Mathematics 4MB3/6MB3 Mathematical Biology

http://www.math.mcmaster.ca/earn/4MB3

2019 ASSIGNMENT 4

Group Name: The Plague Doctors

Group Members: Sid Reed, Daniel Segura, Jessa Mallare, Aref Jadda

This assignment is **due in class** on **Wednesday March 13 2019 at 10:30am**.

```
#Setting the seed for the entire document, for reproducible stochastic simulations
seed = 9
set.seed(seed)
```

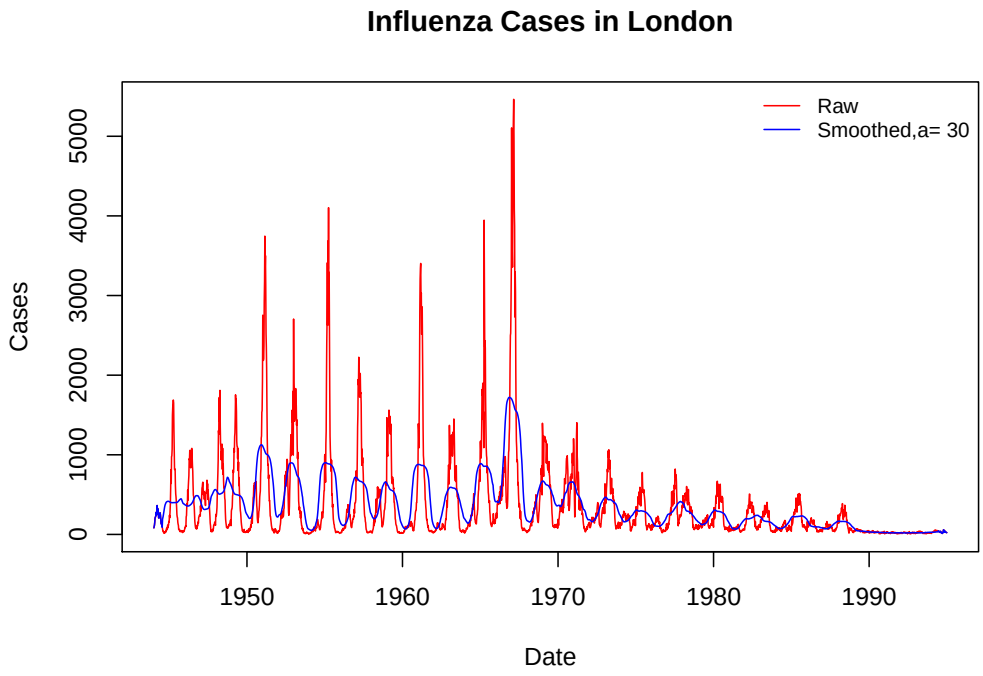
1. (a) You should have received the following data files by e-mail:

`meas_uk_lon_1944-94_wk.csv`
`meas_uk_lpl_1944-94_wk.csv`

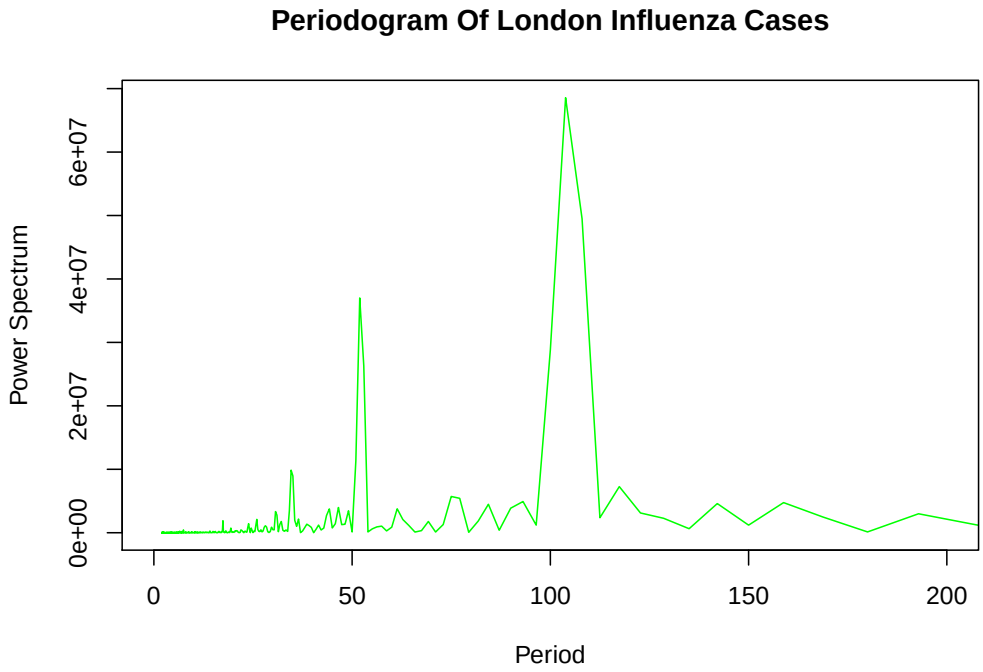
These plain text comma-separated-value files list weekly cases of measles (in London and Liverpool, England, from 1944 to 1994). Depending on which research direction you select, you might receive other files in the same `ymdc` (year,month,day,count) format, where the count column might contain cases or deaths, for example. Write the following R functions:

