Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa
- A Literature Review

Divye Gupta
B.Sc. Student
Saint Mary's University

1. Introduction

EBOLA - One of the world's deadliest viruses! Ebola Virus disease is systematic, attacking every organ and tissue of the human body except the bones and the skeletal muscles. The virus is transmitted from wild animals and spreads through human-to-human transmission. Recently, the largest Ebola outbreak occurred in 2014, claiming more than 3,850 lives in West Africa. Ebola originated from Ebola River in Africa in in northern Democratic Republic of the Congo. In December 2014, Ebola outbreak started in Guinea and later spread Sierra Leone, Liberia and Nigeria. Even after the involvement of public health associations, it was hard to reduce the number of infected cases because the control measures were yet to be determined. Control measures could be determined only by having a Real-time analysis of the number of deaths and infected cases due to EBOV (Ebola Virus).

Author introduces two main parameters to describe the spread of the infection; Basic Reproduction number (R_0) and Effective Reproduction Number (R_e). They are defined as the number of secondary infections generated by an infected index case in the absence and presence of control interventions. To stop the spreading of EBOV, we need to make sure that the R_e drops below 1. Which means that less than 1 person should be infected from a person that is suffering for Ebola to completely stop EBOV from spreading. Ebola has emerged a few times earlier and estimates of R_0 from two outbreaks in Congo (1995) and Uganda (2000) range from 1.3 to 2.74. Estimating the reproduction numbers was very important at that time because that would have made projections of epidemic for the following months and would allow taking precautionary steps accordingly.

The modeling of infectious diseases is a tool which has been used to study the mechanisms by which diseases spread, to predict the future course of an outbreak and to evaluate strategies to control an epidemic. The author describes the Ebola outbreak of 2014 using an SEIR model. In SEIR model, the population of outbreak is divided into four components: susceptible, exposed, infectious and recovered, with the numbers of individuals in a component, or their densities denoted respectively by S(t), E(t), I(t), R(t), that is N = S(t) + E(t) + I(t) + R(t) where N is the total population of outbreak.

The author uses available data (see Table 2) and standard data fitting techniques to obtain values for some of these parameters to give the best fit of the model to the data. The author states (in the Methods section) that he uses an algorithm based on ``maximum likelihood estimates" of the parameters, and that he used an algorithm Nelder & Mead that is implemented in the function optim. Optim is a function for general-purpose optimization based on Nelder–Mead, quasi-Newton and conjugate-gradient algorithms. It includes an option for box-constrained optimization and simulated annealing.

2. Methods

The SEIR model is described by the author using the following differential Equations:

$$\frac{dS}{dt} = -\beta(t)SI/N, \quad (1)$$

$$\frac{dE}{dt} = \beta(t)SI/N - \sigma E, \quad (2)$$

$$\frac{dI}{dt} = \sigma E - \gamma I, \quad (3)$$

$$\frac{dR}{dt} = (1-f)\gamma I. \quad (4)$$

Susceptible (S) Individuals are the people who are likely to get infected by EBOV. Once the virus is transmitted, they enter the Exposed (E) component before they get infected and come in the Infectious (I) category. After they are infected by the Ebola virus, they are either recovered and hence survive (R) or they die.

$$\mathcal{S} \to \mathcal{E} \to \mathcal{I} \to \mathcal{R}$$

The average duration of incubation i.e. the period between which the infection started developing inside and when the symptoms appear is calculated by $1/\sigma$. The average duration of infectiousness is denoted by $1/\gamma$. The fatality rate or the death rate is given by *f.* The rate of transmission of EBOV from one individual to other in absence of the control interventions is constant (since it spreads at a same rate) and is given by β (t) = β. When there are interventions, then β (t) changes with time according to the following equation:

$$\beta(t) = \beta e^{-k(t-\tau)},$$

After the control measures came into action at time $\tau \leq$ t the rate of transmission was assumed to deteriorate exponentially at a rate of k. The time (t) until the transmission rate (β (t)) is at 50% percent of its initial level (β /2) is $t_{1/2}$ = ln(2)/k.
So, while $\beta \geq \beta(t) \geq \beta/2$ □ t = ln(2)/k

