Risk analysis of predicted case numbers and hospitalisations due to an outbreak of the new Omega strain of the SARS-CoV-2 virus

Student 202078239

1 Introduction

Covid-19 (CV-19) is a disease caused by the SARS-CoV-2 virus.[1] It was initially discovered in Wuhan in December 2019, and despite speculation there is still no evidence as to exactly when and where it originated, and how it spread to humans.[2] Despite a city lockdown in Wuhan, the disease had already spread uncontrollably and a global pandemic was declared by the World Health Organisation on 11 March 2020.[3] The virus causes flu like symptoms, with key symptoms for detection being a high temperature, a continuous cough and a loss or change of taste and smell. Some people require hospitalisation and a proportion of those need ventilation. Two vaccines were approved for use in the UK on the 2nd and 30th December, 2020.[4] Vaccine rollout began shortly after on 8th December, 2020.[5]

After stabilising numbers of CV-19 over previous months, there has been a cluster of 20 cases of a new variant found in a group in Scotland who recently travelled from Europe. Little is known about this strain, recently named Omega, although it is believed to be more transmissible. In Europe, where it originated from, the virus symptoms appear to be milder than those of previous strains. However, the virus has so far been found mainly in young adults with typically strong immune systems. Given the threat of increased transmissibility, there is concern as to how the new strain may affect the UK population which has an older demographic and may be more vulnerable to its effects. 80% of the UK population over 12 years have been double vaccinated against CV-19 (by 12th November, 2021),[6] however there are concerns that the current vaccines may not protect as well due to Omega's mutations. Modelling the spread of this new strain and the number of people requiring hospitalisation for it, is therefore crucial for deciding how to manage the new outbreak and if further restrictions will need to be put in place to avoid overwhelming the NHS.

2 Method

To understand the anticipated demands on the NHS, the likely number of infections will be modelled for the first 30 days after outbreak. In the early stages of an epidemic the number of infected cases at time t can be modelled using equation 1.

$$I(t) = I_0 exp(rt) \tag{1}$$

$$r = (R_0 - 1)/\gamma \tag{2}$$

I_0 is the initial number of cases in the population, r (equation 2) is the growth rate of the infection, R_0 is the basic reproductive number which is the expected number of secondary cases that each primary case causes, and γ is the average duration of infectiousness. It is assumed γ=12 days for CV-19.

If $R_0 > 1$, epidemic spread can occur. Data for numbers of secondary cases are available for each of the 20 initial cases, from testing all contacts. A bootstrapping method will be used to quantify a level of uncertainty around the median value of R_0, using the mean of the median values, since the data are skewed (figure 1). The bootstrapping method is a computer intensive method of re-sampling (with replacement) from observed datasets. The resampling gives us B bootstrap estimates of an estimated parameter, θ from which we can form a distribution of uncertainty about the parameter. The R_0 data are not known to follow any particular distribution so a non-parametric bootstrapping function will be used. The dataset is small so non-parametric bootstrapping may under-estimate the variability, especially if the data doesn't capture the tails of the distribution.

A distribution for the predicted number of infected cases, $I(t)$ over the first 30 days will be derived using equation 1 and the distribution of R_0 that is formed from the bootstrapping method. The percentage of CV-19 cases requiring hospitalisation is uncertain and there is no data for this strain, however, medical experts believe that between 25% and 35% may need hospitalisation, and that the most likely figure will be 30%. A beta distribution, fitting these requirements, will be used to model the expected proportion of CV-19 cases requiring hospital beds. Thus a second-order distribution will be modelled using Monte Carlo simulation, which will randomly sample from the beta distribution (quantifying the uncertainty) and input each of these random variables (each representing a random proportion needing hospitalisation) into a binomial distribution (quantifying variation) which models the number of CV-19 cases.

The effect of changing γ, the proportion needing hospitalisation and the bed-threshold values will then be explored to give further insight into how best to manage the new Omega strain.

3 Results

Figure 1: Histogram to show secondary cases from each of the 20 initial cases.

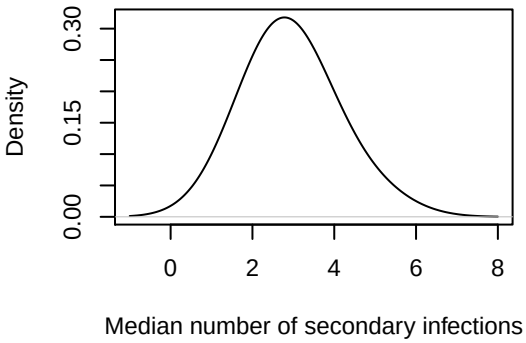

Figure 2: Distribution of median secondary cases from an initial case.

Figure 1 shows that the observed data demonstrates significant individual variability in transmission rates. 7 of the 20 initial cases spread the virus to at least 5 other people and 2 of these cases can be considered super-spreaders, transmitting the virus to 8 and 9 other people respectively. However, 9/20 of the cases spread to 2 or fewer other people.

Figure 2 shows the distribution of the median number of secondary cases from the bootstrapping samples. The mean of the median values of R_0 is 2.95 with a standard deviation of 0.81. The minimum and maximum median values for R_0 are 2.0 and 5.0 and these values are the same as in the 90% credibility interval for the median R_0. To narrow down the range we need to go to the 80% credibility interval, for which we are 80% certain that the median lies within the range of 2 to 4 secondary cases.

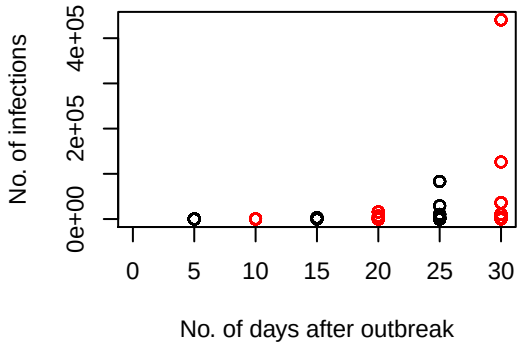

Figure 3: Distribution of infection numbers at 0 to 30 days after outbreak, (median bootstraps).

Table 1: Number of CV-19 infections for 10, 20 and 30 days

Days	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	95% interval
10	46.0	69.8	105.9	133.7	105.9	560.6	(46.0, 560.6)
20	105.9	243.7	560.6	1720.7	560.6	15715.4	(105.9, 15715.4)
30	243.6	850.4	2968.3	35463.5	2968.3	440529.3	(243.6, 440529.3)

Figure 3 shows how the distribution of infection cases changes over time. Table 1 shows summary statistics for the number of infections at 10, 20 and 30 days. Firstly it is noted that the mean and median values differ from each other for each time period, showing the distributions are skewed. Whilst the minimum values at 10, 20 and 30 days are within about 200 cases of each other, the maximum values over the tabled 20 day interval differ hugely, from 561 to 440,529 cases! It is also noted that the 95% credibility intervals increase significantly over the time period from a range of about 515 cases at 10 days to about 16610 at 20 days to about 440,285 cases after 30 days, so our levels of uncertainty are high and increase exponentially over time.

