Modeling Ebola

KAMALALDIN KAMALALDIN [1], AMBER SALOME[1], AND PÉTER ÉRDI [1,2]

[1]Center for Complex Systems Studies, Kalamazoo College, Kalamazoo, MI 49006, USA
[2]Wigner Research Centre for Physics, Hungarian Academy of Sciences, Budapest, Hungary

ABSTRACT

This paper is the combination of a review on the problems of data collection and modeling efforts of the recent Ebola epidemics. After a brief review of data availability the modeling frameworks have been discussed. Both deterministic and stochastic models have been reviewed, and supplemented with the short discussion of some problems of parameter estimation. The methods have been illustrated by a realistic case study. A hint is given for the scope and limits of prediction and control.

Keywords: *deterministic, stochastic, complex systems, modeling, Ebola, epidemiology*

Kamalaldin is a Computer Science and Mathematics major in Kalamazoo College, where he is currently a third year student. While he has a background in chemistry and biology, his true calling is software engineering. When he is not contributing to the growth of Humanex Ventures, a young company in Kalamazoo, Michigan, Kamalaldin enjoys exploring the secrets of the production of the perfect combinations of 1's and 0's that come together to form perfect solutions to complex problems.

Amber Salome is currently a second year student Kalamazoo College. She is studying biology and computer science in hopes of pursuing an interdisciplinary career in these fields. Her interest in data analysis and the application of computer science and math to real life data began at the Kalamazoo Area Mathematics and Science Center with her involvement in a number of research opportunities, one of which was with Dr. Jaqueline Eng at Western Michigan University on the data analysis of joint disease patterns of ancient populations in Eurasia. Her involvement with NCWIT and other women in technology has inspired her to continue research in these fields.

Péter Érdi serves as the Henry R. Luce Professor of Complex Systems Studies at Kalamazoo College. He is also a research professor in his home town, in Budapest, at the Wigner Research Centre of Physics pf the Hungarian Academy of Sciences. In addition, he is the founding co-director of the Budapest Semester in Cognitive science, a study abroad program.

Péter is a Member of the Board of Governors of the International Neural Network Society, of the IEEE Computational Intelligence Society Curriculum Subcommittee, and serves among others as the Editor-in-Chief of Cognitive Systems Research. His books on mathematical modeling of chemical, biological and other complex systems have been published by Princeton University Press, MIT Press, and Springer Publishing House.

Introduction

The Ebola Virus has appeared in human populations since 1976, but never in such a large outbreak as the 2014 Ebola Epidemic[1]. This outbreak began in Guinea on December 26, 2013, however it was not identified until March of 2014[2]. The need for control was high as Ebola quickly spread throughout West Africa with high mortality and transmission rates[3]. Ebola is characterized as a hemorrhagic fever that is transmitted through bodily fluids. It did not originate in the human population, but is thought to come from contact with infected wildlife. There is normally an incubation period of 2-20 days before symptoms appear and the infected person becomes infectious to others. Major symptoms encompass fever, weakness, muscle aches, bleeding, diarrhea, and vomiting[4]. There is no current cure for Ebola, and treatment normally includes fluids to maintain electrolytes as well as monitoring of vitals and infection[4].

The WHO (World Health Organization) has held a primary role in maintaining current situation reports on the epidemic. These reports include the major statistics of the epidemic for each country, with Sierra Leone, Liberia, and Guinea being the most affected of the West African Countries. As of July 5, 2015, the WHO reported a total death toll of 11,246, with 27,573 cumulative cases (15,115 of which have been confirmed)[5]. The outbreak does appear to be declining, with fewer cases occurring and Liberia being declared Ebola free as of May 9, 2015[6]. Yet there have been many warnings against decreasing measures against Ebola, for fear of another outbreak. These cautions are not in vain either, as of June 29, 2015, another Ebola case occurred in Liberia, showing that the fight is far from over[5].

The continual effort to stop this epidemic has been met with a number of difficulties. Ebola is a complex disease to control[7], which has limited the effectiveness of outside help and the amount of data that could be obtained about the epidemic. Poor data collection and underreporting was, and is, a major problem, as outside countries rely on this data to understand how the epidemic progresses[8]. The high level of population mobility has served to increase the spread of Ebola, with many people traveling frequently and uncontrolled. Additionally, cultural differences, such as with burial practices[8,9] made Ebola outbreaks difficult to contain and

diagnose early on. These factors have held an integral role in understanding how to diffuse this epidemic[10].

The role of the public health system was key in controlling the epidemic. One of the critical tools used to combat Ebola is contact tracing, a method of isolating potential infections and monitoring all of those that may have been in contact with an infected person[9].This depth of control can be highly effective, but very difficult with large populations and a lack of resources[8]. However progress is being made in the fight against Ebola, with overall infrastructure changes, increased information acquisition, and better prepared medical facilities[8].

With a limited time frame and resources, understanding the best approach for limiting transmission and how the epidemic will continue to spread is critical. Epidemiological methods for modeling the spread of Ebola can give a somewhat accurate view of how the epidemic will behave in the future under particular circumstances, which can be critical in determining how to proceed with public health efforts[7]. Such an approach has been used before in other studies in an attempt to foresee how an epidemic would effect a community, so that control might be possible. For example, one study looked at the flu cases in a small college town, and was able to model the current spread of the H1N1 influenza epidemic through computational modeling[11]. Similarly, with accurate modeling of the Ebola epidemic, this could prove to be another potentially valuable tool for stopping the spread of this epidemic.

Numerous models have been created with the intent of modeling the current Ebola epidemic as accurately as possible. The multidisciplinary use of mathematics, computer science, and anthropology have come to set the foundation for many successful models. The authors of the article have included their own work, which has been built on examining the different types of models currently used in epidemiology, and experimenting with these models for both educational purposes and potential improvements to the current modeling methodology. Throughout this paper there will be examination of the basics of prominent epidemiological methods and the results of the authors' own work in this field.

Data as a function of time

In the context of a current epidemic, the availability of data is of the utmost importance to modeling efforts. Data, such as the number of cases and in which locations, are used to validate models and optimize parameters. These processes verify whether or not a model is indeed appropriate to simulate an epidemic, and help to find the values for the model parameters (see Parameter Estimation). Therefore, the availability of data is of paramount importance when attempting to model an epidemic, or any system in truth.

The data available on the EVD epidemic in West Africa is limited at best. Recording of EVD cases was done by the governments of each of the countries affected, and tracking of cases is minimalistic. The fact that each country was responsible for publishing its own data meant that each country could have had different criteria for what constitutes as a susceptible EVD case and how it is followed up until confirmation, creating variance in the reports. While one country's data would look promising, it could be true that it is simply underreporting EVD cases, and is sitting on a greater epidemic than those countries that over-reported. For example, earlier in the disease, Liberia reported an "Liberia has reported an unusually high proportion of deaths among patients with suspected (but not probable or confirmed) EVD cases (58% [440 of 754 patients]), as compared with Guinea (13% [4 of 30 patients]) and Sierra Leone (35% [74 of 213 patients])[12].