To solve the equations, author assumes that the outbreak started with one person. That gives us the value of I_o and C_o that is equal to 1. This means that the initial value for S should be N-1 (since there is exactly one infected person in the beginning). The remaining solution components should have initial conditions of zero. The rate of change of infected cases and rate of increase of deaths can be calculated by:

Rate of change of infected cases = Exposed Class / Average of incubation

□ *dC / dt* = σ E

*Rate of increase in deaths= (fatality rate*Number of Infected cases)/average duration of infectiousness*

- $dD/dt = f\gamma I$

Author solved the ODEs in R software environment but in this review we will use SciLab software to test and verify the results of the author. The author has used function *ode* from the package *deSolve* in R to solve the problem. *Ode* function is simply a wrapper around the various ode solvers. It uses a default integrator called *lsoda* to integrate the differential equations. *lsoda* provides an interface to the FORTRAN ODE solver of the same name, written by Linda R. Petzold and Alan C. Hindmarsh. The *ode* function in SciLab solves Ordinary Different Equations defined by:

$$\frac{dy}{dt} = f(t, y)$$
$$y(t_0) = y_0$$

It is an interface to various solvers, in particular to ODEPACK *(collection of FORTRAN solvers for the initial value problem for ordinary differential equation systems).*

In the next paragraph author gives some information about the data obtained/assumed for solving the differential equations (given in Table 2). The data for Guinea, Sierra Leone and Liberia were obtained from the World Health Organization (WHO) and were based on cumulative numbers of reported total cases and total deaths caused in those reported cases. The value of N i.e. the total population size *(N = S+ E +I + R)* was assumed to be 10^6 individuals. Since, the exact number of population doesn't really matter when as long as the components of cases is small than the total population.
In particular,

Value	**Formula**	**Equation**
Basic Reproduction Number	Rate of Transmission * average duration of infectiousness	$R_0 = \beta / \gamma.$
Effective Reproduction Number	Rate of Transmission *Number of Susceptible *average duration of infectiousness	$R_e = \beta(t)S/(\gamma N) \approx \beta(t)/\gamma$

The estimates given in Table 1 were obtained by fitting model to the data, assuming the cumulative numbers of cases and deaths are Poisson distributed. Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time. 95% of confidence intervals (CI) given in the Table 1 *(the values shown in brackets)* were calculated using *mle* (Maximum Likelihood Estimation) function of R. *Mle* function basically sets out to answer the question: what model parameters are most likely to characterise a given set of data.

Table 1. Parameter estimates for the 2014 EBOV outbreak. The basic reproduction number is given by $R_0 = \beta/\gamma$ where $1/\gamma = 5.61$ days is the infectious duration from the study by Chowell et al.[2] The primary case is defined as the index patient that caused the subsequent outbreak. The date of appearance of the primary case in Guinea was set at 2 December 2013.[1] 95% confidence intervals (CI) are shown in brackets. *A Likelihood ratio test showed that treating k as a free parameter does not improve the fit.

Parameter	Guinea	Sierra Leone	Liberia
Basic reproduction number, R_0	1.51 (1.50-1.52)	2.53 (2.41-2.67)	1.59 (1.57-1.60)
Transmission rate, β (per day)	0.27 (0.27-0.27)	0.45 (0.43-0.48)	0.28 (0.28-0.29)
Case fatality rate, f	0.74 (0.72-0.75)	0.48 (0.47-0.50)	0.71 (0.69-0.74)
Rate at which control measures reduce transmission, k (per day)	0.0023 (0.0023-0.0024)	0.0097 (0.0085-0.0110)	0*
Date of appearance of primary case, T	–	23 Apr 2014 (19-25 Apr 2014)	14 April 2014 (11-16 Apr 2014)

