MATHEMATICAL MODELS TO STUDY THE OUTBREAKS OF EBOLA

Jaime Astacio
University of Puerto Rico, Humacao

DelMar Briere
Blackfeet Community College

Milton Guillén
University of California, Santa Cruz

Josué Martínez
University of Texas, Austin

Francisco Rodríguez
California State University, Bakersfield

Noé Valenzuela-Campos
University of California, Davis

BU-1365-M

ABSTRACT

Using S-I-R and S-E-I-R models, it was possible to simulate two Ebola outbreaks: the 1976 outbreak in Yambuku, Zaire and the 1995 outbreak in Kikwit, Zaire. The dynamics of these models are determined by the per-capita death rate of infected individuals and the per-capita effective contact rate of an individual contracting the disease. The basic reproductive number, $\mathcal{R}$, determines the infectiousness of the disease. For Ebola, $1.72 \leq \mathcal{R}_0 \leq 8.60$, and this implies that Ebola is not as infectious as previously postulated. The results of these outbreak simulations will equip scientists in future outbreaks with information that may enable them to minimize potential deaths.

Introduction

The origin of the Ebola virus is somewhat obscure. There have been only three major known outbreaks of the Ebola virus, and all have happened in West Central African countries. The latest major outbreak occurred in Kikwit, Zaire in 1995. The outbreak took the lives of 79 people.

Ebola is a unique member of the ribonucleic acid virus family that has no known natural reservoir. The incubation period of Ebola is 2–21 days, and the infectious period is 4–10 days. The onset of Ebola is characterized by severe headaches, malaise, fever, vomiting, bloody diarrhea, and rash. Severe bleeding and shock are usually followed by death. Diagnosis of Ebola can be difficult, because Ebola is frequently misdiagnosed as typhoid and malaria. Currently there is no treatment of Ebola [3]. The mortality rate of Ebola is anywhere from 50–90%. Ebola is transmitted through primary contact with health workers who are in direct contact with body fluids from the infected. Ebola can also be transmitted through secondary contact by family members caring for the infected. Finally, Ebola can be transmitted where infection control mechanisms are not in practice. These control mechanisms can be as simple as wearing gloves, or as complicated as level-four disease control. Airborne spread has not been proven as a means of transmission.

Our objective is to better understand the mathematical dynamics of a population infected by Ebola when an outbreak occurs. To model this outbreak, we are using systems of differential equations. Several distinct models will be used to study the known data; each model differs in the way the parameters are acquired. From these different models we will choose the model that best fits the data.

Inaccuracies in the model are to be expected since the parameters dictating the behavior of the model are obtained from only a few data points. There have been so few major outbreaks that the amount of data available is limited. The model's precision is dependent on this limitation.

1. Ebola outbreak of 1995 in Kikwit, Zaire.

The object of this part of the project is to model Zaire's 1995 Ebola epidemic, using the Susceptible-Infectious-Recovery (SIR) model (Brauer and Castillo-Chavez, 1994). The dynamics of this system happen in two stages: susceptible to infected, and infected to dead. This is a closed system where those that are susceptible could become infected at some point in time. This

model assumes that the initial population is equal to the population that will eventually be infected. The parameters are μ, the per-capita death rate; and β, the probability that a susceptible host will become infected. The parameter β can vary from a constant function to an exponential function of decay with respect to the number of infected at time t. Once the parameters are found, the system modeling the data is solved using Mathematica 2.22. Mathematica uses the Runge-Kutta method to solve these differential equations.

The population studied is divided into three classes: $S(t)$, the number of susceptible individuals; $I(t)$, the number of individuals infected; and $R(t)$, the number of dead individuals at time t. We will assume that the population studied will be a constant population during the outbreak, meaning there are no deaths due to outside factors and the number of births that occurred are so small that we can essentially ignore them. (This is a valid assumption since the lengths of the Ebola epidemics are not longer than three to four months.) We will denote our total population at time t by N, so at any time t, $N = S(t) + I(t) + R(t)$.

Following a model proposed by Kermack and MacKendrik (1927) to explain the frequent rapid rise and fall of cases observed frequently in epidemics such as the Great Plague in London (1665-1666), the cholera epidemic in London (1865) and the plague in Bombay (1906), we are able to propose a model that approximates the outbreak reasonably well:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI/N, \\ \frac{dI}{dt} &= \beta SI/N - \mu I, \\ \frac{dR}{dt} &= \mu I. \end{aligned}$$

This model takes into consideration the number of people infected due to direct contact with an infected individual at time t: $\beta SI/N$, where $\beta = pc$; p is the probability of successfully getting infected when coming into contact with an infected individual, and c is the per-capita contact rate. The death rate is denoted by μI, where μ is the per-capita death rate. Even though recoveries do occur, we will not return these individuals to the susceptible class since there has never been a person who has recovered from Ebola and contracted the disease again in the same epidemic.