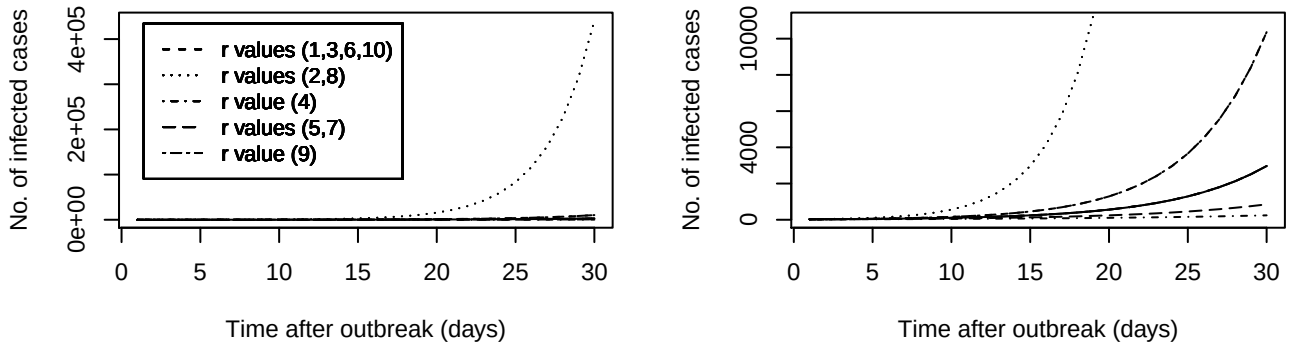

Figure 4: First 10 predicted distributions of CV-19 cases up to 30 days after outbreak (median bootstraps). There are a few with matching distributions. The 2nd plot magnifies the lower values.

Figure 4 represents ten predicted projections of the number of covid cases up to 30 days after outbreak using the first 10 median values from the R_0 distribution. Once again, we can see that the variation of case numbers is small after 10 days but has increased significantly after 20 days and there is a huge variation at 30 days after the outbreak.

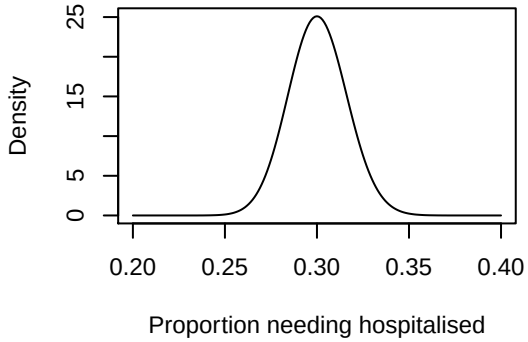

Figure 5: Beta distribution for the proportion of cases needing hospitalisation.

Figure 5 shows the beta distribution for the proportion of CV-19 cases that experts believe may require hospitalisation. The beta distribution was then used in a Monte Carlo simulation, along with a binomial distribution for the number of CV-19 cases. Since the data are skewed, the median value of the number of CV-19 infections at 30 days was chosen for use in the simulation.

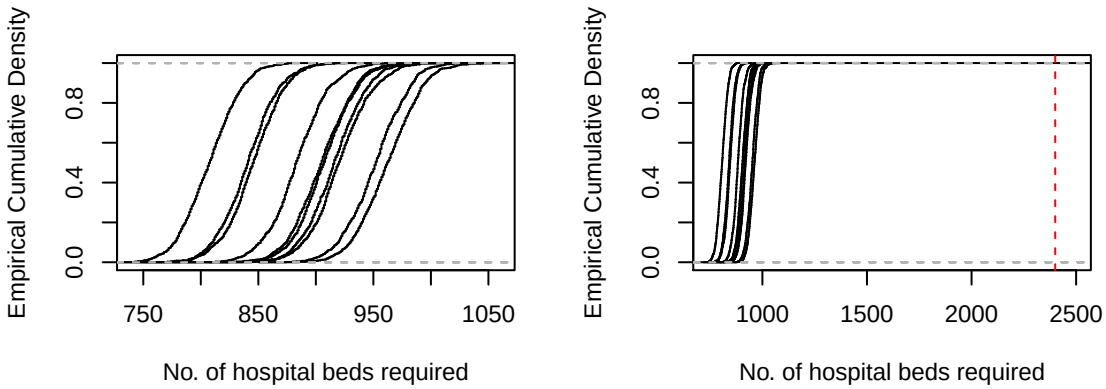

Figure 6: Predicted number of cases requiring hospitalisation 30 days after outbreak. Dashed red line shows bed-threshold of 2400.

Figure 6 shows the simulated number of hospital beds required for CV-19 cases 30 days after outbreak, assuming the median value of 2968 infected cases. Each individual curve represents the variation of predicted hospitalisations and the difference between these curves represents the level of uncertainty we have, which can be reduced if we can access more data. A red dotted line indicates the bed-threshold value, when demand for beds exceeds supply. For the 10 plotted distributions in figure 6, the probabilities of exceeding the bed-threshold value are all 0%. This uncertainty will not impact capacity

planning. However, in figure 4, there is an outlying distribution showing extreme growth in infection cases, simulating super-spreaders with an R_0=5. Consequently, 6.4% of the simulations lead to 440,529 projected cases at 30 days, which is a significant cause for concern, and by taking the median values from our data, these extreme super-spreaders, and likewise the large numbers with very low R_0 values are not being represented at all. Despite having skewed data, it is believed that using the mean for bootstrapping rather than the median would balance out the representation of the more extreme tails of the dataset and provide more conservative estimates of infection cases. Further investigations will therefore use the mean values for bootstrapping.

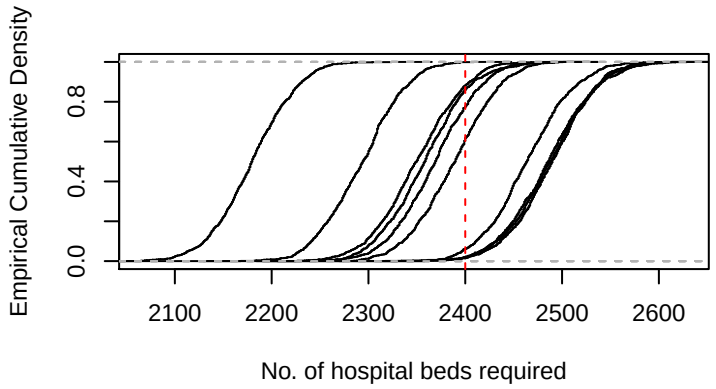

Figure 7: Predicted number of cases requiring hospitalisation 30 days after outbreak, using mean bootstrapping. The red dashed line shows the bed-threshold value of 2400.

When bootstrapping mean values of R_0, a median value of 8068 at 30 days is used in the model. This leads to the distributions of hospital bed requirement shown in figure 7. For the 10 distributions shown, the probabilities of exceeding the bed-threshold of 2400 are 84.2%, 100%, 93.2%, 99.4%, 45.4%, 99.5%, 0%, 68.2%, 67.6% and 77.0%, leading to a mean probability of exceeding the bed-threshold of 73.4% but the 95% credibility interval is 10.2% to 99.9%! This contrasts the 0% probability projected using the median values for bootstrapping. This level of uncertainty is highly likely to impact on capacity planning and the increase in cases in the early days should be very closely monitored to update models. The only way to reduce this uncertainty is to include more observed data in the models as soon as it becomes available.

The γ parameter in the model has been set at 12 days in the study so far. However, γ is difficult to measure and it is uncertain if people become infectious before, after or at the onset of symptoms. It is believed that the value also varies depending on how severe the patient's symptoms are. There are therefore significant variations in the estimates of this parameter from ranging from 7 to 20 days and conflicting reports.[7,8] Current UK guidance asks CV-19 patients to isolate for 10 days. Patients with severe symptoms are likely to be isolated in hospital, reducing further transmission, so simulations using γ values of 7 and 10 days were chosen for comparison.

