Disease Modeling Assignment one submitted to Alfred Ssekagiri

By: Kakembo Fredrick E

5/10/2020

Contents

1. Background of Ebola Virus Disease.

Ebola virus disease (EVD) is a deadly disease caused by the Ebola virus, a virus which is belongs to filoviridae virus family. The disease is spread from one person to another via direct contact with infected bodily fluids, surfaces and materials. The EVD has been known to have a number of outbreaks within the African continent with even the very first reported case in 1976 to have occured near the Ebola river in the Democratic Republic of Congo where the virus got its name.

To date a number of virus species have been named basing on the area where the disease was caused or found such as the Sudan virus, Zaire virus, Tai Forest virus, Bundibugyo virus, Roston and Bombali virus. The virus has been reported to have an average mortality rate of about 50%, meaning on average if 100 people

got infected with the Ebola virus, close to 50 indivuals would probably not survive the EVD. This however has varied between 20% to 90% in some outbreaks in different countries.

The origin of the virus to date has not yet been fully understood though it is believed to have originated from animals (primates in particular such as monkeys). Various approaches have of recent been applied to better understand the transmission dynamics of the Ebola virus disease.

The aim of this report will be to apply a simple SEIR (Susceptible Exposed Infected and Recovered) compartment model framework while incorperating the aspect of Relapsing and Reinfection of previous patients to understand the transmission dynamics of EVD, an aspect that has not been catered for in most of the previous literature published.
We will then fit my model to the **2014-2016 Sierra Leone Ebola outbreak data** to determine how best our model performs on real data and also try different intervention measures and how they could perform based on our model.

2. Model Formulation

For this assignment we will use a 7 compartmental framework model to describe the transmission dynamics of the Ebola virus disease. This model has been modified from the Literature published by Agusto *et al*, 2016 ***found here***. This has been modified from previous literature so as to add aspect of relapses and reinfections.

2.1 Model Assumptions

This model is built on a number of assumptoins, which include;

1. All individuals in the population are very susceptible to be affected by the Ebola Virus disease with exception of only those individuals that have fully recovered from the disease that are assumed to acquire a permanent immunity.
2. When an indivual is exposed to the Ebola virus, they might progress to the disease to show signs and symptoms either early enough after infection or later on after being infected. This is very important since it is very well known that a person becomes infectious when they start showing symptoms of the virus, meaning different stages of the infections/Exposure have different infectious rates.
3. Individuals affected with the virus have 2 fates, either to die from the EVD (which we will term as Disease induced death) or they may be lucky to recover from the disease.
4. Individuals who recover from the virus during early stages remain both habouring the virus hence remain a potential source of infection to other individuals in the community or could stand a chance of having a relapse with the EVD.
   This is because previous studies have shown the virus to be present in fluids of recovered individuals such as semen, sweat, urine for upto 2 to 9 months.
5. Individuals who completely recover from the virus and have no trace of the virus in their body as are assumed to acquire immunity against the virus hence nolonger counted among the susceptible population.
6. Bodies of indivuduals dead from the virus remain a potential source of infection to the rest of the people in the community, to those who get in contact with any of the fluid from the deceased individual.
7. Natural death is accounted for at all the compartments of the model as death arising from any other cause apart from the Ebola virus. This is because the natural deaths have an effect of decreasing the to number of individuals in any given disease state and at any time.
8. Birth rates are however not fully accounted for here since outbreaks are believed to be slightly short lived hence highest chances are new borns are believed to be fully isolated from commnunity hence not a part of susceptible group.

2.2 Flow Diagram of the Ebola Transmission model.

A graphical reprsentation of the model and the disease states are shown below.