(i)
```
## [1] "First 5 rows of the dataframe"
##   year month day cases       date
## 1 1944     1   7    82 1944-01-07
## 2 1944     1  14    98 1944-01-14
## 3 1944     1  21   118 1944-01-21
## 4 1944     1  28   153 1944-01-28
## 5 1944     2   4   206 1944-02-04
## 6 1944     2  11   217 1944-02-11
```

(ii)
```
## [1] "Smoothing data with window size 30"
```

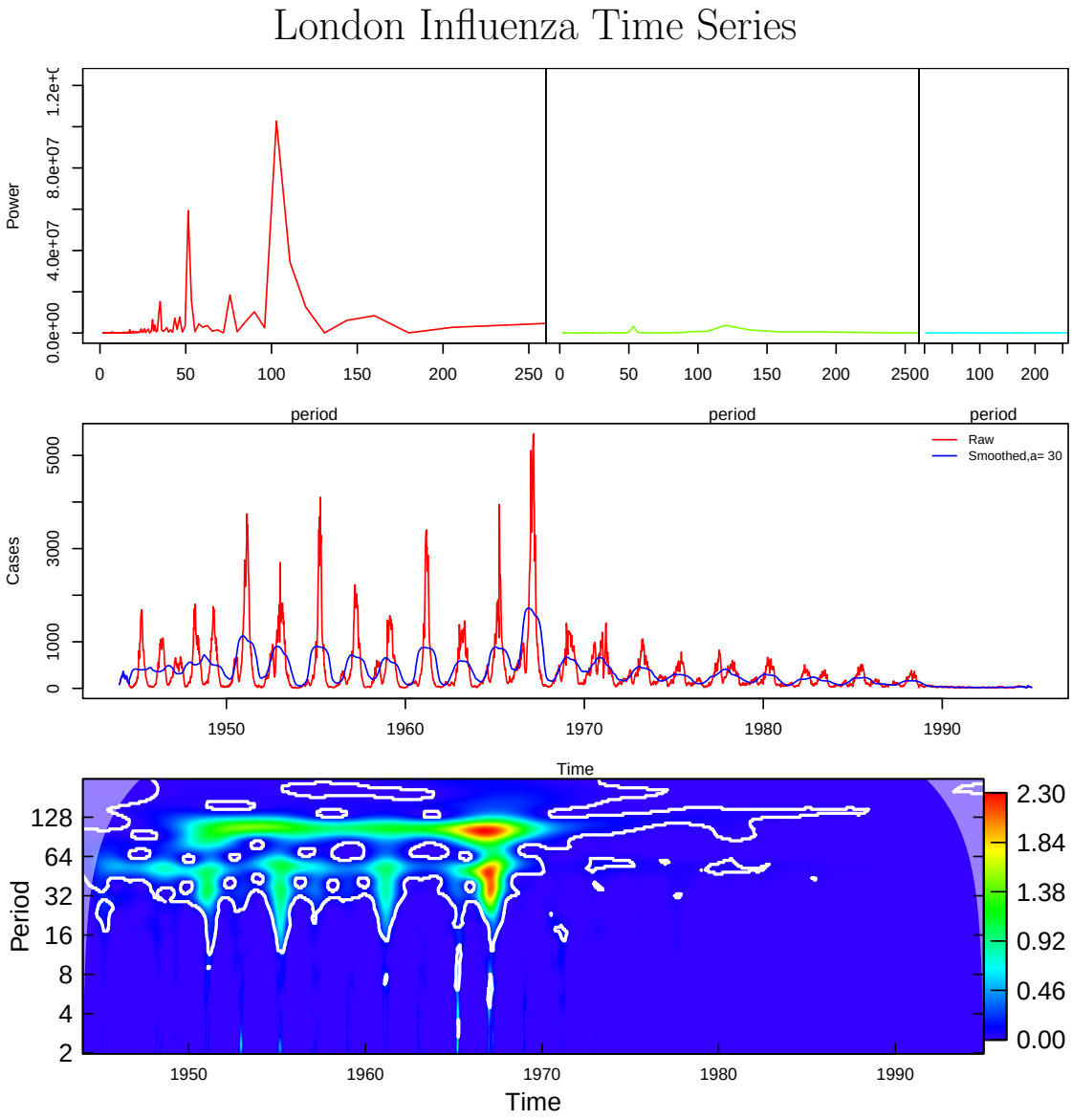


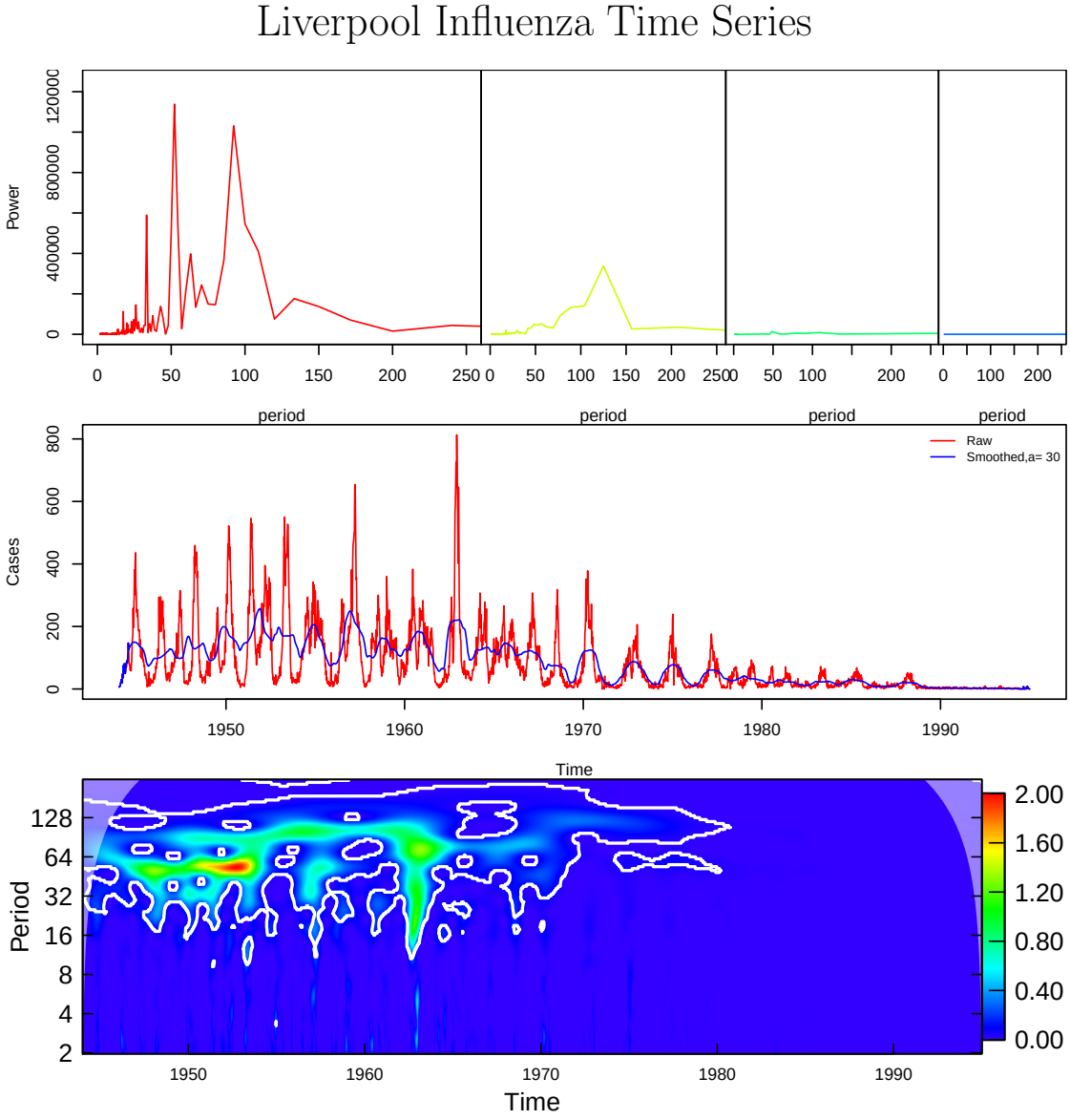
(iii)
```
## [1] "Using pgram estimation method"
```



(b) Using your functions, make a multi-panel plot that clearly shows the temporal pattern of the time series and how its frequency structure changes over time. Think carefully about how to make this multi-panel figure as clear as possible for your-

selves and your readers. Describe your figure, explaining what aspects of your figure you feel are puzzling or interesting and may be possible to understand using mechanistic mathematical modelling. (Repeat this for each of the epidemic time series you are given.)