The data which we are studying is of the number of people that died each day during the outbreak in Kikwit, Zaire in 1995 [4]. See Figures 1 and 2. The second data set is the total number of dead individuals at time t, which can also be interpreted as the integral of the daily death data. From this relation we will be able to estimate the parameter β by solving the second differential equation for small values of t and relating it to the number of dead at time t. Since $dI/dt = \beta SI/N - \mu I$, for small t, $dI/dt \approx \beta I - \mu I$; solving this equation, $I(t) = I(0)\ Exp[(\beta - \mu)t]$, where $I(0) = 1$. Under these conditions, we can assume that $I(t) \propto R(t + 1/\mu)$, because $1/\mu$ is the average *time* for an infected individual to die. So $Exp[(\beta - \mu)t] = kR(t + 1/\mu)$, where $k = 1/0.77$ and 77% of infected people eventually die.

We have data that represents the total number of dead people at time t, cumulative of $R(t)$, so we fit the data with the curve. We take the natural log of the data so that our fit will be a linear fit:

$$(\beta - \mu)t = Ln[1/0.77] + Ln[R'(t + 1/\mu)].$$

The slope of the line which best fits the data is 0.14. By substitution,

$$\begin{aligned} (\beta - \mu)t &= Ln[R(t + 1/\mu)], \\ 0.14t &= Ln[R(t + 1/\mu)], \\ (\beta - \mu)t &= 0.14t, \\ \text{so } \beta &= 0.14 + \mu. \end{aligned}$$

Since the slope of this graph is so sensitive to the number of data points used in the fit, an average of these slopes is used to solve for β. This average slope was taken for fits for 10–20 data points. The average slope is 0.113. With this information we are able to calculate a range for the basic reproductive number $\mathcal{R}_0$, where $\mathcal{R}_0 = \beta/\mu$. For our range of μ between 1/6 and 1/31, $\mathcal{R}_0$ ranges from 1.57 to 5.03. We are now ready to look at the solution of the system of differential equations:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI/N, \\ \frac{dI}{dt} &= \beta SI/N - \mu I, \\ \frac{dR}{dt} &= \mu I. \end{aligned}$$

Since we only have numerical solutions, we can only view the graph of the solutions of each of these equations [see Figure 3]. The solutions plotted there contain the number of susceptible at time t, $S(t)$, the number of dead at time t, $R(t)$, and the number of infected at time t, $I(t)$.

We want to use the relation between $I(t)$ and $R'(t)$ to plot $I(t)$ to fit $R'(t)$ the best. Recall that $R'(t) = \mu I$, so the data being considered is fitted by μI plus a shift of $1/\mu$ to account for the average time from infection to death. We first consider how μ varies: $1/31 < \mu < 1/6$. For the graphs of Figure 4, we have taken $\mu = 1/22$, $1/5$, and $1/30$, and $N(0) = 900$. The respective graphs show exactly what one expects, since by increasing the period that an infected individual lives, $1/\mu$, there is an increase in the number of dead which the model predicts.

The next variable that we have to take into account is $N(0)$. The range for this variable is attained by taking an educated guess as to who the true susceptible individuals are in the population. These individuals are the family members and medical staff that care for the sick. This is a reasonable assumption since the only individuals who are at risk are those that have personal contact with the infected individuals. The lowest possible value of $N(0)$ would be about 250 individuals since only 244 died in this outbreak. A possible top limit for the greatest value of $N(0)$ could be around 900–1000, taking into consideration only the population size of families, health-care workers, and others involved in close and personal contact with infected individuals [refer to Figure 5].

The model, for larger values of $N(0)$, overestimates the number of expected individuals that will die. This observation may give the impression that the model badly represents the data, but in reality this overestimation could be of use to health-care workers who plan for how bad an outbreak may become by knowing statistics about the first 10–20 days of the outbreak.

2. Ebola outbreak of 1976 in Yambuku, Zaire.

We will now model the 1976 outbreak of Ebola in Yambuku, Zaire. The data that we will use for this outbreak was obtained from the Center for Disease Control (CDC) in Atlanta, GA. We will model the total infections that occurred during the outbreak using a modification to the S-I-R model. In this model, we will differentiate between the incubation period and the infectious period of the disease. As before, the number of susceptible indi-

viduals at time t will be denoted as $S(t)$. We will refer to the incubation period of the disease as the latent stage. The number of latent individuals at time t will be denoted by $E(t)$. Individuals that are infected with the disease and are suffering the symptoms of Ebola will be classified as infectious individuals. The number of infectious individuals at time t will be denoted by $I(t)$. Similarly, the number of dead individuals at time t will be denoted by $R(t)$.

The population studied will be a constant population during the outbreak; i.e., the total population at time t will be denoted by N where $N = S(t) + E(t) + I(t) + R(t)$. Our model is:

$$\begin{aligned}
\frac{dS}{dt} &= -\beta S(I + qE)/N \\
\frac{dE}{dt} &= \beta S(I + qE)/N - \delta E \\
\frac{dI}{dt} &= \delta E - \gamma I \\
\frac{dR}{dt} &= \gamma I.
\end{aligned}$$

This model takes into consideration the number of people infected due to direct contact with an infected individual and the number of people infected due to direct contact with a latent individual: $\beta S(I+qE)/N$. In this model, $\beta = pc$ where p is the probability of successfully getting infected when coming into contact with an infected individual, and c is the per-capita contact rate. The parameter q $(0 \leq q \leq 1)$ is a weight factor added to the model since it is known that a susceptible individual has a higher chance of getting infected from an infectious individual than from a latent individual [3]).