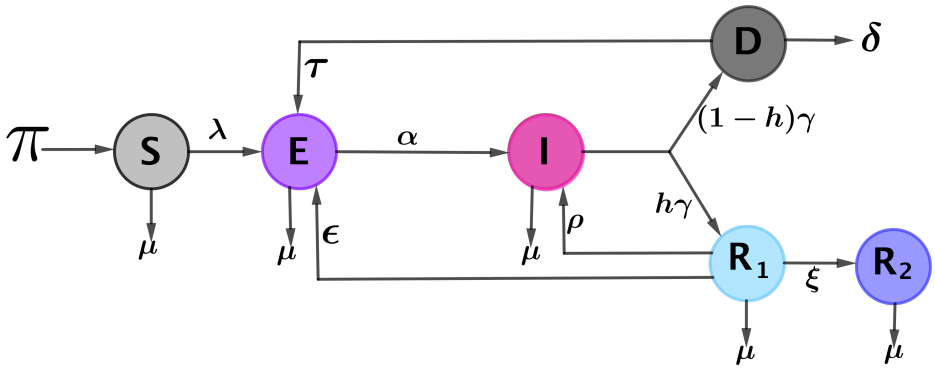


Figure 1: A Flow Diagram of the transmission dynamics model of Ebola virus disease. The model starts with a recruitment rate of individuals into the susceptible Population, S, that are at a risk of being infected with the Ebola Virus disease. Individuals in the **S** compartment are exposed to the virus at a force of infection rate λ, they move into the Exposed group group,E, among which some of these individuals become infected and start showing symptoms of the virus at an infection rate α. Among these infected individuals, a proportion of them recover at a rate γ though could still habour the virus, hence chances of having a relapse of the disease at a rate ρ, or potential source of infection to other susceptible individuals at a rate of ϵ. The individuals habouring the virus could later heal completely at a rate of ξ to become disease free. On the other hand, a proportion of the infected indiduals could die from the disease (disease induced death) at a rate of δ. Dead bodies due to Ebola are a potential source of infection to the susceptible population at a rate of τ. Natural death μ, is accounted for at each state since individuals could die from any other cause apart from the Ebola virus disease, hence reducing the individuals in that state.

2.3 Summary Definition of the States used and parameters.

a.Disease States

- **S** - Susceptible individuals in the population
- **E** - Population of Exposed individuals
- **I** - Individuals infected with Ebola Virus Disease that are showing symptoms
- $\mathbf{R}_1$ - Population of Recovered individuals but are still habouring the vius
- $\mathbf{R}_2$ - Population completely recovered Individuals, free from virus and now immune to the EVD
- **D** - Population of individuals who died from EVD ie Disease induced death

b.Parameters

- π - Rate of Recruitment of individuals into the Susceptible population
- λ - Rate of Exposure of the susceptible individuals. This force of exposure,λ, is defined by [β * I/N], where β is the average number of secondary infections per unit time per infected person.
- α - Rate of progression of Exposed individuals to a symptomatic Inectious stage
- μ - Natural Death Rate
- $\boldsymbol{h}$ - Proportion of individuals who Recovered from EVD

- γ - Rate of recovery of symptomatic individuals
- δ - Death rate caused by Ebola virus disease
- ξ - Rate of progression of recovered individuals to the immune class
- ρ - Infection reactivation rate by the recovered individuals haboring the Ebola virus
- ϵ - Infection Exposure rate from the Recovered indiduals still hobouring the virus
- τ - Infection Exposure rate from the EVD dead individuals

2.4 Model Equations

From the flow diagram above, mathematical partial derivative equations describing the model for the different disease states at a given specified time,t, are given below;

$\mathrm{N}(t) = \mathrm{S}(t) + \mathrm{E}(t) + \mathrm{I}(t) + \mathrm{R}_1(t) + \mathrm{R}_2(t) + \mathrm{D}(t)$

$$\frac{dS}{dt} = \pi N - \lambda S - \mu S \qquad [where..., \lambda = \beta\frac{I}{N} + \epsilon\beta\frac{R_1}{N} + \tau\beta\frac{D}{N}] \tag{1}$$

$$\frac{dE}{dt} = \lambda S - \mu E + \epsilon R_1 + \tau D - \alpha E \qquad [where..., \lambda = \beta\frac{I}{N} + \epsilon\beta\frac{R_1}{N} + \tau\beta\frac{D}{N}] \tag{2}$$

$$\frac{dI}{dt} = \alpha E - \mu I - h\gamma I - (1-h)\gamma I \tag{3}$$

$$\frac{dR_1}{dt} = h\gamma I - \mu R_1 - \rho R_1 - \xi R_1 \tag{4}$$

$$\frac{dR_2}{dt} = \xi R_1 - \mu R_2 \tag{5}$$

$$\frac{dD}{dt} = (1-h)\gamma I - \delta D \tag{6}$$

Note: The infection exposure rates to the Ebola Virus from the dead bodies ,τ, and from the partially recovered inidividuals,ϵ, will only affect the number of exposed indiduals but will not affect (increase or decrease) the number of partially recovered individuals as well as disease induced death individuals.

3. Reproduction Numbers of the model

3.1 Basic Reproduction Number R_o

The Basic Reproduction Number, R_o, is defined as the average number of secondary cases caused by a single infected case/individual in an otherwise susceptible population.