This observation could mean that cases were not being identified at the same level in all of the affected countries.

On November of 2014, the WHO started publishing consolidated forms of the affected countries under the name of situation reports[13]. The situation reports offer a great deal of overall information to parties interested in the current status of the outbreak. However, the information represented in the situation report could not be used by data scientists since it was not presented in computer-readable format. To allay this problem, data scientists started manually extracting the data from situation reports to create data files that can be read by computer simulators[14]. Later on, the WHO started publishing data directly from its database which allowed data scientists to access the raw data instead of having to recreate the data from the situation reports. Although the publication of the raw data was a great leap for collaboration in the scientific community involved in the Ebola epidemic, it revealed that there were inconsistencies between the situation reports and the patient databases their numbers were being pulled from[15].

The database released to the public included death and infection cases, both confirmed and probable, in the countries in which Ebola was being spread. In the case of countries with intense transmission (Liberia, Guinea, and Sierra Leone), the data was further separated by district. This detail in data reporting was helpful for data scientists interested in looking at spatial models for the spread of Ebola between district that could be socially or geographically different[16].

Questioning the consistency of the data, modelers were forced to make more assumptions than they wished for in an attempt to use the data for their models.

However, an important piece of data was not reported by WHO: the number of current infected individuals. While the number of cumulative cases can be important for statistical purposes, it is the instantaneous number of infected individuals that modelers are interested in. While the instantaneous number of cases can be fabricated using the cumulative numbers, the output would only be a rough estimation that does not necessarily reflect the reality of the disease. Which such a gaping hole in the data, our modeling and parameter estimation was limited to using only the cumulative number of cases, not the instantaneous.

As time went by, the reporting of WHO increased in sophistication: more statistics were published, and visual aid was heavily used in WHO's reporting. An interactive map that displays the information from situation reports was published on the WHO's website. The interactive map contains data in written form, charts, bars, and labeled maps. They are updated daily to reflect the current status of the epidemic. Although the reports are concentrated on countries with current high rates of transmission, they also offer insight into countries that neighbor or have strong trade ties with areas of active transmission[17].

Modeling frameworks: From Simplicity to Realistic Models

Structure of the Models, Deterministic Framework

In the simplest situation there are two populations, $I(t)$ denotes the number of already infected (either with biological objects capable of transmit infection or with revolutionary ideas to be transferred to others) individuals, and $S(t)$ is the number of susceptible individuals. It is assumed that who is infected is also infective.

What is now the simplest assumption? The encounter between an infective and a susceptible may imply the transformation of a susceptible to an infective one. Chemical kinetics here also serves as a metalanguage[34] to encode the process:

$$I + S \xrightarrow{r} 2I,$$

where *r* is the effectivity of the encounter, i.e. the infection rate: the larger this rate the more contagious the infection. By adopting (i) the mass action kinetic assumption and (ii) "perfect mixing" (i.e. the space is homogeneous), the kinetic equation is:

$$\dot{S} = -rSI \quad (1a)$$
$$\dot{I} = rSI \quad (1b)$$

Perfect mixing in social applications should be interpreted, as the lack of the existence of any (spatial or social) organization among the participants. If the total population is constant (no birth and death, no immigration from and to the community), i.e. $I(t)+S(t)=N$, then the temporal change of the infected population is described by the equation:

$$\dot{I} = rI(N - I) \quad (2)$$

This is the logistic differential equation. The equilibrium solution $I_{eq}=0$ is unstable. Instability here means that its slightest perturbation implies that $I(t)$ tends to N, the whole population will be infected.

The simple model just presented[33] consists of two subpopulations and it is called as the *SI* model. The classical strategy to defend the susceptible against infection is to remove the infected, so to apply *SIR* models. Kermack and McKendrick defined a deterministic model, by adding a first order removal to equation 1 and so with the structure $S \rightarrow I \rightarrow R$:

$$\dot{S} = -rSI \quad (3)$$
$$\dot{I} = rSI - \gamma I \quad (4)$$
$$\dot{R} = \gamma I, \quad (5)$$

where γ is the removal rate. (Of course, this is the simplest assumption, the removal is proportional with the actual number of the infective ones). The relative removal rate is defined as $\rho = \gamma/r$. A plausible initial condition is that that $S(0)>0$, $I(0)>0$ and $R(0)=0$, nobody is in quarantine before the epidemic breaks out.

The kinetic condition of the outbreak of an epidemic is that the infectious population increases, i.e. $\dot{I} > 0$, i.e. $S(t) > \rho$. The number of susceptible should exceed a **threshold** to have an epidemic outbreak. Since epidemics are threshold phenomenon, one implication is that there is no necessity for full vaccination of the population to avoid the outbreak.

Stochastic models

The stochastic approach to modeling an epidemic incorporates the randomness that pertains to an epidemic and population dynamics. Such randomness creates an uncertainty that is necessary to incorporate in models that are meant to emulate real life, in a way that the deterministic models

cannot match. Stochastic models utilize discrete reaction probabilities as opposed to the static rates of a deterministic model that creates a continuous trajectory[18]. For stochastic simulations with the SIR model, the Gillespie Algorithm is a simple, yet power method for stochastic systems. It is based on both randomness in which reactions take place, and at what time interval the reaction occurs. For this algorithm, the reactions, species types, and population size are necessary.

The algorithm is heavily dependent on reaction probabilities, which factor into the likelihood that a certain reaction will occur at this specific time interval. These reaction probabilities are largely based on population size and initial parameters. The population size is critical, due to the nature of stochastic systems and the spread of disease. Much like a bimolecular chemical reaction, the spread of disease depends on the interaction of two different species, such as susceptible and infected. Once the two species necessary for the reaction to occur come together, the reaction takes place: infection. Consequently, the population size of each species is critical for how many reactions will occur. More reactants means that more reactions will occur, and in the case of an epidemic, the faster the spread of the disease. A similar phenomenon applies with unimolecular reactions, where there is only one reactant. In this case, where the species may not be a reactant, the daughter product from its degradation may be. For this situation, population size matters because more reactants means more products, which may influence another reaction. The Gillespie Algorithm accounts for this, by factoring in the population size into the reaction probabilities. For a unimolecular reaction such as $A \xrightarrow{k} B$ are represented by $A(t)k\Delta t$, where A(t) is the species population size, k is the constant/parameter (more on determining this in the next section), and Δt is the time change, similarly, the probability of a bimolecular reaction $A + B \xrightarrow{k} C$ which involves the interaction of two species is $A(t)B(t)k\Delta t$, where B(t) is the population size of the second species[19].