3. Results

As discussed in previous section, we need the following Equations to calculate our result:

$$\frac{dS}{dt} = -\beta(t)SI/N, \quad (1)$$

$$\frac{dE}{dt} = \beta(t)SI/N - \sigma E, \quad (2)$$

$$\frac{dI}{dt} = \sigma E - \gamma I, \quad (3)$$

$$\frac{dR}{dt} = (1-f)\gamma I. \quad (4)$$

$$dD/dt = f\ \gamma\ I$$

$$dC/dt = \sigma\ E$$

$$\beta(t) = \beta e^{-k(t-\tau)}$$

Various parameters that are included in these equations are:

$\beta(t)$, N, σ, γ, f, β, k, τ.

The author has made some assumptions to solve all of these differential equations. The values of β, f and k were obtained by using mle(maximum likelihood estimation) function in R environment.

We will be verifying all the results in SciLab for each of the effected location (Guinea, Sierra

Leone and Liberia). The initial values for the differential equations are as follows:

Parameter	Guinea	Sierra Leone	Liberia
β (t) (Transmission Rate with interventions)	β (t) = $\beta e^{-k(t-\tau)}$	β (t) = $\beta e^{-k(t-\tau)}$	β (t) = $\beta e^{-k(t-\tau)}$
N (Population)	1,000,000	1,000,000	1,000,000
σ	0.189 (1/5.3)	0.189 (1/5.3)	0.189 (1/5.3)
γ	0.178 (1/5.61)	0.178 (1/5.61)	0.178 (1/5.61)
f (Case Fatality Rate)	0.74 (74%)	0.48 (48%)	0.71 (71%)
β (Transmission Rate)	0.27	0.45	0.28
K (Rate at which control measures reduced transmission)	0.0023	0.0097	0
S (Susceptible)	999,999 (10^6 -1)	999,999 (10^6 -1)	999,999 (10^6 -1)
E (Exposed)	0	0	0
I (Infected)	1	1	1
R (Recovered)	0	0	0
C (Cases)	1	1	1
D (Deaths)	0	0	0

Function used for Guinea, Sierra Leone and Liberia:

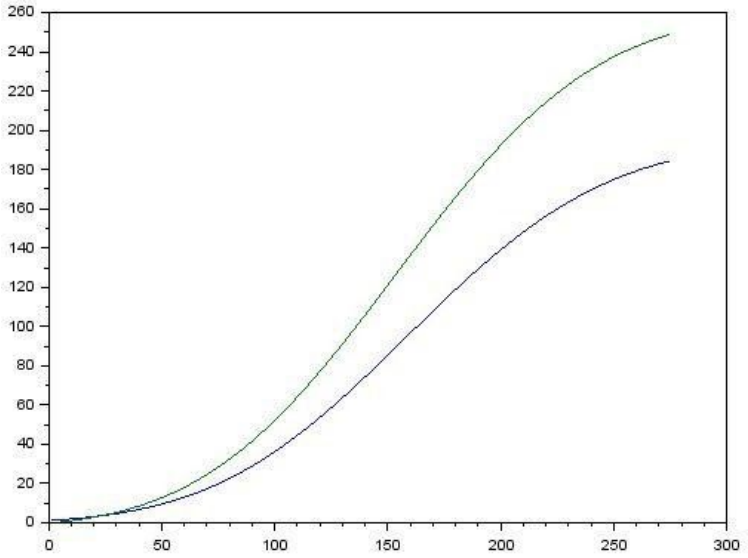
Note that the values of beta, k and fatality are subject to change as per the table given above

```
-->function f = ebola(t,y); N = 1000000; sigma = 1/5.61;
-->gamma = 1/5.3; fatality = 0.74; tau = 0.0; k = 0.0023;  beta = 0.27;
-->betaT = beta*exp(-k*(t-tau));
-->f(1) = - beta*exp(-k*(t-tau)) * y(1) * y(3) / N;
-->f(2) = beta*exp(-k*(t-tau)) * y(1) * y(3) /N - sigma * y(2);
-->f(3) = sigma * y(2) - gamma * y(3);
-->f(4) = (1 -fatality) * gamma * y(3);
-->f(5) = fatality * gamma * y(3); f(6) = sigma * y(2);
-->endfunction

-->y0 = [999999; 0; 1; 0; 1; 0];

-->y0 = [999999; 0; 1; 0; 1; 0];

-->t = [1:10:90];

-->t0 = 0;

-->y = ode(y0, t0, t, ebola);
```