Basically, R_o is derived from 2 major model parameters ie;

- Infectivity rate, β, which is the average number of secondary infections per unit time per person eg if its $2\mathrm{day}^{-1}$, means an infected individual will cause infections per day.
- Recovery Rate, γ, which is the hazard rate of recovery (or death) from the infection eg if its 0.2 meaning the infectious period is 5 days. It is given by 1/(number of infectious days).

Hence from this, the mean number of secondary infections, R_o, is the number of infections per day(β) multiply by the number of days for which people are infcetious (γ), ie

$$R_o = \frac{\beta}{\gamma} \tag{7}$$

From our model formulation we have 3 infectoius states to Ebola virus; ie

i. Infected individuals symptomatic of the Ebola virus disease. These have a rate of infection β to the susceptible population.
ii. Partially recovered individuals still habouring the disease. These have a slightly reduced force of infection by a rate ϵ.
iii. Bodies of individuals who died from the Ebola virus disease. These too have a slightly reduced force of infection by a rate τ.

From this we therefore derive our Basic Reproduction Number, R_o, as;

$$R_o = \frac{\beta + \epsilon\beta + \tau\beta}{\gamma} \tag{8}$$

3.2 Effective Reproduction Number, R_{eff}

Effective Reproduction Number, R_{eff}, is defined as the average number of secondary infections per infected case at a given point in the epidemic.

One main drawback of only focusing on the basic Reproduction Number, R_o, is that it assumes the number of the Susceptible population to be constant throughout the entire period of the epidemic. This is however not always the case since as susceptible individuals become infected and transit into the other states (eg Exposed E, Infected I, Recovered E, or dead D), the number of suceptible individuals in the S compartment gradually decrease depending on the infectivity of the virus.

A more realistic and prefered measure of the state of the infection at any given point in time is the Effective Reproduction Number which takes into account of the different fates of the infection such as immunity or death from the disease.
The Effective Reproduction Number,R_{eff}, is defines as ;

$$R_{eff} = R_o \frac{S}{N} \tag{9}$$

hence at the beginning of the epidemic when the number of Susceptible individuals is almost equal to the total population, N, the Basic Reproduction number is equal to the Effective Reproduction number (ie $R_o = R_{eff}$), but R_{eff} gradually decreases as the number of susceptible population decreases.

4. Application of the model

For this assignment we made use of the **2016 Sierra Leone Ebola outbreak data** downloaded from the World Health Organisation website [**Link**]. This contained all the numbers of confirmed cases during this period of the outbreak. Using this data we shall go ahead to estimate the model parameters that best define this outbreak, after which we shall fit the model with the data. We will conclude by modeling out some intervation measures and determine their effect and efficiency on the outbreak and duration it would last if they well applied.

4.1 Estimation of Model Parameters

There are 2 common methods used in the determination of model parameters given the data. These include;

1. Sum of Squares, SSQ. In this method, we start by providing arbitrary values for the model parameters for which the SSQ algorithm computes the difference between provided parameters and the actual parameters that best fit the data. The aim of the algorithm is vary the input abitrary values of the model parameters while calculating their SSQ value so as to produce a value the parameters that minimizes the value of the Sum of Squares which will correspond to the parameters that best fit the input data.
2. Maximum Likelihood method. This method similar to the Sum of Squares method, will try to modfiy the input abitrary parameter values so as to produce values that best describe the data. This unlike the SSQ method, it calculates the posterior probabilities, and the aim is obtaining the maximum probablity value that given the data, the model parameters returned best describe the outbreak data.

For this assignment we will use the **Least Sum of Squares** approach implemeted in R code and the **deSolve** pacakge to determine the values to the model parameters that descibe the outbreak data.

We cleaned the data to be used in an Excel sheet to contain the values in the format we desired. We will go ahead and load the data here.

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

# Setting Our Working Directory
setwd("~/Desktop/MSBT/Disease Dynamics _ Modelling /Assignment/")

# Reading the data
data <- read.csv("data.csv")

# Removing unwanted variables
data$Country <- NULL
data$Date <- NULL
data$DAY.by.date <- NULL
data$No..of.confirmed.cases <- NULL
data$No..of.confirmed.deaths <- NULL