The individuals in the latent stage eventually show the symptoms of the disease, and pass on to the infectious stage. This is denoted by δE, where δ is the per-capita infectious rate. Then $1/\delta$ becomes the average time for a latent individual to become infectious. This will be denoted by γI, where γ is the per-capita death rate. Then, $1/\gamma$ becomes the average time it takes an individual to die once he/she has entered the infectious stage. As before, death and recovery are taken to be the same, since there has not been a case in which a person who survived Ebola contracts the disease again.

Figure 6 shows the number of people who became infected each day during the outbreak in Yambuku, Zaire in 1976. From this data we can now estimate β using a similar method to the one in the previous model. To do this, we first make three assumptions:

Assumption 1: In the beginning of the epidemic, $N(t) = S(t)$.

Assumption 2: Initially, there is a constant number of individuals infected. Those individuals infect other individuals who become latent. It takes $1/\delta$ days for the latent individuals to become infectious. Therefore, for the first $1/\delta$ days, the rate of change of the infectious individuals is 0 (i.e., $dI/dt = 0$.)

Assumption 3: In order for an individual to become infectious, they must pass through the latent stage. Thus, the data for the latent stage is the same as the data for the infectious stage, the only difference being that the latent stage data occurred $1/\delta$ days before. Since $1/\delta$ is the average time it takes for a latent individual to become infectious, and the latent stage ranges from 2 to 21 days, we choose $1/\delta = 12$. Similarly, since $1/\gamma$ is the average time it takes for an infectious individual to die, and of the infectious stage ranges 4 to 10 days, $1/\gamma = 7$. Thus, we then look at the following equation to estimate β:

$$\begin{aligned} & \frac{dE}{dt} = \beta S(I + qE)/N - \delta E \\ \Rightarrow \quad & \frac{dE}{dt} = \beta(I + qE) - \delta E \qquad (*) \\ & \text{by the first assumption;} \\ \Rightarrow \quad & \frac{dI}{dt} = \delta E - \gamma I = 0 \Rightarrow \delta E = \gamma I \Rightarrow I = \delta E/\gamma \\ & \text{by the second assumption.} \end{aligned}$$

If we substitute I into $*$, then $dE/dt = [\beta(\delta/\gamma - \delta)]E$.

The information for dE/dt is given by the daily infection data; the information for E is the cumulative of the daily infection data. Thus, we have a linear relationship, and we can estimate the slope by doing a linear fit. Using Mathematica and the data for the first 12 days, we obtain the fit shown in Figure 7 where equation of the line is $0.3893t$. Thus, we now have the slope of the best fit line, and $\beta = (0.3893 + \delta)/(\delta/\gamma + q) = 0.567114$, if we take

$q = 0.25$ and the values of δ and γ given above. The choice for q is arbitrary and is picked so that the model best fits the supplied data.

Another important number that needs to be computed is the basic reproductive number, $\mathcal{R}_0$. This number tells us how fast the disease will spread at the beginning of the epidemic. To calculate the value for $\mathcal{R}_0$, we need to find the Jacobian matrix of the system of equations. We then evaluate it at the disease-free state. Since the four- dimensional system can be reduced to a three-dimensional system, only the first three equations need to be considered. It is easy to show that the disease-free state is $(S, E, I, R) = (N, 0, 0, 0)$. Once the Jacobian is evaluated at this point, the determinant and the trace must both be greater than zero to insure that the disease-free state is an unstable fixed point. Once all of this is accomplished, we obtain a value for $\mathcal{R}_0$:

$$\mathcal{R}_0 = (\beta/\gamma)(1 + q\gamma/\delta) = 5.67.$$

All the parameter values are known, and thus we can solve our system of differential equations. The system of differential equations cannot be solved explicitly, so Mathematica is used to solve the system numerically. Figure 8 shows two numerical solutions to the system plotted with the initial infectious data. Note the label of the axes. It reads "Positive Part of dI/dt" because the given data only takes into account the recruitment rate of infectious people. The data does not reflect the infectious individuals that die; therefore, only the positive part of dI/dt is plotted. As can be seen in the plots, the numerical solutions are very good in the first part of the epidemic. After the peak is reached, the model is not very accurate.

3. Ideas for Future Research.

Both of the models that were presented in this research project used a constant effective contact rate, β. This is probably not the best model for β since the probability of contracting the Ebola virus varies as the disease becomes more widespread. People are more careful with whom they have contact, and thus the number of contacts decreases as time elapses or as the number of infected increases. Therefore, it makes sense to have β decrease. Another idea for enhancing the model is to consider quarantine. When infected people are isolated, the number of contacts that can transmit the disease decreases. This is something that could be taken into account in future work with these models.