In a working Gillespie Algorithm, a random time interval is generated, which marks the time between the next "reaction," or change in the number of each "species" in the system. Once a reaction is randomly chosen, the reaction changes the state of the member of the population into a different species, which effects the population size of each group. Consequently, the data structure containing the population sizes changes counts for each species and reaction probabilities are updated, and the simulation is repeated again. This is a bit of a simplification, as the calculations of the probabilities and time involved can be complex. Additionally, in the case of simulations that require a large amount of time, modifications can be used, which increase the efficiency of the model. However there may be tradeoffs involved, as an increase in efficiency can cause a decrease in accuracy. One such model is tau leaping, which simply performs a number of reactions and time intervals in either a set time interval or limit on the number of reactions, and then updates the reaction probabilities and population sizes afterwards, as opposed to after each reaction. This allows for a faster working model, which can be useful in demanding simulations[18].

For a general example of the Gillespie Algorithm in stochastic modeling, suppose that there is a system based on the SIR model with three species (susceptible, infected, removed) and two reactions; infection, which is $S+I \rightarrow 2I$ and removal, $I \rightarrow R$. The starting populations is 50 susceptible, one infected, and none removed. For the first interval, and random value is generated for the time and a reaction is chosen. Since there are so many susceptible and one infected, infection has a much higher probability of occurring than removal, and is likely to be chosen as the first reaction. If chosen, then the susceptible population drops to 49, and infected is at two. Now the probability of infection as the next reaction has slightly decreased, but is still very high,

and the probability of removal has slightly increased. As the simulation progresses and reactions occur, each specie's population changes, as do the reaction probabilities. It is likely that the majority will become infected and then removed as the probabilities change and the simulation continues. The simulation would cease to run when all of the infected have been removed, as they are necessary for either reaction to occur. However if this simulation were to be run again there would be different results, though after many simulations a likely trend would appear. This showcases how a stochastic model is random in nature. The reliance on probabilities, may hint at a general trend, however the exact path that the system plays out is never exact or completely predictable, as a deterministic model would be. The rate constants used in this type of model are a source of parameters that can be adjusted to make the model more reflective of an actual population. This part of modeling relies on the constant adjustment and search for parameters in order to create a better model, a goal that was explored by both the authors and many other researchers in the field of epidemiology[19,20].

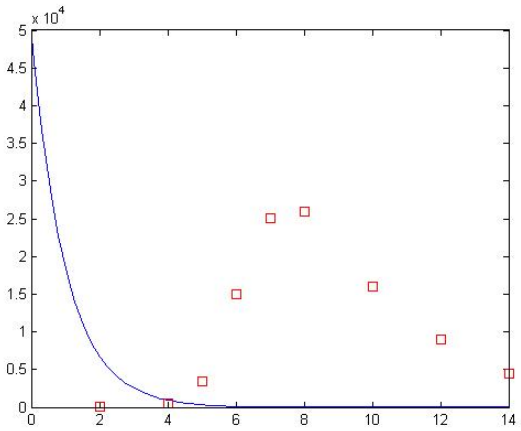
Parameter Estimation

Parameter estimation refers to the process by which modelers attempt to find the constants that would best output a result closest to the observed data. In the case of an epidemic, these constants would be ones like the basic reproductive rate of a disease or the rate of contact between the susceptible and infected populations. In the SIR model,

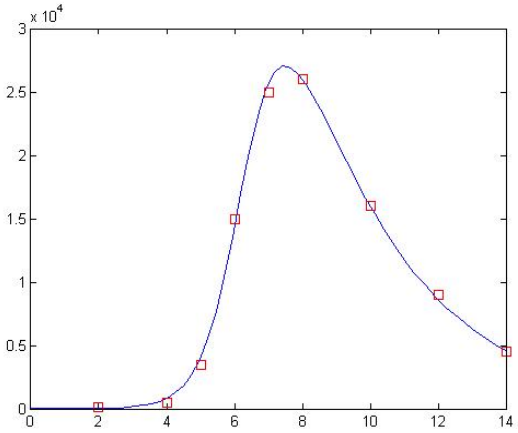
$$\dot{S} = -rSI$$
$$\dot{I} = rSI - \gamma I$$
$$\dot{R} = \gamma I,$$

r and γ are both constant parameters that influence the model's result. Given the same initial conditions, these parameters can be altered to output different results, allowing the model to behave differently. The process of parameter estimation seeks to find the value of those parameters such that the data output from the model matches the observed data (Figure 1)[22].

Finding the model parameters' value is important when contrasting the progression of current epidemics to previous ones. The parameters are also apt to change within the same epidemic: as the epidemic progressed, more data is bound to be published, and optimizing the parameters gives them a new value. These differences in new and old parameters within the same epidemic can also be used to measure the changing dynamic of the epidemic. Examining the difference between past and current model parameters can give an estimation of the degree of severity of the epidemic, and educate the scientific community on how the epidemic is progressing, and therefore inform the public opinion of the degree of necessity for an intervention.



(a) r = 0.000040293800490, γ = 0.329045768752125.



(b) r = 1.004882437649172, γ = 1.000088843687306.}

Figure 1: Changing the parameters in the *SIR* model creates a great shift in how the model operates. The calculated infected population (blue line) modeled against observed infected population (red squares). Initial conditions were 50000 susceptible, 1 infected, 0 removed. The equation was solved using ode45 in Matlab[22].

To set, verify, and update a model, one must determine the parameters of the model by fitting the it to the available data. The idea is that the most-correct parameter are the ones that enable the model to reproduce the closes data set to the observed data. There are various techniques that can be employed to optimize parameters. The simplest of these is the Least Squared Error (LSE) technique. LSE measures the accuracy of the model by measuring the difference between the data produced by the model and the observed data. After this difference is calculated, it is squared to account for negative error (positive and negative errors of the same value are given equal weight), and this value is called the squared error. As the model's output deviates from the observed data, the error grows exponentially. The larger the squared error, the less accurate the model is. Therefore, LSE relies on minimizing this difference in observed data and model data by changing the value of parameters in the model.

This is done by changing a parameter K in one direction of the number value (i.e either subtracting from it or adding to is), and then re-running the model to calculate the squared error again. Since the sample data does not change, any change in the squared error represents the effect of changing parameter K on the model. A decrease in the squared error after changing K represents an enhancement to the parameter and the model, since it represents a closer fit of the model to the sample data. Therefore, if the squared error is reduced, K keeps its new value. Alternatively, an increase in the squared error represents a diminishment in the model's accuracy. Therefore, if the squared error increases, K is set to its old value. The value of K is changed in the other direction of the number line. The aforementioned assertion is done again, and if the squared error increases again, then the value of K is reversed to its previous value, and parameter is considered to be optimized. The assessment is then done for other parameters in a similar fashion.