# Extracting only the data rows
data <- data[1:230, ]

# Viewing the data
kable(data[1:10, ], caption = "First 10 entries for the outbreak data")
```

Table 1: First 10 entries for the outbreak data

Days	Cases	Deaths
0	935	380
7	211	63
10	88	18
14	53	17
18	177	36

Days	Cases	Deaths
20	49	3
24	127	28
26	105	7
28	71	5
33	260	17

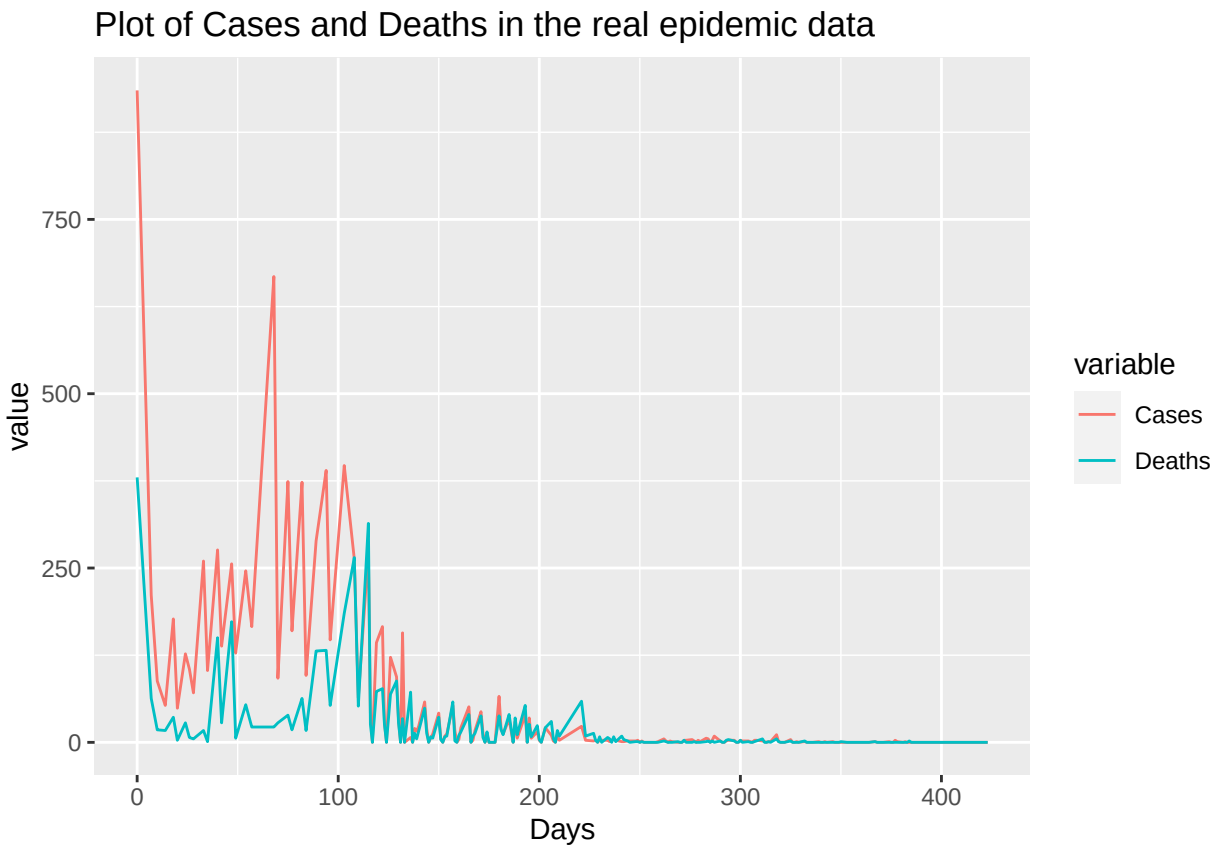
```
kable(data[221:230, ], caption = "Last 10 enties of the outbreak data")
```

Table 2: Last 10 enties of the outbreak data

	Days	Cases	Deaths
221	404	0	0
222	405	0	0
223	406	0	0
224	410	0	0
225	412	0	0
226	413	0	0
227	417	0	0
228	419	0	0
229	420	0	0
230	423	0	0

```
# Plotting the data
data_long <- melt(data, id = "Days")
ggplot(data = data_long, aes(x = Days, y = value, group = variable, color = variable)) +
    geom_line() + labs(title = "Plot of Cases and Deaths in the real epidemic data")
```

```
#Defining model function
SIR_model <- function(time, state, parameters) {
  with(as.list(c(state, parameters)),{
    N = S + E + I  + R1 + R2 + D

    lambda = (beta*I/N) + (epsilon*beta*R1/N) + (tau*beta*D/N)

    dS =  - lambda*S - mu*S
    dE =  lambda*S - mu*E + epsilon*R1 + tau*D - alpha*E
    dI =  alpha*E -mu*I - gamma*I
    dR1 =  h*gamma*I - mu*R1 - rho*R1 -  xi*R1
    dR2 =  xi*R1 - mu*R2
    dD =  (1-h)*gamma*I - delta*D

    return(list(c(dS, dE, dI, dR2, dR2, dD)))
  })
}

SIR_SSQ <- function(parameters, dat) {  # parameters must contain beta & gamma

    # calculate model output using your SIR function with ode()
```


Plot of Cases and Deaths in the real epidemic data
value
750
500
250
0
0
100
200
300
400
Days
variable
Cases
Deaths