More research needs to be conducted to estimate N, the total population. A good number for N is very important, since as it varies, the accuracy of the model also varies, as was seen in the numerical solutions of the models. Information that may prove helpful in estimating N includes the number of staff members in hospitals, family size, and other data that may help determine the total susceptible population at the beginning of an Ebola outbreak.

Research on q is also essential. It is intuitively clear that individuals showing symptoms of the Ebola disease are more infectious than latent individuals who show no symptoms. Therefore, a better value for q would make the model more accurate in predicting the dynamics of a future Ebola outbreak.

4. Conclusions.

The calculated $\mathcal{R}_0$ values ranged from 2.6 to 8.6 for the Yambuku, Zaire outbreak in 1976; meanwhile the range for $\mathcal{R}_0$ for the 1995 epidemic in Kikwit, Zaire was slightly lower: $1.57 \leq \mathcal{R}_0 \leq 5.03$. This makes sense since it shows that Ebola patients were infected more during the first epidemic due to misunderstanding and misdiagnosis of the Ebola virus. The values calculated for the basic reproductive number were lower than one may expect for Ebola, and were lower than for other diseases. For example, $\mathcal{R}_0$ ranged from 16–18 for the measles in England and Wales between 1950-68; and $\mathcal{R}_0$ for HIV in Hampara, Uganda for heterosexuals between 1985-7 was 10–11 (Anderson and May, 1991). This may be due to Ebola's method of transmission and the fear people have of contracting Ebola.

These models are very important because they can put an upper bound on the number of deaths, and thus can help health officials plan for the latter part of an outbreak by calculating the parameters from the data at the start of the epidemic. The number of deaths can also be minimized by altering the environment; i.e., lowering β (the effective contact rate). This can be accomplished by implementing quarantine.

Acknowledgments

The research in this manuscript has been partially supported by grants given by the National Science Foundation (NSF Grant DMS-9600027), the National Security Agency (NSA Grant MDA 904-96-1-0032) and Presidential Faculty Fellowship Award (NSF Grant DEB 925370) to Carlos Castillo-Chavez. Substantial financial and moral support was also provided by the Office of the Provost of Cornell University and by Cornell's College of Agricultural & Life Sciences (CALS) and the Biometrics Unit. The authors are solely responsible for the views and opinions expressed in this report. The research in this report does not necessarily reflect the views and/or opinions of the funding agencies and/or Cornell University.

Thanks to: Carlos Castillo-Chávez, Bonnie Delgado, Carlos Hernández-Suárez, Elion Mboussa, Tamara Parker, Julie M. Seda, Jorge X. Velasco-Hernández, and the Centers for Disease Control (CDC). We would like to specially thank Herbert A. Medina, "El Bañao Número Siete."

References

Anderson, R.M and May, R.M., "Infectious Diseases of Humans," Oxford University Press, 1991.

Brauer, F. and Castillo-Chávez, C., "Basic Epidemiology Models," in *Ecological Time Series*, eds. T.M. Powell and J.H. Steele, Chapman & Hall, New York, pp. 410-447, 1994.

Centers for Disease Control, Atlanta, GA, World Wide Web Page, http:www.cdc.gov/ncidod/diseases/virlfvr/ebolainf.htm.

Kermack, W.O. and McKendrick, A.G., "A contribution to the mathematical theory of epidemics, in *Proceedings of the Royal Society of London, Series A* **115**: 700-721, 1927.

World Health Organization, World-Wide Web Page, http:www.who.ch/programmes/cds/ebupdate.html.

List of Figure Captions

Data of Daily Deaths in Kikwit, Zaire in 1995 Outbreak.
Figure 1

Data of Total Deaths in Kikwit, Zaire in 1995 Outbreak.
Figure 2

Plots of Solutions to the System of Differential Equations with Parameter Values: $\mu = 1/22$, $\beta = 0.113$, $N = 900$.
Figure 3

Graphs of $\mu I(t)$ Compared to Actual Data for $N = 900$, $\beta = 0.113$, and $= 1/22, 1/5, 1/30$, respectively.
Figure 4

Figure 5 (no caption)

Data of Daily Deaths in Yambuku, Zaire in 1976 Outbreak.
Figure 6

Linear Fit to Data to Approximate β.
Figure 7

Figure 8 (no caption).