As parameter K 's value changes continuously, and the squared error is recalculated for each of these values, a table of values arises from which a graph can be drawn. This is called the error function, and to reach an optimized parameter, the parameter must have the value that

corresponds to the minima of that function. In a model where multiple parameters exist, such as the one SEIHFR model, the error function of all the parameters is not two-dimensional, but multi-dimensional. In our model, the parameter error function was 11-dimentional.

Case Studies

Deterministic Studies

Only a few models have been developed with Ebola in mind, one of which is the SEIHFR model developed by Legrand et al. In this model six compartments are taken into consideration:

$$\frac{dS}{dt} = -\frac{\beta_I SI + \beta_H SH + \beta_F SF}{N}$$

$$\frac{dE}{dt} = -\frac{\beta_I SI + \beta_H SH + \beta_F SF}{N} - \alpha E$$

$$\frac{dI}{dt} = \alpha E - (\gamma_h \theta_1 + \gamma_i (1-\theta_1)(1-\delta_1) + \gamma_d (1-\theta_1)\delta) I$$

$$\frac{dH}{dt} = \gamma_h \theta_1 I - \left(\gamma_{dh}\delta_2 + \gamma_{ih}(1-\delta_2)\right) H$$

$$\frac{dF}{dt} = \gamma_d (1-\theta_1)\delta_1 I + \gamma_{dh}\delta_2 H - \gamma_f F$$

$$\frac{dR}{dt} = \gamma_i (1-\theta_1)(1-\delta_1) I + \gamma_{ih}(1-\delta_2)(H) + \gamma_f F$$

S (susceptible) represents those in the population not yet infected, *E* (exposed) those in the population that are infected but not yet infectious (i.e incubation period), *I* (infected) those in the population that are infected and infectious, *H* (hospitalized) those in the population that are infected and in hospital care, *F* (funeralized) those in the population that are infected and have died but not yet buried, and *R* (removed) those in the population that have been either cured or buried. Furthermore, the model incorporates various parameters that will be the topic of another section (table 1).

Parameter Symbol	Parameter meaning
β_I	$Contact\ Rate, Community$
β_H	Contact rate, Hospital
β_F	Contact rate, Funeral
α	1/ Incubation period
γ_h	1/ Time until hospitalization
γ_{dh}	1/ Time from hospitalization to Death
γ_f	1/ Duration of traditional funeral
γ_i	1/ Duration of infection
γ_d	1/ Time from infection to Death
γ_{ih}	1/ Time from hospitalization to recovery
θ_1	Probability a case is hospitalized
δ_1	Case fatality rate, Unhospitalized
δ_2	Case fatality rate, Hospitalized

Table 1. Model parameters for the Ebola Model developed by Legrand et al.

The model has three additional compartments added on top of the SIR model. The Exposed compartment acts as a buffer between the susceptible and the infected population. This compartment is necessary as there has been evidence that infected individuals are no infectious when they are in the incubation period[23]. The hospitalized compartment is included to take into account the different rate of infection and death in the fraction of the population that is treated in hospitals. The funeralized compartment is a particularly special one. Ebola patients can transmit the disease through their bodily fluids, including sweat, even after their death. This means that even after a patient's death, Ebola can be transmitted if susceptible individuals come into contact with the patient's sweat. Unfortunately, this feature of Ebola has significantly affected its transmission, as cultural practices surrounding funerals in West Africa involve extensive physical contact between the dead body and people involved in the funeral. In fact, the genesis of the epidemic is thought to have been in a funeral of an Ebola patient[24].

The SEIHFR model seems to be an appropriate framework to describe realistic situations. Although the SEIRX model was more accurate at predicting the epidemic, it is more recent, and we felt more time for its validation was needed for it to be more reliable than the SEIHFR model. This comes especially after considering the lack of incorporation of a hospitalized compartment, which we believe to be especially important with the medical aid of global nations thought to be playing a major part in the recovery of certain nations but not others.

Stochastic Studies

Legrand's work in the modeling of Ebola gave a critical look at stochastic estimation and modeling, well before the 2014 epidemic[25]. His work was focused on 2 smaller epidemics nearly ten years prior. This model was a dynamic stochastic compartmental model based on the Gillespie algorithm. It utilized the SEIHFR mode of compartmentalization, and parameterization from literature and estimation. Through simulations of hypothetical situations, this model highlighted the importance of hospitalization and transmission control, as many recent models are discovering yet again[25].

Other sources have recognized the need for modeling of the Ebola epidemic, and have used this as a tool to understand the epidemic's progress, as well as the best methods to control the spread. A recent study, "Strategies for containing Ebola in West Africa,"[26] created a stochastic model that looked at multiple parameters involved with the transmission of Ebola, in an effort to evaluate the most effective methods for containment. For their model, subjects were divided into six groups similar to those of Legrand's model, and what was implemented in the authors' models. Their groups were, "susceptible (S), latently infected (E), infected and infectious (I), deceased (F), recovered with sterilizing immunity (R), and buried (D)."[25] In this case, the purpose of the increased complexity was to emphasize the separate effects of the community, funerals, and medical system. Further divisions were also made between these groups, with individual parameters in order to gain a better view of the system. The subsequent model and parameters that researchers made was fit to data from Liberia on a national scale, as well as a local scale in specified counties. It should be noted that in this study, the underreporting of this situation was another (potentially critical) variable to be accounted for in the modeling process. The overall lack of data from the epidemic so far limits the extant and complexity of current models as well.

In accordance with the study's goals, the researchers addressed WHO recommendations for limiting the spread of Ebola, namely changes to burial practices, contact tracing, health worker safety precautions, and case isolation of the infectious and infected. From the modeling efforts, it was clear that individual measures would not be enough; contact tracing and especially burial practices would be critical. An example of their calculated findings showed that, "the combination of case isolation, sanitary burials of hospital deaths, and reduction in nosocomial transmission, each applied with an efficacy of 95%, would reduce the number of daily cases in Liberia to a projected 24 (15 to 41) by December 1."[26] However, as noted in the paper, 95% compliance by the community is likely to be unattainable, at 60% efficacy, "could reduce the number of daily Ebola cases to 7 (2 to 13) by December 1 and to 0 (0 to 3) by March 15 in Liberia." The complexity of the model also allowed the researchers to look at multiple sectors of the population, though not environment, to determine the effect of actions particular to these groups, which allowed them to identify safe burial practices as an important protocol in Ebola containment. From the results of the model, it was made clear that not one, but practicing as many of these measures as possible would be critical for the control of Ebola[26].