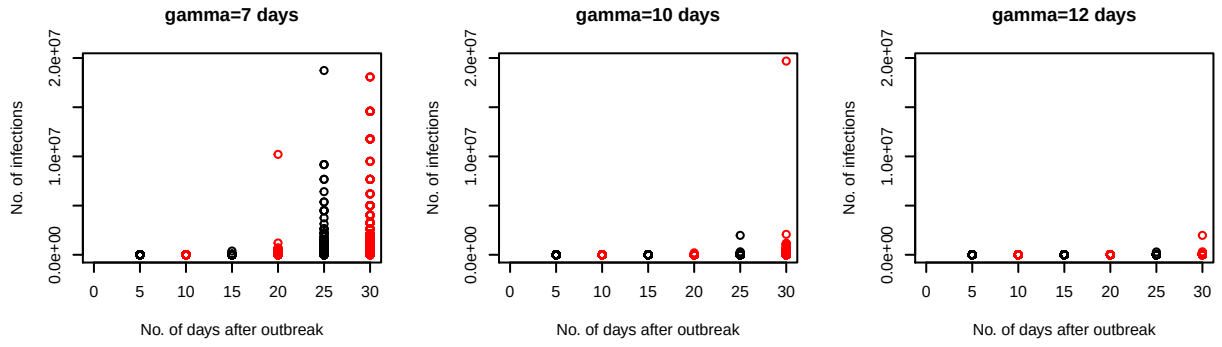

Figure 8: Effect of γ on CV-19 cases. Y-axis is cut at the maximum number of cases for $\gamma = 10$ days.

In equation 2, γ is on the denominator, so as γ increases, the infection's growth rate, r decreases. This is demonstrated by the number of CV-19 cases shown in figure 8.

Table 2: Number of CV-19 infections for 10, 20 and 30 days at γ=7 and γ=10 days

γ	Day	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	95% interval
7	10	89.63	348.23	616.65	811.30	1016.69	14287.79	(137.59, 2763.64)
7	20	402	6063	19013	71611	51683	10207052	(946.60, 381885.10)
7	30	1800	105600	586200	19650000	2627000	7292000000	(6512.29, 52800000)
10	10	57.15	147.78	220.46	249.98	312.85	1989.69	(77.15, 630.01)
10	20	163.3	1092.0	2430.2	4332.2	4893.8	197942.6	(297.59, 19845.49)
10	30	467	8069	26789	119103	76553	19692182	(1147.95, 625140.90)

Table 2 shows the summary statistics and 95% credibility intervals for γ=7 and γ=10. At 30 days after outbreak, the median numbers of infections for γ=(7, 10, 12) are (586200, 26789, 8068).

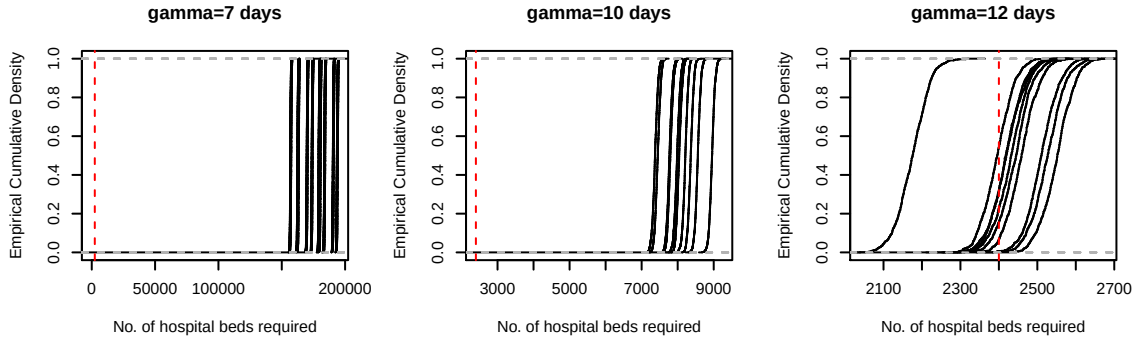

Figure 9: Effect of γ on the number of CV-19 hospitalisations. Dashed red line is bed-threshold of 2400.

Figure 9 shows that at 30 days after outbreak, if γ=10 days, the minimum hospitalisations would be 7148, about 4750 above the bed-threshold of 2400, so all simulated distributions give 100% chance of exceeding it. If γ=7 days, however, hospitalisations could exceed 155,522. In this case, the bed-threshold could be exceeded by day 15. For γ=(7, 10, 12) days, the median number of hospitalisations are (176295, 8013, 2446). At γ=10, the median number of hospitalisations will exceed the bed-threshold of 2400 by 5613 beds on day 30.

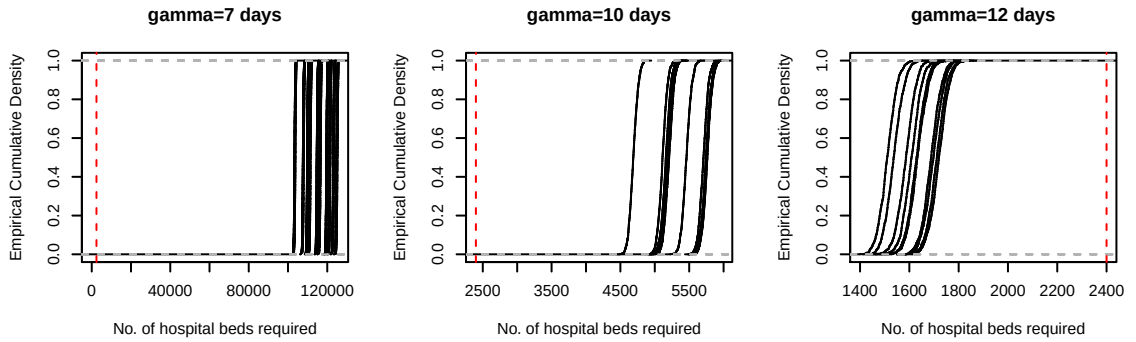

Figure 10: Effect of γ on CV-19 hospitalisations when 15% to 25% require hospitalisation. Dashed red line is bed-threshold of 2400.

Symptoms of the Omega variant are thought to be less severe, so further simulations were run to see how reducing the proportion of cases requiring hospitalisation would affect the number of hospital beds required. A new beta distribution was formed with minimum, modal, and maximum values of 15%, 20% and 25%. Figure 10 shows the hospitalisation numbers. There is now 0% probability of hospitalisations exceeding the 2400 available beds for γ=12 days, but there is 100% probability of exceeding 2400 beds for γ=10 and 7 days. If previous studies are correct, suggesting lower duration of infectiousness for milder symptoms, then the bed-threshold is likely to be exceeded as shown when γ is 10 days or fewer. The median hospitalisations when 15% to 25% of cases need hospital care, for γ=(7, 10, 12) days is (115427, 5232, 1637), so for the median value at γ=10, 2832 more beds would need to be found for CV-19 patients.

Next, bed-threshold values were investigated to manage the likelihood of hospitals being over-run with CV-19 patients. For γ=10 days, with 15% to 25% of infections requiring hospitalisation, the mean probability of exceeding a bed-threshold of 5000 is 89.5%. The mean probability of exceeding a bed-threshold of 5500 is 32.7% and there is 0% chance of exceeding 6000 beds. If γ=7 days, we'd need 20 times as many beds for CV-19 patients to make any impact. In this case, a bed-threshold of 120,000 would have a mean probability of been exceeded of 36.6%. This would impact capacity planning.