Data of Daily Deaths in Kikwit, Zaire in 1995 Outbreak

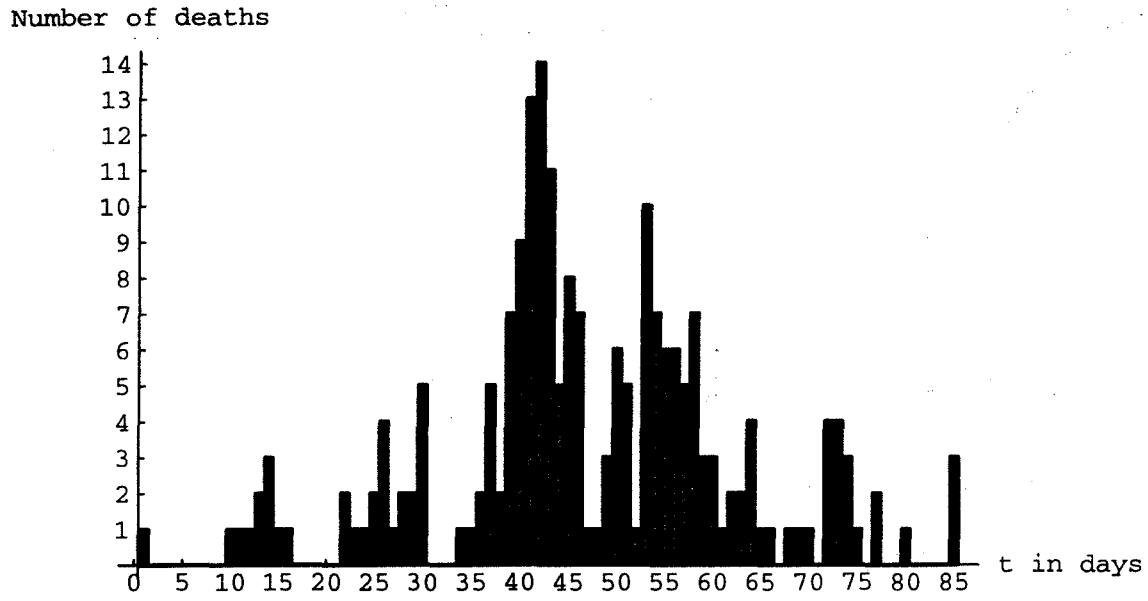


Figure 1

Data of Total Deaths in Kikwit, Zaire in 1995 Outbreak

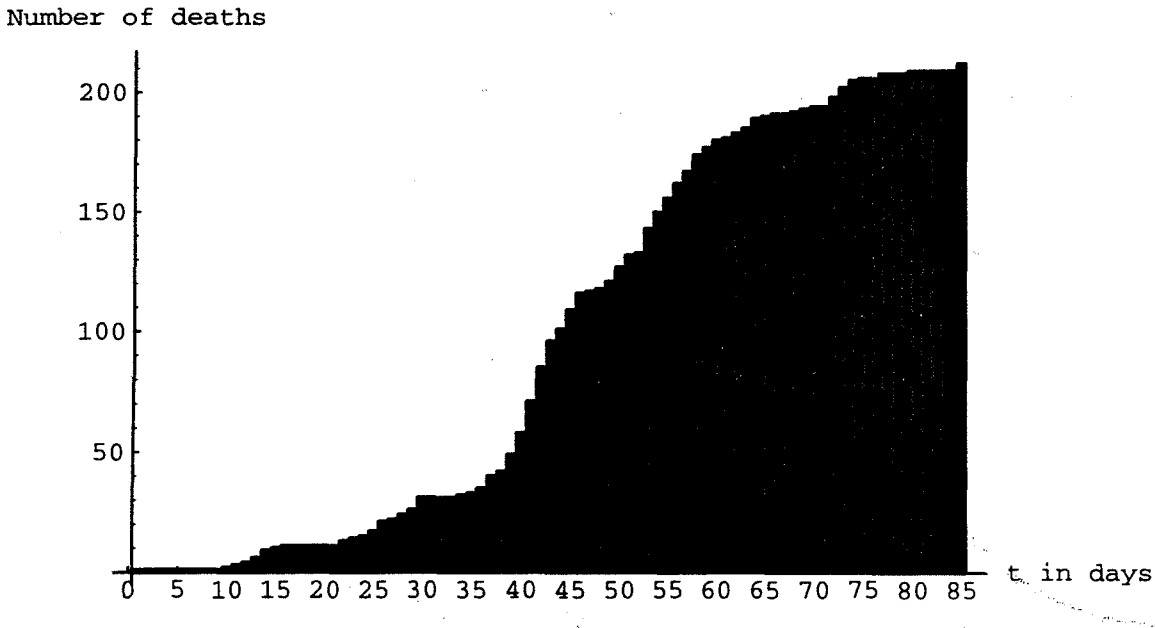
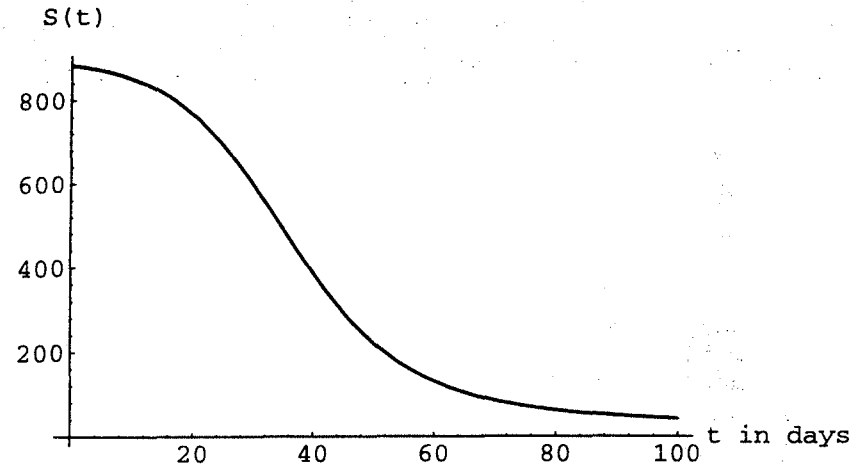
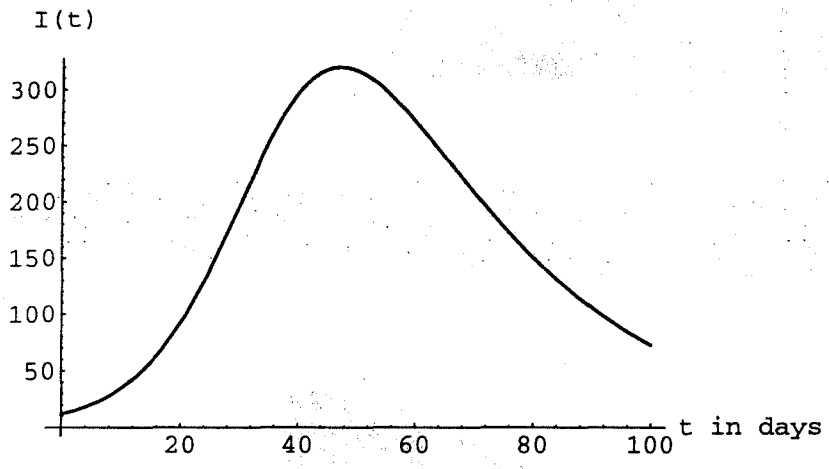