The CDC has also had a role in the modeling effort with their Ebola Response modeling tool[27], primarily based out of Excel. This model is based on stochastic methods, with the use of probabilities and a Markov Chain model. It is divided up into five stages of the disease, "susceptible to disease, infected, incubating, infectious, and recovered" (SIIR model), with the

dead in burial ceremonies considered infectious and the infectious period for 6 days. Three categories were also added to this model, with patients in medical care, a home that could provide isolation, or a home with poor transmission control. This model also built in potential cases leaving or entering the country. Additionally, underreporting was considered a potential variable in the modeling effort, which was accounted for by a calculated correction factor of 2.5. The range created by this correction factor gives a broad range for the uncertainty attributed to the predictions.

As noted before, control of the Ebola epidemic is heavily reliant on patients being in a hospital setting or location where transmission can be controlled, so that transmission per patient is close to zero. Subsequently, this model focuses on the hypothetical situations and parameters involved with patients in or out of these controlled situations (the 3 main categories). With the progression of Ebola as it was back in September 2014, and without any control measures put into effect, the model predicted 8000 total Ebola cases between Liberia and Sierra Leone (16000 with the correction factor). Both estimates are much higher than the actual numbers reported by the WHO[28], with 6000 total infected as of the end of September, 2014. Further extrapolation provides an immense overshot, estimating 550,000 cases (1.4 million corrected) by the end of January in 2015. However, another scenario that tested the effect of placing patients in a setting that provided measures to treat and prevent the transmission of Ebola showed that if 70% of patients were treated this way, the epidemic would almost be over by the end of January, 2015. However there is a cost to delaying these measures, which would increase the length of the epidemic. These results again emphasize how critical the control of Ebola transmission is, without which the infected population would skyrocket[27].

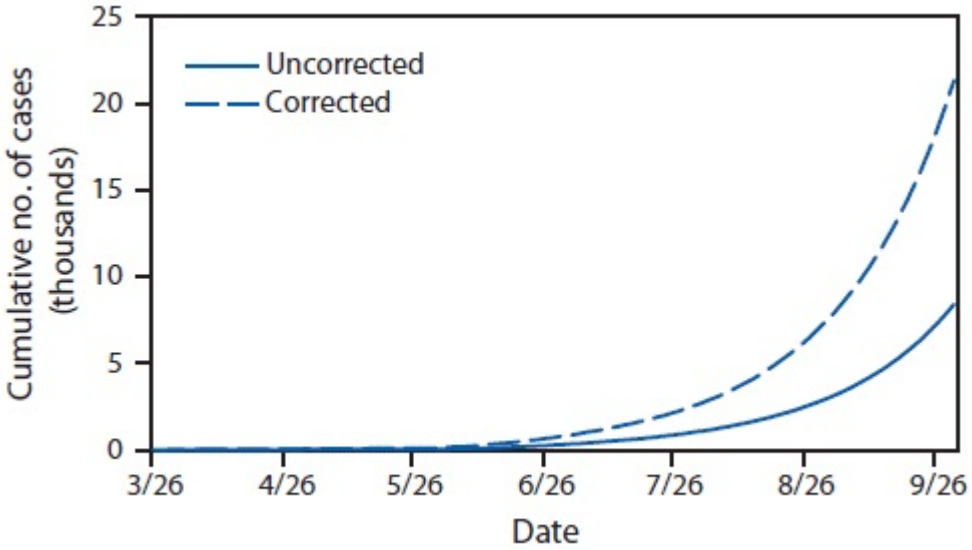


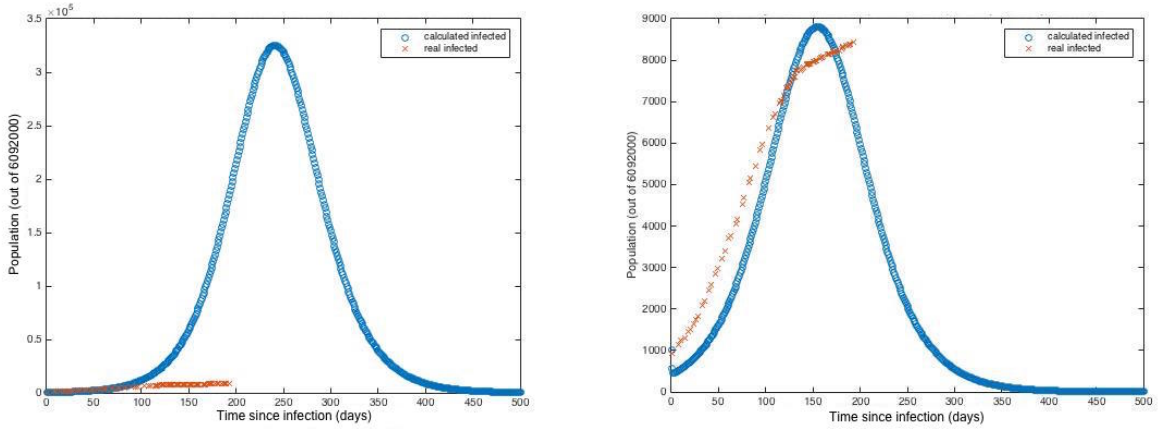
Figure 2. Results from the EbolaResponse modeling tool from the article, "Estimating the Future Number of Cases in the Ebola Epidemic - Liberia and Sierra Leone, 2014-2015," with the Centers for Disease Control. This figure includes corrected and uncorrected estimations for underreporting. Estimation goes until September 30, 2014[27].
(Maltzer, Martin I., Charisma Y. Atkins, Scott Santibanez, Barbara Knust, Elizabeth D. Ervin, Stuart T. Nichol, Inger K. Damon, and Michael L. Washington 2015)
http//www.cdc.gov/mmwr/preview/mmwrhtml/su6303a1.htm?viewType=Print\&viewClass=Print

Simulation Results

Deterministic Models

Our initial step was to recreate the prediction that was bigoted from the recent use of the SEIHFR model in modeling Ebola by Rivers et al. For this, we started with the deterministic model, using MATLAB as a simulation environment. The mathematical models were translated to computational equations, and were differentiated using the ode45 function. Our modeling was limited to Sierra Leone, as its data was the most reliable. Our results predictably showed an exaggeration of the outcome of the epidemic (Figure 3a). This can be attributed to the early attempts of parameter estimation, in which the epidemic seemed to grow exponentially.

Our next goal was to attempt to optimize the parameters with the updated data (See parameter Estimation). While the previous, outdated predictions of the model pointed towards exponential growth, recent data suggests that the epidemic has been following a logistic growth. To better fit the model to this logistic growth, we used the Least Squared Error (LSE) method to optimize the parameters. With these optimizations, the model's predictions were much improved from the outdated model, though still exaggerated (Figure 3b).