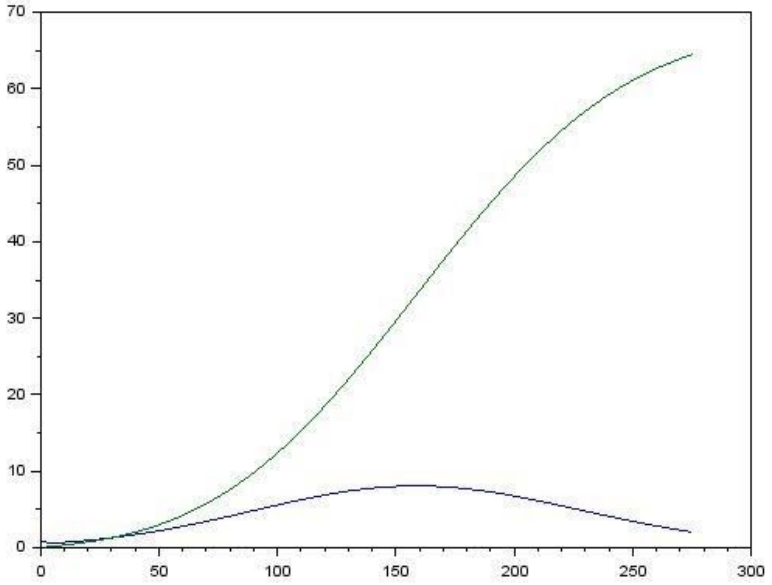
a) Guinea

Total time = 275 days.

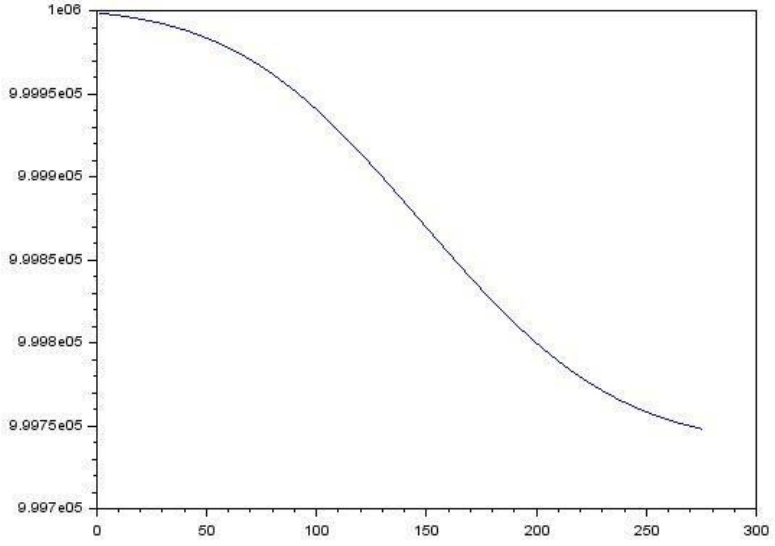
Graph showing number of infected cases (in green) and number of deaths (in blue) in Guinea.