Figure 2

Plots of Solutions to the System of Differential Equations with Parameter Values: $\mu = 1/22$, $ß = 0.113$, $N = 900$





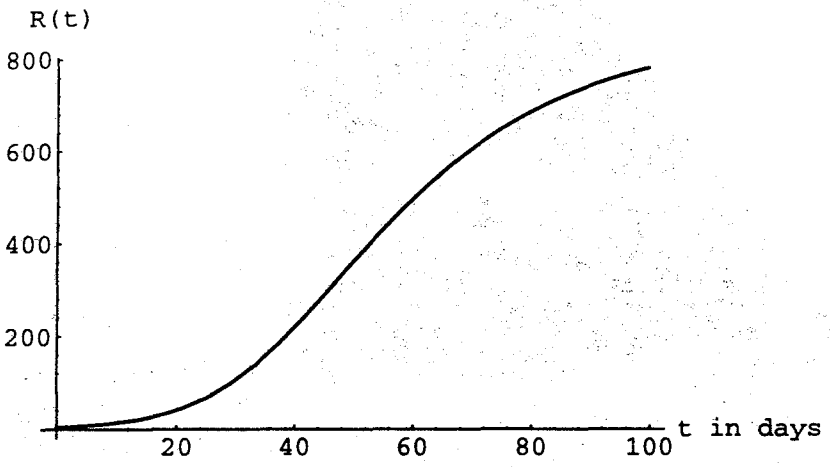


Figure 3

Graphs of μ I(t) Compared to Actual Data for

N = 900, ß = 0.113, **and** μ = 1/22, 1/5, 1/30 respectively

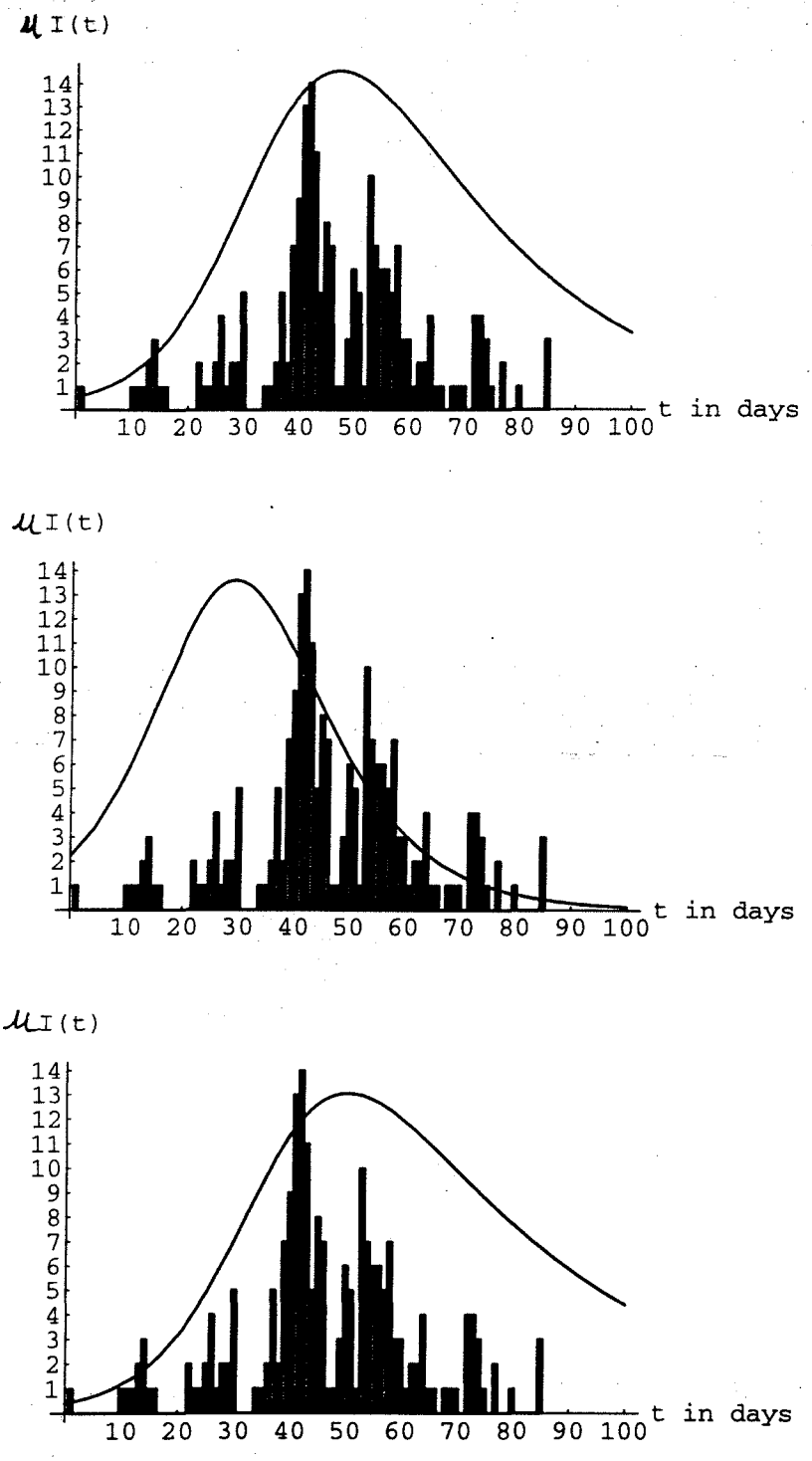