4 Conclusion

Levels of uncertainty for the results is high due to the uncertainty in our parameters. This makes public health planning extremely difficult. Consideration should be made as to whether there are better choices of model to be used: Equation 1 is appropriate for the growth of infected cases during the early stages of an epidemic, but is 30 days still sufficiently early for this formula to be valid? R_0 is defined for people susceptible to the disease, however R_0 values will be different for people who are vaccinated. Also, assumptions for the binomial distribution to model infection numbers are not all valid. It requires trials to be independent and to have the same probability but this is not the case, as disease is likely to spread more in indoor or crowded spaces and some patients are more infectious than others whilst some are more vulnerable to disease.

In this study, trials of different γ values were chosen from the literature to compare simulation outcomes. However, estimates from previous studies vary greatly as it is believed that γ varies according to several factors. It may be better to incorporate γ into equation 2 as a distribution to help account for the spread of estimates rather than using point values.

When investigating super-spreaders, Bi et al found that about 9% of CV-19 patients were responsible for 80% of infections in Shenzhen, China. Sun et al. did a similar study to find that 15% of patients were responsible for 80% of infections in Hunan, China.[9] Whilst these studies used CV-19 data from outbreaks in China rather that the UK, it is clear that targeting super-spreaders and super-spreader events should be a key factor in stemming transmission of the virus and highlights why the extreme values in our distributions should not be ignored, as happened when bootstrapping median values.

In this study both median bootstrapping and mean bootstrapping methods were used for values of R_0, however, they gave very different results. The probability of exceeding the bed-threshold was found to be significantly higher when using the mean. Using mean bootstrapping may appear too conservative but using median bootstrapping misses extreme values. The two models should be tested using data from earlier strains to clarify which method is more accurate. Non-parametric bootstrapping may also under-estimate variability, especially in the tails so a different model for R_0 may be preferred. Researchers have used many different models to predict CV-19 infections, and when the distribution of R_0 is skewed, a negative binomial distribution is often used. Cécile Kremer et al tested 3 different distributions to model R_0 and investigated bias in estimates of variance.[9] They found estimates to be biased when there was large individual variation in R_0, as found in this study. In 2 of their 3 datasets, they found R_0 to be better modelled using a Poisson-lognormal distribution.[9] It would be interesting to test a Poisson-lognormal distribution on this dataset to compare results.

Since the Omega strain is thought to cause milder symptoms, our model was modified to allow for a reduced proportion of patients requiring hospitalisation. However, if milder symptoms lead to a shorter duration of infectiousness as some believe, then the growth rate of the virus may be far greater. This in turn may lead to more patients requiring hospitalisation, so a more benign strain of the virus may cause us significant problems. This study models the number of hospital beds that may be needed due to the outbreak, however, it is crucial that the NHS also has enough appropriately trained staff to look after the CV-19 patients. Patients requiring a ventilator need one for an average of nearly 3 weeks[10] and more staff are required for these patients, so it is critical that there are sufficient numbers of trained staff as well as the necessary beds and equipment. There has also been a significant drain on hospital staff from the increased risk of them catching CV-19 despite their extra personal protective equipment. It would therefore be interesting to extend the study to include the expected number of extra staff required for each new CV-19 hospitalisation case, and any extra staff losses due to sickness.

No matter how good a model is, it will fail to predict correctly without accurate parameters. Our results show both manageable, low growth of the virus, and unmanageable, catastrophic growth. Until we have accurate parameter estimates, we're left to prepare for the worst and hope for the best.

References

[1] URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it.

[2] URL: https://www.nature.com/articles/d41586-021-00502-4.

[3] URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

[4] URL: https://www.instituteforgovernment.org.uk/explainers/coronavirus-vaccine-rollout.

[5] URL: https://www.england.nhs.uk/2020/12/landmark-moment-as-first-nhs-patient-receives-covid-19-vaccination/.

[6] URL: https://coronavirus.data.gov.uk/details/vaccinations.

[7] Andrew William Byrne et al. "Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases". *BMJ Open* 10.8 (2020). ISSN: 2044-6055. DOI: 10.1136/bmjopen-2020-039856. eprint: https://bmjopen.bmj.com/content/10/8/e039856.full.pdf. URL: https://bmjopen.bmj.com/content/10/8/e039856.

[8] Anika Singanayagam et al. "Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020". *Eurosurveillance* 25.32, 2001483 (2020). DOI: https://doi.org/10.2807/1560-7917.ES.2020.25.32.2001483. URL: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.32.2001483.

[9] Cécile Kremer et al. "Quantifying superspreading for COVID-19 using Poisson mixture distributions". *Scientific Reports* 11.1 (2021), p. 14107. DOI: 10.1038/s41598-021-93578-x. URL: https://doi.org/10.1038/s41598-021-93578-x.

[10] URL: https://www.scientificamerican.com/article/not-just-ventilators-staff-trained-to-run-them-are-in-short-supply/.

Appendix
Results from a different run of the code

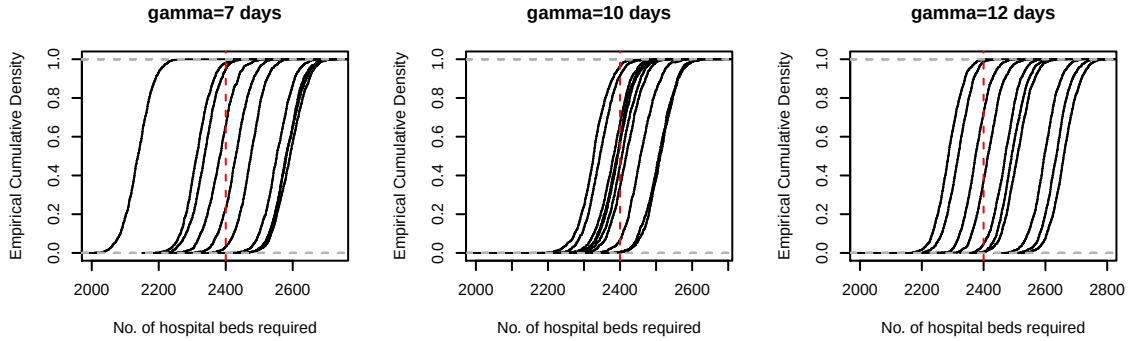

Figure 11: Different computer sampling using mean bootstrapping gave the following results when 25% to 35% require hospitalisation. Dashed red line is bed-threshold of 2400.

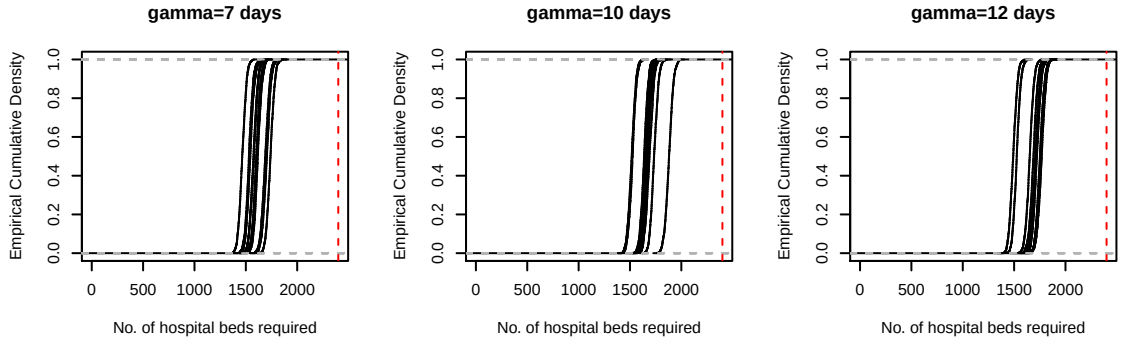

Figure 12: Different computer sampling using mean bootstrapping gave the following results when 15% to 25% require hospitalisation. Dashed red line is bed-threshold of 2400.