Describing The Inflenza Data

The multi-planel figure is a diagram of the time-series, paired with a wavelet power spectrum and multiple period periodograms of time time series. The time points at which the period periodograms were set to begin and end were chosen based on shifts in pattern of the time series as well as dampening of signal power of the wavelet diagram. From the wavelet diagram and the periodperiodograms of the london series, starting at 1944 there is a strong signal at 104 and at 52. As the data is given in weeks, a periodicity of 104 weeks=2 years and a signal at 52 weeks = 1 year. There is a shift in the wavelet diagram ater the year 1970, as the power of the signal is for the biannual and annual cycles is much lower, but are still significant, as illustrated by the white contour lines of the wavelet diagram. THis drop in power can be attributed to the decrease ofin magnitudes of the peaks in the time series diagram at this point in time. Comparing the periodogram from before

1970 to after, it is clear that there is much less pronounced periodicity, with only a small residual peak of the biannual cycle. Furthermore, after the year 1990, the power is no longer significant signal, meaning that there is no longer a significant frequency at the 2 and 1 year cycles. Which can be futher attributed the lack of peaks seen in the period periodogram for this period in time.

For the liverpool time series we see a similar pattern as the one seen in london, with a strong 2 year and 1 year signal from 1944 until 1965 seen in the wavelet diagram and the period periodogram for this time. After 1965 the signal begins to dampen out, as seen in both the wavelelt diagram by the decrease in power of the signal,the time series plot by the decrease in magnitude of the peaks and in the period periodogram of 1965-1977 showing a small peak for the biannual and annual cycles, which are much weaker in power than the peaks of the periodogram from 1944-1965. After 1965 until 1977 the time series begins to dampene down futher, the power of the signal of the wavelet diagram is no longer significant, and the period periodogram of 1997-1990 shows no peaks. This means that there is no particular periodicity of significance for this time period. Similarly to what occurs in the london time series in the year 1990, the time series of the liverpool data dampens down even more, and both the wavelet diagram and the period periodogram are the same as in the 1977-1999 time period, showing that once again, there is no significant period. Some interesting things about the figures are that the the signal of the liverpool data is not as strong as the signal of the london data, as the scale which goes to the maximum power only goes up to 2.00, where as the power scale for the london diagram goes up to 2.3. Similarly, the scale of power of the period periodogram of the london time series goes up to $8 * 10^7$, while it only goes up to $8 * 10^5$ for the liver pool time series, showing that the signal is allot stronger for the london data.

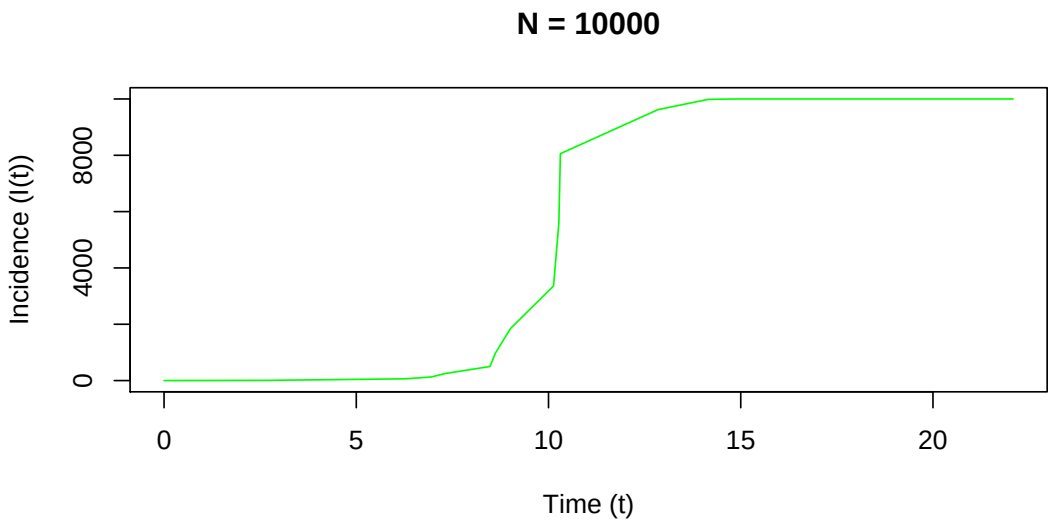
Another interesting thing is how the wavelet diagrams look very similar for both london and liverpool, showing that the disease is functioning similarly at two different population levels. Some things that are puzzling about the figure are why the time series begins to dampen down in liverpool in 1965 where as the dampening begins to ocur in london in 1970. Also, it is of interest that the time series of london begins to dampen right after the largest peak of the time series. Further, we see in the liverpool, in the third segment periodogram there appears to be very little power in that segment, despite the apperance of some oscillation in the time series. This may be an issue of the axis scales, as more clear difference might be observed with a smaller axis, but it is a clear contrast between the observed time series and the period periodograms.