Figure 4

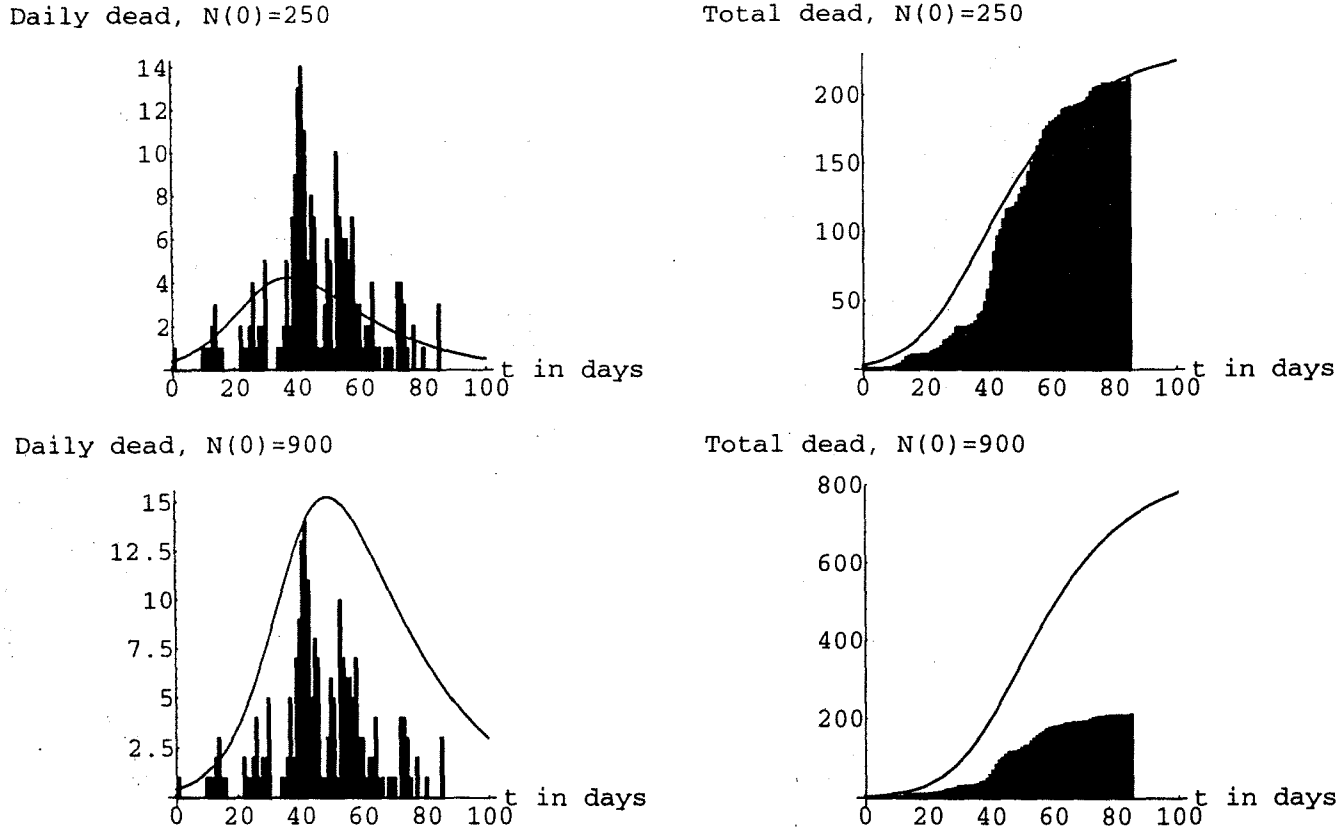

Daily dead, N(0)=250
14
12
10
8
6
4
2
0 20 40 60 80 100
t in days
Total dead, N(0)=250
200
150
100
50
0 20 40 60 80 100
t in days
Daily dead, N(0)=900
15
12.5
10
7.5
5
2.5
0 20 40 60 80 100
t in days
Total dead, N(0)=900
800
600
400
200
0 20 40 60 80 100
t in days

Figure 5