Re-running the code, different values are chosen in the sampling process and the results of the simulations and the predicted outcome of the virus are very different and far less extreme. In this case, impact planning would be tricky if 25-35% of patients are hospitalised, however, if 15-25% required hospitalisation, then the bed-threshold of 2400 has 0% chance of being exceeded. It is interesting to note that in figure 11 the distribution at γ=10 days is narrower, showing lower uncertainty than for γ=7 or 12 days. I suspect this is coincidental. Re-running the model again with more bootstraps and more simulations in the secondary distribution would hopefully reduce the variation in the results.

R code used for analysis

```
### Question 1 ###

# Enter data: Data_R0 is no. of secondary cases from our 20 individuals.
Data_R0 <- c(4,5,2,1,0,5,8,2,2,3,5,9,0,1,3,2,6,2,3,5)
#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/hist.pdf",
#width=3, height=2.5, pointsize=9)
hist(Data_R0, xlab="No. of secondary infections", xlim=c(0,10), ylim=c(0, 5), breaks=11, main = "") #visualise data
#dev.off()

# Non-parametric bootstrapping function (npbs)
npbs<-function(x,b){
  y<-c()
  for(i in 1:b){
    y[i]<-median(sample(x,length(x), replace=TRUE))
  }
  y}

# Run the model for 500 bootstraps
R0_500medians<-npbs(Data_R0,500)   #500 bootstrap samples of median R0

#Plot distribution of R0 medians (with smoothing function)
#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/bootstrap.pdf",
# width=3, height=2.5, pointsize=9)
plot(density(R0_500medians, bw=1), xlab="Median number of secondary infections", main="")
#dev.off()

#Mean of the medians
mean(R0_500medians)
#standard deviation
sd(R0_500medians)
#min & max values
min(R0_500medians); max(R0_500medians)
#Quantiles of the median R0 values
quantile(R0_500medians, probs=seq(0,1,0.025))


### Question 2 ###

##The variables:
# R0_500medians is my distribution of 500 median values of R0 from Q1
t <- 1:30   #Time in days (after outbreak)
gama <- 12  #average duration of infectiousness (days)
r <- (R0_500medians-1)/gama  #500 values of r, the growth rate of the infection
I_0 <- 20   #Initial no. of cases

#Use the parameters to calculate an array of no. of infections at different times after outbreak
I<-array(dim=c(length(t), length(r)))
for (j in 1:length(r)){  # for each of the 500 values of r
for (i in 1:length(t)){  # for each different time
I[i,j] <- I_0*exp(r[j]*t[i])
}}

# No. of Infections, I has 30 rows (time after outbreak, 1-30 days) and 500 columns (r values)
# So we have a distribution of infection nos. for each time after outbreak (rows of data)
head(I[30,], 10)      #quick check of data, day 30
max(I)      #max no. of infections (for the y axis limit)


#y values for graph
t5  <- I[5,]   #infections at 5 days
t10 <- I[10,]  #Infections at 10 days
t15 <- I[15,]  #Infections at 15 days
t20 <- I[20,]  #Infections at 20 days
t25 <- I[25,]  #Infections at 25 days
t30 <- I[30,]  #Infections at 30 days
```

```
#x values for graph
ts5  <- rep(5, 500)
ts10 <- rep(10, 500)
ts15 <- rep(15, 500)
ts20 <- rep(20, 500)
ts25 <- rep(25, 500)
ts30 <- rep(30, 500)

# Graph of the infections
#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/infections.pdf",
# width=3, height=2.5, pointsize=9)
plot(ts10, t10, xlim=c(0, 30), ylim = c(0, max(I)),
     xlab="No. of days after outbreak", ylab="No. of infections", col="red")
points(ts20, t20, xlim=c(0, 30), ylim = c(0, max(I)), col="red")
points(ts30, t30, xlim=c(0, 30), ylim = c(0, max(I)), col="red")
points(ts5, t5, xlim=c(0, 30), ylim = c(0, max(I)))
points(ts15, t15, xlim=c(0, 30), ylim = c(0, max(I)))
points(ts25, t25, xlim=c(0, 30), ylim = c(0, max(I)))
#dev.off()

#Summary stats for infections at 10 days
summary(t10)
quantile(t10, probs=seq(0,1,0.025))

#Summary stats for infections at 20 days
summary(t20)
quantile(t20, probs=seq(0,1,0.025))

#Summary stats for infections at 30 days
summary(t30)
quantile(t30, probs=seq(0,1,0.025))

# Represent no of infections for t=1:30 for first 10 r values
head(I[1,], 10)       #shows the no. of infections first 10 r values
t <- 1:30
#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/predictions.1.pdf
# width=7, height=2.5, pointsize=9)
par(mfrow=c(1,2))
plot(t, I[,1], type="n",
     xlab="Time after outbreak (days)", ylab="No. of infected cases", ylim=c(0, 441000), lwd = 0.6)
for (i in 1:10){
lines(t, I[,i], lty=i+1)
  legend("topleft", inset=0.05, c("r values (1,3,6,10)", "r values (2,8)", "r value (4)", "r values (5,7)",
  "r value (9)"), lty = c(2,3,4,5,6), cex = 0.9)
}
plot(t, I[,1], type="n",
     xlab="Time after outbreak (days)", ylab="No. of infected cases", ylim=c(0, 11000), lwd = 0.6)
for (i in 1:10){
  lines(t, I[,i], lty=i+1)
}
#dev.off()

#Above graph gives one extreme value of infection cases.... Why?
hist(R0_500medians) #Distribution of R0
max(R0_500medians) # Max R0 in model is 5 secondary infections
sum(ifelse(R0_500medians==5, 1, 0))/500   #6.4% cases give R0=5

### Question 3 ###

#Expert opinion of proportions needing hospital bed
minHosp <- 0.25
maxHosp <- 0.35
modal <- 0.3

#Distribution of possible proportions needing beds
prop.range <- seq(0.2,0.4,0.001)
```

```
a <- 250
calc.b <- function(a, mode){
  b <- 2-a+(a-1)/mode
  print(b)
}
calc.b(a=a, mode=modal)
b <- 582
prop.Hosp <- dbeta(prop.range, a, b) #Distribution of values for proportion that need hospital

#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/HospProp.pdf",
# width=3, height=2.5, pointsize=9)
plot(prop.range, prop.Hosp, xlab="Proportion needing hospitalised", type="l", ylab="Density")
#dev.off()