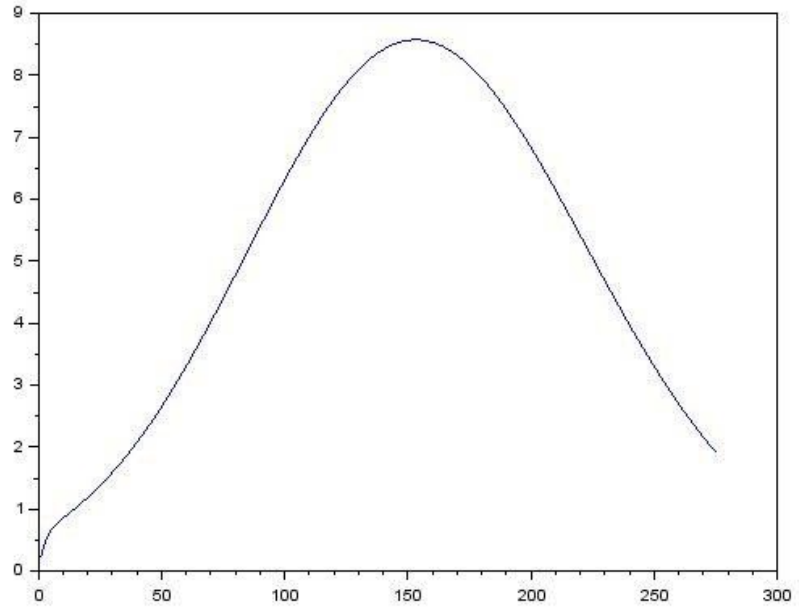
Now we will plot the graph for infected (green) and recovered (blue) cases in Guinea:



Graph for susceptible individuals

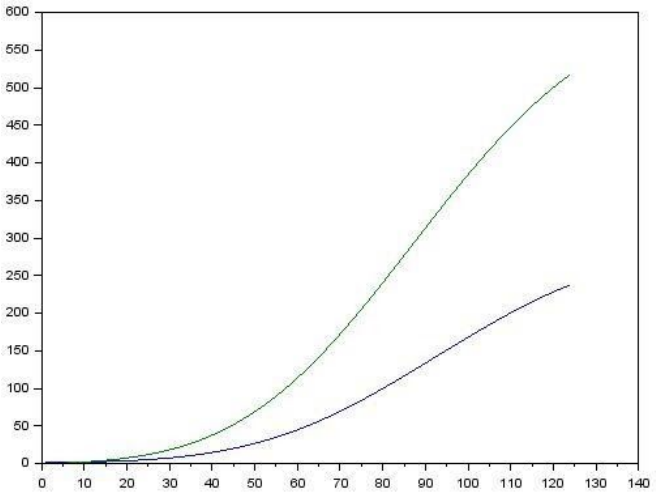


Graph for Exposed individuals

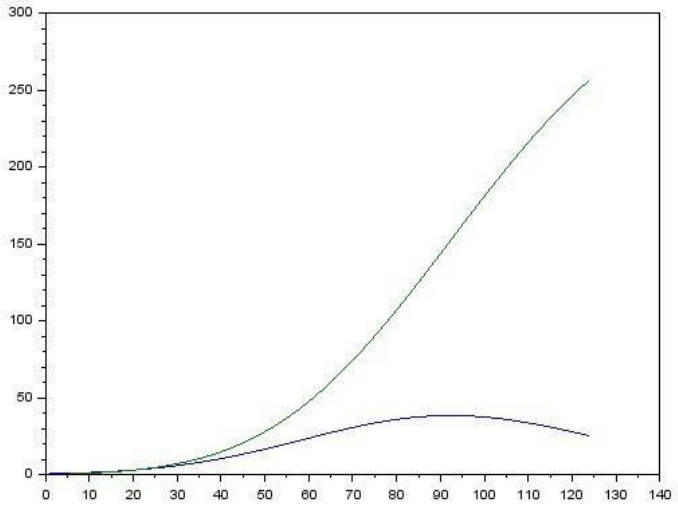


b) Sierra Leone
Total time = 124 days.