```
    result <- as.data.frame(ode(y = initial_state_values  # vector of initial state
                                                          # values, with named elements
                                , times = times           # vector of times
                                , func = SIR_model           # your predefined SIR function
                                , parms = parameters)     # the parameters argument
                                                          # entered with SIR_SSQ()
    )

    # SSQ calculation: needs the dat argument (the observed data you are fitting to)
    # assumes the data you are fitting to has a column "I"

    # select only complete cases, i.e. rows with no NAs, from the dataframe
    dat <- na.omit(dat)

    # select elements where results$time is in dat$Days
    # And the compute sum of SSQ between the infected and the dead with that of real data
    deltas2 <- ((result$I[result$time %in% dat$Days] ) - dat$Cases)^2 +
                (result$D[result$time %in% dat$Days] - dat$Deaths)^2

    SSQ   <- sum(deltas2)

    return(SSQ)

}
```

After us defining the model function and the SSQ function, we then choose arbitary values of the model parameters that we shall initialize our model with, so as to come to the most optimal solution. To automate this process, we shall use the **Optim()** function from R that will run the SSQ function a number of times to determine which values are a better parameters in representing the model data.

To do this, we shall assume a population size of 1M individuals in which we have one infected individual, and no deaths so far.

```
#Initial parameter values
initial_state_values <- c(S = 10000-1, E=0, I=1, R1=0, R2=0, D=0)

# choose values to start our optimisation
beta_start  <- 1
gamma_start <- 0.5
epsilon_start <- 0.2
tau_start <- 0.4
pi_start <- 1
mu_start <- 0.001
alpha_start <- 1
h_start <- 0.5
xi_start <- 0.8
rho_start <- 0.15
delta_start <- 0.5

# times - dense timesteps for a more detailed solution
times <- seq(from = 0, to = 400, by = 0.1)

# Running optim and assigning output to optimised
```

```
optimised <- optim(par = c(beta = beta_start,
                           epsilon = epsilon_start,
                           tau = tau_start,
                           pi = pi_start,
                           mu = mu_start,
                           alpha = alpha_start,
                           h = h_start,
                           gamma = gamma_start,
                           xi = xi_start,
                           rho = rho_start,
                           delta = delta_start)

                         , fn = SIR_SSQ
                         , dat = data
      )