#Using the median no. of infections/cases at 30 days, [median(t30) = 2968 = n],
#generate random variables from binomial distribution with p=prop.Hosp in (binom(n,p))
#to simulate the no. of beds required in hospital.

#Function to generate values from binomial dist for no. of beds required
n <- 2968       #median no. of infections at 30 days
#n1 is no. of values sampled from beta dist (uncertainty)
#n2 is no. of values sampled from binomial dist (variability)

simul <- function(n1, n2){
  y <- array(dim=c(n1, n2));
for(i in 1:n1){
  p <- rbeta(1,a,b);     #Generate 1 value from our beta distribution (a=250, b=582)
  for (j in 1:n2){
    y[i,j] <- rbinom(1, n, p)
  }
}
y
}

#Run function for no. of beds required
beds <- simul(10,1000)

min(beds)   #check for min for x-axis
max(beds)   #check for max for x-axis

#Plot distributions
#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/HospCases.pdf",
# width=6, height=2.5, pointsize=9)
par(mfrow=c(1,2))
plot(ecdf(beds[1,]), do.p=FALSE, verticals=TRUE, main="",
     xlab="No. of hospital beds required", ylab="Empirical Cumulative Density", lty=1, xlim=c(min(beds), max(beds)))
for (i in 2:10)
  lines(ecdf(beds[i,]), verticals=TRUE, do.p=FALSE)
abline(v = 2400, col = "red", lty=2)
plot(ecdf(beds[1,]), do.p=FALSE, verticals=TRUE, main="",
     xlab="No. of hospital beds required", ylab="Empirical Cumulative Density", lty=1, xlim=c(min(beds), 2500))
for (i in 2:10)
  lines(ecdf(beds[i,]), verticals=TRUE, do.p=FALSE)
abline(v = 2400, col = "red", lty=2)
#dev.off()

#Probability of threshold being exceeded
cumulprob1 <- ecdf(beds[1,])
1-cumulprob1(2400)
cumulprob2 <- ecdf(beds[2,])
1-cumulprob2(2400)
cumulprob3 <- ecdf(beds[3,])
1-cumulprob3(2400)
cumulprob4 <- ecdf(beds[4,])
1-cumulprob4(2400)
cumulprob5 <- ecdf(beds[5,])
1-cumulprob5(2400)
```

```
cumulprob6 <- ecdf(beds[6,])
1-cumulprob6(2400)
cumulprob7 <- ecdf(beds[7,])
1-cumulprob7(2400)
cumulprob8 <- ecdf(beds[8,])
1-cumulprob8(2400)
cumulprob9 <- ecdf(beds[9,])
1-cumulprob9(2400)
cumulprob10 <- ecdf(beds[10,])
1-cumulprob10(2400)

### Question 4 ###
#### Use MEAN bootstrapping ####

# Non-parametric bootstrapping function (npbs2)
npbs2<-function(x,b){
  y<-c()
  for(i in 1:b){
    y[i]<-mean(sample(x,length(x), replace=TRUE))
  }
  y}

# Run the model for 500 bootstraps
R0_500means<-npbs2(Data_R0,500)   #500 bootstrap samples of mean R0

#Plot distribution of R0 means (with smoothing function)
#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/bootstrap2.pdf",
# width=3, height=2.5, pointsize=9)
plot(density(R0_500means, bw=1), xlab="Mean number of secondary infections", main="")
#dev.off()

#Mean of the means
mean(R0_500means)
#standard deviation
sd(R0_500means)
#min & max values
min(R0_500means); max(R0_500means)
#Quantiles of the mean R0 values
quantile(R0_500means, probs=seq(0,1,0.025))


### (Q. 2) ###

##The variables:
# R0_500means is my distribution of 500 mean values of R0 from Q1
t <- 1:30   #Time in days (after outbreak)
gama <- 12  #average duration of infectiousness (days)
gama7 <- 7
gama10 <- 10
r2 <- (R0_500means-1)/gama  #500 values of r, the growth rate of the infection
# r for different gammas
r_gama7 <- (R0_500means-1)/gama7
r_gama10 <- (R0_500means-1)/gama10

I_0 <- 20   #Initial no. of cases

#Use the parameters to calculate an array of no. of infections at different times after outbreak
I2<-array(dim=c(length(t), length(r)))
for (j in 1:length(r2)){  # for each of the 500 values of r
  for (i in 1:length(t)){  # for each different time
    I2[i,j] <- I_0*exp(r2[j]*t[i])
  }}
# I for gamma 7
Ig7<-array(dim=c(length(t), length(r_gama7)))
for (j in 1:length(r_gama7)){  # for each of the 500 values of r
  for (i in 1:length(t)){  # for each different time
    Ig7[i,j] <- I_0*exp(r_gama7[j]*t[i])
```

```
  }}

# I for gamma 10
Ig10<-array(dim=c(length(t), length(r_gama10)))
for (j in 1:length(r_gama10)){  # for each of the 500 values of r
  for (i in 1:length(t)){  # for each different time
    Ig10[i,j] <- I_0*exp(r_gama10[j]*t[i])
  }}

# No. of Infections, I has 30 rows (time after outbreak, 1-30 days) and 500 columns (r values)
# So we have a distribution of infection nos. for each time after outbreak (rows of data)
head(I2[30,], 10)      #quick check of data, day 30
max(I2)      #max no. of infections (for the y axis limit)


#y values for graph
t5m  <- I2[5,]   #infections at 5 days
t10m <- I2[10,]  #Infections at 10 days
t15m <- I2[15,]  #Infections at 15 days
t20m <- I2[20,]  #Infections at 20 days
t25m <- I2[25,]  #Infections at 25 days
t30m <- I2[30,]  #Infections at 30 days

#y for gamma 7
t5g7  <- Ig7[5,]   #infections at 5 days
t10g7 <- Ig7[10,]  #Infections at 10 days
t15g7 <- Ig7[15,]  #Infections at 15 days
t20g7 <- Ig7[20,]  #Infections at 20 days
t25g7 <- Ig7[25,]  #Infections at 25 days
t30g7 <- Ig7[30,]  #Infections at 30 days

#y for gamma 10
t5g10  <- Ig10[5,]   #infections at 5 days
t10g10 <- Ig10[10,]  #Infections at 10 days
t15g10 <- Ig10[15,]  #Infections at 15 days
t20g10 <- Ig10[20,]  #Infections at 20 days
t25g10 <- Ig10[25,]  #Infections at 25 days
t30g10 <- Ig10[30,]  #Infections at 30 days

#x values for graph
ts5  <- rep(5, 500)
ts10 <- rep(10, 500)
ts15 <- rep(15, 500)
ts20 <- rep(20, 500)
ts25 <- rep(25, 500)
ts30 <- rep(30, 500)

# Graph of the infections
#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/infections2.pdf",
# width=3, height=3, pointsize=9)
plot(ts10, t10m, xlim=c(0, 30), ylim = c(0, max(I2)),
     xlab="No. of days after outbreak", ylab="No. of infections", col="red")
points(ts20, t20m, xlim=c(0, 30), ylim = c(0, max(I2)), col="red")
points(ts30, t30m, xlim=c(0, 30), ylim = c(0, max(I2)), col="red")