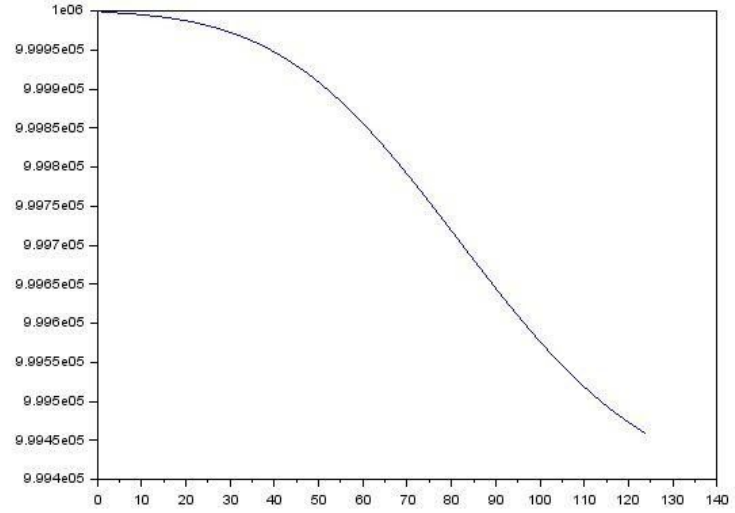
Graph showing number of infected cases (in green) and number of deaths (in blue) in Sierra Leone.



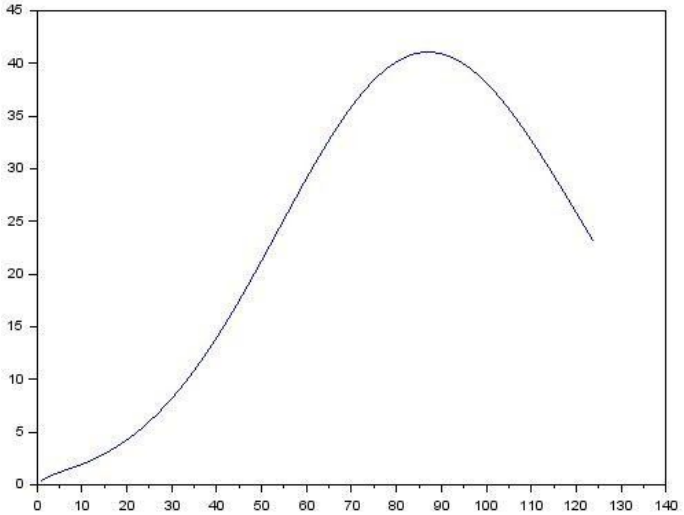
Now we will plot the graph for infected (green) and recovered (blue) cases in Sierra Leone:



Graph for susceptible individuals

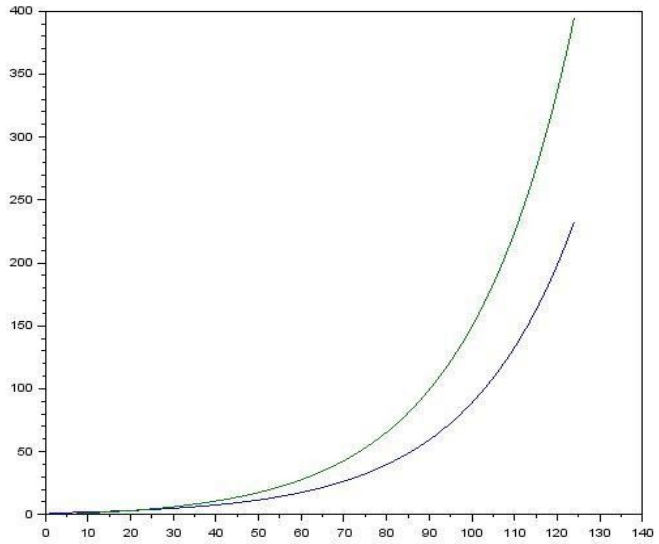


Graph for Exposed individuals

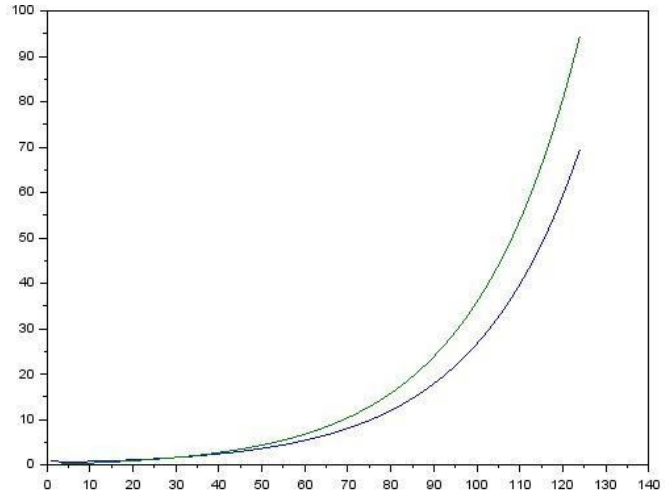


c) Liberia
Total time = 124 days.

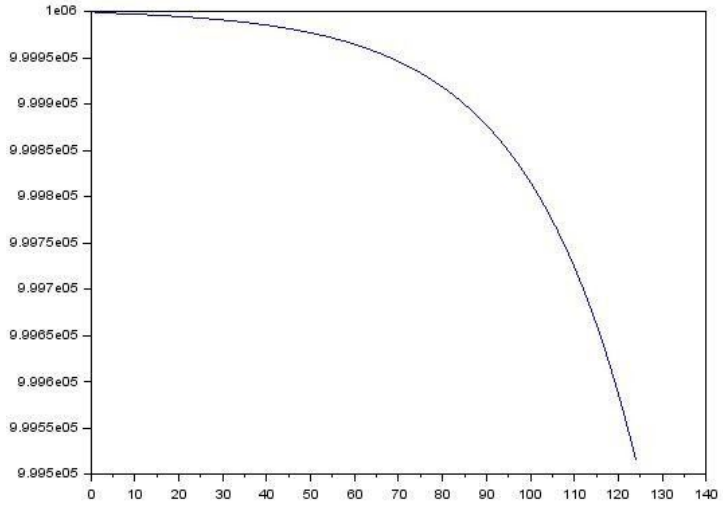
Graph showing number of infected cases (in green) and number of deaths (in blue) in Liberia.



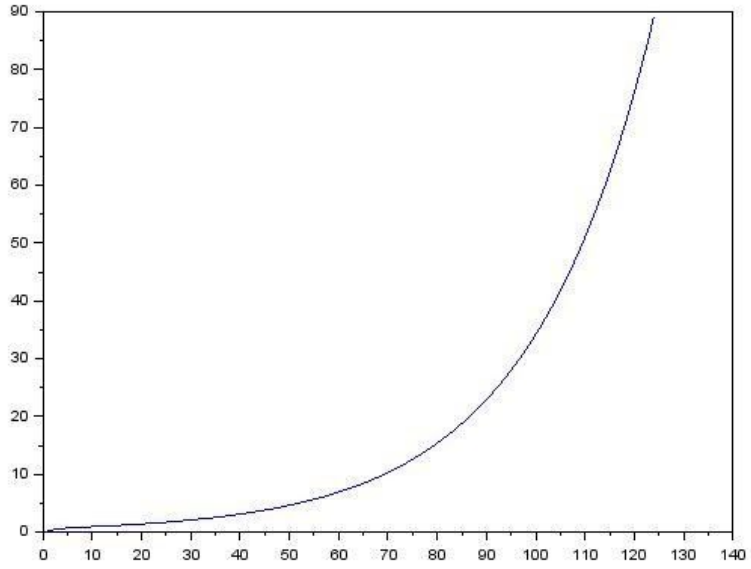
Now we will plot the graph for infected (green) and recovered (blue) cases in Liberia:



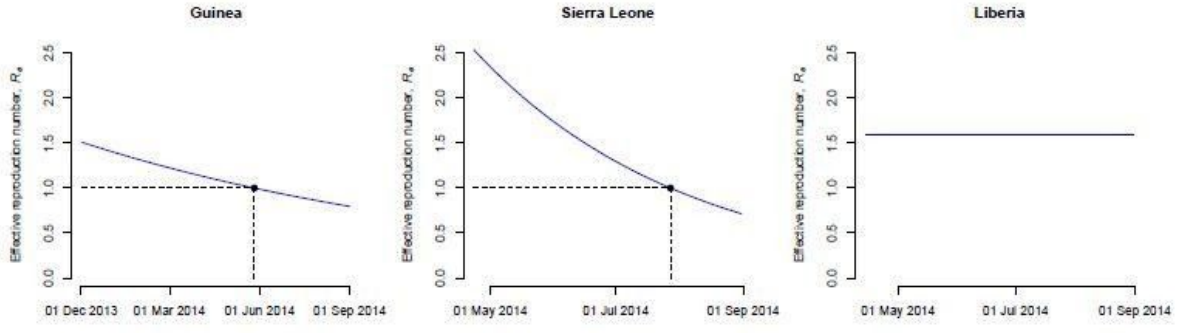
Graph for susceptible individuals



Graph for Exposed individuals



By these graphs we can see that the gap of number of cases and deaths widens in Guinea and Sierra Leone but if you see the graph for Liberia, it seems to be constant. Hence, the Effective reproduction number in Liberia remained constant until 1st September 2014.



4. Discussion

The study provides us the basic and efficient estimate of reproduction umbers of Ebola outbreak of 2014 by using mathematical modelling. The maximum estimates of R_0 were 1.51 for Guinea, 2.53 for Sierra Leone, and 1.59 for Liberia. Recall that the estimates for EBOV outbreak in Congo (1995) and Sudan virus (SUDV) in Uganda (2000) were 1.3 to 2.7. Are current values are quite similar to the ones in Congo and Sudan. These previous estimates were a source of information for the mathematical model the author used to provide a good description of the current epidemic. The basic reproduction number in Sierra Leone seems to be significantly higher than in Guinea and Liberia. That could have been because of the difference in population of the three cities. The population of Guinea (11.75 million) is much higher than the population of Sierra Leone (6.092 million). So the reason could have been because of the human mobility.

Sierra Leone and the most recent ancestor of genome analysis of EBOV outbreak share the same estimated date of occurrence. What we learnt mostly is that the results need caution to be analysed with optimum results. The major limitation of this model is that we assumed the transmission rate decays exponentially (β (t) = β e-k(t- τ)) once the control measures come into action after the first infectious case. Hence we cannot imply this model for the fluctuations faced in the number of cases in Guinea. These fluctuations could be result of the control interventions adopted in different parts of Guinea. With more information the modelling aspect could have given a better analysis as to what data fits better in which part of the country and the estimates of R_o and R_e could have been more accurate.

Although the current effectiveness remains unknown, the current facts clearly state that improvement of control measures is required (especially in Liberia). With more information not only better model could have chalked out but we would accumulate better estimate of epidemic projections and uncertainty around them which would act as a helping hand in predicting the evolution of this EBOV epidemic. Every possible measure for avoiding human contact was encouraged along with WHO prescribing drugs which were unregistered in order to cater the unfortunates. The study tells that the control interventions might have been successful in Sierra Leone and Guinea by the end of May and July 2014, respectively. In Liberia, a lot of improvement in cautions was yet to be implemented since the effective reproduction number remained constant at 1.59.
Here are the last recorded values of deaths and cases in these three cities.

City	Total Suspected cases	Laboratory-Confirmed Cases	Total Deaths
Guinea	3524	3096	2337
Sierra Leone	12170	8559	3842
Liberia	9862	3151	4408

Source: http://wwwnc.cdc.gov/travel/notices/warning/ebola-liberia

This shows that most of the deaths were recorded in Liberia. Many of the cases were not even reported in the Laboratory and people died without any medication.