(a) Ebola model prediction with outdated parameters obtained by Rivers et al.

(b) Ebola model prediction with optimized parameters.

Figure 3. Parameter optimization of the SEIHFR model yields more realistic results that better fit the observed data. LSE method was used to estimate the parameters.}

Using the LSE method maintained the exponential growth of the model, which could either mean that the SEIHFR model cannot adapt to a logistic growth or that the LSE method was simply not sufficient to optimize the parameters to fit a logistic growth in this particular model.

Stochastic Models

Stochastic models were explored by the authors as an extension of the framework of the deterministic version of Legrand's SEIHFR model[25]. The aim was to implement a stochastic

interpretation of previous deterministic work in modeling of Ebola, specifically a comparative study with Caitlin Rivers's research[29].

As the SEIHFR model was the focus of this project, the implementation focused on what parameters (stochastic rate constants) were to be used for this model, and any way to improve them. For unimolecular reactions (transitions) the deterministic rate constant is analogous to the stochastic rate constant. Certain unimolecular parameters that represented transitions in units of time still had to be translated to an official stochastic parameter, by taking the inverse to determine the rate constant. However bimolecular reactions are more complicated as they rely on the collision of two different species. The rate constant for the stochastic bimolecular parameter is the volume dependent version of the deterministic model, however for our purposes in modeling the population of an indeterminate amount of space for a large population, we directly used the deterministic rate constants for these reactions as well[30]. In reality, the parameters for stochastic and deterministic models are different in how they are calculated and represented. Note that the purpose of these rate constants was to find the probabilities in order to implement the Gillespie model, as opposed to the direct use by the deterministic model.

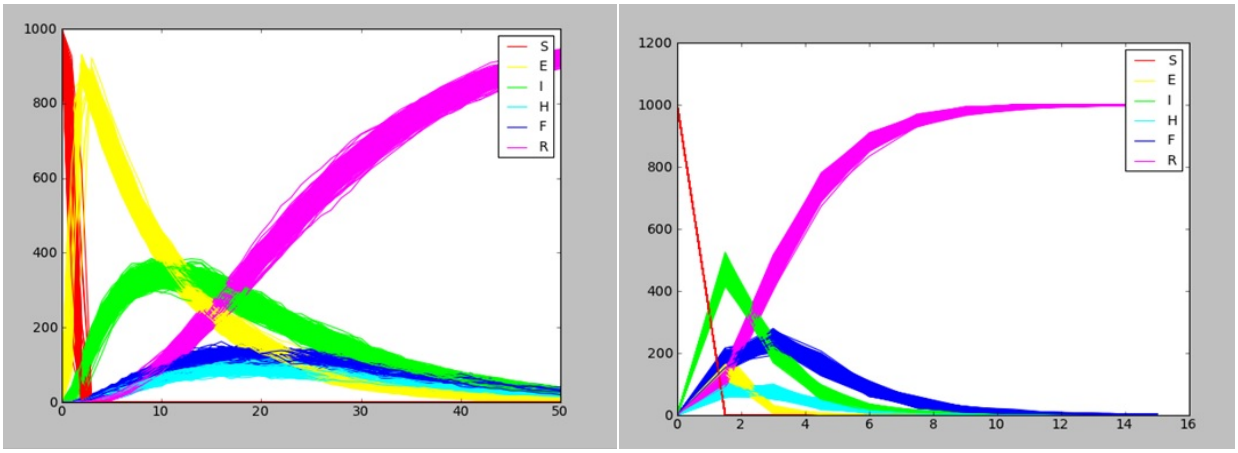
Consequently, the models were developed through the manipulation of the deterministic rate constants. The stochastic model was set up similarly to the deterministic model, with species and reactions. The SEIHFR model framework[25] was the basis for the stochastic modeling as well, as it is more complex than the SIR model, and was used for our comparative study with Caitlin Rivers's work in developing the deterministic SEIHFR model. In this model, the species are: susceptible (S), exposed (E), infected (I), funeralized (F), hospitalized (H), removed (R). These behave according to the subsequent interactions, or"reactions":

Original Parameter[29]	Updated Parameter	Reaction	Description
0.128	0.190	$S+I \rightarrow E+I$	General infection
0.111	0.668	$S+F \rightarrow E+F$	Susceptible infected by funeralized subject
0.080	0.641	$S+H \rightarrow E+H$	Susceptible infected by hospitalized subject
0.100	1.555	$E \rightarrow I$	Transition period: exposed becomes symptomatic
0.048	0.285	$I \rightarrow H$	Infected individuals going to hospital
0.058	0.419	$I \rightarrow F$	Infected individuals dying and being funeralized
0.120	0.838	$H \rightarrow F$	Hospitalized patients dying and being funeralized
0.222	0.726	$F \rightarrow R$	Transition of funeralized being removed
0.010	0.085	$I \rightarrow R$	Infected being removed
0.016	0.344	$H \rightarrow R$	Hospitalized being removed

Table 1. Comparison of the original parameters of the deterministic Ebola SEIHFR model from the parameters determined by Caitlin Rivers[29] with the newly determined deterministic parameters in their stochastic representation.

For the stochastic modeling of this epidemic, CAIN software was used[21]. Our simulations were based on the Gillespie algorithm, which requires defined "species" and “reactions," and our calculated reaction probabilities, as represented by represented by corresponding parameters for the model. Our simulations, performed 1000 realizations per simulation for various population sizes in order to realize the dynamics. Each simulation started out with one infected in the population.

The first stochastic implementation used the parameters used by Caitlin Rivers[29] and validated by recreating the deterministic model. The stochastic model used a smaller population size (1000 total individuals), and found results with significant stochastic variation (Figure 4a) however notable trends are apparent in the path of the data, despite internal variation between trajectories This process was repeated with the new parameters that were determined for the stochastic model. All factors were kept constant, except for the change in parameters (Figure 4b).