points(ts5, t5m, xlim=c(0, 30), ylim = c(0, max(I2)))
points(ts15, t15m, xlim=c(0, 30), ylim = c(0, max(I2)))
points(ts25, t25m, xlim=c(0, 30), ylim = c(0, max(I2)))
#dev.off()

#gamma plots
#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/gammas.pdf",
# width=6.5, height=2, pointsize=9)
par(mfrow=c(1,3))
plot(ts10, t10g7, xlim=c(0, 30), ylim = c(0, max(Ig10)),
     xlab="No. of days after outbreak", ylab="No. of infections", col="red", main="gamma=7 days")
points(ts20, t20g7, xlim=c(0, 30), ylim = c(0, max(Ig10)), col="red")
points(ts30, t30g7, xlim=c(0, 30), ylim = c(0, max(Ig10)), col="red")
```

```
points(ts5, t5g7, xlim=c(0, 30), ylim = c(0, max(Ig10)))
points(ts15, t15g7, xlim=c(0, 30), ylim = c(0, max(Ig10)))
points(ts25, t25g7, xlim=c(0, 30), ylim = c(0, max(Ig10)))

plot(ts10, t10g10, xlim=c(0, 30), ylim = c(0, max(Ig10)),
     xlab="No. of days after outbreak", ylab="No. of infections", col="red", main="gamma=10 days")
points(ts20, t20g10, xlim=c(0, 30), ylim = c(0, max(Ig10)), col="red")
points(ts30, t30g10, xlim=c(0, 30), ylim = c(0, max(Ig10)), col="red")
points(ts5, t5g10, xlim=c(0, 30), ylim = c(0, max(Ig10)))
points(ts15, t15g10, xlim=c(0, 30), ylim = c(0, max(Ig10)))
points(ts25, t25g10, xlim=c(0, 30), ylim = c(0, max(Ig10)))

plot(ts10, t10m, xlim=c(0, 30), ylim = c(0, max(Ig10)),
     xlab="No. of days after outbreak", ylab="No. of infections", col="red", main="gamma=12 days")
points(ts20, t20m, xlim=c(0, 30), ylim = c(0, max(Ig10)), col="red")
points(ts30, t30m, xlim=c(0, 30), ylim = c(0, max(Ig10)), col="red")
points(ts5, t5m, xlim=c(0, 30), ylim = c(0, max(Ig10)))
points(ts15, t15m, xlim=c(0, 30), ylim = c(0, max(Ig10)))
points(ts25, t25m, xlim=c(0, 30), ylim = c(0, max(Ig10)))
#dev.off()

#Summary stats for infections at 10 days
summary(t10m)
quantile(t10m, probs=seq(0,1,0.025))

#Summary stats for infections at 20 days
summary(t20m)
quantile(t20m, probs=seq(0,1,0.025))

#Summary stats for infections at 30 days
summary(t30m)
quantile(t30m, probs=seq(0,1,0.025))

# for gamma 7
summary(t10g7)
quantile(t10g7, probs=seq(0,1,0.025))
summary(t15g7)
summary(t20g7)
quantile(t20g7, probs=seq(0,1,0.025))
summary(t30g7)
quantile(t30g7, probs=seq(0,1,0.025))

# for gamma 10
summary(t10g10)
quantile(t10g10, probs=seq(0,1,0.025))
summary(t20g10)
quantile(t20g10, probs=seq(0,1,0.025))
summary(t30g10)
quantile(t30g10, probs=seq(0,1,0.025))

# Represent no of infections for t=1:30 for first 10 r values
head(I2[30,], 10)       #shows the no. of infections at t=30 first 10 r values
t <- 1:30
#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/predictions2.pdf"
# width=7, height=3, pointsize=9)
par(mfrow=c(1,2))
plot(t, I2[,1], type="n",
     xlab="Time after outbreak (days)", ylab="No. of infected cases", ylim=c(0, 52650))
for (i in 1:10){
  lines(t, I2[,i], lty=i+1)
#  legend("topleft", c("", "", "", "", ""), lty = i+1)
}
plot(t, I2[,1], type="n",
     xlab="Time after outbreak (days)", ylab="No. of infected cases", ylim=c(0, 12000))
for (i in 1:10){
  lines(t, I2[,i], lty=i+1)
}
```

```
#dev.off()

### for gamma 7
head(Ig7[30,], 10)       #shows the no. of infections at t=30 first 10 r values
t <- 1:30

par(mfrow=c(1,2))
plot(t, Ig7[,1], type="n",
     xlab="Time after outbreak (days)", ylab="No. of infected cases", ylim=c(0, 14588327.40))
for (i in 1:10){
  lines(t, Ig7[,i], lty=i+1)
}
plot(t, Ig7[,1], type="n",
     xlab="Time after outbreak (days)", ylab="No. of infected cases", ylim=c(0, 12000))
for (i in 1:10){
  lines(t, Ig7[,i], lty=i+1)
}

### for gamma 10
head(Ig10[30,], 10)       #shows the no. of infections at t=30 first 10 r values
t <- 1:30

par(mfrow=c(1,2))
plot(t, Ig10[,1], type="n",
     xlab="Time after outbreak (days)", ylab="No. of infected cases", ylim=c(0, 17081.1753))
for (i in 1:10){
  lines(t, Ig10[,i], lty=i+1)
}
plot(t, Ig10[,1], type="n",
     xlab="Time after outbreak (days)", ylab="No. of infected cases", ylim=c(0, 12000))
for (i in 1:10){
  lines(t, Ig10[,i], lty=i+1)
}

### Q. 3 ###

#Expert opinion of proportions needing hospital bed
minHosp2 <- 0.15
maxHosp2 <- 0.25
modal2 <- 0.2

#Distribution of possible proportions needing beds
prop.range <- seq(0.1,0.3,0.001)
a2 <- 250
calc.b <- function(a2, mode){
  b <- 2-a2+(a2-1)/mode
  print(b)
}
calc.b(a=a2, mode=modal2)
b2 <- 997
prop.Hosp2 <- dbeta(prop.range, a2, b2) #Distribution of values for proportion that need hospital

#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/HospProp2.pdf",
# width=3, height=3, pointsize=9)
plot(prop.range, prop.Hosp, xlab="Proportion needing hospitalised", type="l", ylab="Density")
#dev.off()

#Using the median no. of infections/cases at 30 days, [median(t30) = 8068 = n],
#generate random variables from binomial distribution with p=prop.Hosp in (binom(n,p))
#to simulate the no. of beds required in hospital.

#Function to generate values from binomial dist for no. of beds required
n <- 8068       #median no. of infections at 30 days -----> change this as appropriate
#n1 is no. of values sampled from beta dist (uncertainty)
#n2 is no. of values sampled from binomial dist (variability)

simul <- function(n1, n2){
```

```
  y <- array(dim=c(n1, n2));
  for(i in 1:n1){
    p <- rbeta(1,a,b);     #Generate 1 value from our beta distribution (a=250, b=582)
    for (j in 1:n2){
      y[i,j] <- rbinom(1, n, p)
    }
  }
  y
}

# For 2nd beta distribution for proportion being hospitalised
simul2 <- function(n1, n2){
  y <- array(dim=c(n1, n2));
  for(i in 1:n1){
    p <- rbeta(1,a2,b2);     #Generate 1 value from our beta distribution (a2=250, b2=997)
    for (j in 1:n2){
      y[i,j] <- rbinom(1, n, p)
    }
  }
  y
}