2. Consider the SI model,

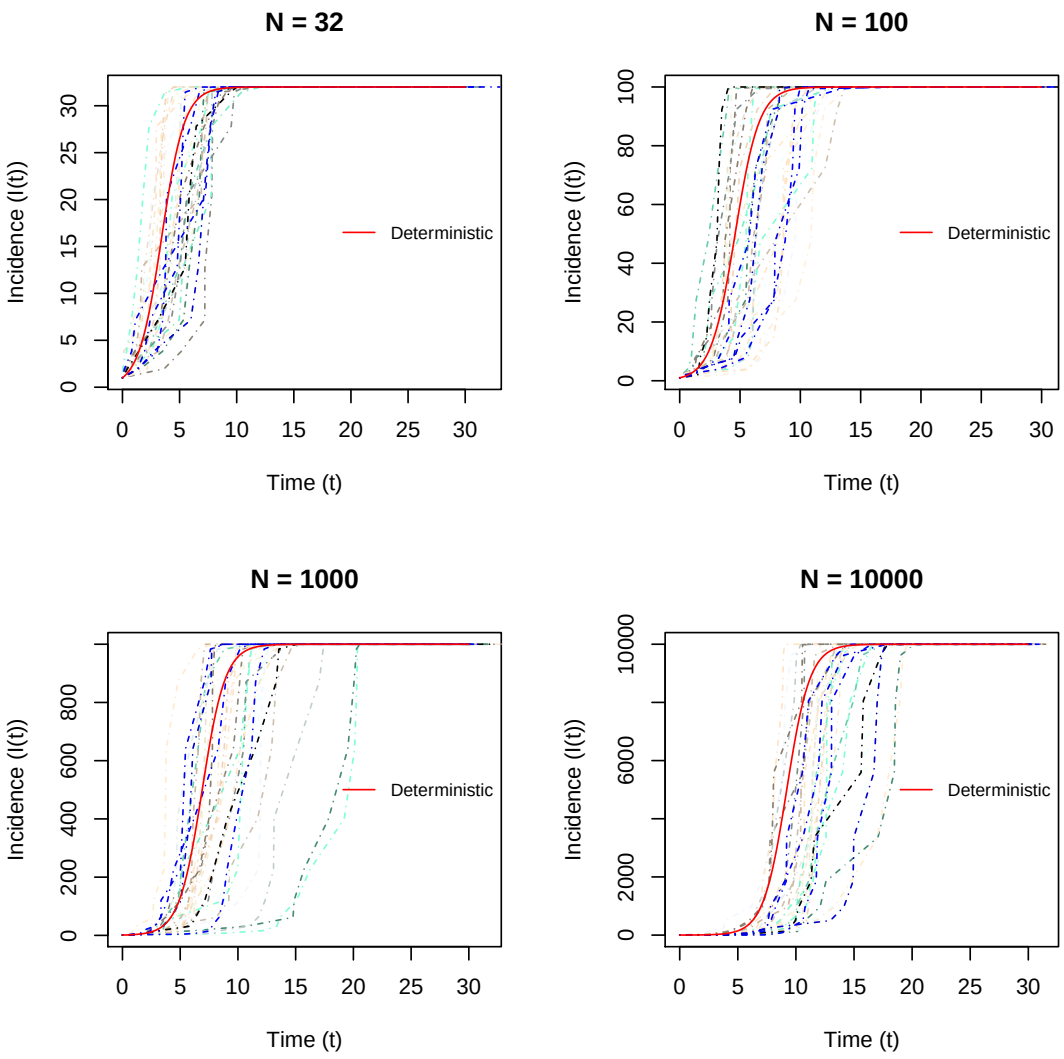
$$\frac{dI}{dt} = \beta(N - I)I\,, \qquad I(0) = I_0, \tag{1}$$

where β is the transmission rate, N is the population size and $I(t)$ is the number of infected individuals at time t.