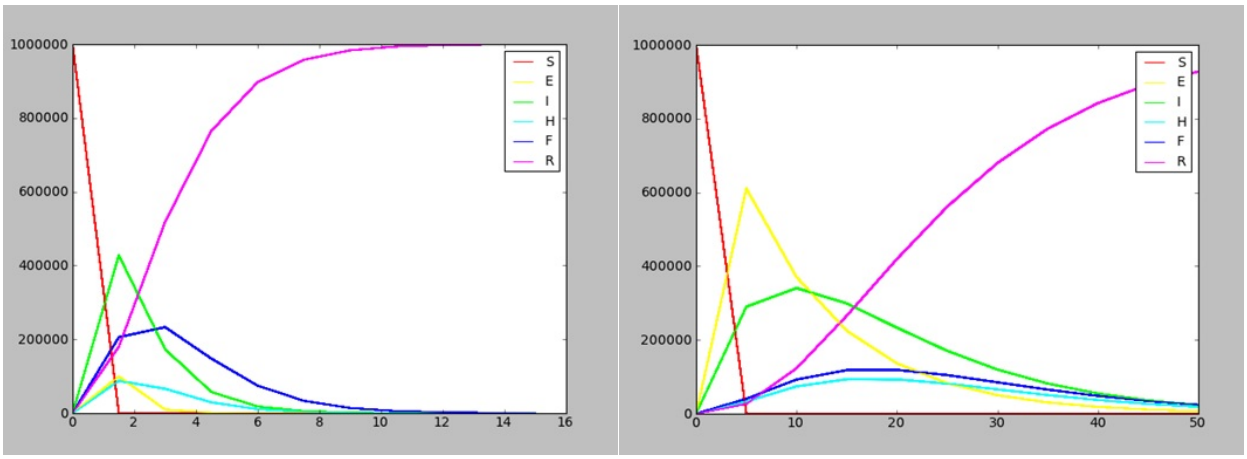


(a) (b)

Figure 4. Results from 1000 trajectories of the stochastic simulation of modified SEIHFR model with Tau Leaping and susceptible population size of 1000. Figure A was based on original parameters[29] and figure B was based on our updated parameters.

The next step of the implementation was to work with a larger population size in the model. This was done with a population size of 1,000,000, and the same rate constants. This model showed a decrease in variation of simulation trajectories from the previous model. For both stochastic models that were simulated at 1,000 and 1,000,000 individuals, it appeared that there was a decrease in variation as the population size increased (Figure 5). This is related to the thermodynamic limit, as both the number of particles (individuals) and volume approach infinity (where the ratio of particles to volume [population density], remains constant), we find that the stochastic model approaches the deterministic model.[30, 35]

Implementation for the stochastic versions of both Caitlin Rivers's model and the updated parameter model gave an interesting comparison. In both models, the susceptible population rapidly decreased to zero. This means that the entire population would have become infected in a short amount of time, which is not likely to happen in an actual epidemic. Simulation with the new, updated parameter changed the population much faster than the parameters based off of Rivers's model, showing similar results for 15 unit time as opposed to the previous 50 unit time (figures 4 and 5). Additionally, there was less observed variation in the populations with the updated parameters, especially for the population of smaller size. The updated parameters do not appear to give much improvement to Caitlin Rivers's original parameters, even in light of the stochastic interpretation. Along with examination of previous work, these results are just the beginning of our exploration into the field of stochastic modeling and parameter estimation, giving an interesting glimpse into the process and results of stochastic simulation.



(a) (b)

Figure 5. Results from 1000 trajectories of the stochastic simulation of modified SEIHFR model with Tau Leaping and susceptible population size of 1000. Figure A was based on original parameters[29] and figure B was based on our updated parameters.

Prediction and Control

Further research in modeling of epidemics may be fruitful in gaining understanding of the process and parameter determination. However there are a number of limitations that come along with this modeling. The SEIHFR model may be too complex and limited in its ability to accurately represent the various facets of the population. There are many compartments to investigate in this model, and the data that is obtained from the actual epidemic only accounts for the new cases and deaths, which at times was not accurate enough to be reliable. The identifiability of some compartments cannot be affectively measured at this time using available databases. Furthermore, their contribution to the results may not critical to determining the path of Ebola, when compared to the standard SIR model. Consequently, there may be a question to its validity, and a possible reason for reverting back to the SIR model, as it is simpler and may be just as useful and accurate. Additionally, with the stochastic modeling exhibited in this paper, there appears to be a dire need for separate parameter determination, or a different process of extracting stochastic parameters from the more easily calculated deterministic parameters. There is much to be done in this area of work.

Deterministic and Stochastic modeling can be applied to many more systems than just the current Ebola epidemic. Applying these techniques and future development are critical in the use of epidemiology to fight the spread of disease. The scope of these application extends far beyond our work, and the basics of modeling. Epidemiology and the use of modeling and complex systems offers a rich field to explore and connect to real world topics and crises.

Acknowledgements

We would like to thank a number of people who aided us in our research, their help was truly invaluable. Dr. Gabor Lente at University Debrecen, Hungary was a great help with our parameter estimation and how the deterministic parameters related to stochastic counterparts. Dr. Marisa Eisenberg provided direction and assistance in learning parameter estimation theories. We are also incredibly grateful for Dr. Caitlin Rivers's research and deterministic models, which gave us direction and were an inspiration for the start of our project. PE thank the Henry Luce Foundation to let him to serve as a Henry R Luce Professor.

References

1. CDC. "Risk of Exposure." *Centers for Disease Control and Prevention*. Centers for Disease Control and Prevention, 22 July 2015. Web.
2. WHO, Origins of the 2014 Ebola epidemic. (2015). Retrieved June 3, 2015, from http://www.who.int/csr/disease/ebola/one-year-report/virus-origin/en/
3. Wilson, C. (2015, April 12). Ending Ebola: The 2014 Ebola Outbreak. Retrieved June 3, 2015, from http://thebrainbank.scienceblog.com/2015/04/12/ending-ebola-the-2014-ebola-outbreak/
4. CDC. "Signs and Symptoms." *Centers for Disease Control and Prevention*. Centers for Disease Control and Prevention, 02 Nov. 2014. Web.
5. WHO. "Ebola Situation Report - 8 July 2015." *Ebola Situation Report*. World Health Organization, 8 July 2015. Web. http://apps.who.int/ebola/current-situation/ebola-situation-report-8-july-2015
6. WHO. "Ebola Situation Report - 27 May 2015." Ebola Situation Report. World Health Organization, 27 May 2015. Web. http://apps.who.int/ebola/en/current-situation/ebola-situation-report-27-may-2015
7. Network Dynamics and Simulation Science Laboratory. "Ebola." Ebola. Virginia Bioinformatics Institute, 2015. Web. 23 Aug. 2015. http://www.vbi.vt.edu/ndssl/ebola
8. Butler, Declan. "Six Challenges to Stamping out Ebola." Nature.com. Nature Publishing Group, 24 Feb. 2015. Web. http://www.nature.com/news/six-challenges-to-stamping-out-ebola-1.16964
9. O'Carroll, Lisa. "Charities Fear Ebola Resurgence after Spike in Cases." The Guardian. The Guardian, 11 June 2015. Web. http://www.theguardian.com/world/2015/jun/11/ebola-charities-fear-resurgence-spike-infections
10. Valdez, L. D., H. H. Aragão Rêgo, H. E. Stanley, and L. A. Braunstein. "Predicting the Extinction of Ebola Spreading in Liberia Due to Mitigation Strategies." Sci. Rep. Scientific Reports 5 (2015): 12172. Web. http://www.nature.com/srep/2015/150720/srep12172/full/srep12172.html
11. Schwartz, E., Choi, B., & Rempala, G. (2015). Estimating epidemic parameters: Application to H1N1 pandemic data. Mathematical Biosciences. doi:doi:10.1016/j.mbs.2015.03.007
12. WHO Ebola Response Team. "Ebola Virus Disease in West Africa --- The First 9 Months of the Epidemic and Forward Projections." *World Health Organization*. N.p., 2014. Web. <http://dx.doi.org/10.1056/NEJMoa1411100>.
13. WHO. "Situation Reports with Epidemiological Data: Archive." *World Health Organization*. WHO, 2015. Web. 16 Oct. 2015. <http://apps.who.int/ebola/en/ebola-situation-reports-archive>.
14. Rivers, Caitlin. "How to Run a Crowdsourced Data Repo" /." *Caitlin Rivers*. N.p., 1 Apr. 2015. Web. 16 Oct. 2015. <http://www.caitlinrivers.com/blog/how-to-run-a-crowdsourced-data-repo>.
15. Rivers, Caitlin. "Ebola Data from Situation Reports vs Patient Database." *Caitlin Rivers*. N.p., 20 Nov. 2014. Web. 16 Oct. 2015. <http://www.caitlinrivers.com/blog/ebola-data-from-situation-reports-vs-patient-database>.
16. OCHA ROWCA. "Sub-national Time Series Data on Ebola Cases and Deaths in Guinea, Liberia, Sierra Leone, Nigeria, Senegal and Mali since March 2014." -*Humanitarian Data Exchange*. HDX Beta, Mar. 2014. Web. 16 Oct. 2015. <https://data.hdx.rwlabs.org/dataset/rowca-ebola-cases>.
17. WHO. "Ebola Situation Report Map." *World Health Organization*. WHO, 14 Oct. 2015. Web. <http://maps.who.int/MapJournal/>.
18. University of Zurich. "Stochastic Simulation of a Simple Epidemic."Stochastic Simulation of a Simple Epidemic. University of Zurich, n.d. Web.

