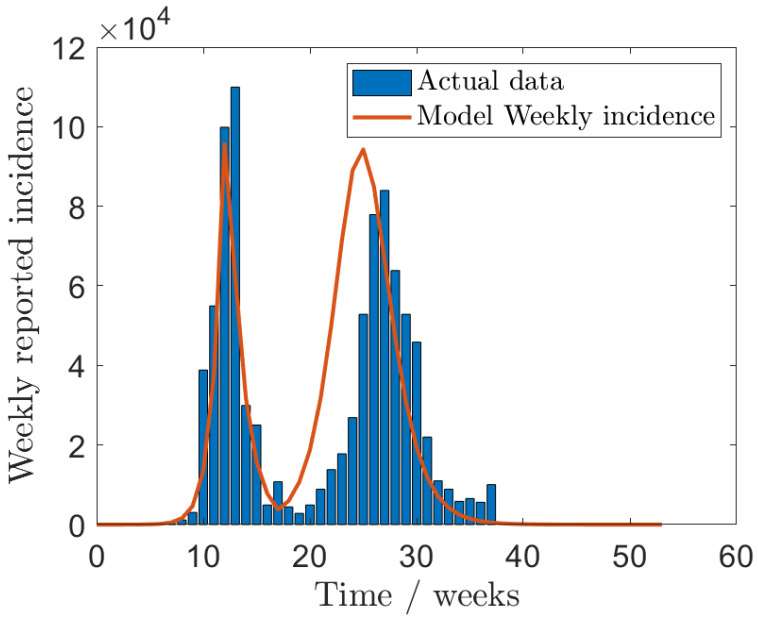

Figure 9: Model's weekly reported incidence compared to collected data.

There are clear differences between the data and the model. The first peak happens at about the same time for both, although the data has a larger peak. The second peak happens too early in the model and overestimates the amount of cases. There could be many reasons for this, but it is most likely that reporting cases was delayed by various factors. Towards the end of the second peak, the data and the model seem to fit each other a bit better. This may be because case reporting may have become more efficient at some point around the second peak. A reason that the second peak might be lower than the model could be that a significant proportion of people decided to self isolate, which would reduce β. The first peak matches up quite well, this is likely because generally, no extra precaution was taken by individuals and so β was closer to the estimated value in the model.

A general problem with the SIR model is that β and γ are assumed to be constant for long periods of time, but many factors can affect these, ranging from people deciding to self isolate, to people wearing masks, more frequently washing their hands, etc. Furthermore, different people will have different reactions to the disease and birth and death rates are not accounted for. It is highly likely that some people could decide not to report when they have symptoms and this would reduce the recorded reported number of cases in the data. This could be more likely around the time of the second peak because people may be fed up and not care about reporting if they are sick. During the first peak, people could be in panic and some people could have a different disease with similar symptoms and that could be reported as a case, this could be why the first peak in the data is higher than the model.

Question 4

(a)

The herd immunity threshold (HIT) is the proportion of the population required to be immune to a disease for the spread of the disease to decline. It is given by:

$$\text{HIT} = 1 - \frac{1}{R_0}$$

In questions 1, 2 and 3 (when not in holiday), HIT $\approx 0.38 = 38\%$ (of N). The proportion of the infected population above the HIT is the "overshoot". This threshold can also be calculated from the point where the total number of infected people has plateaued, we can find $S = \frac{N}{R_0}$ here. It is also the point at which $R_e = 1$. If 38% of the population was immune at the start of the epidemic, then there would be no outbreak.

The difference between the models in questions 1 and 2 is the 13% initial immunity in question 2 (this is besides the discrepancy between reported and unreported cases).

If we plot the population immunity as a percentage for the cases in questions 1 and 2, we can see the difference in the overshoot.

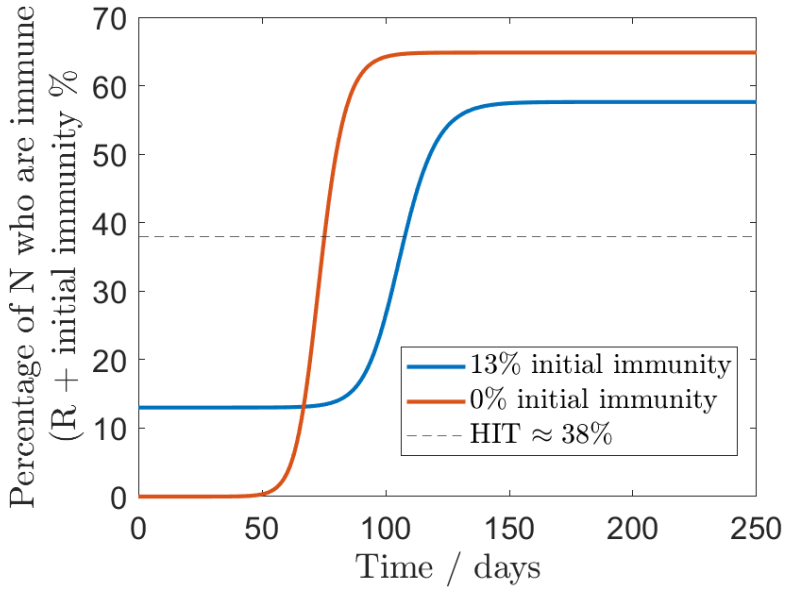

Figure 10: Immune proportion of population with and without initial immunity.

With a 13% initial immunity, there is about a 20% overshoot, with no initial immunity, there is about a 27% overshoot. Our model assumes a recovered individual is immune. Having initial immunity reduces the total proportion of the population that is immune and will lead to fewer infections.

If we now consider the holiday mentioned in question 3, we can produce the following plot:

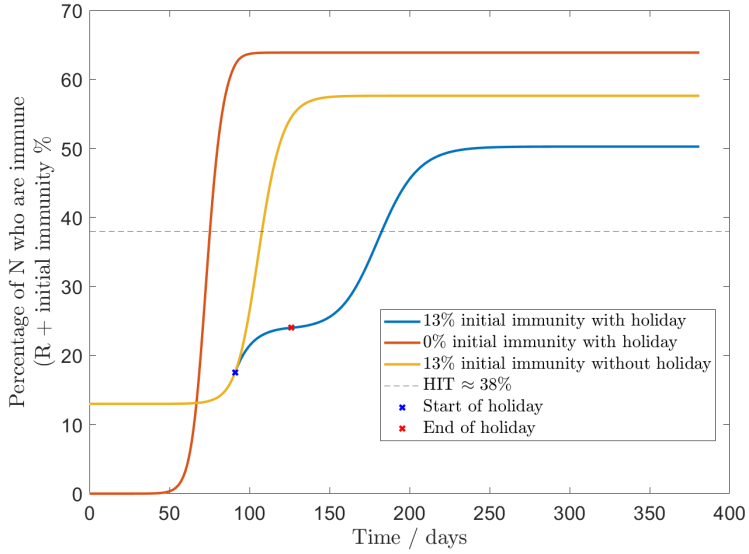

Figure 11: Immune proportion of population considering various factors.

We can see the effect of the holiday on the proportion of the population that is immune. The overshoot is now close to 12%, a quite significant reduction. This is given that there is still the initial 13% immunity. However, without this initial immunity, the overshoot is about 26%, a reduction of 1% from when there is no holiday.

Therefore it can be concluded that a reduction in contact rate is most effective at reducing overshoot with a somewhat significant proportion of the population already being immune.