(a)
```
source('gillespie.R')
#simulation vars
beta = 1
N = 10000
I0 = 1
tmax = 20
realizations = 30
result <- SI.Gillespie(beta,N,I0,tmax)
plot(result[[1]], result[[2]], col="green", type="l", xlab="Time (t)",
     ylab="Incidence (I(t))", main=paste("N =",N))
```



(b)
```
source('gillespie.R')
#simulation vars
beta = 1
ns = c(32,100,1000,10000)
I0 = 1
tmax = 30
realizations = 30
colors <- colors()
multipanel(realizations,beta,ns,I0,tmax,colors=colors)
```



3. The natural history of smallpox is shown in Figure 1. The US Centers for Disease Control and Prevention (CDC) has recently discovered that a group of bioterrorists plans to reintroduce smallpox to the United States. The CDC has reason to believe that the terrorists are also bioengineers and have successfully altered the virus so that it causes the early rash stage to be twice as long as it was when the virus was last circulating naturally in the 1970s. Moreover, the existing smallpox vaccine apparently provides no protection against the altered virus. The CDC wants your opinion on how the alterations to the virus will affect $\mathcal{R}_0$ and the expected final size of an epidemic if the planned attack is successful.

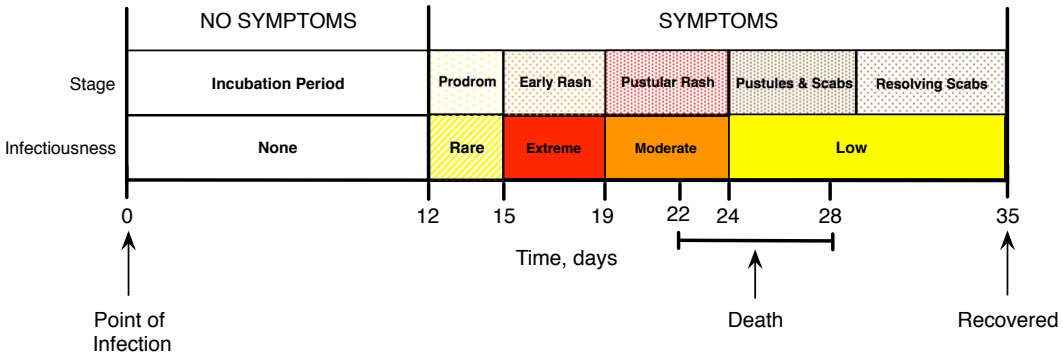
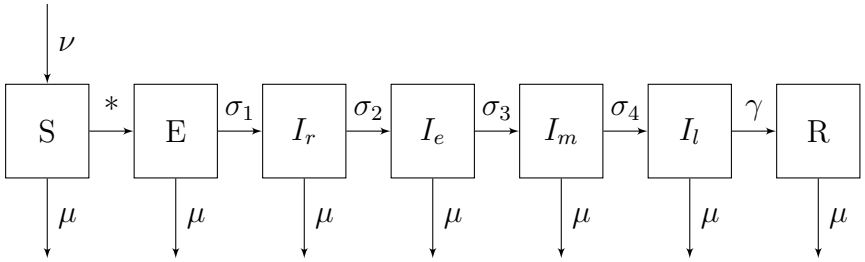


Figure 1: The natural history of smallpox infection. The prodrom stage begins with fever but the patient is very rarely contagious. Early rash is the most contagious stage, when the rash develops and transforms into bumps. During the pustular rash stage bumps become pustules, which then turn into scabs during the pustules and scabs stage and fall off during the resolving scabs stage. The infected person is contagious until the last scab falls off. (*This is Figure 3.4 from page 82 of Olga Krylova's 2011 McMaster University PhD thesis.*)

(a)



Note: $* = \beta_r I_r + \beta_e I_e + \beta_m I_m + \beta_l I_l$
Each I_i compartment represents a disease stage (prodrom, early rash, etc.) that has a unique rate of infectiousness (rare, extreme, moderate, and low) and transmission rate β_i. The parameters ν and μ represent the "birth" and natural death rates, respectively.

$$
\begin{aligned}
\frac{dS}{dt} &= \nu - \beta_r I_r S - \beta_e I_e S - \beta_m I_m S - \beta_l I_l S - \mu S \\
\frac{dE}{dt} &= \beta_r I_r S + \beta_e I_e S + \beta_m I_m S + \beta_l I_l S - \sigma_1 E - \mu E \\
\frac{dI_r}{dt} &= \sigma_1 E - \sigma_2 I_r - \mu I_r \\
\frac{dI_e}{dt} &= \sigma_2 I_r - \sigma_3 I_e - \mu I_e \\
\frac{dI_m}{dt} &= \sigma_3 I_e - \sigma_4 I_m - \mu I_m \\
\frac{dI_l}{dt} &= \sigma_4 I_m - \gamma I_l - \mu I_l \\
\frac{dR}{dt} &= \gamma I_l - \mu R
\end{aligned}
$$

(b) In this case the susceptible population can come into contact with individuals in 4 different stages of infectiousness. Each ith term in $\mathcal{R}_0$ corresponds to the number of new infectious individuals per individual that stays in compartment i. For example, the first term in $\mathcal{R}_0$ correspond to the number of secondary infectious individuals per individual in the prodrom stage of the disease.