optimised #Lets have a look at the model output
```

```
## $par
##         beta      epsilon          tau           pi           mu        alpha
##  0.674783351  0.079486107  0.334117689  1.023556203  0.002019265  0.918331361
##            h        gamma           xi          rho        delta
##  0.502841836  0.742562278  1.002756603 -0.022289310  0.850545613
##
## $value
## [1] 2259206
##
## $counts
## function gradient
##      502       NA
##
## $convergence
## [1] 1
##
## $message
## NULL
```

The `optim` function gave us the following parameters to be the most optimum values for our to SEIR model to better represent the Sierra Lion data. The parameter values are:

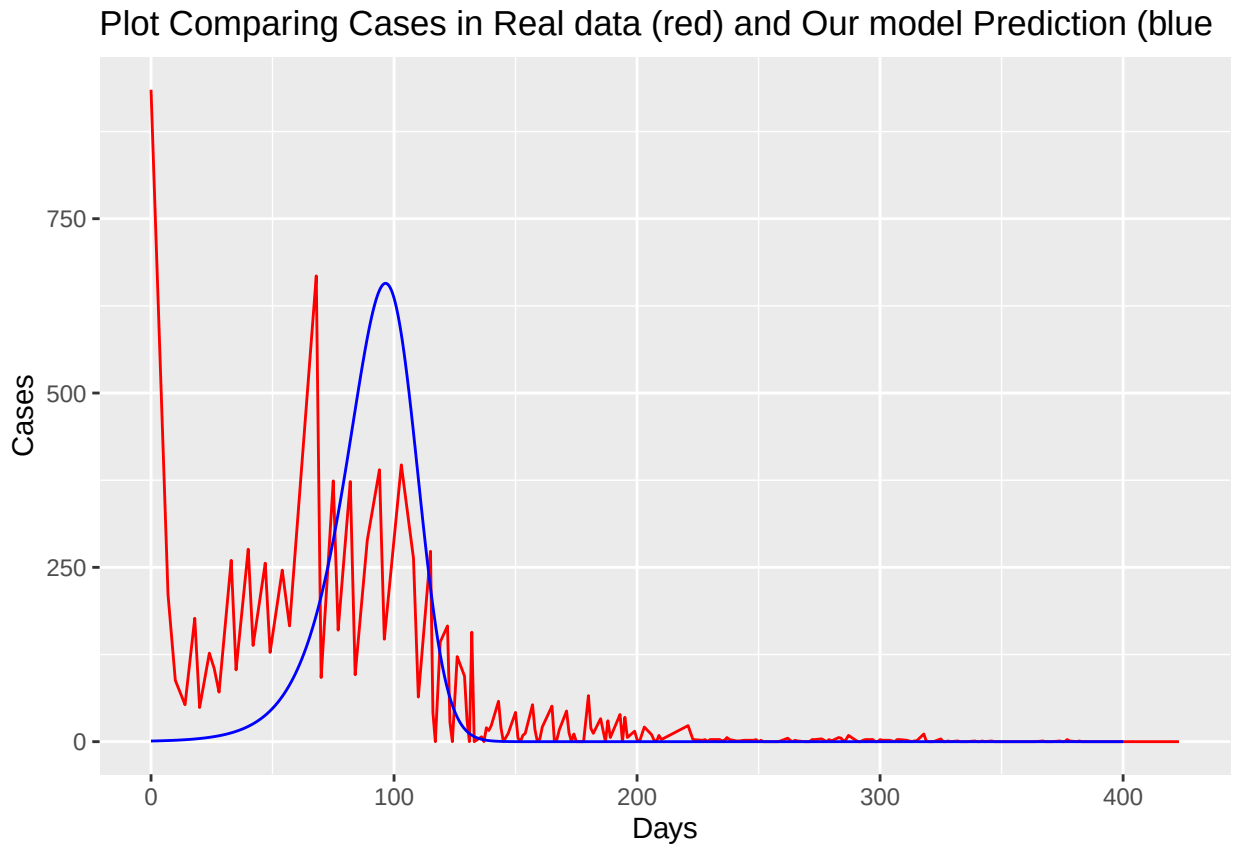
```
# Parameters produced
optimised$par
```

```
##         beta      epsilon          tau           pi           mu        alpha
##  0.674783351  0.079486107  0.334117689  1.023556203  0.002019265  0.918331361
##            h        gamma           xi          rho        delta
##  0.502841836  0.742562278  1.002756603 -0.022289310  0.850545613
```

We will now use these parameters to generate a plot against the input data and have a visual how best our model represents the data.

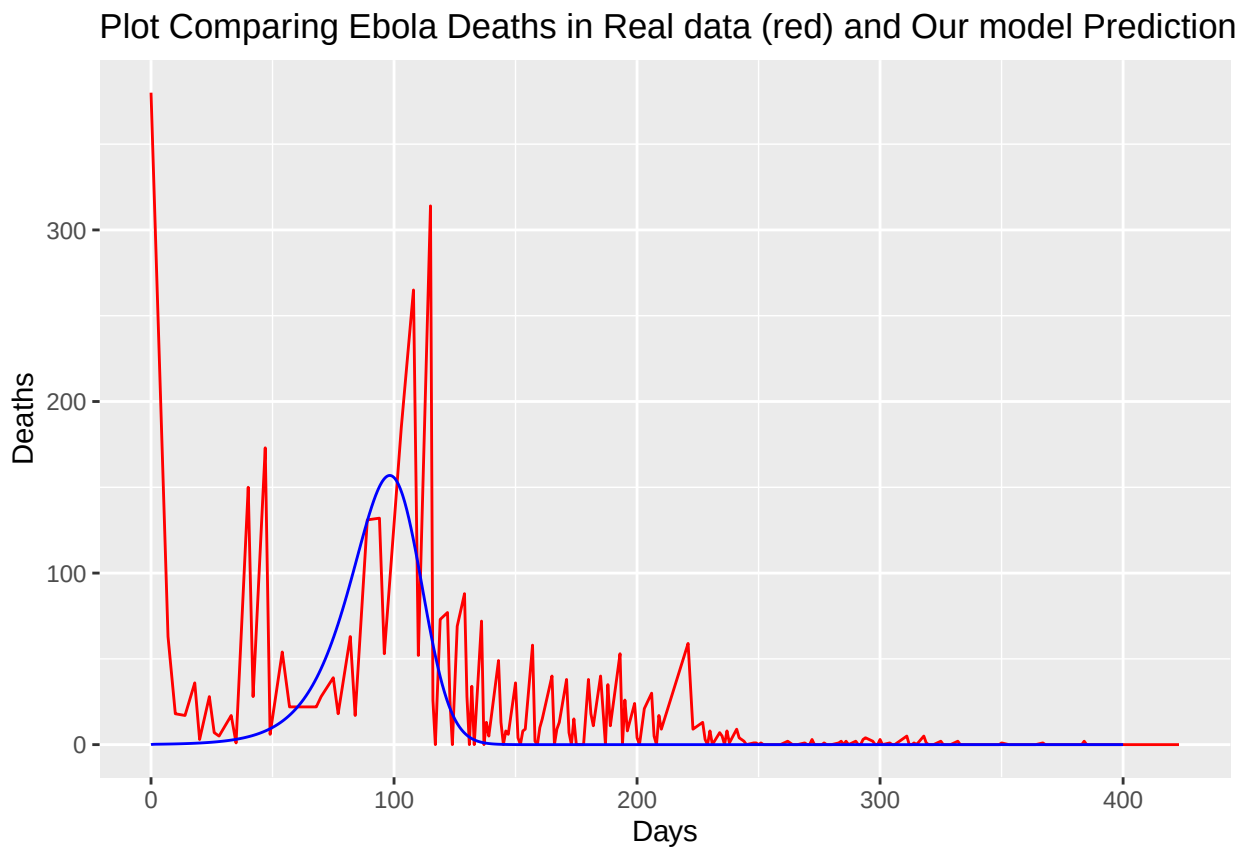
4.2 Plot our model and compare against the input data