19. Erban, Radek, Jonathan Chapman, and Philip Maini. "A practical guide to stochastic simulations of reaction-diffusion processes." *arXiv preprint arXiv:0704.1908* (2007). https://people.maths.ox.ac.uk/erban/Education/StochReacDiff.pdf.
20. Karig, David. "Introduction to Stochastic Simulation with the Gillespie Method." Princeton, n.d. Web. 18 Apr. 2015. <https://www.cs.princeton.edu/picasso/seminarsS05/Karig_slides.pdf>.
21. Cain: Stochastic Simulations for Chemical Kinetics. (n.d.). Retrieved June 3, 2015, from http://cain.sourceforge.net/
22. Levin, Drew. "Matlab Geeks." *Matlab Geeks*. Wordpress, 24 Dec. 2011. Web. 16 Oct. 2015. <http://matlabgeeks.com/tips-tutorials/modeling-with-odes-in-matlab-part-4a>.
23. WHO. "Frequently Asked Questions on Ebola Virus Disease." *World Health Organization*. N.p., 2015. Web. 16 Oct. 2015. <http://www.who.int/csr/disease/ebola/faq-ebola/en/>.
24. Gire, Stephen K., and Et Al. "Genomic Surveillance Elucidates Ebola Virus Origin and Transmission during the 2014 Outbreak." *Science AAAS*. American Assocation for the Advancement of Science (AAAS), 28 Sept. 2014. Web. <http://www.sciencemag.org/content/345/6202/1369.full>.
25. LEGRAND J, GRAIS RF, BOELLE PY, VALLERON AJ, FLAHAULT A. Understanding the dynamics of Ebola epidemics. *Epidemiology and Infection*. 2007;135(4):610-621. doi:10.1017/S0950268806007217.
26. Pandey, A., K. E. Atkins, J. Medlock, N. Wenzel, J. P. Townsend, J. E. Childs, T. G. Nyenswah, M. L. Ndeffo-Mbah, and A. P. Galvani. "Strategies for Containing Ebola in West Africa." Science 346.6212 (2014): 991-95. Science AAAS. American Assocation for the Advancement of Science (AAAS), 30 Oct. 2014. Web. <http://www.sciencemag.org/content/346/6212/991.full>. http://www.sciencemag.org/content/346/6212/991.full.
27. Maltzer, Martin I., Charisma Y. Atkins, Scott Santibanez, Barbara Knust, Elizabeth D. Ervin, Stuart T. Nichol, Inger K. Damon, and Michael L. Washington. "Estimating the Future Number of Cases in the Ebola Epidemic — Liberia and Sierra Leone, 2014–2015." Centers for Disease Control and Prevention. Centers for Disease Control and Prevention, 07 Oct. 2014. Web. http://www.cdc.gov/mmwr/preview/mmwrhtml/su6303a1.htm?viewType=Print&viewClass=Print
28. WHO. "WHO: Ebola Response Roadmap Situation Report 24 September 201." World Health Organization. World Health Organization, n.d. Web. 24 Sept. 2015. http://www.who.int/hac/crises/cod/sitreps/ebola_roadmap_sitrep_24september2014.pdf
29. Rivers CM, Lofgren ET, Marathe M, Eubank S, Lewis BL. Modeling the Impact of Interventions on an Epidemic of Ebola in Sierra Leone and Liberia. PLOS Currents Outbreaks. 2014 Nov 6. Edition 2. doi: 10.1371/currents.outbreaks.4d41fe5d6c05e9df30ddce33c66d084c.
30. Erdi, P., & Lente, G. (2014). Stochastic chemical kinetics: Theory and (mostly) systems biological applications. New York, New York: Springer.
31. Butler, Declan. "Six Challenges to Stamping out Ebola." *Nature.com*. Nature Publishing Group, 24 Feb. 2015. Web.http://www.nature.com/news/six-challenges-to-stamping-out-ebola-1.16964
32. CDC, 2014 Ebola Outbreak in West Africa. (2015, May 22). Retrieved June 3, 2015, from http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/
33. *Kermack, W. O. and McKendrick, A. G. "A Contribution to the Mathematical Theory of Epidemics." Proc. Roy. Soc. Lond. A 115, 700-721, 1927.*
34. Érdi, Péter, and J. Tóth. "Mathematical Models of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models." *Mathematical Models of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models*. Princeton, NJ: Princeton UP, 1989. Print.
35. Kurtz, Thomas J. "The Relationship between Stochastic and Deterministic Models for Chemical Reactions." *The Journal of Chemical Physics* 57, 2976 (1972): *The Relationship between Stochastic and Deterministic Models for Chemical Reactions*. Web. 14 Nov. 2015. <http://dx.doi.org/10.1063/1.1678692>.