Data of Daily Deaths in Yambuku, Zaire in 1976 Outbreak

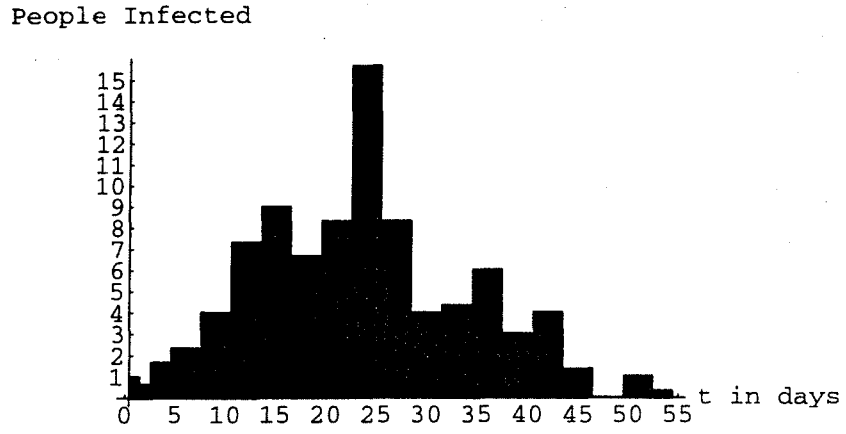


Figure 6

Linear Fit to Data to Approximate β

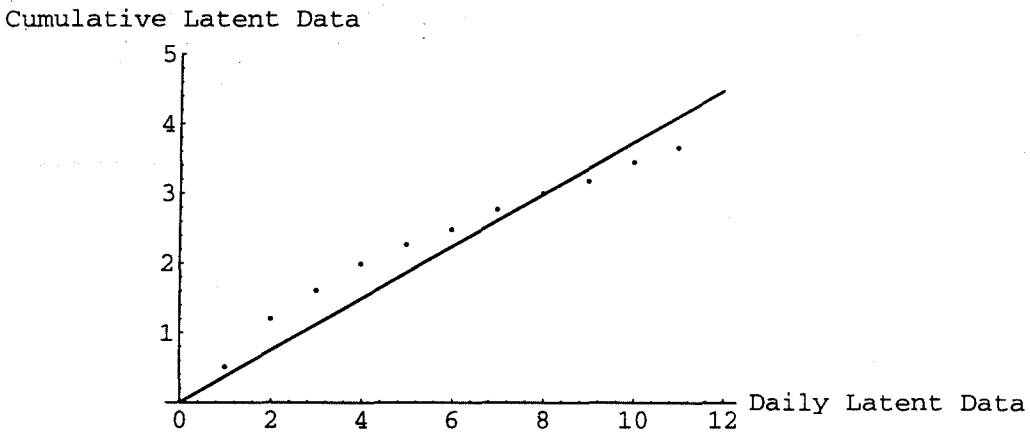
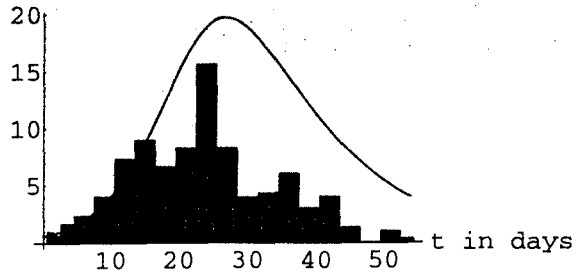


Figure 7

Positive Part of di/dt. N=600



Positive Part of di/dt. N=700

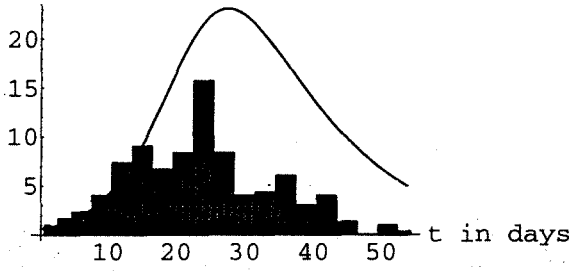


Figure 8