```
## plotting our optimised model output, with the epidemic data using ggplot

# i. For the cases Using Parameters to generate model data
opt_mod <- as.data.frame(ode(y = initial_state_values  # named vector of initial state values
,
    times = times  # vector of times
, func = SIR_model  # our predefined SIR function
, parms = optimised$par))

opt_plot <- ggplot()
opt_plot <- opt_plot + geom_line(aes(x = Days, y = Cases), colour = "red", data = data)

opt_plot <- opt_plot + geom_line(aes(x = time, y = I + E), colour = "blue", data = opt_mod) +
    labs(title = "Plot Comparing Cases in Real data (red) and Our model Prediction (blue")

opt_plot
```




```
# ii. For the Deaths

# colors = c('Model_deaths' = 'blue', 'Real_deaths' = 'red')
opt_plot_death <- ggplot()
```

```
opt_plot_death <- opt_plot_death + geom_line(aes(x = Days, y = Deaths), colour = "red",
    data = data)

opt_plot_death <- opt_plot_death + geom_line(aes(x = time, y = D), colour = "blue",
    data = opt_mod) + labs(title = "Plot Comparing Ebola Deaths in Real data (red) and Our model Predic

opt_plot_death
```



Introducing a control Measure: Quarantine

For this model we shall consider a control measure of quarantening all infected individuals who have transited from the exposed category but now show actaul signs and symptoms hence higher spreading rates. By quarantinning all infected indiviaduals with symptoms, we intend to reduce to infectivity rate of the virus by reducing the ppossible transmission sources and only be left with the exposed individuals who are not yet showing any symptoms of the disease hence a much reduce infectivity rate. We assume there is a **viable (effective) treatment option** for the disease where all quarantined individuals will be fully treated and completely recover by the time quantining duration is done.

Lets now introduce γ_q the rate of removal of infected individuals from the population into a quarantine period until a period they are completely healed (ie transit directly to the R_2 compartment).

In so doing, our new model equations reasulting from a modification of equations 3 and 5 will be as follows;

$$\mathrm{N}(t) = \mathrm{S}(t) + \mathrm{E}(t) + \mathrm{I}(t) + \mathrm{R}_1(t) + \mathrm{R}_2(t) + \mathrm{D}(t)$$

$$\frac{dS}{dt} = \pi N - \lambda S - \mu S \qquad [where..., \lambda = \beta\frac{I}{N} + \epsilon\beta\frac{R_1}{N} + \tau\beta\frac{D}{N}] \tag{10}$$

$$\frac{dE}{dt} = \lambda S - \mu E + \epsilon R_1 + \tau D - \alpha E \qquad [where..., \lambda = \beta\frac{I}{N} + \epsilon\beta\frac{R_1}{N} + \tau\beta\frac{D}{N}] \tag{11}$$

$$\frac{dI}{dt} = \alpha E - \mu I - h\gamma I - (1-h)\gamma I - \gamma q I \tag{12}$$

$$\frac{dR_1}{dt} = h\gamma I - \mu R_1 - \rho R_1 - \xi R_1 \tag{13}$$

$$\frac{dR_2}{dt} = \xi R_1 - \mu R_2 + \gamma q I \tag{14}$$

$$\frac{dD}{dt} = (1-h)\gamma I - \delta D \tag{15}$$

Lets assume at any time, our control measure can only remove half of the total infected indivuals from the population into quarantine. This means our quarantine rate, γ_q will be given by;

$$\gamma_q = \frac{0.5 * I}{N} \tag{16}$$

Lets now see the effect of quarantine on the pandemic the disease outbreak with a graph.