More specifically, each term in $\mathcal{R}_0$ is the product of the transmission rate of an individual in stage i (each β_i term), the proportion of individuals that survives to stage i, and the average time each individual that enters the ith stage stays in the ith stage.

For example, the second term in $\mathcal{R}_0$ corresponds the extreme infectivity stage, where the transmission rate of an individual in this stage is β_e, the proportion of individuals that survives to and enters this stage is $\frac{\sigma_1\sigma_2}{(\sigma_1+\mu)(\sigma_2+\mu)}$, and the average time each individual that enters the extreme stage stays in this stage is $\frac{1}{\sigma_3+\mu}$.

Ergo, each transmission possibility with the infectious individuals in all 4 stages should be calculated similar to the $SEIR$ model, and they can be summed up to return the R_0 value. Putting the 4 equations together we get:

$$
\begin{aligned}
\mathcal{R}_0 = &\frac{\beta_r\sigma_1}{(\sigma_1+\mu)(\sigma_2+\mu)} + \frac{\beta_e\sigma_1\sigma_2}{(\sigma_1+\mu)(\sigma_2+\mu)(\sigma_3+\mu)} \\
&+ \frac{\beta_m\sigma_1\sigma_2\sigma_3}{(\sigma_1+\mu)(\sigma_2+\mu)(\sigma_3+\mu)(\sigma_4+\mu)} + \frac{\beta_l\sigma_1\sigma_2\sigma_3\sigma_4}{(\sigma_1+\mu)(\sigma_2+\mu)(\sigma_3+\mu)(\sigma_4+\mu)(\mu+\gamma)}
\end{aligned}
$$

(c) Assuming that $\mathcal{F}$ = inflow of new infecteds to infected compartments, and $\mathcal{V}$ =

outflow from infected compartments minus inflow of non-new infecteds we have:

$$\mathcal{F} = \begin{pmatrix} \beta_r I_r S + \beta_e I_e S + \beta_m I_m S + \beta_l I_l S \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \mathcal{V} = \begin{pmatrix} \sigma_1 E + \mu E \\ -\sigma_1 E + \sigma_2 I_r + \mu I_r \\ -\sigma_2 I_r + \sigma_3 I_e + \mu I_e \\ -\sigma_3 I_e + \sigma_4 I_m + \mu I_m \\ -\sigma_4 I_m + \gamma I_l + \mu I_l \end{pmatrix}$$

Let F = linearization of $\mathcal{F}$ at DFE.
Let V = linearization of $\mathcal{V}$ at DFE. Calculating F and V we have:

$$F = \begin{pmatrix} 0 & \beta_r & \beta_e & \beta_m & \beta_l \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$V = \begin{pmatrix} (\sigma_1 + \mu) & 0 & 0 & 0 & 0 \\ -\sigma_1 & (\sigma_2 + \mu) & 0 & 0 & 0 \\ 0 & -\sigma_2 & (\sigma_3 + \mu) & 0 & 0 \\ 0 & 0 & -\sigma_3 & (\sigma_4 + \mu) & 0 \\ 0 & 0 & 0 & -\sigma_4 & (\gamma + \mu) \end{pmatrix}$$

The next generation matrix is FV^{-1}. Calculating FV^{-1} using a symbolic manipulation software we get:

$$FV^{-1} = \begin{pmatrix} m_1 & m_2 & m_3 & m_4 & m_5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where

$$m_1 = \frac{\beta_r\sigma_1}{(\sigma_1+\mu)(\sigma_2+\mu)} + \frac{\beta_e\sigma_1\sigma_2}{(\sigma_1+\mu)(\sigma_2+\mu)(\sigma_3+\mu)} + \frac{\beta_m\sigma_1\sigma_2\sigma_3}{(\sigma_1+\mu)(\sigma_2+\mu)(\sigma_3+\mu)(\sigma_4+\mu)} + \frac{\beta_l\sigma_1\sigma_2\sigma_3\sigma_4}{(\sigma_1+\mu)(\sigma_2+\mu)(\sigma_3+\mu)(\sigma_4+\mu)(\mu+\gamma)}$$

$$m_2 = \frac{\beta_r}{\sigma_2+\mu} + \frac{\beta_e\sigma_2}{(\sigma_2+\mu)(\sigma_3+\mu)} + \frac{\beta_m\sigma_2\sigma_3}{(\sigma_2+\mu)(\sigma_3+\mu)(\sigma_4+\mu)} + \frac{\beta_l\sigma_2\sigma_3\sigma_4}{(\sigma_2+\mu)(\sigma_3+\mu)(\sigma_4+\mu)(\mu+\gamma)}$$