#Run function for no. of beds required
beds2 <- simul(10,1000)
bedsg7 <- simul(10, 1000)
bedsg10 <- simul(10, 1000)

beds12b2 <- simul2(10,1000) #beds required: gamma=12, 2nd beta distribution
beds10b2 <- simul2(10,1000) #beds required: gamma=10, 2nd beta distribution
beds7b2 <- simul2(10,1000) #beds required: gamma=7, 2nd beta distribution

min(beds2)   #check for min for x-axis
median(beds2)#median no. of beds required
max(beds2)   #check for max for x-axis
min(bedsg7)
median(bedsg7)
max(bedsg7)
min(bedsg10)
median(bedsg10)
max(bedsg10)

#2nd beta distribution data
min(beds12b2)
max(beds12b2)
median(beds12b2)
min(beds10b2)
max(beds10b2)
median(beds10b2)
min(beds7b2)
max(beds7b2)
median(beds7b2)

#Plot distributions
pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/HospCases2.pdf",
# width=4, height=2.5, pointsize=9)
par(mfrow=c(1,1))
plot(ecdf(beds2[1,]), do.p=FALSE, verticals=TRUE, main="",
     xlab="No. of hospital beds required", ylab="Empirical Cumulative Density", lty=1, xlim=c(min(beds2), max(beds2)
for (i in 2:10)
  lines(ecdf(beds2[i,]), verticals=TRUE, do.p=FALSE)
abline(v = 2400, col = "red", lty=2)
#plot(ecdf(beds2[1,]), do.p=FALSE, verticals=TRUE, main="",
#     xlab="No. of hospital beds required", ylab="Empirical Cumulative Density", lty=1, xlim=c(min(beds2), 2500))
#for (i in 2:10)
#  lines(ecdf(beds2[i,]), verticals=TRUE, do.p=FALSE)
#abline(v = 2400, col = "red", lty=2)
#dev.off()
```

```
#gammas
#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/
#HospCasesGammas.pdf", width=6, height=2, pointsize=9)
par(mfrow=c(1,3))
plot(ecdf(bedsg7[1,]), do.p=FALSE, verticals=TRUE,
     xlab="No. of hospital beds required", ylab="Empirical Cumulative Density", lty=1, xlim=c(0, max(bedsg7)),
 main="gamma=7 days")
for (i in 2:10)
  lines(ecdf(bedsg7[i,]), verticals=TRUE, do.p=FALSE)
abline(v = 2400, col = "red", lty=2)

plot(ecdf(bedsg10[1,]), do.p=FALSE, verticals=TRUE,
     xlab="No. of hospital beds required", ylab="Empirical Cumulative Density", lty=1, xlim=c(2400, max(bedsg10)),
  main="gamma=10 days")
for (i in 2:10)
  lines(ecdf(bedsg10[i,]), verticals=TRUE, do.p=FALSE)
abline(v = 2400, col = "red", lty=2)

plot(ecdf(beds2[1,]), do.p=FALSE, verticals=TRUE,
     xlab="No. of hospital beds required", ylab="Empirical Cumulative Density", lty=1, xlim=c(min(beds2), max(beds2)
     main="gamma=12 days")
for (i in 2:10)
  lines(ecdf(beds2[i,]), verticals=TRUE, do.p=FALSE)
abline(v = 2400, col = "red", lty=2)
#dev.off()

#gammas and 2nd beta distriburtion
#pdf(file = "/Users/sarah/Desktop/MSc\ Applied\ Statistics\ in\ Health\ Sciences/MM925/Assessment/
#HospCasesGammasBeta2.pdf", width=6, height=2, pointsize=9)
par(mfrow=c(1,3))
plot(ecdf(beds7b2[1,]), do.p=FALSE, verticals=TRUE,
     xlab="No. of hospital beds required", ylab="Empirical Cumulative Density", lty=1, xlim=c(0, max(beds7b2)),
     main="gamma=7 days")
for (i in 2:10)
  lines(ecdf(beds7b2[i,]), verticals=TRUE, do.p=FALSE)
abline(v = 2400, col = "red", lty=2)

plot(ecdf(beds10b2[1,]), do.p=FALSE, verticals=TRUE,
     xlab="No. of hospital beds required", ylab="Empirical Cumulative Density", lty=1, xlim=c(2400, max(beds10b2)),
      main="gamma=10 days")
for (i in 2:10)
  lines(ecdf(beds10b2[i,]), verticals=TRUE, do.p=FALSE)
abline(v = 2400, col = "red", lty=2)

plot(ecdf(beds12b2[1,]), do.p=FALSE, verticals=TRUE,
     xlab="No. of hospital beds required", ylab="Empirical Cumulative Density", lty=1, xlim=c(min(beds12b2), 2400),
      main="gamma=12 days")
for (i in 2:10)
  lines(ecdf(beds12b2[i,]), verticals=TRUE, do.p=FALSE)
abline(v = 2400, col = "red", lty=2)
#dev.off()
###

#Probability of threshold being exceeded
cumulprob1 <- ecdf(beds2[1,])
1-cumulprob1(2400)
cumulprob2 <- ecdf(beds2[2,])
1-cumulprob2(2400)
cumulprob3 <- ecdf(beds2[3,])
1-cumulprob3(2400)
cumulprob4 <- ecdf(beds2[4,])
1-cumulprob4(2400)
cumulprob5 <- ecdf(beds2[5,])
1-cumulprob5(2400)
cumulprob6 <- ecdf(beds2[6,])
1-cumulprob6(2400)
```

```
cumulprob7 <- ecdf(beds2[7,])
1-cumulprob7(2400)
cumulprob8 <- ecdf(beds2[8,])
1-cumulprob8(2400)
cumulprob9 <- ecdf(beds2[9,])
1-cumulprob9(2400)
cumulprob10 <- ecdf(beds2[10,])
1-cumulprob10(2400)

## Probabilities for beds10b2
cumulprob1 <- ecdf(beds10b2[1,])
1-cumulprob1(6000)
cumulprob2 <- ecdf(beds10b2[2,])
1-cumulprob2(6000)
cumulprob3 <- ecdf(beds10b2[3,])
1-cumulprob3(6000)
cumulprob4 <- ecdf(beds10b2[4,])
1-cumulprob4(6000)
cumulprob5 <- ecdf(beds10b2[5,])
1-cumulprob5(6000)
cumulprob6 <- ecdf(beds10b2[6,])
1-cumulprob6(6000)
cumulprob7 <- ecdf(beds10b2[7,])
1-cumulprob7(6000)
cumulprob8 <- ecdf(beds10b2[8,])
1-cumulprob8(6000)
cumulprob9 <- ecdf(beds10b2[9,])
1-cumulprob9(6000)
cumulprob10 <- ecdf(beds10b2[10,])
1-cumulprob10(6000)

## Probabilities for beds7b2
cumulprob1 <- ecdf(beds7b2[1,])
1-cumulprob1(120000)
cumulprob2 <- ecdf(beds7b2[2,])
1-cumulprob2(120000)
cumulprob3 <- ecdf(beds7b2[3,])
1-cumulprob3(120000)
cumulprob4 <- ecdf(beds7b2[4,])
1-cumulprob4(120000)
cumulprob5 <- ecdf(beds7b2[5,])
1-cumulprob5(120000)
cumulprob6 <- ecdf(beds7b2[6,])
1-cumulprob6(120000)
cumulprob7 <- ecdf(beds7b2[7,])
1-cumulprob7(120000)
cumulprob8 <- ecdf(beds7b2[8,])
1-cumulprob8(120000)
cumulprob9 <- ecdf(beds7b2[9,])
1-cumulprob9(120000)
cumulprob10 <- ecdf(beds7b2[10,])
1-cumulprob10(120000)

# Get summary stats for probability of exceeding threshold
z = c(0.842, 1, 0.932, 0.994, 0.454, 0.995, 0, 0.682, 0.676, 0.77)
summary(z)
quantile(z, probs=seq(0,1,0.025)) #Quantiles for the 10 graphed distributions of hospitalisation cases

z2 = c(0,0,0,0,0,0,0.272, 1,1,1)
mean(z2)

z3 = c(.991, .995, 1,0,1,1,0.968, .998, 1,1)
mean(z3)

z4 = c(0,0,1,0,0,1,0,0,1,0.66)
mean(z4)
```