```
# Re-Defining model function
SIR_model_qtine <- function(time, state, parameters) {
    with(as.list(c(state, parameters)), {
        N = S + E + I + R1 + R2 + D

        lambda = (beta * I/N) + (epsilon * beta * R1/N) + (tau * beta * D/N)
        # gammaq = (0.5*I)/N

        dS = -lambda * S - mu * S
        dE = lambda * S - mu * E + epsilon * R1 + tau * D - alpha * E
        dI = alpha * E - mu * I - gamma * I - gammaq * I
        dR1 = h * gamma * I - mu * R1 - rho * R1 - xi * R1
        dR2 = xi * R1 - mu * R2 + gammaq * I
        dD = (1 - h) * gamma * I - delta * D

        return(list(c(dS, dE, dI, dR2, dR2, dD)))
    })
}
```

```
## plotting our model with quarantine influence

# Defining initial parameters
initial_state_values <- c(S = 10000 - 1, E = 0, I = 1, R1 = 0, R2 = 0, D = 0)


# Using Parameters to generate model data
opt_mod_qtine <- as.data.frame(ode(y = initial_state_values  # named vector of initial state values
,
    times = times  # vector of times
```

```
, func = SIR_model_qtine  # our predefined SIR function
, parms = c(optimised$par, gammaq = 0.5)))

head(opt_mod_qtine)
```

```
##   time        S          E         I         R1         R2          D
## 1  0.0 9999.000 0.00000000 1.0000000 0.00000000 0.00000000 0.00000000
## 2  0.1 9996.917 0.06186946 0.8857815 0.04954679 0.04954679 0.03328164
## 3  0.2 9994.841 0.11377769 0.7898193 0.09878824 0.09878824 0.06014461
## 4  0.3 9992.770 0.15736938 0.7091897 0.14857453 0.14857453 0.08170492
## 5  0.4 9990.704 0.19403484 0.6414463 0.19970369 0.19970369 0.09889122
## 6  0.5 9988.642 0.22495491 0.5845357 0.25294017 0.25294017 0.11247862
```

```
tail(opt_mod_qtine)
```

```
##       time          S             E             I            R1            R2
## 3996 399.5 0.01723509 1.505566e+173 6.134484e+172 3.603716e+174 3.603716e+174
## 3997 399.6 0.01717664 1.665447e+173 6.785926e+172 3.986407e+174 3.986407e+174
## 3998 399.7 0.01711840 1.842307e+173 7.506548e+172 4.409738e+174 4.409738e+174
## 3999 399.8 0.01706035 2.037948e+173 8.303695e+172 4.878024e+174 4.878024e+174
## 4000 399.9 0.01700250 2.254365e+173 9.185493e+172 5.396038e+174 5.396038e+174
## 4001 400.0 0.01694484 2.493764e+173 1.016093e+173 5.969063e+174 5.969063e+174
##                  D
## 3996 1.217701e+172
## 3997 1.347013e+172
## 3998 1.490057e+172
## 3999 1.648291e+172
## 4000 1.823329e+172
## 4001 2.016955e+172
```

```
opt_plot_qtine <- ggplot()
opt_plot_qtine <- opt_plot_qtine + geom_line(aes(x = time, y = I + E), colour = "blue",
    data = opt_mod_qtine)
```

After it appears that after applying a control measure of quarantining the infected indivials and treating them separately will lead to a slight epidemic in the beginning 13th day of the start, however this is very shortlived compared to the case when we have no control measure in place.

This implies that quarantining of infected individuals is a good strategy to reduce spread of a disease befero it can become a serious epidemic. To further strengthen the control strategy, a vaccine could be looked into to ensure total eradication of the disease, however an ebola vaccine as we know is not currently not available hence we can not really include it among the possible viable options at the moment.

References

1. https://www.cdc.gov/vhf/ebola/about.html
2. https://github.com/imdevskp/ebola_outbreak_dataset/blob/master/ebola_data_cleaning.ipynb

_