$$m_3 = \frac{\beta_e}{\sigma_3+\mu} + \frac{\beta_m\sigma_3}{(\sigma_3+\mu)(\sigma_4+\mu)} + \frac{\beta_l\sigma_3\sigma_4}{(\sigma_3+\mu)(\sigma_4+\mu)(\mu+\gamma)}$$

$$m_4 = \frac{\beta_m}{\sigma_4+\mu} + \frac{\beta_l\sigma_4}{(\sigma_4+\mu)(\mu+\gamma)}$$

$$m_5 = \frac{\beta_l}{\mu+\gamma}$$

$\mathcal{R}_0$ can be calculated as the spectral radius of the matrix FV^{-1} (or $\rho(FV^{-1})$). Therefore, as the matrix is upper-triangular:

$$\mathcal{R}_0 = \rho(FV^{-1}) = \frac{\beta_r\sigma_1}{(\sigma_1+\mu)(\sigma_2+\mu)} + \frac{\beta_e\sigma_1\sigma_2}{(\sigma_1+\mu)(\sigma_2+\mu)(\sigma_3+\mu)} + \frac{\beta_m\sigma_1\sigma_2\sigma_3}{(\sigma_1+\mu)(\sigma_2+\mu)(\sigma_3+\mu)(\sigma_4+\mu)} + \frac{\beta_l\sigma_1\sigma_2\sigma_3\sigma_4}{(\sigma_1+\mu)(\sigma_2+\mu)(\sigma_3+\mu)(\sigma_4+\mu)(\mu+\gamma)}$$

(d) The early rash stage of the altered virus is expected to be twice as long as the original strain, meaning that the mean infectious period in the early rash stage ($\frac{1}{\sigma_3}$) has doubled. In other words, σ_3 is divided in half. So the $\mathcal{R}_0$ for the new strain can be written as:

$$\mathcal{R}_0 = \frac{\beta_r\sigma_1}{(\sigma_1+\mu)(\sigma_2+\mu)} + \frac{\beta_e\sigma_1\sigma_2}{(\sigma_1+\mu)(\sigma_2+\mu)((\sigma_3/2)+\mu)} + \frac{\beta_m\sigma_1\sigma_2(\sigma_3/2)}{(\sigma_1+\mu)(\sigma_2+\mu)((\sigma_3/2)+\mu)(\sigma_4+\mu)} + \frac{\beta_l\sigma_1\sigma_2(\sigma_3/2)\sigma_4}{(\sigma_1+\mu)(\sigma_2+\mu)((\sigma_3/2)+\mu)(\sigma_4+\mu)(\mu+\gamma)}$$

Since β_e is the highest infectiousness rate (extreme), this change in σ_3 will have the most significant impact on the second fraction of the equation, $\frac{\beta_e\sigma_1\sigma_2}{(\sigma_1+\mu)(\sigma_2+\mu)((\sigma_3/2)+\mu)}$, almost doubling its value.
The changes in the last 2 fractions in $\mathcal{R}_0$ are very minor and can be ignored. Since natural death rate (μ) is much smaller than σ_3, the changes in the numerator almost cancel out the changes in the denominator completely. Thus, $\mathcal{R}_0$ can be expected to increase significantly in the new virus.

(e) If the bio-terrorists were able to double the early rash stage of smallpox, this will greatly increase the reproductive ratio $\mathcal{R}_0$ by nearly doubling what was calculated

for the unaltered strain of the disease ($\mathcal{R}_0 \simeq 5$). Increasing $\mathcal{R}_0$ will increase the critical proportion of vaccinated individuals in the population from 80% to 90% in order to prevent an epidemic. Using the final size formula $Z = 1 - epx(-\mathcal{R}_0 * Z)$ for the unaltered disease yields a final size of $Z = Z = 1 - epx(-5 * Z)$, so without intervention the final size of an epidemic caused by the unaltered strain is estimated to be 99.3%. With the altered strain,with $Z = 1 - exp(-10 * Z)$ the final size of the epidemic without any intervention will be similar but much closer to 100%. Potentially creating a viable pre-epidemic vaccination to the susceptible population will be effective in decreasing $\mathcal{R}_0$ and moreover will decrease the estimated final size of an epidemic. If an effective vaccine can be mass implemented to reach 20% of the United States population, then the reproductive ratio can be decreased to 80% of the original estimated $\mathcal{R}_0$ and reduce the final size estimation accordingly. Additionally, if an outbreak happens, isolating individuals when they enter the extremely infectious (early rash) stage of the disease will have greatest impact on decreasing $\mathcal{R}_0$. Thus it is imperative to implement vaccination, isolation, and quarantine intervention methods as soon as possible.

— END OF ASSIGNMENT —

Compile time for this document: March 12, 2019 @ 23:55